Analysing Vehicular Mobility Structure under a Temporal Perspective

Gabriel Ribeiro Diniz · Felipe Domingos da Cunha · Antonio A. F. Loureiro

Received: 05 April 2021

Abstract In recent years, Vehicular Networks have received much attention because they can contribute to solving several problems of urban scenarios. An interesting perspective to characterize these urban problems is to study vehicles’ mobility records to increase our knowledge about them. Thus, an alternative is to model these records as graphs enabling us to apply algorithms and graph theory to design new solutions for vehicular networks. This work explores three different strategies to model vehicular networks and analyzes two real well-known traces from Rome and San Francisco. We perform a study on the network aspect, measuring metrics that portray some network characteristics and proprieties. We highlight the advantages and limitations of each approach evaluated. Also, we present some applications and future directions for the knowledge extracted with this analysis.

Keywords VANET’s Analysis · Mobility · Reliability · Time-Varying Graphs
1 Introduction

Vehicular Networks (VANETs) are special ad hoc networks based on wireless communication (Moustafa and Zhang, 2009; Cunha et al., 2014a). This network is composed of vehicles and roadside units and is considered the fundamental key to intelligent transportation systems. In VANETs, the communication may occur by direct contact among the vehicles, by contact between vehicles and Roadside Units (infrastructure), or with a hybrid system. The Roadside Units are a gateway between OBUs and the communications infrastructure, especially when the network is sparse. We can find VANETs in urban centers, highways, or rural areas.

In the last decades, the increase of the vehicular fleet has caused several problems in vehicular scenarios like traffic jams, pollution, and fuel consumption. VANETs could help solve these problems by assisting drivers during their journeys, sending alerts to avoid collisions, and providing services that can save fuel, decrease pollution, reduce traffic jams, and provide good entertainment to the passengers in vehicles (Cunha et al., 2014a).

One of the crucial challenges in VANETs is the network analysis used to identify patterns that can help us understand how mobility flows. Thus, several researchers have been conducted in the literature to analyze and identify features that could improve the traffic experience. As a result of this analysis, it would be possible to create protocols, applications, and services to explore the obtained knowledge improving the network performance (Luo et al., 2010).

An interesting approach to study the vehicles’ mobility is to model a VANET using graphs: a mathematical tool that portrays the relationship among vehicles (nodes) (Christensen and Albert, 2007). These structures enable us to apply graph theory metrics, exploring the vehicles’ connectivity, which provides a more refined analysis. The model used to build a graph of a VANET is a critical design issue that directly impacts the network analysis. Moreover, another essential factor to consider in this investigation is its temporal aspect. When we analyze only a single instant (snapshot), we could lose much information that the temporal dimension naturally reveals.

A Literature overview reports three graphs used to model a vehicular network: instantaneous, aggregate, and temporal. The first approach is the graph of a single snapshot of the role period is analyzed. An instance of a graph is constructed from this single shot, and the analysis is performed. Secondly, in the Aggregate graph, all the contacts between vehicles are aggregate in one single and process study. Finally, in the Temporal graphs, the temporal aspect is considered, and the research is conducted to capture the network’s dynamic over time.

This article presents an analysis of two real well-known data sets of vehicular mobility from Rome, Italy, and San Francisco, USA, which portray real vehicles’ mobility data. In particular, we perform this analysis using three different graph models: instantaneous, aggregated, and time-varying. We focus on verifying which approach can deliver the highest reliability and accuracy. More specifically, our contributions include:
We compare the three different approaches used to analyze vehicular mobility, showing each one’s advantages and disadvantages, evaluating metrics to extract knowledge about the vehicular mobility properties.

We discuss the different metrics that can reveal the network’s real properties, highlighting each one’s advantages and applications.

We describe a group of applications in vehicular networks where we can explore our analysis results. Also, we present some works to explain how this methodology has been used.

The remainder of this work is organized as follows. In Section 2, we describe the primary studies from the literature that have used and characterized the existing mobility traces. The data traces analyzed in this work, and a summary of the approaches are described in Section 3. The characterization traces are presented in Section 4, together with a discussion about how the different methods could manifest different results. Section 5 demonstrates how we can apply our results. Finally, in Section 6, we offer our conclusions and future remarks.

2 Literature Overview

In recent years, the volume of vehicles has become increasingly more significant. This increase added to the lack of urban planning resulted in several traffic problems. In the past years, several studies have been published in the literature that model and study vehicular mobility (Zhu et al., 2011) to solve these problems. Usually, in these studies, it is common to model the vehicular network using graphs. By the use of these structures, we can apply some specific algorithms to analyze the mobility traces. In the literature, there are different ways to build these graphs. Usually, the three different approaches employed to generate vehicular graphs are instantaneous graphs (models a specific instant in time), aggregate graphs (models a particular interval in time), and time-varying graphs (models a scenario that evolves) best explained in Section 3.

An instantaneous graph captures a snapshot of the network at a given moment. In this way, it is impossible to know what has happened before in the network and how it will evolve. Naboulsi and Fiore (2013) used the instantaneous topology behavior of Cologne (Uppoor and Fiore, 2011) to identify limitations in connectivity, availability, reliability, and navigability of the topology in-network and city component levels. The vehicles tend to group into small clicks, leading to believe that store-carry-forward protocols and implementation of Roadside Units (RSUs) could improve the connectivity. Moreover, Naboulsi and Fiore (2017) analyzed the Cologne topology in terms of connectivity, availability, reliability, and navigability and unveiled how the underlying structure of the vehicular network is composed of vehicles gathered into small clicks connected in a weak, intermittent fashion.

Hou et al. (2016) also modeled the impact of mobility on the VANET connectivity using instantaneous graphs. They analyzed the instantaneous graphs
of the network every 10 minutes. As a result, the authors discovered that connectivity is affected by mobility. Also, they used instantaneous graphs to investigate the relationship between the communication rate and the giant component size. To quantify the quality of the routing protocols, Pallis et al. (2009) studied the connectivity in a synthetic trace of the city of Zurich. It is essential to observe that the instantaneous graph ignores the temporal element, which usually loses essential information about network behavior evolution.

An aggregate graph represents all mobility events in a given time, i.e., it groups the mobility information of a given interval into a single graph. Guerber et al. (2018) analyzed the contacts in a trace of Rome (Bracciale et al., 2014) considering the degree, intermediation, proximity, and cohesion measures. They generated an aggregate graph to create a GeoSocial protocol for message routing.

Cunha et al. (2014b) created aggregate graphs for Zurich and San Francisco’s mobility datasets using a time window of 1 hour to understand their social analysis. In Zurich’s data, they noticed the small world phenomenon, where the degree of distribution follows a Power Law that describes the network as a free-scale network. In San Francisco data, no social properties were found.

Comparing how gaps in traces affect the analysis of the contact duration and network capacity, Cunha et al. (2016) aggregated one day of each analyzed dataset into a single graph. Also, Diniz et al. (2017) performed a classification of vehicular mobility in the city of Rome. The dataset was aggregated every hour, and the analysis was performed within these time intervals. In the same direction, Qiu et al. (2018) used a taxi trace of a city in China and grouped the data into 4-hour groups separated on 21 different days. The objective was to analyze how vehicles’ speed, the relative velocity between communicating vehicles, and link duration can significantly affect V2V communications and the network. When we use the aggregate graph approach, unreliable data is generated because it considers that one edge remains during the whole time, which leads the analysis to interpret that this contact exists during the entire time.

When we analyze traces, the temporal aspect is essential. Thinking about that, Qiao et al. (2017) evaluated the temporal structural characteristics involved with time-ordered pathways, accessibility, and connectivity in a dataset of Beijing. The authors used a time-varying graph model where a graph $G = (V,E)$ where $V$ represents all vehicles, and $E$ represents a link between vehicles where the distance is created. Euclidean distance between the vehicles is smaller than the radius of the communication. Each edge is formed by a quadruple $(u,v,t,\lambda)$ where $u$ and $v$ are the vehicles, $t$ is the time the contact occurred and $\lambda$ is the duration of the contact.

Glacet et al. (2015) analyzed Bologna and Cologne’s cities and divided their study into two steps. First, a sequence of snapshot graphs was generated, such that $G_t = (V,E)$ represents the connection graph at the instant $t$. The authors defined itineraries as a sequence of tuples, where each tuple is made up of a pair of nodes and timestamps. If this itinerary is doable at
analysed trace data. In the analysis, it was possible to identify a relationship
that highlights the suitability of temporal communication for vehicular envi-
ronments and can provide valuable guidelines for evaluating and predicting
the performance of vehicular networks.

After analyzing all these works, we summarize all of the main features in
one single Table 1. It will be easier to compare the studies and identify what
each research did. Each work analyzes different metrics for complex networks,
which suggests the need to centralize all of the metrics in a single work, explor-
ing the impact these different modeling approaches have on complex networks’
metrics.

3 Modeling

In this section, we discuss the methodology used to perform this work. We
present details about the datasets, the graph models used to capture the trace
interactions, and the metrics applied to evaluate the results.

3.1 Graph Models

Aiming to perform our analysis, we choose three approaches based on the
most used in the literature, as shown in Figure 1. The instantaneous and the
aggregate graphs are frequent by researchers. Still, the time-varying graph is
an unconventional approach, and it is harder to find works that use these
approaches considering time as a factor.

**Instantaneous Graph:** an instantaneous graph is a representation of the
network in time \( t \), we formalize a graph like \( G_I = (V,E) \). \( V(t) = \{v_i(t)\} \) is
a set of nodes that represent the vehicles traveling through the streets in the
time that the graph was generated, and \( E(t) = \{e_{ij}(t)|v_i(t),v_j(t) \in V, i \neq j\} \)
is a set of edges that represent vehicle \( i \) and \( j \) are within a predefined limit of
communication.

**Aggregate Graph:** an aggregate graph is a representation of the network
during an interval. First, we define the begin (\( t_b \)) and the end (\( t_e \)) of an
interval. Second, we define \( G_A = (V,E) \), we \( V \) is the set of nodes that contain
all the nodes that make up the network during the period, and \( E \) is the set
of edges that includes all the edges that appeared in the network during the
time interval. In this work, a period of 15 minutes was used to generate the
aggregate graph. This time was chosen because it is often used in related work.

**Time-Varying Graph:** we defined a time-varying graph as a set of graphs
\( G_t = (G_0,G_1,...,G_n) \) and every element of \( G_t \) represent the network in a time
\( t \). We create a time-varying graph of a hole day in both traces resulting in a
set of 86,400 graphs representing the 24 hours of the day.
3.2 Vehicular Mobility Traces

With the progress of technology and the widespread use of vehicles equipped with GPS devices, several traces have been become available, mainly taxis, since they can easily share their location.

We analyze two well-known large-scale traces of two big cities, Rome in Italy and San Francisco in the USA. Catches of Rome took place between February 1, 2014 and March 2, 2014, totaling 30 days of 24 hours. The records were taken every seven seconds by GPS installed within 320 taxis. The file contains approximately 1.5 GB with 21.8 million lines. Data collected have the unique vehicle identifier, date, time of reading, and vehicle’s geographical position (latitude, longitude). The San Francisco trace is a real dataset with GPS information of 551 taxis in San Francisco’s city for four weeks. This information is captured from GPS devices in each vehicle, at each minute, over its trajectory. To conduct our work, we select a day to perform the analyses.
in both traces. It is worth mentioning that both traces are calibrated (Cunha et al., 2016).

3.3 Macroscopic Metrics

The macroscopic metrics represent the network global state measures, portraying all vehicles’ general behavior and the graphs’ evolution over time.

**Diameter:** it is the longest distance between any pair of nodes. To find the graph’s diameter, find the shortest path between each pair of nodes. The largest of paths is the diameter of the graph. In this work, we calculate the diameter of the number of hops.

**Density:** the density is 0 for a graph without edges and 1 for a complete graph. This metric is essential that reveals network connectivity, and higher values portray a network as more connected.

**Edges Number:** if there is a distance between two nodes, which is less than a pre-established maximum communication limit, then there is an edge between these nodes. The maximum communication limit defined in this work is 200 meters.

**Number Connected Components:** in an undirected graph, a connected component is a subgraph that a path connects all nodes that belong to this subgraph.

**Number Nodes in Largest Component:** this metric represents the number of nodes in the component with the higher number of nodes.

3.4 Microscopic Metrics

The microscopic metrics define individual values for the node. We compute the metrics for all the nodes, and we present the average value over time.

**Average Node Degree:** node degree is the number of edges connected to the node. We calculate the average of all node’s degrees in the network.

**Average Betweenness Centrality:** Compute the shortest-path betweenness centrality for nodes. Betweenness centrality of a node $v$ is the sum of the fraction of all-pairs shortest paths that pass through $v$ Freeman (1977):

$$C_B = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)}$$  \hfill (1)

where $V$ is the set of nodes, $\sigma(s,t)$ is the number of shortest $(s, t)$-paths, and $\sigma(s,t|v)$ is the number of those paths passing through some node $v$ other than $s, t$. If $s = t$, $\sigma(s,t) = 1$, and if $v \in s,t$, $\sigma(s,t|v) = 0$.

**Average Closeness Centrality:** closeness centrality (Freeman, 1979) of a node $u$ is the reciprocal of the sum of the shortest path distances from $u$ to all $n - 1$ other nodes. Since the sum of distances depends on the number of nodes in the graph, closeness is normalized by the sum of the minimum possible distances $n - 1$. 

where \( d(v, u) \) is the shortest-path distance between \( v \) and \( u \), and \( n \) is the number of nodes in the graph.

**Average Clustering Coefficient:** for unweighted graphs, the clustering of a node \( u \) is the fraction of possible triangles through that node that exists,

\[
C_u = \frac{2T(u)}{\text{deg}(u)(\text{deg}(u) - 1)}
\]

where \( T(u) \) is the number of triangles through node \( u \) and \( \text{deg}(u) \) is the degree of \( u \). The value of \( C_u \) is assigned to 0 if \( \text{deg}(u) < 2 \).

### 4 Numerical Results

This section presents the results obtained in this study. The purpose is to compare the influence among different graphs used in the modeling have on the results. A shown in Section 2, the three approaches analyzed are found in the literature in several works. However, our study explores an analysis comparing the metrics measured in the three approaches.

#### 4.1 Traffic Characterization

First of all, we explore a traffic flow analysis throughout the day in two selected databases. Thereby, it is likely to comprehend the traffic trend, and we present a better understanding of the results in the following sections.

In San Francisco, it is possible to notice by Figure 2-(a) that the traffic of vehicles has a different pattern of movement from the city of Rome; it is heavy around midnight. Since that time, vehicles' flow has started to reduce to 7am, reaching the lowest number of cars in transit. Then, the traffic increases again, and nearby 12pm, it hits its flow peak once more and remains like that until midnight.

Figure 2-(b) shows the mobility of vehicles in Rome during the day. It is possible to note that between 12pm and 4am, a few vehicles are driving around the city. Afterward, the flow of vehicles begins to increase, and around 1pm, the network flow achieves its peak. After these peak hours, the flow slows, and it gets calm between 4pm and 8pm. After that, the movement in the network declines again until the end of the night.

#### 4.2 Comparing Approaches

This section introduces the numerical results for the three approaches: Instantaneous graphs, Aggregate graphs, and Temporal graphs for the Rome
Although the Instantaneous graph’s purpose is to analyze a mere instant of the total time, we decided that it would be more just that it was generated every 15 minutes. Thus, the comparison with the Aggregate and Temporal graphs would be fairer. At first, we will discuss the number of edges. By observing the Figures 3-(a) and 4-(a), it is possible to see the difference in the three approaches in both cities.

The results identified how the aggregate graph presents higher edge values than the other two approaches during the day. These values also happen when investigating network density, which depends directly on the graph’s number of edges. It becomes possible to confirm it though the Figures 3-(b) and 4-(b).

In the Instantaneous graph, the values are closer to reality. But, there is still an information loss every 15 minutes, since a simple second is analyzed in the period.

Remarkably, the Temporal graph is likely to examine and continue the metrics’ evolution throughout the day. With a high number of edges, the graph becomes more connected, and the number of components is smaller because the nodes are bundled into large components. It happens when analyzing the connected components in the aggregate graph in Figures 3-(c) and 4-(c).

In the Instantaneous graph, it is possible to verify the evolution of the related components more realistically. Only in the Temporal graph can it be possible to understand what truly occurs in every moment during the day. Since there are fewer components in the aggregate metric, each of them has more nodes, and consequently, the number of nodes in the largest component increases. This can be observed in Figures 3-(d) and 4-(d).

Another metric directly affected by the number of edges is the average rate of nodes. Analyzing the Figures 3-(e) and 4-(e), we can see that in the Aggregate graph, the number of nodes is superior when compared to the other two graphs. It occurs due to the current connections that are added in a single graph every 15 minutes. It is fundamental to point out that the component’s
number and nodes’ analyses reveal how connected the network is. In this way, it is possible to identify how to spread solutions and deliver content on the network.

Therefore, Temporal analysis can portray these patterns more realistically. Other metrics that show the impacts of different modeling types are the Clustering Coefficient, the Betweenness Centrality, and the Closeness Centrality. By observing Figures 3-(g), 4-(g), 3-(h), 4-(h), 3-(i), and 4-(i), these three metrics present higher values in the Aggregate graph when compared to the Instantaneous and Temporal graphs. This behavior is related to the greater number of edges in the Aggregate graph.

In a graph with more edges and fewer components, nodes tend to create groups because the possibility of sharing the same purpose grows. Consequently, the option of participating in the shortest path between two other nodes also increases since a higher number of shorter paths will exist in another component.

Finally, the distance from one node to all others (in the number of hops) also tends to be higher. A component’s node has more nodes to calculate the distance between them. Finally, the diameter of the network will interpret. When analyzing the result in the two databases in Figures 3-(f) and 4-(f), it is clear that the impact in the city of Rome was more significant than in the city of San Francisco. In Rome, the Aggregate approach has larger diameter results than in the Temporal.

In contrast, in San Francisco, this difference is not significant. In the instantaneous approach, it is possible to see the evolution curve of the metric. However, in the Temporal graph, the diameter’s development is perceiving more precisely throughout the day. The diameter is a metric which portrays the network’s connectivity and how many hops are necessary to disseminate a message to reach the whole network.

4.3 Temporal Edge Correlation

Temporal correlation is a method that measures the probability of an edge persists on two consecutive instants in a temporal graph. By analyzing the metric, it is likely to understand better the chance of remaining contacts between vehicles during the day. If a connection tends to exist, it means these vehicles establish a relation among themselves, like a social network.

At first, we discuss the analysis held in San Francisco. In Figure 5-(a), it is possible to see that the temporal correlation in San Francisco is high and tends to keep constant throughout the day. That also means, in its data set, vehicles create lasting connections throughout the day.

Concerning the Figure 5-(b), it is noticed that the contact is likely to be constant, which occurred during the day tends to be stable. This characteristic shows that the vehicle’s contacts are long-lasting and that an automobile remains connected with its social circle. At night, this chance declines a little but remains high, with values close to 0.8. It happens because it is at the end of a
working day, and taxis begin to leave in directions to different neighborhoods, which leads to a no longer existence of some edges.

The temporal correlation can support the improvement of applications for ITS (Intelligence Transportation System) once it allows understanding the behavior of the edges (connectivity) in a graph. When it comes to vehicular networks, as it becomes necessary to make the best use of the connection between vehicles, we can design customized applications, protocols, and services for each situation and location.

4.4 Frequency vs. Duration

In this section, we explore the relation of frequency and duration of contact between vehicles. Thus, we can better learn how the exchange of information among vehicles works during the day. For this analysis, we use the Temporal graph approach because, in a previous study, it was noticed that the temporal graph tends to get closer to reality than the other two models.
Initially, we analyzed the behavior of vehicle connections in San Francisco. As observed in Figure 6-(a), the average of contacts occurred only once an hour. Their connected period is closer to 20 seconds for most of the day. This feature reveals that vehicles do not meet often and do not keep long-lasting contacts. Additionally, it is possible to see that during the night, the contacts last longer than during the day. Probably these are periods where the cabs remain at the final stop. This type of information suggests that different solutions/strategies should be developed/considered for several moments.

In Rome, Figure 6-(b), in the period between 12am at night and 6am in the morning, the more lasting contacts are less frequent. This situation means that vehicles tend to encounter for a longer time, but they rarely encounter again during this period. From 6am, the scenario changes, and contacts have a shorter duration. However, the encounters are more frequent. As such, vehicles tend to stay less time closer; however, they tend to have more intermittent contacts.

In Rome, contacts remain active for a longer time, making it possible to exchange more data between a pair of nodes. On the other hand, it does not
happen in San Francisco, where contacts should be better used since they do not occur frequently, and there is a short duration. Customized solutions will probably be necessary once vehicles have different behavior in each city.

4.5 Metrics Correlation

Aiming to explore the metrics’ correlation and investigate whether there is a dependency among them, we compute the Pearson Correlation. Regarding the Figure 7-(a), we can observe many correlations among the metrics for San Francisco City. Firstly, we discuss the strong direct correlations. There are five metrics, which verify strong and direct connections. These metrics are the Number of Edges, the Average Rate of the nodes, the number of nodes in the Largest Component, the Closeness Centrality, and the Network Density. This behavior is understandable once these metrics are causally linked.
For instance, if the number of edges in a graph rises, the average number
of nodes and the network density also increase. More edges in the graph also
mean it tends to be more connected. As a result, the number of nodes also
becomes higher in the component. In the same case, in each component, more
nodes lead to closeness, which causes a growth in the Closeness Centrality.

In inverse correlations, as one variable increases its value to another, it
decreases. It is observed that a single metric has a strong inverse relationship
with other metrics: the Network Diameter. Regarding the metrics mentioned
above, we can see they are correlated. Because the graph gets connected, its
vertices tend to be closer to each other, explaining the inverse correlation in
the network diameter. The more connected a graph is, the smaller the diameter
will be since the node will have fewer hops to reach the network’s other vertices.

Analyzing Rome’s correlations, it is possible to verify strong direct con-
nections (values between 0.9 and 1 and between 1 and 0.9) and strong inverse
correlations (values between 0.8 and 0.9 and between −0.9 and −0.8). A di-
rect correlation means that as the evaluated variable’s value increases, the
correlated variables also increase its value.

Observing Figure 7-(b), we can identify two strong direct correlations. The
first is the Clustering Coefficient and the average rate of nodes. It portrays
that when a node’s number rises, the possibility of being involved in a group
with similar purposes also increases. This behavior was observed because the
node increases its circle of contacts, increasing the chance of having the same
goals.

Another analyzed the direct correlation is between the network density and
the closeness centrality. This result was expected since the network density
grows, so do the edges in the graph, thus causing the nodes to become closer
to each other. Another observed correlation is between the density and the
number of components. This one is considered inverse and strong because, as
the network density rises, more edges are part of the graph connecting more
nodes, consequently decreasing the number of components.

4.6 Aggregate Time Interval’s Variation

After performing all of these analyzes, we realized that the aggregate time
interval could greatly impact the traces’ analysis. We saw that the different
works that applied the aggregate graph to analyze the trace used other ag-
gregation times through the literature review. Given this fact, we decided to
analyze how the different aggregation times used in the analysis will impact
complex network metrics results.

The aggregation times used in the related works were 15 minutes, 30 min-
utes, 1 hour, and 4 hours. These were also the times we used to create the
aggregated graphs and generate the analyzes. After developing all the ana-
lyzes, we plotted a graph where it is possible to compare the metrics’ results
in each aggregation time used for Rome and San Francisco.
First, let us look at the behavior of the results in San Francisco. It was already predicted that as the aggregation time increases, the metrics will undergo changes in the values measured. As we saw in the previous section, the edge number is a metric related to all other metrics, so let us look at it first. Analyzing Figure 8-(a), we can see that as we increase the aggregation time,
more edges are added to the graph. An interesting factor is a significant increase in this metric when we use the aggregation time of 240 minutes. With a high aggregation time, we can see that the value of the network analysis tries to distance itself a lot from reality.

As a result of the raised edges, the number of nodes has also increased considerably. This behavior occurred when we used the aggregation time of 240 minutes. It is interesting to note Figure 8-(e) that the variation in the degree of the nodes with the aggregation time of 240 minutes is not so great as the day goes on, which shows that vehicles have a maximum standard of contacts per day. As the aggregation time passes and this value approaches the top, it tends to maintain.

Another unusual behavior that we were able to identify was the large increase or decrease in some metrics’ values in the 12-hour time. After analyzing, we assume that this variation identifies a gap in the data collection of some taxis, which would have impacted these metrics considerably. We can see that the diameter (Figure 8-(f)) of the network feels a large increase in the 12-hour time, in addition to the closeness centrality (Figure 8-(i)) and the number of nodes in the largest component (Figure 8-(d)) suffers a sharp drop in their values.

In Rome’s city, it is possible to see that the aggregation time increase also affects the metrics’ results. Let’s start with metrics that are behaving differently than seen in the city of San Francisco. If we look at the betweenness centrality (Figure 9-(h)), the clustering coefficient (Figure 8-(g)), and the diameter (Figure 8-(f)). It is possible to note that these metrics reach their maximum value when the aggregation time is 30 minutes, differently from the behavior seen in San Francisco. This behavior can be determined for many peculiarities of Rome, such as the street’s layout, the length of the trips realized, and some mobility patterns.

Another aspect that we observed was that, even with the number of edges (Figure 9-(a)) growing when we increase the aggregation time from 15 to 30 minutes, the degree of the nodes (Figure 9-(e)) does not follow this behavior and remains almost the same. We can note that the number of edges has not increased considerably, contributing to the node degree not changing during the time observed.

Finally, it is possible to observe that the number of components (Figure 9-(c)) after aggregation time of 30 minutes does not variate, which means that this metric probably reaches its saturated value in the time of 30 minutes. Even if the aggregation time increases, the metric value will not suffer considerable changes.

5 Knowledge Applied

Many are the applications for the knowledge extracted with our analysis. In literature, we found many works and new branches to explore our work. In this section, we present these works and some challenges in applying our
results. We divided this section into five subsections. In these subsections, we will present relevant topics in vehicular networks and what impact our studies have on the work of different areas.

5.1 Network Communication

Ad-hoc networks are wireless networks that do not require a common access point to the nodes connected to them. All network devices work as if they were a router, sending information from neighboring devices in community (Corrêa et al., 2006). Vehicle networks are ad hoc networks composed of vehicles and may also contain fixed infrastructures along the roads. Communication in this network type is often complicated due to vehicular networks’ characteristics, such as a frequent change in network topology, node speed, and signal attenuation.

The authors of the work (Meneguette et al., 2014), aiming to improve communication in vehicular networks, developed a new algorithm to disseminate geographic data that considers the network partitions for urban environments.
According to the authors, the proposed algorithm eliminates the broadcast storm problem. It maximizes the ability to disseminate data in networks with frequent disconnections at the cost of a low delay and a low overhead. The results show that the algorithm is efficient in terms of coverage, sending rate, delay, and maintaining the sending and receiving pattern at different network densities.

Similarly, the authors of the paper SOUSA (2017) present a new congestion signaling protocol for vehicular networks with low communication overhead. The protocol uses only V2V communication; in other words, it is independent of external infrastructures. The purpose of the protocol is to reduce vehicles’ average travel time, ensuring a low communication overhead in VANETs. The proposed protocol presented the best performance in terms of signaling overhead. In scenarios with higher traffic demand, the number of messages generated by the proposed protocol was lower than other related protocols.

The first way to validate a new data dissemination protocol is to run simulations of protocols to compare them. For this, it is necessary to explore the communication among them. As we saw in our results, the type of graph used
to model mobility can directly influence vehicles’ communication. Therefore, our study reinforces the need to perform careful data modeling, so avoiding the simulations favor a specific protocol and explore a scenario closer to reality. Also, it is very suggested to use real vehicular traces to improve the accuracy of the evaluation.

5.2 Understanding City Mobility

City mobility has been studied in recent years. Identifying mobility patterns can help to identify points of improvement in traffic. Consequently, these improvements mean a better quality of life for the population. The advantage of studying mobility is understanding the mobility patterns of a city and taking advantage of these patterns to make progress in the city issues.

The objective of this study Azolin and Silva (2019) is to insert public transport into a strategy for assessing resilience in urban mobility in the face of an eventual restriction in the supply of fuel. In the case study carried out in São Carlos, it was observed that the insertion of public transport, even in conditions of minimal operation, provided a considerable gain in resilience: 21.4% in the most pessimistic scenario for active modes.

The authors of work Lessa et al. (2019) analyze the levels of accessibility by bus in the city of Belo Horizonte. In this work, indicators of bus accessibility and population mobility are proposed. The results showed that some regions present a more significant discrepancy between accessibility and mobility levels, especially those located in peripheral areas where access to the bus transport system is lower than expected by the methodology. The results, although essentially exploratory, can help to minimize possible distortions in the accessibility offered.

As seen in the works cited above, studying mobility can bring improvements to the city. Our study also briefly characterizes the towns of Rome and San Francisco. Despite not being the scope of this work, it was possible to notice some patterns exhibited in the two cities to help develop new advances in the city’s mobility.

5.3 Design New Services

Several studies are being done in vehicular networks every day. Some of these articles are developing new services that benefit the driver, the road’s mobility, or even communication on the network. Some services will generate entertainment for people in traffic, making their experience in traffic more interesting.

The authors Ruiz et al. (2019) proposed a comparative analysis of the performance of the IEEE 802.11p and IEEE 802.11g standards in the transmission of videos. The work’s objective was to identify the behavior of both in a V2V (Vehicle-to-Vehicle) scenario aiming at a better cost-benefit and a new
connectivity option. The work results showed that While the IEEE 802.11p standard practically suffered no loss of quality in the video transmission, the IEEE 802.11g standard could not transmit the video completely.

Another type of service that is well studied is that which aims to reduce drivers’ travel time. In Gomides et al. (2019), they proposed a fully distributed algorithm able to minimize these impacts by reorganizing the vehicular flow. The algorithm developed by the authors showed several improvements such as reducing travel time, time in congestion, in addition to increasing the average speed achieved with low impact on the number of messages transmitted to enable a good performance of the proposed system.

One of the ways that our study can help in the generation of new services is through validation. One part of generating a new service is comparing it to existing services to show better performance. Our study provides three types of modeling that allow you to assess whether the service created can be better in some ways. This is a way to validate a new service and its performance in front of related solutions on literature.

5.4 Identifying Routines

Identifying routines in city mobility is essential. When we classify the repeated routines, we can take advantage of predictability to develop solutions that benefit this knowledge. There are more and more studies that study the behavior of environments. These researches are necessary because each area has a unique behavior, influenced by interests and cultural habits. For this reason, we must analyze each scenario separately.

One of these works that identify routines in an environment is proposed in Diniz et al. (2017). In this work, the authors characterize the mobility of the city of Rome. In the study, it was possible to identify that the vehicle routines are repeated over the weeks and then on the weekends. They also analyzed which categories of points of interest were most visited in the city over the days. With this type of knowledge, it is possible to develop services that take benefit of these routines. For instance, these patterns can help us develop an application to send advertisements or warnings to people close to a particular place at a specific time.

The routines characterization in urban mobility is receiving increasing attention. A great benefit in identifying these habits is to take advantage of this knowledge in favor of new services. This analysis can also help us use prediction to act in advance to solve various problems of urban mobility. Our study shows that, depending on the modeling used in urban mobility, this identification of routines may not represent the environment. Therefore, care must be taken when choosing the type of mobility modeling.
5.5 Smart-city Planning

In recent years, technology has evolved considerably. This made it increasingly possible to develop the concept of Smart Cities. Several applications have already appeared in this area, such as road accident warning, vehicle routing to avoid congestion, checking which places need cleaning, and identifying problems on the road. These applications can help managers to make more conscious decisions in the city administration.

The author Leite et al. (2013) has developed an application that aims to improve the experience of those using public transport. The developed system provides real-time information about the buses available in the city. For this, they used geoprocessing techniques based on public transport users’ locations, points of interest, and information on buses’ routes.

The authors of the work in Gallo (2016) proposed a scheme that focused on active collaborative monitoring carried out from mobile applications. The application servers receive, validate, and allow the visualization of the collected data, facilitating new applications. The authors carried out an experiment using such a platform to map accessibility conditions of the streets and sidewalks of a region in São Paulo.

Nowadays, people are trying to optimize all the most straightforward tasks to have more time to perform other activities. Also, with smart city systems, it is possible to optimize some tasks to spend less time in traffic. There are yet other benefits of reducing CO2 emissions, increasing the population’s quality of life, and facilitating the streets’ cleaning.

Our study may favor the development of smart cities and the design of new services for this scenario. For instance, considering the graph mobility during the day, it is possible to provide a good time to clean the streets and realize new road construction, identifying a moment of less traffic on the roads to reduce the traffic impact.

6 Conclusion

The Vehicular Networks subject is gaining more and more space among researchers. New technologies and studies are being carried out to improve people’s quality of life concerns to traffic. Reliability is a fundamental issue to implement efficient tools to enhance Vehicular Network performance. In this work, we present important analyzes of the different VANET’s modeling approaches. After reviewing the literature, we found three different types of modeling: Instantaneous, Aggregate, and Temporal.

We present a discussion about how the metric results can be affected by the topological modeling approach. Moreover, we carry out in-depth analyzes of micro and macroscopic aspects in two well-known large-scale traces, the San Francisco and the Rome traces. We identified choosing which approach to use should be made cautiously and planned, as this choice can dramatically impact the studies’ results.
There are still opportunities to investigate in the future. First, we want to study different traces of different types of vehicles, such as taxis, buses, cars, and bicycles, which allow us to have a broader view of the impact of other models on the city’s mobility as a whole. Also, we intend to use more statistical tests, tools, techniques and define mobility models to explore mobility patterns. Second, we will investigate other factors that can impact these three different approaches. Besides, we will examine these results as an entry for new services and protocols in vehicular networks.

7 Acknowledgements

This work has been partially supported by CAPES and CNPq.

8 Availability of data and materials

Not applicable.

9 Competing interests

Not applicable.

10 Funding

Not applicable.

11 Authors’ contributions

Not applicable.

References

Azolin LG, Silva A (2019) O transporte público em uma estratégia de avaliação da resiliência na mobilidade urbana. In: 33o Congresso de Pesquisa e Ensino em Transporte da ANPET

Bracciale L, Bonola M, Loreti P, Bianchi G, Amici R, Rabuffi A (2014) Crawdad dataset roma/taxi (v. 2014-07-17). See http://crawdad.org/roma/taxi/20140717

Christensen C, Albert R (2007) Using graph concepts to understand the organization of complex systems. International Journal of Bifurcation and Chaos 17(07):2201–2214
Corrêa U, Pinto A, Codas A, Ferreira D, Montez C (2006) Redes locais sem fio: Conceito e aplicações. Florianópolis: Universidade Federal de Santa Catarina

Cunha FD, Boukerche A, Villas L, Viana AC, Loureiro AA (2014a) Data communication in vanets; a survey, challenges and applications. Ad Hoc Networks

Cunha FD, Vianna AC, Mini RA, Loureiro AA (2014b) Is it possible to find social properties in vehicular networks? In: 2014 IEEE Symposium on Computers and Communications (ISCC), IEEE, pp 1–6

Cunha FD, Silva FA, Coles C, Maia G, Ruiz LB, Andrade RM, Mini RA, Boukerche A, Loureiro AA (2016) Communication analysis of real vehicular calibrated traces. In: 2016 IEEE International Conference on Communications (ICC), IEEE, pp 1–6

Diniz GR, Loureiro AA, Cunha FD (2017) On the characterization of vehicular mobility. DIVANet '17 Proceedings of the 6th ACM Symposium on Development and Analysis of Intelligent Vehicular Networks and Applications pp 23–29

Fiore MF, Glacet C, Bedogni L (2018) Temporal reachability in vehicular networks. INFOCOM 2018-IEEE Conference on Computer Communications pp 1–9

Freeman L (1979) Centrality in networks: I. conceptual clarifications. Social Network

Freeman LC (1977) A set of measures of centrality based on betweenness. Sociometry pp 35–41

Gallo DS (2016) Monitoramento colaborativo para cidades inteligentes. PhD thesis, Universidade de São Paulo

Glacet C, Fiore M, Gramaglia M (2015) Temporal connectivity of vehicular networks: The power of store-carry-and-forward. In: 2015 IEEE Vehicular Networking Conference (VNC), IEEE, pp 52–59

Gomides TS, Lourencô M, de Souza PH, Guidoni DL (2019) Sgtd: Sistema de gerenciamento de tráfego distribuído para redes veiculares. In: Anais Principais do XXXVII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos, SBC, pp 1042–1055

Guerber CR, Gomes EL, Fonseca M, Munaretto A (2018) Classificação geosocial de contatos para disseminação de dados em redes veiculares oportunistas. In: Anais do XXXVI Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos, SBC

Hou X, Li Y, Jin D, Wu DO, Chen S (2016) Modeling the impact of mobility on the connectivity of vehicular networks in large-scale urban environments. IEEE Trans on Vehicular Tech 65(4):2753–2758

Leite D, Rocha J, Baptista C (2013) Busão: um sistema de informações móvel para auxílio à mobilidade urbana através do uso de transporte coletivo. In: Anais Principais do IX Simpósio Brasileiro de Sistemas de Informação, SBC, pp 170–181

Lessa DA, Lobo C, Cardoso L (2019) Accessibility and urban mobility by bus in belo horizonte/minas gerais–brazil. Journal of Transport Geography
Luo Y, Zhang W, Hu Y (2010) A new cluster based routing protocol for
vanet. In: 2nd Int’l Conf. on Networks Security, Wireless Comm. and
Trusted Computing, IEEE, pp 176–180
Meneguette RI, Maia G, Madeira ER, Pinto AR, Loureiro A, Villas LA
(2014) Um algoritmo autônomo para disseminação de informações em
redes veiculares. Anais do Simpósio Brasileiro de Redes de Comuta-
dores e Sistemas Distribuídos (SBRC’14)
Moustafa H, Zhang Y (2009) Vehicular networks: techniques, standards, and
applications. Auerbach publications
Naboulsi D, Fiore M (2013) On the instantaneous topology of a large-
scale urban vehicular network: the cologne case. Proc of the Int’l
Symp on Mobile Ad Hoc Networking and Computing (MobiHoc) DOI
10.1145/2491288.2491312
Naboulsi D, Fiore M (2017) Characterizing the instantaneous connectivity
of large-scale urban vehicular networks. IEEE Transactions on Mobile
Computing 16(5):1272–1286
Pallis G, Katsaros D, Dikaiakos MD, Loulloudes N, Tassiulas L (2009) On
the structure and evolution of vehicular networks. In: IEEE Int’l Symp.
on Modeling, Analysis & Simulation of Computer and Telecommunica-
tion Systems, pp 1–10
Qiao L, Shi Y, Chen S (2017) An empirical study on the temporal structural
characteristics of vanets on a taxi gps dataset. IEEE Access 5:722–731
Qiu Z, Zhang S, Zhou W, Yu S (2018) Empirical study on taxi’s mobility na-
ture in dense urban area. In: IEEE INFOCOM 2018-IEEE Conference on
Computer Communications Workshops (INFOCOM WKSHPS), IEEE,
pp 232–237
Ruiz I, Ademilton F, Carvalho T, Jailton J (2019) Análise comparativa de
transmissão de vídeo em redes veiculares utilizando padrões ieee 802.11
ge ieee 802.11 p. Anais do Encontro Anual de Tecnologia da Informação
9(1):14–14
SOUZA RSd (2017) Protocolo de sinalização de congestionamento para re-
des veiculares. Repositório Institucional da UFPI
Uppoor S, Fiore M (2011) Large-scale urban vehicular mobility for network-
ing research. In: 2011 IEEE Vehicular Networking Conference (VNC),
IEEE, pp 62–69
Zhu K, Niyato D, Wang P, Hussain E, In Kim D (2011) Mobility and handoff
management in vehicular networks: a survey. Wireless Comm and Mobile
Computing 11(4):459–476
| Reference                  | Approach       | Datasets          | Metrics Analyzed                                                                 |
|---------------------------|----------------|-------------------|----------------------------------------------------------------------------------|
| Naboulsi and Fiore (2013) | Instantaneous  | Cologne           | Components Number, Giant Component, Node Degree, Betweenness Centrality           |
| Hou et al. (2016)         | Instantaneous  | Shanghai          | Component Speed, Component Size                                                  |
| Naboulsi and Fiore (2017) | Instantaneous  | Cologne           | Density, Vehicles Speed, Components Number, Components Size Number of Nodes, Number of Nodes in Largest Component |
| Pallis et al. (2009)      | Instantaneous  | Zurich            | Node Degree, Contact Duration, Lobby Index, Diameter, Closeness Centrality, Betweenness Centrality, Bridging Centrality, Clusters Number, Clustering Coefficient, Communities Number |
| Guerber et al. (2018)     | Aggregate      | Rome              | Node Degree, Betweenness Centrality, Closeness Centrality                        |
| Cunha et al. (2014b)      | Aggregate      | Zurich            | Distance, Diameter, Density, Persistence of Edges, Node Degree, Clustering Coefficient, Closeness Centrality |
| Cunha et al. (2016)       | Aggregate      | San Francisco, Rome, Shanghai | Contact Duration, Inter Contact Time, Network Capacity                       |
| Diniz et al. (2017)       | Aggregate      | Rome              | Interest Point Analyse                                                            |
| Qiu et al. (2018)         | Aggregate      | Beijing           | Linear Interpolation, Distance, Contact Duration                                 |
| Qiao et al. (2017)        | Temporal       | Beijing           | Temporal Path, shortest time path, Influence Set, Source Set, Average Range Radius, Closeness Centrality, Components Number |
| Glacet et al. (2015)      | Temporal       | Cologne e Bologna | Components Number, Node Degree, Packet Loss                                      |
| Fiore et al. (2018)       | Temporal       | -                 | Transient Closing Density, Accessibility, Convergence Time, Density              |