On the Topology of a Large-scale Urban Vehicular Network

Diala Naboulsi, Marco Fiore

On the Topology of a Large-scale Urban Vehicular Network

Diala Naboulsi, Marco Fiore

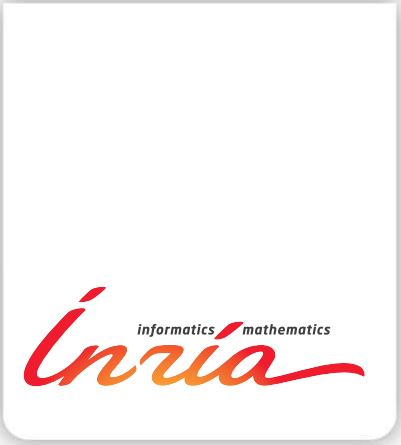
RESEARCH REPORT

N° 8112

September 2012

Project-Team URBANET

ISSN 0249-6399 ISRN INRIA/RR--8112--FR+ENG



On the Topology of a Large-scale Urban Vehicular Network

Diala Naboulsi*†, Marco Fiore* †

Project-Team URBANET

Research Report n° 8112 — September 2012 — 19 pages

Abstract: Despite the growing interest in a real-world deployment of vehicle-to-vehicle communication, the topological features of the resulting vehicular network remain largely unknown. We lack a clear understanding of the level of connectivity achievable in large-scale scenarios, the availability and reliability of connected multi-hop paths, or the impact of daytime. In this paper, we adopt a complex network approach to provide a first characterization of a realistic large-scale urban vehicular ad hoc network. We unveil the low connectivity, availability, reliability and navigability of the network, and exploit our findings to derive network design guidelines.

Key-words: VANET, network connectivity, complex networks, network topology, large-scale urban networks

* INSA Lyon
† INRIA

RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Sur la Topologie d'un Réseau Véhiculaire à Grande Echelle

Résumé : Malgré l'intérêt croissant au déploiement de communication entre véhicules dans le monde réel, les caractéristiques topologiques du réseau véhiculaire résultant restent largement inconnus. Il nous manque une compréhension claire du niveau de la connectivité réalisable dans des scénarios à grande échelle, la disponibilité et la fiabilité des chemins multi-sauts, ou l'impact de l'heure de la journée. Dans cet article, nous adoptons une approche de réseaux complexes pour fournir une première caractérisation d'un réseau ad hoc véhiculaire réaliste à grande échelle dans un environnement urbain. Nous dévoilons la faible connectivité, disponibilité, fiabilité et navigabilité du réseau, et exploitons nos découvertes pour en tirer des lignes directrices pour la conception de réseaux.

Mots-clés : VANET, connectivité des réseaux, réseaux complexes, topologie des réseaux, réseaux urbains à grande échelle

1 Introduction

More than a decade after the allocation of a dedicated frequency bandwidth in the USA, vehicular communication networks have finally abandoned their long-standing status of fundamental research exercise and have started developing into real-world systems. On-going standardization efforts, including IEEE 802.11p, IEEE 1609, OSI CALM-M5 and ETSI ITS, jointly with the growing interest of the automobile industry, have played a major role in speeding up the process over the last few years. Thus, it will be soon time to field test the vast amount of vehicular networking solutions proposed by the scientific community during the last decade, whose scope ranges from medium access to transmission power control, from data rate adaptation to mobile address management, from multi-hop routing to reliable data transfers.

However, the real-world performance of many of such protocols risk to fail the expectations, for the simple reason that they will be confronted to a network they were not designed for. As a matter of fact, dedicated solutions have – and are – being proposed for vehicular networks whose major topological features remain largely unknown. In urban environments in particular, even basic questions stay unanswered, such as: *is the vehicular ad hoc network well connected or highly partitioned? Which size can clusters of multi-hop connected vehicles attain? Which is the internal structure of such clusters? How sparse or dense are single-hop communication neighborhoods? How do all these network connectivity features vary in time? How do they depend on the geographical location?*

The responses to these questions directly determine the strengths, weaknesses and overall capabilities of a spontaneous vehicular network, and shall thus be among the main drives to the design of dedicated protocols. Moreover, they are the key to quantifying the *availability* and *reliability* of the network, i.e., the main concerns of car manufacturers when it comes to vehicle-to-vehicle multi-hop communication.

The aim of this paper is to shed some light on the major topological features of a large-scale urban vehicular network, providing qualitative and quantitative answers to the questions above and highlighting their impact on vehicular networking solutions. To that end, after a presentation of the literature in Sec. 2, we leverage a specific mobility dataset, presented in Sec. 3, that is representative of a typical middle-sized European city and yields unprecedented realistic microscopic and macroscopic mobility features. We study the dynamics of the vehicular network topology in such an urban scenario through a *complex network* approach [1–3], by modeling the dataset as a set of instantaneous connectivity graphs, as from Sec. 4, and analyzing their features from a communication network perspective, in Sec. 5, 6 and 7. Finally, conclusions are drawn in Sec. 8.

2 Related work

The characterization of the instantaneous topology of a vehicular ad hoc network is a subject that, for its own nature, demands knowledge of the exact position of all the vehicles circulating in a large region. This makes direct experimental evaluations, that have high economic costs and are extremely time-consuming, impractical. It also renders the available datasets of recorded real-world vehicular mobility, that are limited to specific subsets of vehicles such as buses [4] or taxis [5], too incomplete for a thorough topological analysis. As a consequence, studies on the the instantaneous connectivity of vehicular ad hoc networks mainly rely on simulative or analytical tools.

Highway environments are simpler to analyze in terms of connectivity, since they result in unidimensional road traffic flows. The studies on the topology of vehicular ad hoc networks on highways aim at determining which combinations of vehicular density, absolute and relative car speed, technology penetration rate and communication range are required to achieve a full network connectivity. However, the results, be they based on synthetics traces of highway traffic [6, 7] or mathematical models [8–10], do not apply to the more complex urban scenarios we are interested in.

As a matter of fact, characterizing the topology of urban vehicular networks is a significantly harder

process. Not only the problem becomes bidimensional, but realistic road layouts can be extremely complex and include heterogeneous restrictions, such as speed limits and one-way rules. Even worse, road intersections are regulated by non-trivial mechanisms, such as traffic lights, roundabouts, stop and yield signs.

Analytical studies of the vehicular connectivity in urban areas are therefore based on strong simplifying assumptions, namely Poisson distribution of cars and regular-grid road layouts [11–14]. These make the problem mathematically tractable through, e.g., percolation theory, but dramatically reduce the realism and interest of the results.

As far as simulation-based analyses of the connectivity of urban vehicular networks are concerned, seminal works have focused on the impact of microscopic mobility models [15, 16]. These studies unveiled the high bias that unrealistic random and pseudo-random car mobility can induce on the network topology, and the importance of properly modeling road signalization. However, they do not study the network from a macroscopic point of view as done in this paper.

The impact on the vehicular network connectivity of macroscopic traffic parameters, such as the vehicular density, the arrival rates and the routes traveled by drivers, is instead assessed in [17, 18]. However, these works assume a small-scale regular-grid road layout and aim at defining the generic connectivity behavior of the urban vehicular network in idealized and controllable settings, rather than at studying a large-scale realistic scenario as we do.

Large urban areas have been considered in studies of the vehicular network connectivity in Porto, Portugal [19], and Zurich, Switzerland [20]. In the first case, the authors describe the evolution of the average network degree for around 5 minutes, assuming that 10% of the vehicles participate in the network. Our analysis provides a more complete picture, as it accounts for a broad range of metrics and spans over 24 hours. The authors of [20] analyze instead a region of 25 km^2, during 3 hours that cover the morning traffic peak. Despite the thorough evaluation, the trace employed only includes major road arteries in heavy traffic conditions, and the mesoscopic simulator used to generate the vehicle movement yields a rather uniform distribution of cars over the road layout. These factors contribute to a vehicular network significantly biased towards unrealistically high connectivity [21].

3 Vehicular mobility dataset

The vehicular mobility dataset we employ is part of the TAPASCologne project, that aims at reproducing, with the highest level of realism possible, the car traffic in the greater urban area of Köln, Germany. The dataset has been synthetically generated by coupling state-of-art tools dedicated to specific aspects of vehicular traffic modeling. A short summary is provided next, we refer the reader to [21] for further details.

Road topology. The street layout of the Köln urban area is extracted from the OpenStreetMap (OSM) database. The OSM initiative provides freely exportable maps of cities worldwide, contributed and updated by a vast user community. The OSM road information is generated and validated by means of satellite imagery and GPS traces, and is commonly regarded as the highest-quality map data publicly available.

Microscopic mobility. The microscopic mobility of vehicles is simulated with the Simulation of Urban Mobility (SUMO) software. SUMO is an open source tool developed by the German Aerospace Center (DLR), capable of accurately modeling the behavior of individual drivers, accounting for car-to-car and car-to-road signalization interactions. The level of detail of the simulation and the high scalability of the environment make of SUMO the most advanced open-source microscopic vehicular mobility generator available today.

Macroscopic mobility. At the macroscopic level, large-scale car flows across the urban area are determined by coupling travel demand and traffic assignment models. The former determine the locations

Figure 1: Snapshot of the vehicular mobility dataset in the Köln urban region.

where each vehicle starts and ends its trip; the latter compute the exact path between such locations.

The travel demand is derived through the Travel and Activity PAtterns Simulation (TAPAS) methodology [22]. This techniques generates the vehicular trips by exploiting information on (i) the population, i.e., home locations and socio-demographic characteristics, (ii) the points of interests in the urban area, i.e., places where working and free-time activities take place, and (iii) the time use patterns, i.e., habits of the local residents in organizing their daily schedule. Applying the TAPAS technique on real-world data collected in the Köln region allows to obtain a faithful representation of daily activities of the area residents.

The traffic assignment is computed via the relaxation technique proposed by Gawron [23]. This method takes as inputs the car trips and the street topology, computes the fastest route for each vehicle, and then assigns to each road segment a cost reflecting the intensity of traffic over it. By iteratively moving part of the traffic to alternate, less congested paths, and recomputing the road costs, the scheme finally achieves a so-called dynamic user equilibrium. The coupling of the TAPAS-generated travel demand and Gawron's traffic assignment results in an unprecedented level of realism in the macroscopic modeling of vehicular mobility.

Dataset summary. The mobility dataset spans over a whole day and encompasses 4500 km of roads in an area of 400 km^2, with per-second information on the position and speed of vehicles involved in more than 700.000 trips. This is the largest and most complete vehicular mobility dataset freely available to date. A snapshot of the vehicular mobility in the dataset is provided in Fig. 1, where the position of vehicles at 7:00 (each dot represents one car, its color corresponding to its speed) is superposed to the road topology.

4 Network model and metrics

Our analysis is technology- and protocol-independent, as it targets the physical topology of the vehicular network. To that end, we borrow tools from complex network theory, that have been successfully employed to characterize a number of large-scale real-world networks such as those of Internet routers, World Wide Web pages or interacting social species [1–3]. Next, we provide a description of how we

model the vehicular network topology into a dynamic graph, and we formally define the metrics that will be used in our study.

We sample the vehicular mobility dataset with a fixed frequency[1], and, at each sampled time instant t, we model the vehicular network topology as a graph. We study the *instantaneous* connectivity of the network, by considering each of such graphs separately. Therefore, temporal aggregation of the network connectivity graphs or temporal network modeling through tensors are out of the scope of this paper, although we plan to address those in the future, given their interest for, e.g., delay-tolerant network approaches.

We refer to the instantaneous graph at time t as $G(\mathbb{V}(t), \mathbb{E}(t))$. There, $\mathbb{V}(t) = \{v_i\}$ is a set of vertices (or nodes, as the two terms will be used interchangeably in the following) v_i, each representing a vehicle i traveling in the road scenario at time t, and $\mathbb{E}(t) = \{e_{ij}(t) \,|\, v_i, v_j \in \mathbb{V},\, i \neq j\}$ is the set of edges $e_{ij}(t)$, modeling the availability of a communication link between vehicle i and vehicle j at t.

The number of nodes in the network varies with time and is referred to as $\mathcal{N}(t) = \|\mathbb{V}(t)\|$, where $\|\cdot\|$ denotes the cardinality of the included set. The set $\mathbb{E}(t)$ depends instead on the RF signal propagation model adopted, which in this work we will assume to be a simple unit disc model. We acknowledge that, despite being a common practice in the related works presented in Sec. 2, this is a drastic simplification of the reality. However, deterministic propagation models (e.g., ray-tracing ones) do not scale well to our large-scale vehicular scenario, as they require expensive re-computations at each movement of every network node. As for stochastic models, they introduce a random noise that would force us to evaluate a large set of instances for each sampled time instant, again an unbearable task given the size of our scenario. The unit disc model allows to drastically reduce the computational complexity of the analysis, and yet to capture the average behavior of the network. In the following, we denote the unit disc communication range as R. We stress that, as a consequence of the simple propagation model adopted, only bidirectional links are present in the graph model, i.e., $G(\mathbb{V}(t), \mathbb{E}(t))$ is undirected and $e_{ij}(t) = e_{ji}(t)$, $\forall i, j, t$.

We also denote the shortest multi-hop communication path between any two cars i and j at time t as the ordered sequence of vertices in the shortest path from v_i to v_j, i.e., $\mathbb{p}_{ij}(t) = \{v_i, v_{k_1}, v_{k_2}, \ldots, v_j\}$. The length of such a shortest path is denoted as $\mathcal{L}_{ij}(t) = \|\mathbb{p}_{ij}(t)\| - 1$ and maps to the number of transmission hops separating vehicles i and j. If no shortest path exists between i and j at time t, then $\mathbb{p}_{ij}(t) = \emptyset$ and $\mathcal{L}_{ij}(t) = \infty$. If multiple equivalent shortest paths are available between the same node pair, one is randomly picked. Note than, by definition, $\mathbb{p}_{ii}(t) = \{v_i\}$ and $\mathcal{L}_{ii}(t) = 0$, $\forall i, t$.

The graph model allows us to introduce several metrics, useful to characterize the connectivity properties of the whole network, of subsets of the same, and of individual vertices.

Component. Let us associate to each vertex v_i at time t a subset of the vertices $\mathbb{V}_i(t) = v_i \cup \{v_j \,|\, \mathbb{p}_{ij}(t) \neq \emptyset\}$, and a subset of the edges $\mathbb{E}_i(t) = \{e_{jk}(t) \mid v_j, v_k \in \mathbb{V}_i(t)\}$. We define the subgraph $C_i(t) = G(\mathbb{V}_i(t), \mathbb{E}_i(t))$ as the component within which vertex v_i lies at time t. In other words, the component $C_i(t)$ is the graph representing the portion of the network that vehicle i can reach via multi-hop communication at time t. The size of the component is the number of nodes that belong to it, i.e., $\mathcal{S}_i(t) = \|\mathbb{V}_i(t)\|$. By construction, the subgraph is the same for all vehicles in the same component, i.e., $C_i(t) = C_j(t)$ iff $\mathbb{p}_{ij} \neq \emptyset$, or, from the opposite perspective, no communication path exists between two different components, i.e., $\mathbb{p}_{ij} = \emptyset$ iff $C_i(t) \neq C_j(t)$. Since all node pairs in a same component are connected, given a generic component $C_i(t)$, it holds that $\mathcal{L}_{jk} < \infty$, $\forall v_j, v_k \in C_i(t)$. We can thus define the average shortest path between any two nodes in the component as

$$\mathcal{L}_i(t) = \frac{2}{\mathcal{S}_i(t)\,(\mathcal{S}_i(t) - 1)} \sum_{v_j, v_k \in C_i(t), j<k} \mathcal{L}_{jk}.$$

[1] In the following, unless stated otherwise, we will employ a sampling periodicity of 10 seconds, that was found to yield results identical to those obtained with higher sampling frequencies, at a lower computational cost.

Finally, we refer to the set of the components present in the network at time t as $\mathbb{C}(t) = \{C_i(t) | C_i(t) \cap C_j(t) = \emptyset,\ \forall i < j\}$. The number of components in the network at time t is then $\mathcal{C}(t) = \|\mathbb{C}(t)\|$.

Giant component. A component $C_i(t)$ is said to be a giant component if $\mathcal{S}_i(t) \geq 0.1 \cdot \mathcal{N}(t)$. That is, a giant component at time t has a size of the same order of magnitude as the whole network at the same time. We also denote the largest component in the network at time t as $C_{max}(t)$, and its size as $\mathcal{S}_{max}(t)$. If the largest component $C_{max}(t)$ is a giant component, we denote its size as $\mathcal{S}_{giant}(t) = \mathcal{S}_{max}(t) \geq 0.1 \cdot \mathcal{N}(t)$. We remark that a giant component is not necessarily present in the network, i.e., it can be that $\mathcal{S}_{max}(t) < 0.1 \cdot \mathcal{N}(t)$.

Degree. Let us consider a subset $\mathbb{V}_i^1(t) = \{v_j | \exists e_{ij}(t)\}$ of vertices and a subset $\mathbb{E}_i^1(t) = \{e_{jk}(t) | v_j, v_k \in \mathbb{V}_i^1(t)\}$ of edges. We define the subgraph $C_i^1(t) = G(\mathbb{V}_i^1(t), \mathbb{E}_i^1(t))$ as the one-hop communication neighborhood of vertex i at time t. The degree of a vertex v_i is then $\mathcal{D}_i(t) = \left\|\mathbb{V}_i^1(t)\right\|$, and corresponds to the number of one-hop communication neighbors of vehicle i.

Betweenness centrality. The betweenness centrality of a vertex v_i at time t is

$$\mathcal{B}_i(t) = \frac{\|\{\mathbb{p}_{jk}(t) \mid v_i \in \mathbb{p}_{jk}(t),\ i \neq j, k\}\|}{(\|\mathbb{V}(t)\| - 2)\,(\|\mathbb{V}(t)\| - 1)},$$

where the denominator is the maximum number of shortest paths in the network that do not start or end at v_i. The betweenness centrality represents how frequently a vehicle i is part of the shortest path between any two other cars, and ranges from zero (no shortest path passes by i) to one (i lies within all the shortest paths of the network).

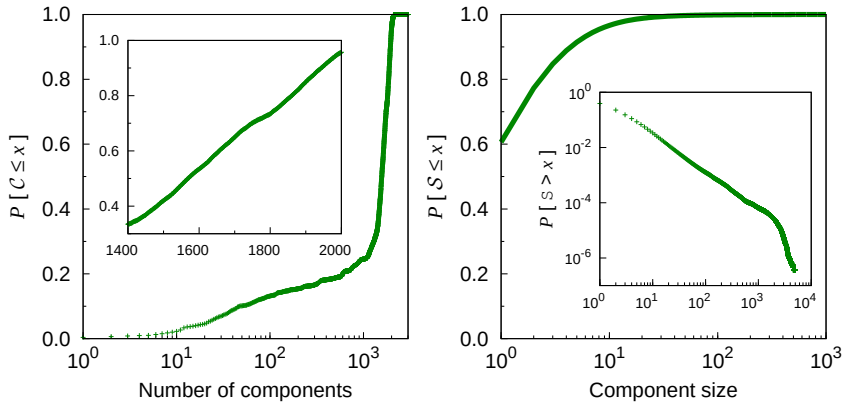
For the sake of clarity, in the following we will imply that all measures are time dependent and drop the time notation. We will thus denote, for a generic time instant, the number of network nodes as $\mathcal{N}$ and the number of components as $\mathcal{C}$. Also, for the vehicle-dependent metrics, we will assume that they refer to a generic vehicle i and drop the per-vehicle notation. Therefore, the size of a generic component will be denoted by $\mathcal{S}$, its average path length as $\mathcal{L}$, and the degree and betweenness centrality of a generic node as $\mathcal{D}$ and $\mathcal{B}$.

5 Network-level analysis

We start our analysis by considering the vehicular network as a whole and discussing its global properties in terms of connectivity. For the sake of clarity, we initially limit our analysis to one particular value of the communication radius R. To that end, we employ R = 100 m, as this is the value referenced by field tests as a typical distance for reliable DSRC vehicle-to-vehicle communication [24, 25]. More precisely, the extensive experimental analysis in [25] shows that a distance of 100 m allows around 80% of the packets to be correctly received in urban environments, under common power levels (15-20 dBm) and with robust modulations (3-Mbps BPSK and 6-Mbps QPSK). We will later generalize our study to R = 50 m, experimentally identified as the largest distance at which vehicle-to-vehicle communication has a packet delivery ratio close to one [24,25], and R = 200 m, i.e., the maximum distance granting a reception ratio above 0.5 [25]. We consider wireless links losing more than 50% of the packets hardly exploitable by the network.

Analysis for R = 100 m. The level of connectivity of the whole vehicular network is mainly characterized through two metrics: the number of components $\mathcal{C}$, that is an index of the level of network fragmentation, and the component size $\mathcal{S}$, which describes the heterogeneity of the fragmentation.

The Cumulative Distribution Function (CDF) of $\mathcal{C}$, aggregated over all the vehicular network graphs extracted from the whole 24 hours of road traffic, is portrayed in Fig. 2(a). In 80% of cases the vehicular network has more than 1000 disconnected components, and the inset plot shows that most of the probability is concentrated in a linear growth between 1000 and 2000 components. This suggests that the vehicular network is typically highly partitioned into thousands of separate node groups unable to communicate with each other.



(a) (b)

Figure 2: Distributions of (a) the number of components and (b) the component size, when aggregating all the samples of $\mathcal{C}$ and $\mathcal{S}$ over the whole day.

The distributions of the component size $\mathcal{S}$ helps us clarify whether we face many components of similar size or a heterogeneous network of both large and small components. The CDF in Fig. 2(b) shows that the network is largely made of very small components, with 60% of vertices being *singletons*, i.e., isolated vehicles, and 95% of the components comprising 10 vehicles or less. However, by looking at the Complementary CDF (CCDF), portrayed in the inset plot, we can appreciate the heavy tail of the distribution, appearing as linear on a log-log scale. There is thus a non-negligible probability that the aforementioned small components coexist with components that include up to 2000 vehicles connected through direct links or multi-hop routes. After such a component size, the CCDF has an exponential decay, and components as large as 4500 nodes appear with significant lower probability.

Such a general view of a partitioned and heterogeneous network is however aggregate over time and space. In order to unveil the impact of the daytime and the differences between geographical areas, we show in Fig. 3 the instantaneous vehicular network fragmentation measured in the Köln region at different hours. In each plot, every circle corresponds to one component, its diameter and color mapping to the size of the component it represents: broader circles map to larger components, while the color code is reported on the bars at the right of the plots. The resulting images give a rough, yet intuitive, idea of the behavior of the network connectivity evolution: early in the morning, i.e., before 6:00, the network is very partitioned and only small components of 40 vehicles or less are present. The morning traffic peak, between 7:00 and 8:00, has very positive impact on the network topology, with the appearance of very large components of thousands of vehicles and a diffuse presence medium-sized components of several tens of cars. That effect disappears later on, and large components do not reappear until the afternoon traffic peak, between 16:00 and 18:00, although a slightly increased connectivity is observed around noon. Moreover, large components mostly appear in the city center, where the traffic is denser. We can infer that both time and space are paramount factors in the characterization of the vehicular network topology, which appears to be generally highly fragmented, with larger components only appearing at specific locations during the traffic peak hours.

Generalization to different R's. We now study how different communication ranges impact the observations above. Fig. 4(a) portrays the CDF of the number of components $\mathcal{C}$ when the transmission range R varies from 50 to 200 m. The probability distribution for low $\mathcal{C}$ does not vary with R: since we observed in Fig. 3 that the number of components is consistently high throughout the whole day, we can now conjecture that these values of $\mathcal{C}$ map to hours when no more than a few hundreds of isolated vehicles are present in the area, i.e., before 5:00 and after 22:00. When $\mathcal{C}$ is instead higher than 500, a higher R shifts the probability concentration to the left, or, in other words, a higher communication radius

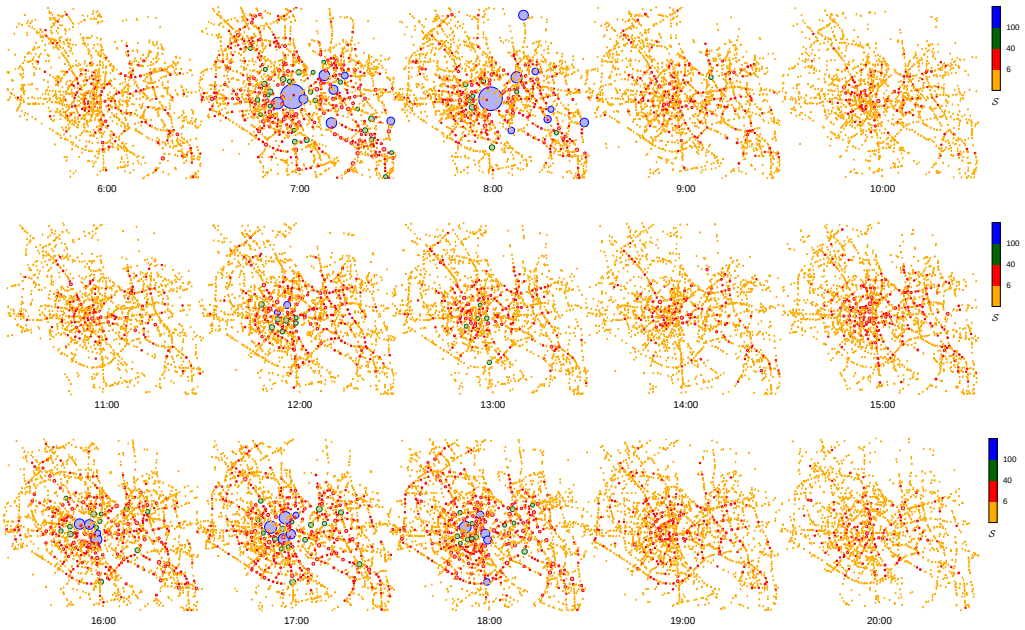


Figure 3: Geographical representation of the network components over the significant hours of the whole day. This figure is best viewed in colors.

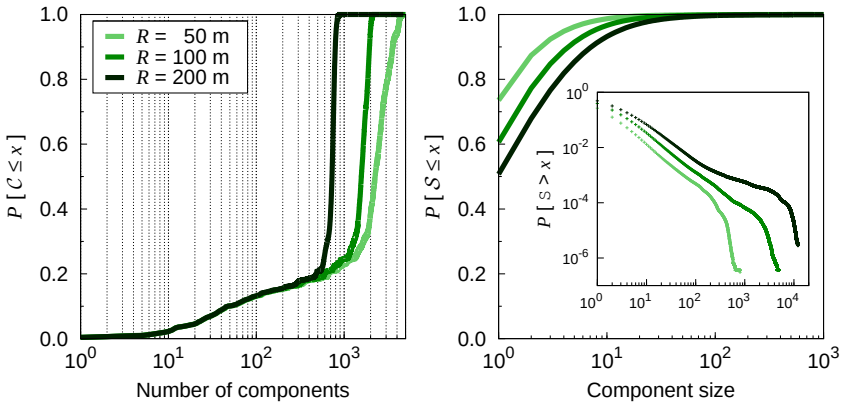
bounds the maximum number of components to a lower value. As a result, we never have more than 1000 components when R = 200 m, while we can have up to 5000 components when R = 50 m. Clearly, this is a symptom that R can significantly enhance or disrupt the network-level connectivity.

A confirmation comes from the component size distributions, in Fig. 4(b). There, higher values of R result in larger components at all levels: both small components, in the CDF, and large ones, in the inset CCDF, are positively affected by R. However, it is interesting to observe that the network is still significantly partitioned even when R = 200 m, as 90% of the components comprise no more than 10 cars. Conversely, reducing R to 50 m severely disrupts the network connectivity, and no component with more than 1000 vehicles is observed.

A space-time view of the results above is provided by Fig. 5, that displays the geographical distribution of components in low (11:00) and high (7:00) road traffic conditions, when R is set to 50 m (left) and 200 m (right). It is evident that a reduced transmission range yields a completely fragmented network, where multi-hop connectivity is hard to spot even during the traffic peak hours. A higher R leads instead to a network that is diffusely more connected (remark the medium- and large-sized components even at 11:00) and allows for very large components, including more than 10000 vehicles to form in presence of intense road traffic. Once more, we remark the critical impact of both time and space, with the largest components always emerging in downtown Köln for all R's.

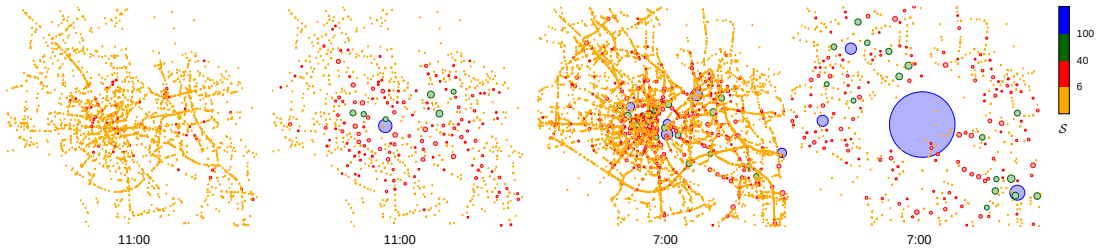
Relationship to $\mathcal{N}$. The strong time dependence of the network outlined by the previous results pushes us to verify if a more rigorous relationship can be found between the individual time instants and the vehicular network topology they yield. Following a complex network common practice, we observe the correlation of the connectivity metrics with the number $\mathcal{N}$ of nodes, i.e., vehicles, present in the network.

Fig. 6 shows the standard deviation of the number of components $\mathcal{C}$ computed among graphs that have a similar $\mathcal{N}$, which in turn varies along the x axis. The deviation is expressed as a percentage of the average $\mathcal{C}$ over the same set of graphs. We observe that the standard deviation is extremely low, typically within 5% from the average value, for any significant $\mathcal{N}$. This behavior, that holds for any R, implies that all the graphs that have a similar number of vertices also have a very similar number of clusters.



(a) (b)

Figure 4: Distributions of (a) the number of components and (b) the component size, when aggregating all the daily samples of $\mathcal{C}$ and $\mathcal{S}$, for different R's.



(a) R = 50 m (b) R = 200 m (c) R = 50 m (d) R = 200 m

Figure 5: Geographical maps of clusters at 11:00 (top) and 7:00 (bottom), when R is 50 (left) and 200 (right) meters. This figure is best viewed in colors.

Fig. 7 shows the distributions of the component size $\mathcal{S}$ measured at different sample hours, and compares them to distributions obtained by aggregating the data from all the network graphs with a similar number of vertices $\mathcal{N}$. We can observe that for both small (CDF, left) and large (CCDF, right) components, as well as for any value of the transmission range R, the distributions at different daytime with similar $\mathcal{N}$ overlap among them as well as with the aggregate one for that $\mathcal{N}$.

We can conclude from these results that it is not the absolute time that drives the vehicular network properties, but the number of vehicles present in the urban region. This is a non-trivial conclusion: it implies that the network behavior in, e.g., the morning and the afternoon is the same, provided that we consider two instants with similar $\mathcal{N}$. Also, it accomunates vehicular networks to many other complex networks, for which the network size $\mathcal{N}$ is the reference parameter driving the evolution of the network [1–3]. In the remainder of the paper, we will thus study the vehicular topology properties at the light of the characterizing network parameter $\mathcal{N}$.

Key networking insights. From a networking viewpoint, we can remark that *the large-scale vehicular network is severely and consistently partitioned*, even when considering idealized RF signal propagation and the maximum reliable vehicle-to-vehicle communication ranges identified by experimental DSRC evaluations. Multi-hop routing through the whole network is impossible, and *carry-and-forward techniques are mandatory in order to reach all the nodes*. Finally, completely different connectivity prop-

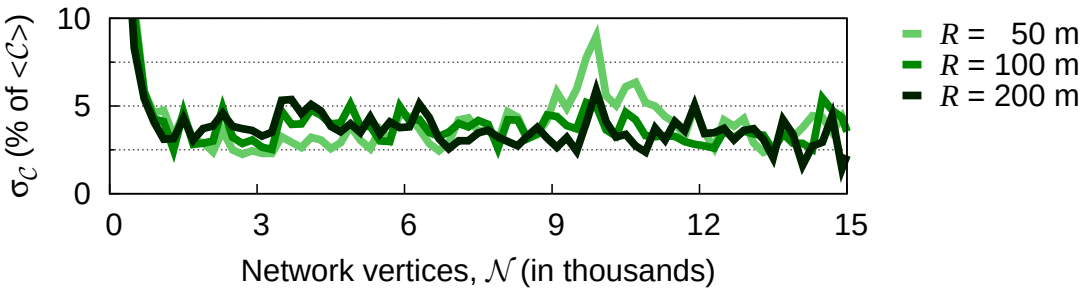


Figure 6: Standard deviation of the component number $\mathcal{C}$, expressed as a percentage of the corresponding average $\mathcal{C}$, versus $\mathcal{N}$ and for different R's.

erties are observed at diverse geographical location and daytimes: the common practice of *evaluating vehicular network protocols in small-scale scenarios characterized by arbitrary vehicular densities may be harmful*.

6 Component-level analysis

The highly fragmented network topology motivates us to study more deeply the internal structure of the individual components. In particular, our interest lies in the large components we have observed to appear in the network, since it is within them that multi-hop communication can take place and vehicular ad hoc network protocols can mainly operate.

Component dynamics versus $\mathcal{N}$. Fig. 8 portrays the evolution of the largest network component, C_{max}, in terms of its size $\mathcal{S}_{max}$, versus the network size $\mathcal{N}$ and for different values of the transmission range R. The color of the points refers to the number of components in the network, $\mathcal{C}$, whose landmark values are also pointed out by the arrows.

When R is set to 50 m, the size of C_{max} only slightly increases even for large $\mathcal{N}$'s, implying that large components are never observed in the system. Also, $\mathcal{C}$ monotonically grows with $\mathcal{N}$, suggesting that, in any vehicular traffic condition, new nodes joining the network have high chances of being isolated, forming new singleton components.

As R is set to 100 m, the behavior significantly differs. The largest component has a negligible size until $\mathcal{N} \sim 6000$, and then grows up to around 4500 nodes. Although there is some variability in the values of $\mathcal{S}_{max}$ for a same $\mathcal{N}$, the positive correlation between the two is evident. Interestingly, the number of components $\mathcal{C}$ grows during the first phase, up to $\mathcal{N} \sim 6000$, and then remains constant at $\mathcal{C} \sim 2000$. What happens is that, once a critical network density is achieved, new nodes joining the network are not isolated anymore, but join existing components: it is not anymore the number $\mathcal{C}$ of components to grow, but their size $\mathcal{S}$, as for $\mathcal{S}_{max}$.

When R = 200 m, the behavior changes once more. The critical threshold at which the $\mathcal{S}_{max}$ starts to grow shifts down to $\mathcal{N} \sim 3000$, and is followed by a neat linear relationship between the largest component size and $\mathcal{N}$. Components as large as 12000 vertices can be observed in the network when $\mathcal{N}$ reaches its maximum. The evolution of $\mathcal{C}$ is also notable, as the number of components concurrently present in the network initially grows up to 900 at $\mathcal{N} \sim 5000$, and then starts decreasing. Therefore, when R = 200 m, new nodes joining the network after the critical vehicular density has been reached not only are not isolated, but tend to bridge existing components, increasing their size and reducing their number.

Although the dispersion of points at a given $\mathcal{N}$ only allows to individuate rough approximations of the critical vertex density thresholds, the values identified above are qualitative indicators of the network

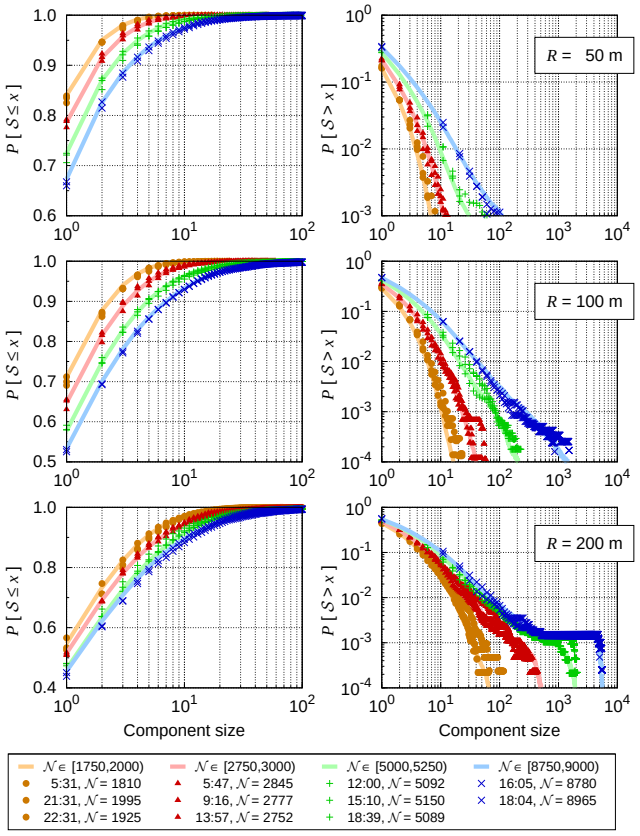


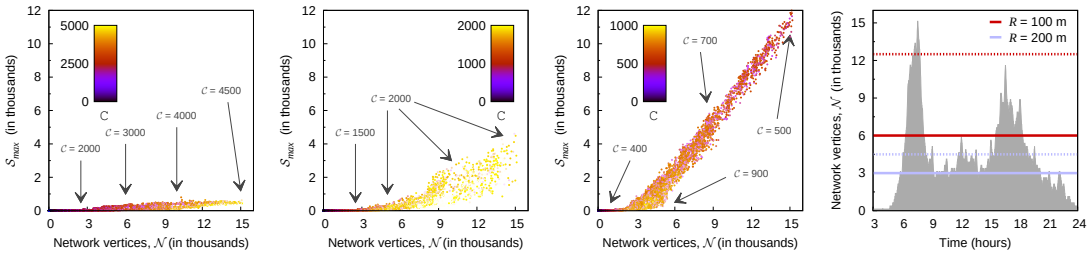
Figure 7: CDF (left) and CCDF (right) of the component size $\mathcal{S}$, for different values of R. Each plots portrays the distributions at several daytimes, compared with those aggregated over all the snapshots with similar $\mathcal{N}$.

availability, i.e., the possibility of leveraging vehicle-to-vehicle multi-hop communication to a meaningful extent. These thresholds are indicated as solid lines in Fig. 8(d), where they overlap to the evolution of $\mathcal{N}$ over time. Clearly, large clusters can be expected to appear in the vehicular network mainly during the morning and afternoon rush hours. This maps to an availability of around 4 hours/day and 8 hours/day when R = 100 m and 200 m, respectively. The network is instead never available when R = 50 m.

Spatio-temporal dynamics of C_{max}. Not only the availability, but also the *reliability* of large components is a critical factor for the design of vehicular network solutions. In order to analyze this second aspect, we studied the one-second variability of the largest component size $\mathcal{S}_{max}$, and observed that: (i) when R = 100 m, $\mathcal{S}_{max}$ undergoes significant variations at each second, its value doubling or halving every few seconds; (ii) when R = 200 m, $\mathcal{S}_{max}$ variations still involve thousands of nodes albeit the largest cluster $\mathcal{C}_{max}$ appears more stable due to its larger size, that limit fluctuations to 15% of $\mathcal{S}_{max}$.

The reason behind these major size changes of $\mathcal{C}_{max}$ lies in its internal dynamics. Although a rigorous discussion is not possible due to space limitations, we provide an intuitive example of such dynamics in Fig. 9. The two plots show the vertices in $\mathcal{C}_{max}$ at two subsequent seconds: the internal structure is highlighted in terms of betweenness centrality, the size and color of each nodes indicating its own $\mathcal{B}$ value.

The largest cluster in the left image comprises over 1.600 vehicles, and we can note the very high



(a) R = 50 m (b) R = 100 m (c) R = 200 m (d) Evolution of $\mathcal{N}$ over time

Figure 8: (a), (b), (c): scatterplots of the largest component size $\mathcal{S}_{max}$ versus the network size $\mathcal{N}$, for different values of the communication range R. Colors and arrows indicate the number of components in the network $\mathcal{C}$. (d): evolution of $\mathcal{N}$ over the daytime.

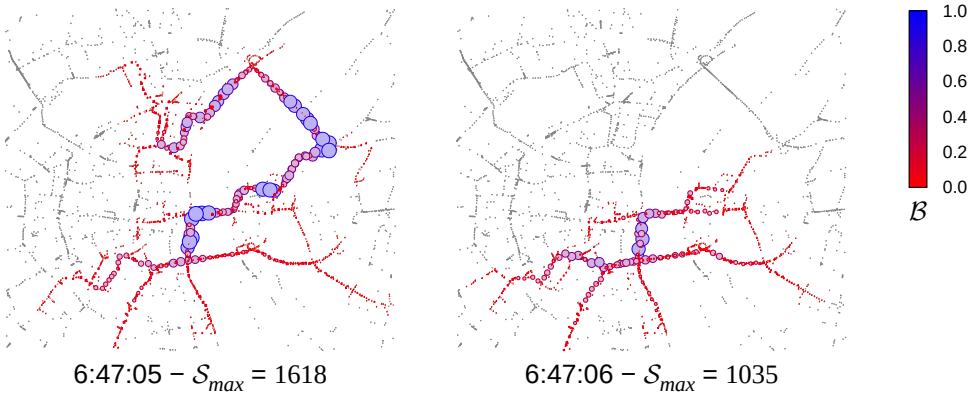
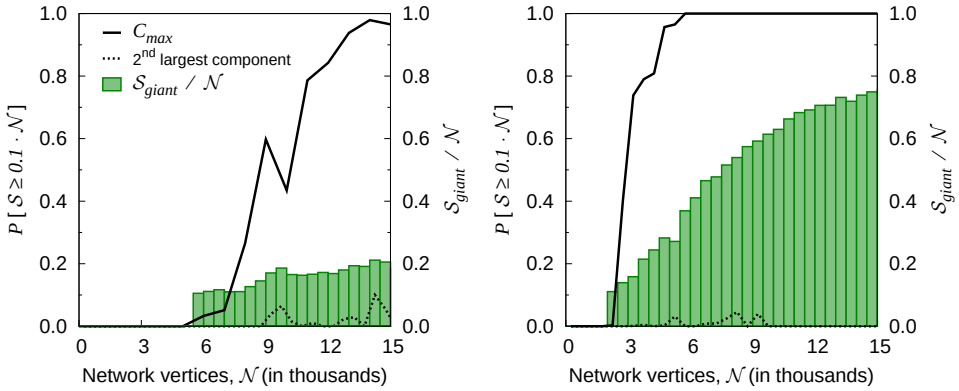


Figure 9: Internal structure of $\mathcal{C}_{max}$ at two subsequent time seconds, when R = 100 m. Vertices color and size indicate the node's betweenness centrality.

betweenness centrality of nodes in the northeastern region (where a bridge joins the two sides of the Rhine river). This implies that vertices in that area are part of a very large number of shortest paths, or, in other words, they branch together large groups of vehicles. Indeed, at the next time second, in the right plot, a shift of a few vehicles in that same region is sufficient to break the connectivity over the bridge, disconnecting the whole group of vehicles located in the northern area. As a result, $\mathcal{S}_{max}$ suddenly drops to around 1.000 nodes.

These dynamics are very frequent in the vehicular network, where very large components are the result of vehicles traveling between smaller components, branching the latter together: however, such vehicles only build *weak ties* that break as soon as the bridging vertex moves away. We can conclude that, when R = 100 m, large components suffer from low reliability, in the sense that they undergo a continuous merge-and-split process that makes long multi-hop paths appear and disappear at a high frequency. When R = 200 m, the sheer size of large clusters reduces the impact of these variations: however, we recall that the unreliability moves to the single-hop level, with wireless links that are significantly more loss-prone. A tradeoff thus appears between unreliabilities at the network and link level, depending on the communication range adopted.

Giant components. The previous results look at the largest component in general, neglecting whether



(a) R = 100 m (b) R = 200 m

Figure 10: Probability of existence of a first and second giant components and relative size of the first giant component versus $\mathcal{N}$, for different R's.

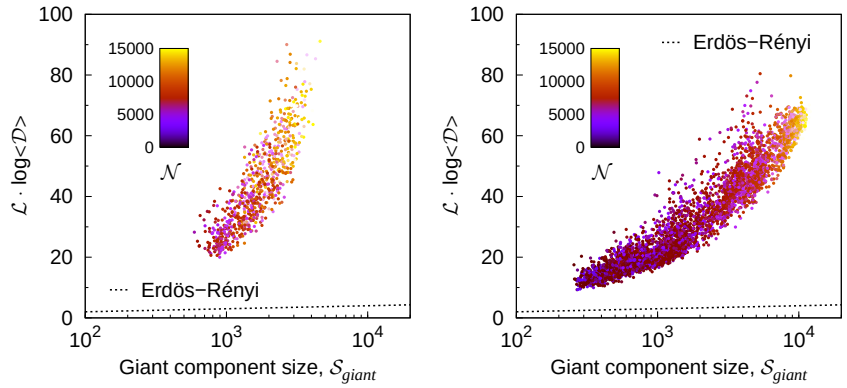
C_{max} is actually a "large enough" component. We now leverage the rigorous notion of giant component introduced in Sec. 4 to tell apart components that group a meaningful number of nodes, and study their appearance and structure. Fig. 10 displays the probability that the first (C_{max}) and the second largest components in the network are giant components, as $\mathcal{N}$ varies. We focus on R equal to 100 and 200 m only, since no giant component ever appears in the network for R = 50 m. In both cases, giant components start to appear at around the approximate critical threshold previously identified, i.e., $\mathcal{N} \sim$ 6000 and 3000 respectively; however, they do so with low-to-medium probability. In order to observe a giant component with a high probability, e.g., 90%, significantly higher values of $\mathcal{N}$ are to be considered, namely, 12500 for R = 100 m and 4500 for R = 200 m. These new thresholds, reported as dashed lines in Fig. 8, evidence that, when R = 100 m, persistent giant components are basically unavailable in the network. When R = 200 m, a persistent giant component is available, although only during the morning and afternoon rush hours.

Fig. 10 also reports the size of the giant component $\mathcal{S}_{giant}$, expressed as a fraction of the overall network size $\mathcal{N}$. When R = 100 m, not only the giant component is less probable to appear, but it never includes more than 20% of the network vertices. Conversely, when R = 200 m, $\mathcal{S}_{giant}$ monotonically grows with $\mathcal{N}$, arriving to include 75% of the network nodes during the morning traffic peak.

A second giant component rarely appears in the network, and a third giant component is never observed. Therefore, when it is available, the giant component maps to C_{max} and suffers from the large rapid variations previously described for C_{max}, reproposing the fundamental tradeoff between network-level and link-level unreliability under diverse values of R.

We further explore the internal structure of giant components by studying the average length $\mathcal{L}$ of the shortest paths between any two vertices within the component, in Fig. 11. The value of $\mathcal{L}$ is lower for small giant components, i.e., when $\mathcal{S}_{giant}$ is 500 or smaller. However, it then quickly grows as $\mathcal{S}_{giant}$ increases. This behavior holds for both R = 100 m and R = 200 m, and confirms that larger giant components are not denser and thus better connected, only geographically wider as a result of the (possibly temporary) merge of smaller components. Thus, giant components become more difficult to *navigate* as they grow in size [2, 3].

Fig. 11 also compares the vehicular network graph with an Erdös-Rényi random graph, a typical example of *small world* network where each pair of nodes is separated by a low number of hops [1]. Since $\mathcal{L} \approx \frac{log\mathcal{N}}{log<\mathcal{D}>}$ in an Erdös-Rényi graph, the latter is maps to a line in the plots. In fact, the line stays much lower than all the vehicular network points, a clear indication that vehicular networks are not small



(a) R = 100 m (b) R = 200 m

Figure 11: Small world property of the vehicular network versus $\mathcal{N}$ for different R's. The plots also reports the result for an Erdös-Rényi random graph.

worlds.

Key networking insights. Our analysis allows us to observe that very large network components do exist in the network, and that *multi-hop connected communication among several thousands vehicles is possible*. However, these very large components are affected by low availability, only appearing during a few hours each day. Moreover, their reliability is poor at any communication radius R, due a fundamental *tradeoff between network- and link-level reliability*: one has to choose between a merge-and-split network of stable links and a better connected network of loss-prone links. These observations evidence the difficulty of routing data within a purely ad hoc vehicular network: only limited time windows are exploitable, and, even then, designing an efficient protocol may be impossible, due to the complex and fast network topology dynamics or the unreliable vehicle-to-vehicle links. Therefore, we conclude that *carry-and-forward transfer paradigms are compulsory also to route data within large connected components*.

From a different perspective, the low availability and reliability of the large components, jointly with their peculiar internal structure, let us argue that *roadside unit (RSU) deployment is critical to achieve large components of connected vehicles that are reliable at both network and link level*. In particular, RSUs should be deployed where weak ties tend to appear, so as to act as permanent bridges among small but stable ad hoc connectivity islands.

Assuming the presence of bridging RSUs, large vehicular network components have proven not to be easy to navigate, as they are sparse non-small worlds. This results in most vehicles in these component being connected by a single long multi-hop path, passing through bridging nodes, which allows us to derive important rules for the design of efficient vehicular ad hoc routing protocols. First, *geographic information is indispensable to build and maintain routes* along the complex long multi-hop paths that characterize the connected components. Second, practical protocols should consider that *bridging vertices risk to be burdened with excessive forwarding loads*. Third, in all cases, due the length of multi-hop paths, *non-negligible end-to-end latency cannot be avoided*.

7 Node-level analysis

We conclude our study by descending at the individual node level. Our interest focuses on the characterization of the vertex neighborhood at one and two-hop distance.

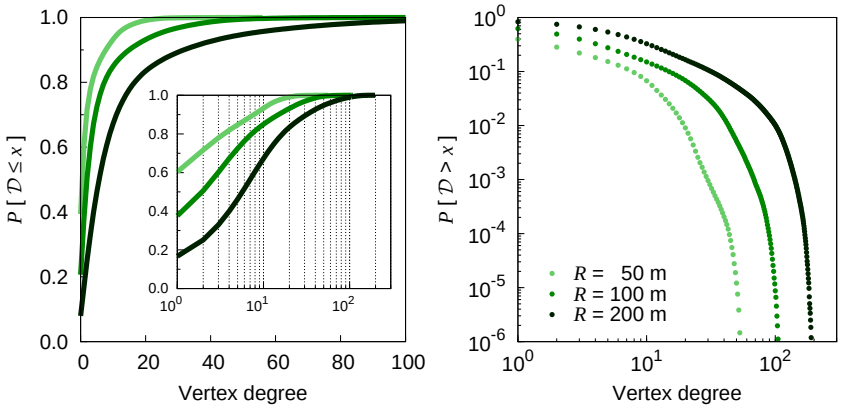


Figure 12: CDF (left) and CCDF (right) of the vertex degree $\mathcal{D}$, when aggregating all the samples of over the whole day and for different R's.

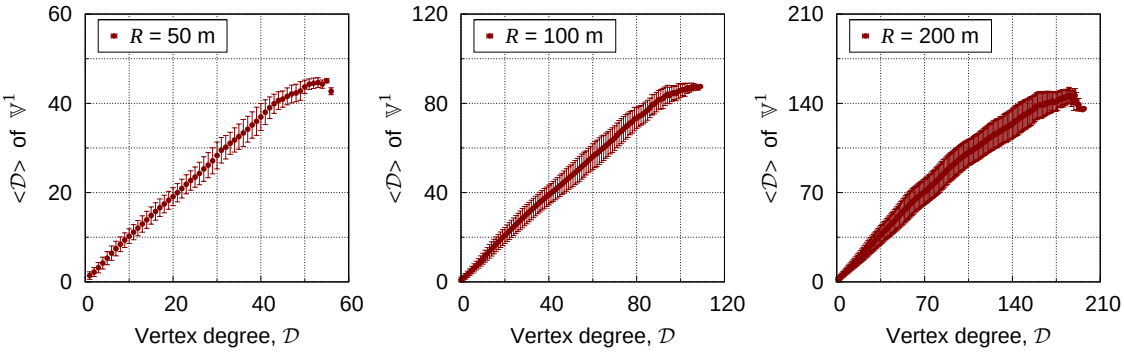


Figure 13: Assortativity for different values of the communication range R.

One-hop neighborhood. Fig. 12 shows, for each value of R, the CDF (left) and CCDF (right) of the node degree $\mathcal{D}$ in the vehicular network, when aggregating all samples over 24 hours. The plots point out how common small (e.g., $\mathcal{D} < 5$) one-hop communication neighborhoods are for any R, although less probable under larger communication ranges. Higher R's also result in an increased heterogeneity of the one-hop neighborhood: when R = 200 m, isolated nodes and vertices with 170 neighbors can coexist in the network.

The degree distribution does not follow a power law for any R, and we can thus conclude that vehicular networks are not *scale-free* in the degree distribution [3]. This means that the maximum one-hop neighborhood size is constrained (in our case, by geographical restrictions on the vertices positions) to sizes that are small with respect to the network size $\mathcal{N}$. In other words, although 170 neighbors may seem a lot, they only appear for high $\mathcal{N}$'s and represent in fact a mere 1% of the whole network. This is again a sign of poor navigability.

Two-hop neighborhood. The characterization of two-hop neighborhoods can be performed by observing the *assortativity* of the network, i.e., the correlation between the degree $\mathcal{D}$ of a node and the average degree of its one-hop neighbors in $\mathbb{V}^1$. Fig. 13 shows that the vehicular network is highly assortative: a strong linear relationship exists between these two measures, with a low standard deviation as from the errorbars.

This points out that high-degree vertices are typically connected with high-degree vertices, and low-degree ones with other low-degree vertices. The combination of lack of scale-free property and high assortativity is a proof that, unlike many other real-world networks, vehicular networks do not show a backbone of high-degree *hub nodes* interconnecting low-degree leaf nodes. Rather, the network is structured around weakly tied cliques of vehicles with similar degree.

Key networking insights. We observed vehicular communication neighborhoods to be heterogeneous and assortative. Apart from validating our previous remarks on the scarce navigability of connected components, these results imply that vehicles can move in a few tens of seconds from quasi-isolation to being part of cliques of hundreds of cars. Then, *medium access control, data rate adaptation and power control algorithms will play a more important role than expected*, as their rapidity to adapt to the varying network conditions may really make the difference between an efficient network and a useless one.

8 Discussion and conclusions

We presented a first study of the instantaneous topological features of a spontaneous vehicular network in a realistic urban environment. Our large-scale complex network analysis allowed us to draw conclusions on the significant limitations of the topology in terms of connectivity, availability, reliability and navigability, at both network and component levels; we also unveiled the degree heterogeneity and assortativity of vehicular neighborhoods. Our observations let us advocate the adoption of carry-and-forward transfer paradigms and/or the deployment of RSUs at weak tie locations, as well as the key role of protocols for channel access and rate/power control.

Concerning the limitations of our work, the study is constrained to one specific scenario due to the lack of other mobility datasets that are publicly available and yield a high level of realism. However, recent topological studies on urban road layouts have concluded that cities worldwide can be classified into macroscopic groups that feature very similar structural properties [26]. We thus conjecture that the single scenario we consider may be representative of many other urban regions of similar nature and size.

We also remark that we only considered the instantaneous vehicular network connectivity, and thus our results are not conclusive for disruption-tolerant networks (DTNs) based on carry-and-forward communication paradigms. Moreover, we limited the analysis to the average RF signal propagation behavior granted by a computationally-feasible disc model. The characterization of the topology of DTNs and the impact of more realistic propagation remain open problems that we plan to address in the future.

References

[1] R. Albert, A. Barabási, “Statistical Mechanics of Complex Networks”, *Reviews of Modern Physics*, 74(1):47–97, Jan. 2002.

[2] M.E.J. Newman, “The Structure and Function of Complex Networks”, *SIAM Review*, 45(2):167–256, 2003.

[3] S Boccaletti, V. Latora, Y. Moreno, M. Chavez, D. Hwang, “Complex Networks: Structure and Dynamics”, *Physics Reports*, 424(4-5):175–308, 2006.

[4] M. Doering, T. Pögel, W.-B. Pöttner, L. Wolf, “A new mobility trace for realistic large-scale simulation of bus-based DTNs,” *ACM CHANTS*, Chicago, IL, USA, Sep. 2010.

[5] J. Yuan, Y. Zheng, X. Xie, G. Sun, “Driving with knowledge from the physical world,” *ACM SIGKDD*, San Diego, CA, USA, Aug. 2011.

[6] M.M. Artimy, W. Robertson, W.J. Phillips, "Connectivity in Inter-vehicle Ad Hoc Networks". *CCECE*, Niagara Falls, NY, USA, May 2004.

[7] H. Füssler, M. Torrent-Moreno, M. Transier, R. Krüger, H. Hartenstein, W. Effelsberg, *Studying Vehicle Movements on Highways and their Impact on Ad-Hoc Connectivity*, ACM MC2R, 10(4):26–27, Oct. 2006.

[8] S. Yousefi, E. Altman, R. El-Azouzi, M. Fathy "Analytical Model for Connectivity in Vehicular Ad Hoc Networks", *IEEE Transactions on Vehicular Technology*, 57(6):3341–3356, Nov. 2008.

[9] M. Khabazian, M.K. Ali, "A Performance Modeling of Connectivity in Vehicular Ad Hoc Networks", *IEEE Transactions on Vehicular Technology*, 57(4):2440–2450, Jul. 2008.

[10] G.H. Mohimani, F. Ashtiani, A. Javanmard, M. Hamdi, "Mobility Modeling, Spatial Traffic Distribution, and Probability of Connectivity for Sparse and Dense Vehicular Ad Hoc Networks", *IEEE Transactions on Vehicular Technology*, 58(4):1998–2006, May 2009.

[11] S. Shioda, J. Harada, Y. Watanabe, T. Goi, H. Okada, K. Mase, "Fundamental Characteristics of Connectivity in Vehicular Ad Hoc Networks", *IEEE PIMRC*, Cannes, France, Sep. 2008.

[12] Y. Zhuang, J. Pan, L. Cai, "A Probabilistic Model for Message Propagation in Two-Dimensional Vehicular Ad-Hoc Networks", *ACM VANET*, Chicago, IL, USA, Sep. 2010.

[13] I. Ho, K.K. Leung, J.W. Polak, "Stochastic Model and Connectivity Dynamics for VANETs in Signalized Road Systems", *IEEE Transactions on Networking*, 19(1):195–208, Feb. 2011.

[14] X. Jin, W. Su, Y. Wei, "Quantitative Analysis of the VANET Connectivity: Theory and Application", *IEEE VTC-Spring*, Budapest, Hungary, May 2011.

[15] I. Ho, K.K. Leung, J.W. Polak, R. Mangharam, "Node Connectivity in Vehicular Ad Hoc Networks with Structured Mobility", *IEE LCN*, Dublin, Ireland, Oct. 2007.

[16] M. Fiore, J. Härri "The Networking Shape of Vehicular Mobility", *ACM MobiHoc*, Hong Kong, PRC, May 2008.

[17] M. Kafsi, P. Papadimitratos, O. Dousse, T. Alpcan, J.-P. Hubaux, "VANET Connectivity Analysis", *IEEE AutoNet*, New Orleans, LA, USA, Dec. 2008.

[18] W. Viriyasitavat, F. Bai, O.K. Tonguz, "Dynamics of Network Connectivity in Urban Vehicular Networks", *IEEE Journal on Selected Areas in Communications*, 29(3):515–533, Mar. 2011.

[19] H. Conceicão, M. Ferreira, Joaõ Barros, "On the Urban Connectivity of Vehicular Sensor Networks", *DCOSS*, Santorini, Greece, Jun. 2008.

[20] G. Pallis, D. Katsaros, M. D. Dikaiakos, N. Loulloudes, L. Tassiulas, "On the Structure and Evolution of Vehicular Networks", *IEEE/ACM MASCOTS*, London, UK, Sep. 2009.

[21] S. Uppoor, M. Fiore, "Large-scale Urban Vehicular Mobility for Networking Research", *IEEE VNC*, Amsterdam, Holland, Nov. 2011.

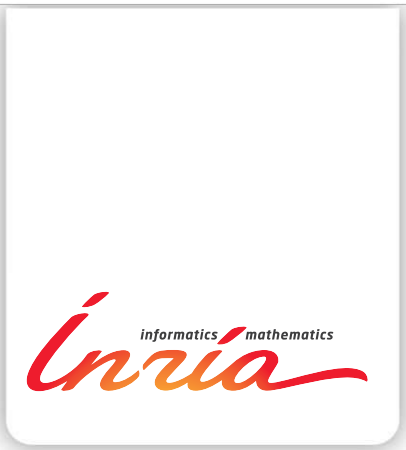
[22] G. Hertkorn, P. Wagner, "The application of microscopic activity based travel demand modelling in large scale simulations", *World Conference on Transport Research*, Istanbul, Turkey, Jul. 2004.

[23] C. Gawron, "An Iterative Algorithm to Determine the Dynamic User Equilibrium in a Traffic Simulation Model", *International Journal of Modern Physics C*, 9(3):393–407, 1998.

[24] D. Hadaller, S. Keshav, T. Brecht, S. Agarwal, "Vehicular Opportunistic Communication Under the Microscope", *ACM MobiSys*, San Juan, Puerto Rico, USA, Jun. 2007.

[25] F. Bai, D.D. Stancil, H. Krishnan, "Toward Understanding Characteristics of Dedicated Short Range Communications (DSRC) From a Perspective of Vehicular Network Engineers", *ACM MobiCom*, Chicago, IL, USA, Sep. 2010.

[26] P. Crucitti, V. Latora, S. Porta, "Centrality measures in spatial networks of urban streets", *Physics Review E*, 73(3), Mar. 2006.



RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399