Gramaglia, Marco; Trullols-Cruces, Óscar; Naboulsi, Diala; Fiore, Marco; Calderón, María. (2016). Mobility and connectivity in highway vehicular networks: a case study in Madrid. *Computer Communications,* 78, pp. 28-44.

DOI: https://doi.org/10.1016/j.comcom.2015.10.014

Mobility and connectivity in highway vehicular networks: a case study in Madrid

Marco Gramaglia[a], Oscar Trullols-Cruces[b], Diala Naboulsi[c], Marco Fiore[a,e], Maria Calderon[d]

[a] *CNR-IEIIT, 10129 Torino, Italy.*
[b] *UPC, 08034 Barcelona, Spain.*
[c] *INSA Lyon and Inria, 69621 Villeurbanne, France.*
[d] *UC3M, 28911 Madrid, Spain.*
[e] *Inria, 69621 Villeurbanne, France.*

Abstract

The performance of protocols and architectures for upcoming vehicular networks are commonly investigated by means of computer simulations, due to the excessive cost and complexity of large-scale experiments. Dependable and reproducible simulations are thus paramount to a proper evaluation of vehicular networking solutions. Yet, we lack today a reference dataset of vehicular mobility scenarios that are realistic, publicly available, heterogeneous, and that can be used for networking simulations straightaway. In this paper, we contribute to the endeavor of developing such a reference dataset, and present original synthetic traces that are generated from high-resolution real-world traffic counts. They describe road traffic in quasi-stationary state on three highways near Madrid, Spain, for different time-spans of several working days. To assess the potential impact of the traces on networking studies, we carry out a comprehensive analysis of the vehicular network topology they yield. Our results highlight the significant variability of the vehicular connectivity over time and space, and its invariant correlation with the vehicular density. We also underpin the dramatic influence of the communication range on the network fragmentation, availability, and stability, in all of the scenarios we consider.

Keywords: Vehicular networks, highway traffic, synthetic traces, vehicle-to-vehicle communication, connectivity, complex networks.

1. Introduction

A key enabling technology of future Intelligent Transportation Systems (ITS), vehicle-to-vehicle (V2V) communication is envisioned to interconnect vehicles into distributed, self-organized networks. The latter are expected to complement today's mobile access architecture, and support services such as cooperative awareness, collision avoidance, or data dissemination.

The emergence of large-scale vehicular networks requires that a large fraction of vehicles is equipped with dedicated radio interfaces. Such a pervasive deployment of V2V communication is closer than one would imagine: standards for V2V communication, such as IEEE 802.11-2012 [1], IEEE 1609 [2],OSI CALM-M5 [3] and ETSI ITS-G5 [4] are now finalized, and regulators in the USA plan to enforce V2V radio interfaces on all new vehicles by 2017 [5]. Early large-scale field tests are also in progress, e.g., within the sim^{TD} project in Germany, or the Ann Arbor Safety Pilot in Michigan, USA.

This notwithstanding, experimental trials of vehicular networking solutions remain an exception, due to their costs and complexity. The vast majority of applications, protocols and architectures for upcoming vehicular networks is evaluated via computer simulation. The dependability of results is then conditional on the level of realism of the models assumed, and the representation of the mobility of individual vehicles is often the single feature that introduces the largest bias [6].

For that reason, during the past decade, significant efforts have been made to gather real-world road traffic data [7, 8], develop effective tools for the simulation of vehicular movement [9–12], and generate realistic synthetic mobility traces [13–15]. Still, a reference set of realistic, publicly shared, heterogeneous road traffic scenarios for networking simulation is not yet available. This situation, originated by a manifest scarcity of mobility traces featuring the required level of realism and spatiotemporal granularity, is raising questions on the dependability and reproducibility of research results [16]. Within such a context, this paper puts forward several major contributions, as follows.

First, we take a step forward in the direction of dependable and reproducible vehicular networking research, by providing to the community multiple novel realistic highway traffic traces for network simulation. The traces are based on real-world traffic count measurements that feature an unprecedented level of detail, and are representative of heterogeneous motorway segments and road traffic conditions, as discussed in Sec. 2.

Second, we outline a detailed methodology to generate synthetic mobility traces of unidirectional highway traffic starting from road traffic counts. The traces model road traffic in quasi-stationary conditions, where macroscopic features such as the average vehicular density, speed, and out-flow observed on each highway lane are invariant over the full span of the simulated road segment. To that end, we leverage inherent properties of the real-world data for the per-vehicle calibration of well-known car-following and lane-changing microscopic models. Details are provided in Sec. 3.

Third, we characterize the vehicular network connectivity resulting from the proposed synthetic traces. To that end, we perform a network protocol-independent study, by adopting an instantaneous topology model, as discussed in Sec. 4. We investigate the impact of a wide range of parameters, including time (i.e., hour of the day, day of the week), highway settings (i.e., number of lanes, speed limits), road traffic conditions (i.e., free flow or synchronized traffic), and V2V communication range. Our results, presented in Sec. 5 underscore, in all of the scenarios we considered, the following properties: (i) the dramatic impact that relatively small communication range variations have on the network structure; (ii) the prevalent role of the vehicular density in driving network connectivity via three-phase dynamics; (iii) the limited availability and stability of long-range multi-hop vehicular networks, (iv) the fact that the highway vehicular network is difficult to navigate.

Finally, a comparative review of the related literature is provided in Sec. 6, before we draw conclusions in Sec. 7.

(a) Highway locations (b) M30 - Aerial view (c) M40 - Aerial view (d) A6 - Aerial view

Figure 1: (a) Geographical location of the measurement points on the three highways considered in our study, near Madrid, Spain: M30 (A), M40 (B) and A6 (C). (b,c,d) Close-by views of measurement points on M30, M40 and A6.

2. Source measurement data

The synthetic traces we present in this paper are based on empirical data that comes from real-world measurements carried out in the region of Madrid, Spain. The data, kindly provided to us by the Spanish office for the traffic management (Dirección General de Tráfico, DGT) and the Madrid City Council, details the vehicular traffic conditions on the following three arterial highways.

M30. With an average distance of 5.17 Km from the city center, M30 is the inner part of the Madrid city beltway system, which also comprises the outermost M40 and M50. The data employed in this study comes from measurements along the northbound direction, close to the junction with the A-2 Motorway and marked as A in Fig. 1a. There, M30 features 4 lanes in the main carriageway, as it can be observed in the aerial view of Fig. 1b. The speed limit along M30 is 90 Km/h.

M40. Motorway M40 is part of the intermediate layer of the Madrid city beltway system. It has an average distance of 10.7 Km from the city center, and traverses both the most peripheral areas of the municipality as well as several surrounding minor cities. The measurement point, marked as B in Fig. 1a, is at the 12.7-Km milepost, where M40 traverses the suburb of San Blas and the town of Coslada. The measures cover the southbound carriageway, in Fig. 1c, which includes 3 lanes with a speed limit of 100 Km/h.

A6. Autovía A6 is a motorway that connects the city of A Coruña to the city of Madrid. A6 enters the urban area from the northwest, collecting the traffic demand of the conurbation built along it. The data collection point is placed around the 11-Km milepost in the Madrid direction, depicted with a C in Fig. 1a, where A6 features 3 lanes, as per Fig. 1d. The speed limit is 120 Km/h.

2.1. Collecting fine-grained traffic count data

The sensors deployed on the three highways are induction loops, i.e., loops of wires buried under the concrete layer and creating a magnetic field. When a vehicle transits on the vertical axis of the loop, it induces a variation in the magnetic field. If two loops are placed close to each other, other metrics, e.g., the vehicle speed and length, can be also determined.

Usually, these devices are programmed to supply coarse-grained data, since public transportation authorities are generally interested in aggregate measures on, e.g., the number of vehicles transiting on a road, their average speed, or the percentage of heavy vehicles[1], so as to detect major alterations of traffic conditions [17, 18]. The loops used in this work are normally configured to supply data averaged over 60 seconds, but their setup was changed specifically for our study, so as to provide fine-grained information on each transiting vehicle.

Not only the level of detail, but also the timing and duration of the measurements are critical aspects of the data collection. Indeed, vehicular traffic presents significant daily variability, and rush hours yield diverse traffic conditions than off-peak hours, especially on main arterial roads like those we consider. In order to capture such temporal heterogeneity, and compatibly with the limitations imposed by the dedicated setup needed at the induction loops, we collected the following datasets.

[1] As an example, Dirección General de Tráfico provides elaborations of the traffic count data via the Infocar web service at `http://infocar.dgt.es`, with visualizations of the historical aggregate data at the observation points.

One day-long dataset, collected on M30 during 24 hours of a typical weekday in May 2010. This dataset features variable conditions, from very sparse traffic at night to heavy congestion during the morning rush hours. It thus provides a rather complete view of the possible traffic scenarios met on a real-world highway.

Sixteen 30-minute datasets, collected on M40 and A6. These datasets were recorded on multiple weekdays of May 2010, during the morning traffic peak (from 8:30 a.m. to 9 a.m.), and during off-peak hours (from 11.30 a.m. to 12 p.m.). The rationale for these shorter datasets is that they allow us to generalize our study, by investigating the effects induced by different roads (e.g., number of lanes, speed limits and proximity to the city center) and different weekdays.

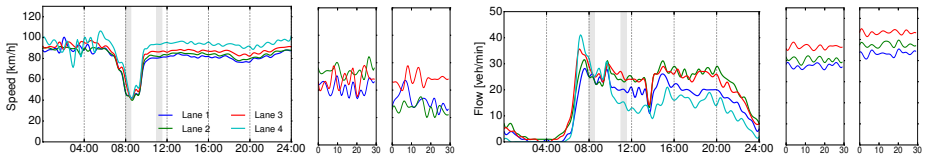
Overall, these traffic count datasets provide a comprehensive view of heterogeneous traffic conditions, and they do so at a high level of detail. Their unprecedented combination of precision and completeness makes them an ideal input to the microscopic simulation of highway traffic, enabling the generation of realistic mobility traces that are representative of many and varied traffic situations.

2.2. Understanding the data

Each traffic count dataset entry records one vehicle transiting at the measurement point, and includes:

- **Timestamp**: the time at which the vehicle transit was recorded by the induction loop. The precision of the time reference is 100 milliseconds.
- **Speed**: the vehicle speed, in Km/h.
- **Lane**: the lane on which the vehicle transited.

An overview of the traffic count data is provided in Fig. 2. The day-long time series of the vehicular speed and in-flow on M30 are portrayed separately for each lane in Fig. 2a and Fig. 2d, respectively. The in-flow is the number of vehicles transiting by the measurement point per minute, and us typically

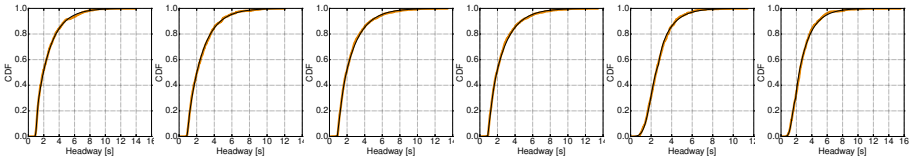


(a) M30 speed (b) M40 (c) A6 (d) M30 flow (e) M40 (f) A6

Figure 2: Traffic count data overview. Per-lane speed (a) and in-flow (d) recorded during a full day on M30, and during two sample 30-minute intervals (highlighted as gray-shaded in day-long plots) on M40 (b,e) and A6 (c,f).

used as a measure of road traffic intensity. We remark the very low in-flow at night, i.e., from midnight to around 7.30 a.m., where speeds also tend to be the highest. Early morning, from 7.30 a.m. to 10 a.m. is characterized by a significant increase of in-flow and reduction of speeds – a clear symptom of congestion. Once the morning rush hours have passed, the traffic is quite regular over the rest of the day, with the notable exception of some flow reduction at around 2 p.m., i.e., lunch time in Spain. On a per-lane basis, the speed of the rightmost lane is typically the lowest, while that of the leftmost lane is normally the highest: this is expected, since overtaking is only allowed to the left in Spain, which pushes faster vehicles to travel on left lanes. Also, we observe that traffic tends to be the thickest in the central lanes, at least in standard, non-congested situations: again, this is the common behavior in Spain, with the rightmost lane left to heavy trucks and the leftmost one used for overtaking only.

From a traffic flow theoretical standpoint, the diverse combinations of speed and in-flow present in the M30 dataset fall into two different road traffic states. So-called *free flow* traffic [19], characterized by neatly separated speeds on different lanes, dominates most of the dataset. This is especially evident from 10 a.m. onward, as beforehand the traffic is either too sparse to be statistically significant, or too thick to be in free flow. The latter situation, i.e., thick traffic leading to congestion, is observed during the early morning, between 8 a.m. and 10 a.m. During this period, the traffic is in so-called *synchronized* state [19], where the density is such that all lanes are equally jammed: indeed, we can



(a) M40, lane 1(b) M40, lane 2 (c) A6, lane 2 (d) A6, lane 3 (e) M30, lane 3(f) M30, lane 4

Figure 3: Inter-arrival time CDF measured on May 12, 2010. Each plot refers a lane on M40 at 8:30 a.m. (a, b), A6 at 11:30 a.m. (c, d), and M30 at 11:30 a.m. (e,f). Solid black lines represent the mixture model for each distribution.

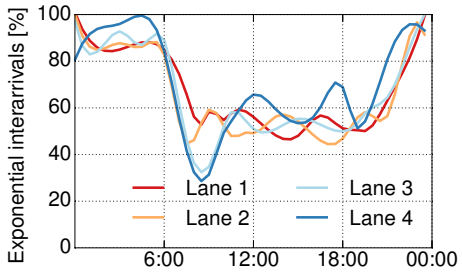


Figure 4: Time series of the percentage of road traffic entering M30 with exponential inter-arrivals. Curves refer to different lanes of the highway.

remark the distinctive slower, homogeneous speeds on all lanes.

As far as the 30-minute datasets collected on M40 and A6 are concerned, the speed and in-flow yielded by two sample excerpts are shown in the remaining plots of Fig. 2. Their time-spans are highlighted in the day-long M30 plots as gray-shaded intervals, so as to give a better perception of how their duration compares to that of the M30 data. Throughout all these datasets, road traffic is mostly in a free flow state, but for rare and episodic spontaneous local perturbations that rapidly disappear.

2.3. Interarrival times analysis

The analysis of vehicle inter-arrival times in the traffic count datasets we collected on M30, M40 and A6 shows that a mixture Gaussian-exponential model yields an excellent approximation of the empirical data. Fig. 3 shows the match between the mixture model and the experimental data on multiple combinations

of highway, lane, day and hour.

The mixture model also provides valuable information on drivers' behavior. On the one hand, the Gaussian part of the distribution captures *bursty* arrivals of vehicles that travel close to each other at similar speeds, a behavior typical of congested road traffic. On the other hand, the exponential part of the distribution models *isolated* vehicles whose movement is less constrained by that of other cars, which is normally observed in pure free flow traffic conditions.

An intuitive representation of the mixture of the two road traffic behaviors is depicted in Fig. 4. There, we portray the percentage of road traffic measured on M30 during the whole day that exhibits exponential inter-arrivals. The value on the y axis is expressed as the percentage of vehicles that show an isolated behavior (i.e., exponential inter-arrivals); clearly, the residual percentage is made of vehicles traveling in bursts (i.e., with Gaussian inter-arrivals). Results are divided by lane.

We observe that inter-arrivals are never purely exponential. In fact, the Poisson arrival assumption may be a somehow decent approximation at night, between 11 p.m. and 6 a.m. However, throughout the rest of the day, all lanes are characterized by an even mixture of bursty and isolated arrivals. In fact, we even remark the prominence of the first type of arrivals on the leftmost lanes (i.e., lanes 3 and 4) between 8 a.m. and 9 a.m., i.e., during the morning traffic peak.

Some differences also emerge among lanes. Inter-arrivals on the leftmost lane, denoted as lane 4 in the plot, tend to have a more exponential behavior in the general case: as shown by Fig. 2d, this lane is typically less trafficked than the others, and vehicles traveling on it are more isolated. However, during the morning rush hours, traffic on the leftmost lanes increases significantly, and the high speed of vehicles traveling on such lanes forces drivers to keep very similar safety distances: ultimately, this results in very homogeneous traffic and low-variance Gaussian inter-arrivals.

Interestingly, all the results above invalidate, in the case of our target scenarios, the common assumption of exponential or even uniform distribution of

the time headway between subsequent vehicles on each lane.

For additional details on the modeling of inter-arrival times in our datasets, we refer the reader to the discussions in [20, 21].

3. Vehicular mobility traces

Our objective is to generate road traffic traces that are representative of unidirectional highway traffic in *quasi-stationary* state, i.e., such that traffic conditions are comparable between the in-flow and out-flow boundaries of the simulated road segments. Quasi-stationarity is a common assumption in vehicular networking research, see, e.g., [17, 22–28]. It provides a controlled environment where ungoverned road traffic phenomena (e.g., continuous road traffic variations due to in- and out-ramps, unpredictable drivers' behaviors, or accidents) do not bias the evaluation of network solutions. Although it does not model macroscopic perturbations induced by the aforementioned phenomena, quasi-stationarity still allows a full-fledged representation of the microscopic dynamics of real-world road traffic (including, e.g., varying vehicle speed due to acceleration or deceleration, lane changes, overtakes).

In this section, we feed the real-world traffic count data presented in Sec. 2 to a microscopic vehicular mobility simulator[2], based on state-of-the-art car-following and lane-changing models (Sec. 3.1) that are purposely calibrated (Sec. 3.2) so as to derive our trace (Sec. 3.3).

3.1. Microscopic models

The car-following and lane-changing microscopic mobility models implemented by our simulator are IDM and MOBIL. Both models have been validated by the transportation research community, and are widely adopted for the simulation of vehicular networks.

The Intelligent Driver Model (IDM) [29] characterizes the behavior of the driver of a vehicle i through the instantaneous acceleration $dv_i(t)/dt$, calculated

[2] Available at `http://www.it.uc3m.es/madrid-traces`.

as

$$\frac{dv_i(t)}{dt} = a\left[1 - \left(\frac{v_i(t)}{v_i^{max}}\right)^4 - \left(\frac{\Delta x_i^{des}(t)}{\Delta x_i(t)}\right)^2\right], \tag{1}$$

$$\Delta x_i^{des}(t) = \Delta x^{safe} + \left[v_i(t)\Delta t_i^{safe} - \frac{v_i(t)\Delta v_i(t)}{2\sqrt{ab}}\right]. \tag{2}$$

In (1), $v_i(t)$ is the current speed of vehicle i, v_i^{max} is the maximum speed its driver would like to travel at, and $\Delta x_i^{des}(t)$ is the so-called *desired dynamical distance*, representing the distance that the driver should keep from the leading vehicle. The latter is computed in (2) as a function of several measures taken with respect to the car in front of vehicle i: the minimum bumper-to-bumper distance Δx^{safe}, the speed difference $\Delta v_i(t)$, and the minimum safe time headway, i.e., the time the driver needs in order to react to sudden braking by the front vehicle and avoid an accident, denoted as Δt_i^{safe}. In both equations, a and b denote the maximum absolute acceleration and deceleration, respectively. When combined, these formulae return the instantaneous acceleration of the car, as a combination of the desired acceleration on an empty road, i.e., the term $[1-(v_i(t)/v_i^{max})^4]$, and the braking deceleration induced by the preceding vehicle, i.e., the term $(\Delta x_i^{des}(t)/\Delta x_i(t))^2$.

The Minimizing Overall Braking Induced by Lane-changes (MOBIL) model [30] builds on a game theoretical approach, and lets the driver of a vehicle i move to an adjacent lane if the advantage in doing so is greater than the disadvantage of the trailing car j in the new lane. The (dis)advantage is measured in terms of acceleration, which translates into the inequality

$$\left|\frac{dv_i(t)}{dt}\right|_L - \frac{dv_i(t)}{dt} + a_L \geq p\left(\frac{dv_j(t)}{dt} - \left|\frac{dv_j(t)}{dt}\right|_L\right) + k \cdot a, \tag{3}$$

where the notation $|\cdot|_L$ denotes accelerations computed as if vehicle i were traveling on the lane to its left rather than in the current one. In (3), $p \in (0, 1]$ is a politeness factor that models the selfishness of the driver with respect to the new back vehicle j, $k \cdot a$ is a hysteresis threshold that prevents lane hopping, and a_L is a bias acceleration that can be used to favor or limit movements to left. An identical formulation can be used for right-hand-side lane changes, and the

respective advantages can be compared to determine the final lane movement, if any. Note that, in Spain, road traffic regulations enforce drivers to travel on the rightmost lane whenever possible: we thus expect $a_R > a_L$ and $a_R > 0$, i.e., right-hand-side movements to be favored over left or no movement, if equivalent conditions are present on all lanes.

3.2. Model parameter calibration

In order to obtain quasi-stationary traffic conditions over the simulated highway segment, some calibration of the IDM and MOBIL parameters are necessary. Specifically, for the acceleration a, deceleration b, politeness factor p and minimum bumper-to-bumper distance Δx^{safe} the default values suggested in [29, 30] work well. The other parameters have instead to be adapted to the specificities of the road traffic scenarios we considered, as summarized in Tab. 1. We remark that ours is the first work integrating fine-grained traffic counts in a microscopic vehicular mobility generator; in this context, the calibration presented below is mandatory in order to avoid instability in the synthetic road traffic[3].

Maximum desired speed. Vehicles are introduced in the simulation at the time and with the speed defined by the real-world traffic count dataset. However, we need to determine the maximum desired speed v_i^{max} of each vehicle i, i.e., the cruise velocity that its driver would keep if alone on the highway [29]. We proceed as follows.

First, we recall that, according to traffic flow theory, vehicles in a free flow state have limited interactions, which allows them to travel at velocities close to

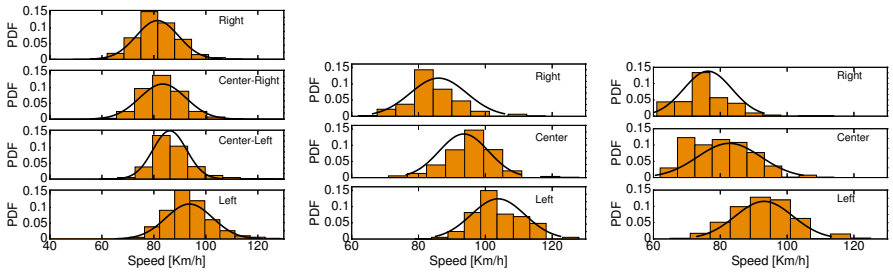
[3]Specifically, we recorded a significant amount of dangerous driving behaviors in the real-world traffic count data, leading to inter-distances that are incompatible (i.e., too small) with the speed difference (too high) among subsequent vehicles. In such a scenario, letting vehicles move at constant speed, or choosing desired speeds and safe time headway from non-calibrated distributions, leads to continuous accidents or extremely slow traffic. Also, removing misbehaving vehicles is not an option, since their number is not negligible, and discarding them would limit data realism. Our parametrization can accommodate such dangerous but realistic situations in a synthetic mobility trace.

Table 1: IDM and MOBIL parameter settings

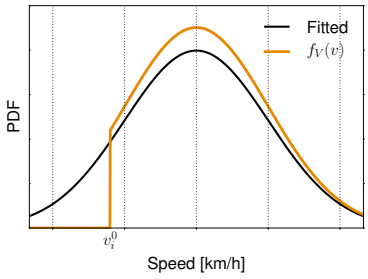
Model	Parameter	Meaning	Value
IDM	a	Maximum acceleration	1 m/s^2
IDM	b	Maximum (absolute) deceleration	2.5 m/s^2
IDM	v_i^{max}	Maximum desired speed	$\sim f_V(v)$
IDM	Δx^{safe}	Minimum distance	1 m
IDM	Δt_i^{safe}	Minimum safe time headway	$\sim f_T(\Delta t)$
MOBIL	p	Politeness factor	0.5
MOBIL	a_L	Bias acceleration (left)	0 m/s^2
MOBIL	a_R	Bias acceleration (right)	0.2 m/s^2
MOBIL	k	Hysteresis threshold factor	0.3

their maximum desired speed. We thus assume that real-world ingress speeds in the free flow zone can be used as a baseline for the derivation of the desired speeds. We identify the free flow zone in each traffic count dataset: in the M30 dataset, as discussed in Sec. 2.2, free flow characterizes the hours from the start of the day and 6 a.m. (when synchronized traffic first appears), and from 10 a.m. (once synchronized traffic dissolves) to midnight; in the M40 and A6 datasets, we can safely consider that road traffic is consistently in free flow.

Second, we extract the free flow speed distributions for each road, on a per-lane basis. The corresponding Probability Density Functions (PDF) are shown in Fig. 5a, Fig. 5b, and Fig. 5c, for M30, M40 and A6, respectively. In the latter two cases, the empirical distributions overlap for all combinations of day and hour, and are thus aggregated. The PDFs are separated by lane, as drivers traveling on different lanes tend to have dissimilar maximum desired speeds. Interestingly, all distributions have Gaussian shapes, which let us model the maximum desired speeds as a Gaussian-distributed random variables, whose fitted PDFs are portrayed as solid lines in Fig. 5. Clearly, the mean $\mu_{h,l}$ and standard deviation $\sigma_{h,l}$ of the fitted distributions vary depending on the highway h and lane l considered: there is a neat trend for lanes towards the left to yield higher velocities than those towards the right, in all scenarios.



(a) M30 (b) M40 (c) A6



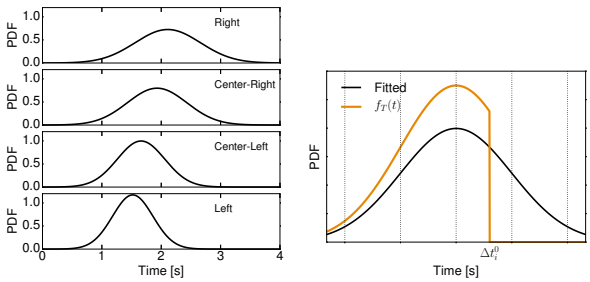
(d) Final distribution example

Figure 5: Calculation of the maximum desired speed v_i^{max}. (a,b,c) Empirical and fitted distributions of the free flow speed on each lane of M30, M40 and A6, respectively. (d) Example of per-vehicle truncation and normalization of the fitted distribution, so that only values larger than the initial speed v_i^0 are considered for v_i^{max}, $\forall i$.

As a third step, we adapt the final lane-dependent v_i^{max} distribution on a per-vehicle basis, as

$$f_V(v) = \begin{cases} 0, & v < v_i^0 \\ \frac{\sqrt{2}\, exp(-(v-\mu_{h,l})^2/2\sigma_{h,l}^2)}{\sigma_{h,l}\sqrt{\pi}\left[1+erf((v_i^0-\mu_{h,l})/\sigma_{h,l}\sqrt{2})\right]}, & v \geq v_i^0. \end{cases} \quad (4)$$

The expression in (4) truncates and re-normalizes the Gaussian distribution at the speed v_i^0 recorded in the real-world traffic count data for vehicle i. This is graphically explained in Fig. 5d. This way, the initial velocity of i, i.e., v_i^0, becomes the lower bound to v_i^{max}, which guarantees that the maximum desired speed of a vehicle i is never lower than v_i^0. The opposite would be unrealistic, for two reasons: first, it would imply that i enters the simulation at a speed higher than the maximum velocity it targets, which hardly makes sense; second, it



(a) M30 (b) Final distribution example

Figure 6: Calculation of the minimum safe time headway Δt_i^{safe}. (a) Reference distributions of the typical safe time headway on each lane of M30, as inferred by the experimental flow, speed, and inter-arrival information contained in the traffic count dataset. (b) Example of per-vehicle truncation and normalization of the reference distribution, so that only values smaller than the initial inter-arrival time Δt_i^0 are considered for Δt_i^{safe}, $\forall i$.

would force an immediate braking according to the IDM model in (1), slowing down the following vehicles and introducing an unrealistic queuing perturbation in the highway traffic.

Minimum safe time. The minimum safe time headway Δt_i^{safe} is known to vary across real-world scenarios. In [29], the default value is 1.5 s. However, drivers in different countries prefer diverse safe times, from 0.9 s in Germany [31] to 3 s in some States of USA [32].

In order to determine the correct per-vehicle Δt_i^{safe} for our scenario, we follow a similar approach as that taken for the calculation of the maximum desired speed. In this case, however, extracting the baseline empirical distributions is less straightforward, and we opt for a mixed analytical-empirical approach, as follows.

From the dataset, we can measure the inter-arrival times between vehicles, which can be directly related to the Δt_i^{safe} values. However, as discussed in Sec. 2.3, the mixture Gaussian-exponential shape of inter-arrivals is known to aggregate bursty as well as isolated arrivals [20]. The latter are generated by vehicles that travel far away from each other: in this case, drivers are not influenced by the behavior of nearby vehicles, and thus isolated arrivals are not

representative of actual safety distances. As a result, we need to exclude them from the Δt_i^{safe} estimation, and preserve bursty arrivals that refer to thick traffic, where drivers actually keep a minimum safe time headway with respect to their front vehicle.

We resort to traffic flow theory to perform the operation above, on a per-lane basis. On a highway h, the vehicular density ρ on lane l can be expressed as

$$\rho_{h,l} = \frac{1}{L + \Delta t_{h,l}^{safe} v_{h,l}} \tag{5}$$

where L is the average length of the vehicles, $v_{h,l}$ is the average speed, and $\Delta t_{h,l}^{safe}$ is the average safe time headway [33]. From density $\rho_{h,l}$, we can compute the vehicular flow $q_{h,l} = \rho_{h,l} \cdot v_{h,l}$, which results in

$$\Delta t_{h,l}^{safe} = \frac{1}{q_{h,l}} - \frac{L}{v_{h,l}}. \tag{6}$$

Expression (6) directly relates $\Delta t_{h,l}^{safe}$ to the maximum value of the flow $q_{h,l}$ and average speed $v_{h,l}$. The maximum flow $q_{h,l}$ can be inferred by identifying in the M30 dataset the time interval at which the speed breakdown occurs on each lane in Fig. 2. The average speed $v_{h,l}$ is easily computed as the average velocity of vehicles in free flow conditions. Considering $L = 4$ m as the vehicle length, we obtain typical values of $\Delta t_{h,l}^{safe}$ on each lane of every highway. In the M30 dataset, we have 2.11, 1.93, 1.66 and 1.52 s for lanes from the rightmost to the leftmost, respectively. Interestingly, these values are well aligned with those found in the literature [29, 31, 32].

The reference Gaussian distribution of safe time headway is then assigned a mean $\Delta t_{h,l}^{safe}$. The standard deviation $\sigma_{h,l}$ is set such that the minimum inter-arrival time recorded in the real-world traffic count dataset, i.e., 0.3 s, represents the 0.99 quantile of the distribution, i.e., three standard deviations. Formally, $\sigma_{h,l} = (\Delta t_{h,l}^{safe} - 0.3)/3$. The resulting per-lane distributions are plotted in Fig. 6a for the M30 case[4].

[4] An equivalent analysis is not possible for M40 and A6, since the associated traces do not

As a final step, similar to what done for the maximum desired speed, a per-vehicle distribution is to be determined from the lane-dependent reference ones. In this case, the final Δt_i^{safe} distribution is

$$f_T(\Delta t) = \begin{cases} \frac{\sqrt{2}\, exp(-(\Delta t - \Delta t_{h,l}^{safe})^2 / 2\sigma_{h,l}^2)}{\sigma_{h,l}\sqrt{\pi}\left[1+erf((\Delta t_i^0 - \Delta t_{h,l}^{safe})/\sigma_{h,l}\sqrt{2})\right]}, & t \leq \Delta t_i^0 \\ 0, & t > \Delta t_i^0, \end{cases} \tag{7}$$

where Δt_i^0 is the initial inter-arrival time of vehicle i recorded in the traffic count dataset. Again, (7) yields transformations that truncate and re-normalize the reference distribution, as graphically shown in Fig. 6b. In this case, Δt_i^0 becomes the upper bound to Δt_i^{safe}, ensuring that no vehicle enters the simulation with an inter-arrival time that is lower than its minimum safe time headway. Such a situation would in fact lead to sudden braking, and possibly to accidents.

Lane change bias and hysteresis threshold. In our highway scenarios, the default MOBIL settings result in a traffic that is highly skewed towards the left lane, which thus suffers from unrealistic congestion. We ran a comprehensive campaign to identify the combination of right (a_R) and left (a_L) lane change bias, and lane change hysteresis threshold factor (k) that grants quasi-stationary traffic over the different lanes. Such consistent ingress and egress per-lane properties were obtained for $a_R = 0.2$ m/s^2, $a_L = 0$ m/s^2, and $k = 0.3$. Interestingly, the lane change bias favor movements to the right in absence of a clear preference among lanes, which is in compliance with road regulation in Spain.

3.3. Synthetic mobility traces

The final synthetic traces are composed of one day-long trace describing road traffic over the four lanes of M30, and sixteen 30-minute traces of vehicular mobility along M40 and A6, for different day and hour combinations[5]. The

feature congestion periods. We assume that drivers on M40 and A6 have minimum safe time headway values comparable to those computed for M30, and reuse the same distributions.

[5] Available at `http://www.it.uc3m.es/madrid-traces`.

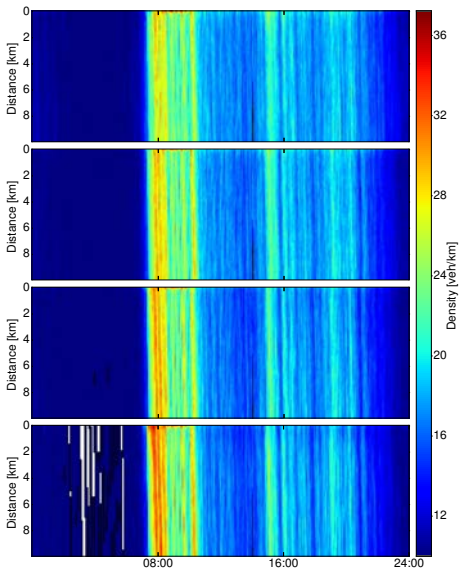


Figure 7: Vehicular density heatmap, day-long M30 trace. Plots refer to lanes from right to left (bottom to top). Figure best viewed in color.

traces record the position of each vehicle at every 500 ms, over a 10-Km road stretch[6].

As mentioned earlier, all traces are representative of quasi-stationary road traffic. This clearly emerges in Fig. 7, which shows heatmaps of the vehicular density on each lane of M30 over 24 hours: density variations at the beginning of the trace (i.e., at distance equal to 0 Km) reflect throughout the whole length of the road, up to the end of the simulated segment (i.e., at distance equal to 10 Km). The slight slope is normal, and due to the time required for vehicles to traverse the highway segment. White stripes in the bottom plot indicate occasional absence of traffic on the leftmost lane at night.

The unprecedented combination of source data granularity, temporal duration, and road heterogeneity makes these traces the current state-of-the-art for

[6]The road segment span is a configurable parameter in our simulator. We opted for a 10-km distance since it is a common choice in the literature that allows evaluating the performance of most networking solutions.

vehicular networking studies in highway environment. This is supported by the comparative analysis of our datasets with respect to synthetic mobility traces used in the networking literature, as discussed in Sec. 6.

Finally, we underscore that the original methodology presented in Sec. 3.2 can be used to calibrate any model of microscopic vehicular mobility. Thus, it is fully compatible with the models implemented by popular road traffic simulators used by the networking research community, such as SUMO [9] or VanetMobiSim [10].

4. Vehicular network model

We consider the mobility traces presented in Sec. 3, and analyze them from a vehicular networking perspective. Specifically, we are interested in investigating the connectivity properties of spontaneous vehicular networks that emerge from the mobility traces. The rationale for such an approach is that network connectivity is the base upon which solutions at all network layers are built. Thus, a connectivity study is, by its own nature, protocol-independent. Moreover, connectivity analyses have been shown to unveil the availability, stability and internal structure of the network – all of which are paramount notions to the sensible design of vehicular networking solutions [34].

As a preliminary step to our analysis, we present in this section the network model that we assume (Sec 4.1). We then leverage this model to formally define the connectivity metrics used in our study (Sec 4.2).

4.1. Instantaneous connectivity graph

Our analysis focuses on the instantaneous connectivity of spontaneous vehicular networks. Therefore, at each time instant t, we represent the network as an undirected graph $G(\mathbb{V}(t), \mathbb{E}(t))$, where $\mathbb{V}(t) = \{v_i(t)\}$ is a set of vertices [7] $v_i(t)$, each mapping to a vehicle i in the network at that time. $\mathbb{E}(t) = \{e_{ij}(t)\}$ is

[7] Or nodes – the two terms will be used interchangeably.

the set of edges $e_{ij}(t)$, connecting $v_i(t)$ and $v_j(t)$ if a direct V2V communication link exists, at time t, between vehicles i and j.

We adopt a unit disc model to represent the radio-frequency signal propagation. Hence, an edge $e_{ij}(t)$ exists if vehicles i and j are separated by a distance of at most R meters at time t, where R is the communication range. We employ this simple model due to the fact that deterministic (based on, e.g., ray tracing techniques) and stochastic (based on, e.g., statistical approaches) propagation models do not scale to the large mobile scenarios we consider, composed of tens of thousands instantaneous graphs, each including hundreds of vehicles. Instead, the unit disc model is computationally inexpensive, and fully captures the connectivity dynamics induced by vehicular mobility, which occur at timescales in the order of seconds.

In order to make our study as general as possible, we repeat all of our analyzes for several significant values of R. Despite physical layer standards for vehicle-to-vehicle Dedicated Short-Range Communication (DSRC) claiming up to 1-Km ranges [35], independent experimental studies demonstrated that acceptable packet delivery ratios are constrained to much lower distances [36–39]. Extensive experimental analyses in [37] show that a distance of 100 m allows around 80% of the packets to be correctly received in urban environments, when using common power levels (15-20 dBm) and robust modulations (3-Mbps BPSK and 6-Mbps QPSK). Under similar settings, $R = 50$ m is experimentally identified as the largest distance at which vehicle-to- vehicle communication attains packet delivery ratios close to one [36, 37]. Conversely, $R = 200$ m is the maximum distance granting a reception ratio above 0.5 [37]. The propagation conditions appear to be even worse in pure highway environments, where $R = 50$ m is found to be the threshold beyond which the packet delivery ratio drops, on average, below 50% [38]. This occurs even when transmissions are performed at 21 dBm, i.e., the maximum power allowed in Europe (where the tests were performed), and using the lowest coding rate with BPSK modulation, corresponding to a data rate of 6 Mbps with standardized 20-Mhz channel bandwidth. Finally, extensive field trials on 35 highways in the United States,

Germany, Austria, Italy, and Australia confirmed that reliable vehicle-to-vehicle communication is achieved, in the vast majority of cases, at distances ranging from 46 to 229 m [39]. In the light of all these results, in our analysis we will consider $R \in [50,200]$.

4.2. Connectivity metrics

We use the graph model to define the metrics of interest to our connectivity study. First of all, we denote the **number of nodes** in the graph (i.e., the number of vehicles in the road scenario) at time t as $\mathcal{N}(t) = \|\mathbb{V}(t)\|$.

We name a **component** $C_m(t) = G(\mathbb{V}_m(t), \mathbb{E}_m(t))$ a subgraph of $G(\mathbb{V}(t), \mathbb{E}(t))$, such that $\mathbb{V}_m(t)$ is a subset of $\mathbb{V}(t)$ including all and only the vertices mapping to vehicles that can communicate via direct or multi-hop V2V links at time t. Similarly, $\mathbb{E}_m(t) \subseteq \mathbb{E}(t)$ includes all edges mapping to communication links among vehicles whose corresponding vertices are in $\mathbb{V}_m(t)$. We denote as $\mathcal{S}_m(t) = \|\mathbb{V}_m(t)\|$ the **size of the component** $C_m(t)$.

By definition, components are disjoint, i.e., a vertex belongs to one and only one component at each time instant. We thus use $\mathbb{C}(t) = \{C_m(t)\}$ to refer to the **set of components** appearing in the network at time t, and $\mathcal{C}(t) = \|\mathbb{C}(t)\|$ to indicate the **number of components**. As a result, the **average size of components** appearing at time t is referred to as $S_{avg}(t) = \mathcal{N}(t)/\mathcal{C}(t)$.

We denote $C_{max}(t) = C_m(t)$, s.t. $m = \arg_n \max \mathcal{S}_n(t)$, as the **largest component** appearing in the network at time t. As $C_{max}(t) = G(\mathbb{V}_{max}(t), \mathbb{E}_{max}(t))$, we also use $S_{max}(t) = \|\mathbb{V}_{max}(t)\|$ to represent the **size of the largest component** at the same time instant.

With reference to the internal structure of a given component, we can identify, for each pair of vertices $v_i(t)$ and $v_j(t)$ belonging to a same component $C_m(t)$ at time t, a **shortest path** of length $p_{ij}(t)$, which corresponds to the sequence of vertices in $C_m(t)$ that connect vehicles i and j at minimum communication hop cost. We can thus define the **average shortest path** of the component $C_m(t)$ as $l_m(t) = \sum_{(i,j),i\neq j} p_{ij}(t)/(\mathcal{S}_m(t) \cdot (\mathcal{S}_m(t) - 1))$.

Finally, we name **vertex degree** the number of nodes directly connected

Table 2: Notation employed in the vehicular network connectivity analysis. All metrics refer to the instantaneous topology of the vehicular network.

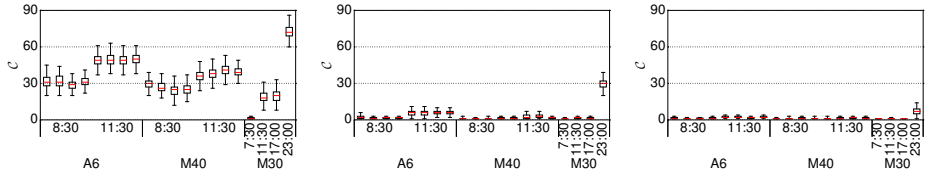
Parameter	Meaning
R	Vehicle-to-vehicle radio-frequency communication range
$\mathcal{N}$	Number of network nodes
$\mathcal{C}$	Number of network components
$\mathcal{S}_{avg}$	Average size of a (generic) component
$\mathcal{S}_{max}$	Size of the largest component
l	Average shortest path within a (generic) component
k	Degree of a (generic) node

to a given vertex $v_i(t)$ at time t, formally $k_i(t) = \|\{v_j(t)\ s.t.\ \exists\, e_{ij}(t)\}\|$. The degree of vertex $v_i(t)$ thus maps to the number of direct V2V communication neighbors of vehicle i.

For the sake of simplicity, we drop the time notation in the rest of the paper, and we refer to all metrics at a generic time instant. Similarly, we consider generic clusters or nodes, and drop the cluster and node indices. Then, $\mathcal{N}$ represents the number of vertices in the network, $\mathcal{C}$ the number of components, $\mathcal{S}_{avg}$ the average size of a component, and $\mathcal{S}_{max}$ the largest component size. Equivalently, l is the average shortest path of a component, and k is the node degree of a generic vertex. Tab. 2 summarizes the notation introduced above and used throughout Sec. 5 below.

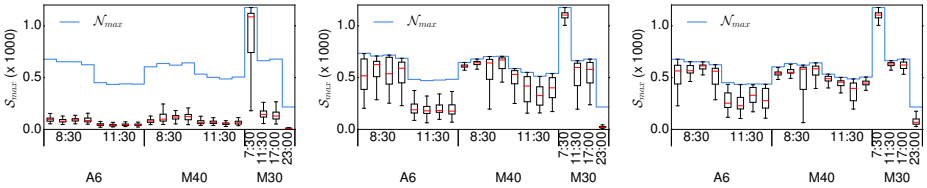
5. Vehicular network connectivity

Our study of the connectivity of vehicular networks considers a variety of highway scenarios (M30, M40, A6) and road traffic conditions (sparse overnight traffic, daytime free flow traffic, congested traffic during rush hours). It is organized by focuses. We will first address network-wide connectivity features (Sec. 5.1), and then study how they depend on the vehicular density and communication range (Sec. 5.2). The availability and stability of the network are then discussed (Sec. 5.3). Finally, we investigate the internal structure of the



(a) $R = 50$ m (b) $R = 100$ m. (c) $R = 200$ m.

Figure 8: The distribution of the number of components, $\mathcal{C}$, for different mobility traces, and under varying R values.



(a) $R = 50$ m (b) $R = 100$ m. (c) $R = 200$ m.

Figure 9: The distribution of the largest component size, $\mathcal{S}_{max}$, for different mobility traces, and varying R values.

highway vehicular network, so as to assess its navigability (Sec. 5.4). We summarize our discussion by providing networking insights (Sec. 5.5).

5.1. Network-wide connectivity

We start by studying the global connectivity properties of the network at each time instant. Thus, we focus on the distributions of the number of components $\mathcal{C}$ and of the size of the largest component $\mathcal{S}_{max}$. Indeed, $\mathcal{C}$ is a measure of how fragmented the network is, while $\mathcal{S}_{max}$ is the maximum number of nodes that can be reached via multi-hop communication at a given time instant. Therefore, the lower $\mathcal{C}$ and the larger $\mathcal{S}_{max}$, the better connected the vehicular network.

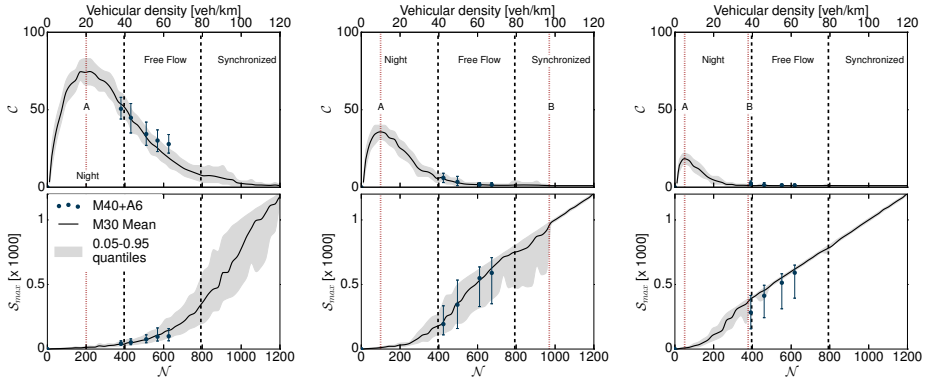
In Fig. 8 and Fig. 9, we present the distributions of $\mathcal{C}$ and $\mathcal{S}_{max}$, respectively. In both figures, each plot refers to a different value of the communication range R. Within every plot, each candlestick summarizes the distribution for one

(subset of) mobility trace, and is obtained by aggregating the $\mathcal{C}$ or $\mathcal{S}_{max}$ metrics computed in all instantaneous graphs observed at every 500 ms during a 30-minute timespan. In the M40 and A6 cases, 30 minutes match the whole duration of each trace, whereas in the M30 scenario we selected four representative 30-minute subsets of the day-long trace, i.e., at 7.30 a.m. (traffic peak time), 11.30 a.m. and 5 p.m. (free flow traffic comparable to that encountered in the M40 and A6 cases), and 11 p.m. (very sparse traffic).

Each box extends from the lower to the upper quartile of the distribution, with a line at the median. The whiskers pinpoint the minimum and maximum values. Also, the step function in the plots of Fig. 9 is the maximum value $\mathcal{N}_{max}$ of $\mathcal{N}$ observed throughout the whole 30-minute interval. It thus represents the upper bound, and an important benchmark value, to $\mathcal{S}_{max}$: the closer $\mathcal{S}_{max}$ to $\mathcal{N}_{max}$, the nearer the vehicular network to a fully connected single component.

Communication range. When observing the plots, the most striking result is the dramatic impact of the communication range R, whose value can dramatically improve or disrupt the network-wide connectivity. For $R = 50$ m, there are, on average, between 20 and 50 disconnected components throughout all datasets – excluding a few outlying situations that we will discuss later in detail. As R grows, however, the network fragmentation is reduced, and more nodes join the largest component: e.g., when $R = 100$ m, $\mathcal{C}$ typically drops below 10; when $R = 200$ m, almost all vehicles belong all the time to one single component. We conclude that *the communication range is the first and foremost parameter controlling the vehicular network connectivity*, as it can induce variations in $\mathcal{C}$ and $\mathcal{S}_{max}$ that are typically much larger than those imputable to the many and varied road traffic conditions encountered throughout the 20 datasets in Fig. 8 and Fig. 9.

Vehicular density. Still, some diversity is noted across the different traces, and, in a couple of cases, the impact of the road traffic scenario attains levels comparable to those induced by communication range variations. Although the relative performance of each 30-minute (sub-)trace tends to be consistent throughout all values of R, such diversity is perhaps best observed for $R = 50$



(a) $R = 50$ m (b) $R = 100$ m (c) $R = 200$ m

Figure 10: $\mathcal{C}$ and $\mathcal{S}_{max}$ versus the number of nodes $\mathcal{N}$ for the M30, M40 and A6 datasets, for different R values. Blue points represent the average $\mathcal{C}$ with the 5th and 95th percentile when M30, M40 and A6 have valid $\mathcal{N}$ values.

m, in Fig. 8a and Fig. 9a. There, both $\mathcal{C}$ and $\mathcal{S}_{max}$ show three major behaviors. The first is that of traces referring to free flow traffic conditions, i.e., all M40 and A6 traces, plus M30 traces at 11.30 a.m. and 5 p.m.: these present a comparable fragmentation, as the network is separated into 20-50 small components. The second and third behaviors correspond instead to outliers. On the one hand, the M30 trace at 7.30 a.m. yields a vehicular network that consists of a single connected component, as $\mathcal{C} \sim 1$ and $\mathcal{S}_{max} \sim \mathcal{N}_{max}$. On the other hand, the M30 trace at 11 p.m.results in extremely poor connectivity, with 70 or more components of a few nodes each. As these outlying behaviors correspond to rush hours and sparse overnight traffic, respectively, we speculate that *the vehicular density is the second key parameter that drives vehicle-to-vehicle network-wide connectivity.*

5.2. Laws of vehicular connectivity

We investigate whether some general law exists that can explain the fluctuations of vehicular network connectivity as a function of the different system parameters. To that end, we model the network-wide connectivity metrics, i.e.,

$\mathcal{C}$ and $\mathcal{S}_{max}$, as functions of the factors that appear to influence them the most in the analysis of Sec. 5.1, i.e., R and the vehicular density. We map the latter to the number of nodes $\mathcal{N}$, following a common practice in network science analyses [40].

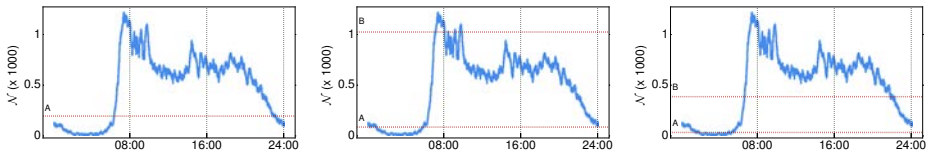
Three-phase connectivity in $\mathcal{N}$. Fig. 10 portrays the evolution of $\mathcal{C}$ and $\mathcal{S}_{max}$ versus $\mathcal{N}$. As the latter is a proxy for the vehicular density, we also report that measure, expressed in vehicles/km, on the top x axis. For the sake of clarity, at this time we limit our analysis to the mean behavior recorded in the day-long M30 trace, denoted by solid black lines in all plots of Fig. 10. We will introduce the other elements of the plots in due time.

The dynamics of both $\mathcal{C}$ and $\mathcal{S}_{max}$ are strongly dependent on $\mathcal{N}$. The largest component size, in the bottom plots, features a clear positive correlation with $\mathcal{N}$. The number of components, in the top plots, displays instead a skewed bell shape. Comparing the plots, *the instantaneous vehicular connectivity appears to be characterized by three phases, or behavioral regions, as a function of $\mathcal{N}$, under any R.*

I. Initially, for low $\mathcal{N}$, $\mathcal{S}_{max} \sim 1$ and $\mathcal{C}$ grows linearly with $\mathcal{N}$. This means that the network is very sparse, and increasing the number of vehicles $\mathcal{N}$ just means to introduce additional isolated nodes: as these nodes are not connected with each other, they become new components (of one node each).

II. Once a first critical $\mathcal{N}$ threshold is reached (denoted by the leftmost red dotted vertical line "A" in the plots), a second behavior ensues. Namely, $\mathcal{S}_{max}$ grows super-linearly with $\mathcal{N}$, and $\mathcal{C}$ decreases sub-linearly with $\mathcal{N}$. Beyond this first critical vehicular density, new cars are not isolated anymore, but tend to be connected to each other. Thus, they either join existing components or even bridge them into larger ones.

III. The third region is attained after a second $\mathcal{N}$ threshold (denoted by the rightmost red dotted vertical line "B" in the plots) is surpassed. There, $\mathcal{S}_{max} \sim \mathcal{N}$ and $\mathcal{C} \sim 1$, i.e., the vehicular network becomes fully connected



(a) $R = 50$ m (b) $R = 100$ m. (c) $R = 200$ m.

Figure 11: Number of nodes $\mathcal{N}$, for different values of R. Critical thresholds are marked with A and B.

into a single component whose size matches the number of vehicles on the highway segment. Additional vehicles necessarily end up in the giant component and increase its size.

The qualitative three-phase behavior above is invariant across different values of the communication range R. The impact of R is on the critical $\mathcal{N}$ thresholds that trigger phase changes: the "A" and "B" critical densities are shifted to the left (i.e., intervene at lower vehicular density) for larger values of R. This naturally induces better connectivity for higher values of R, for a fixed $\mathcal{N}$.

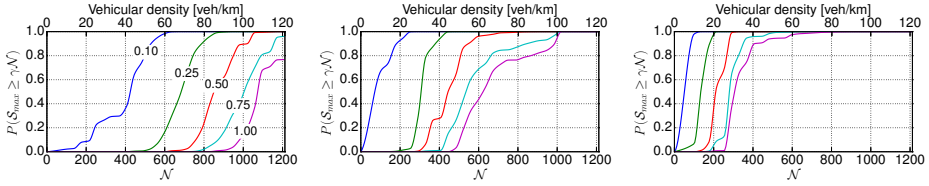
Some interesting considerations emerge from the mapping of the critical density thresholds above to the time series, shown in Fig. 11. We remark that when $R = 50$ m, the vehicular network never reaches the third phase. Indeed, it remains in the first phase at night, and in the second phase for the rest of the day, i.e., 7 a.m. to 11 p.m. The second phase also dominates when $R = 100$ m, as the network spends just a few hours in the first (2 a.m. to 6 a.m.) and third (7 a.m. to 8 a.m.) phases. The network behavior changes radically for $R = 200$ m, where the third phase spans over most of the day (7 a.m. to 10 p.m.), and the rest of the time is spent in the second phase.

These results let us comment that *attaining the third phase, i.e., persistent full connectivity, in highway vehicular networks cannot be taken for granted*, as it requires either elevate communication ranges, or significant traffic congestion conditions. In all cases, common values of highway V2V communication range, e.g., 50-100 m [38, 39], seldom allow reaching this phase.

Impact of other road traffic parameters. In all plots of Fig. 10, the light gray region around the mean shows how the 0.05-0.95 quantile range of the $\mathcal{C}$ and $\mathcal{S}_{max}$ metrics varies as a function of $\mathcal{N}$. We observe that such a range is fairly small throughout all plots, which means that the network-wide connectivity dynamics we discussed above are statistically consistent, i.e., yield a moderate variability. This is an important remark, since it implies that other parameters characterizing the road traffic do not have a significant impact on the vehicular network connectivity. In other words, factors such as the specific daytime or day of the week, the number of lanes of the highway, or the speed limits are only responsible for minor variability around the connectivity dynamics dictated by R and $\mathcal{N}$. Another way to read the same conclusion is that *considering one single road traffic parameter, i.e., $\mathcal{N}$, is enough to properly characterize the vehicular connectivity in all situations encountered during a typical working day.*

On a related point, the precise conditions of road traffic do not appear to be directly related to the connectivity of the vehicular network. In all plots of Fig. 10, black vertical dashed lines separate the different regions (in the $\mathcal{N}$ space) characterized by diverse traffic conditions. Specifically, these thresholds roughly identify $\mathcal{N}$ ranges corresponding to sparse overnight traffic (left region), typical daytime free flow traffic (middle region), and synchronized congested traffic (right region). By confronting these $\mathcal{N}$ thresholds with those that denote connectivity phase changes ("A" and "B"), we do not observe any significant overlap. Thus, *no direct correspondence can be established between the sole road traffic state and the vehicular network connectivity.*

Comparison across different traces. The plots in Fig. 10 also include the $\mathcal{C}$ and $\mathcal{S}_{max}$ recorded for the sixteen M40 and A6 traces. These are represented as filled circles in the plots, where dots represent the mean values recorded for different values of $\mathcal{N}$, and are obtained by aggregating all traces showing a similar vehicular density. Error-bars represent the 0.05 and 0.95 quantiles. These dots do not cover the whole $\mathcal{N}$ range, since the M40 and A6 traces only capture 30 minutes of traffic, mostly in free flow conditions, and thus only provide a partial view of the connectivity dynamics. Still, the majority of M40



(a) $R = 50$ m (b) $R = 100$ m. (c) $R = 200$ m.

Figure 12: Vehicular network average γ-availability versus the number of nodes $\mathcal{N}$, for different values of R.

and A6 fall very close to the mean behavior observed in the M30 case, and their 0.05-0.95 quantile ranges tend to correspond to those of M30. Therefore, *we conclude that the same three-phase connectivity dynamics in $\mathcal{N}$ holds for all of the highway scenarios we consider.* Moreover, the impact of R on the network connectivity is equivalent in all such scenarios.

5.3. Availability and stability

As prominent connectivity factors, R and $\mathcal{N}$ control two key network properties, i.e., availability and stability. We now quantify these very features, and investigate how they depend on the communication range and vehicular density.

Network availability. The *availability* maps to the probability that vehicle-to-vehicle communications build a network that can be actually exploited for basic services such as multi-hop cooperative awareness, content dissemination, or data aggregation. Formally, we say that the system has a level of availability γ if a component of size at least equal to $\gamma\mathcal{N}$ is present in the network.

Fig. 12 portrays the level of availability one can expect from the vehicular networks in our reference mobility scenarios, as a function of the chief factors R and $\mathcal{N}$. The three plots refer to different communication ranges, and each plot illustrates the average probability that a level of availability γ is attained at a given vehicular density. For instance, the leftmost curve in Fig. 12a shows that, for $R = 50$ m and $\mathcal{N} = 400$, the network is 0.1-available (i.e., there exists a component that includes 10% of the nodes or more) with a probability of 30%.

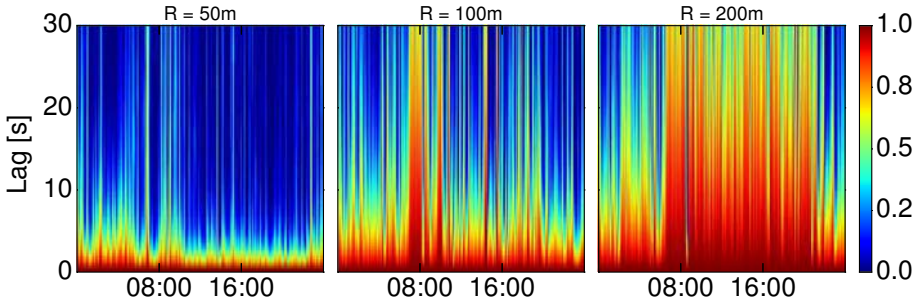
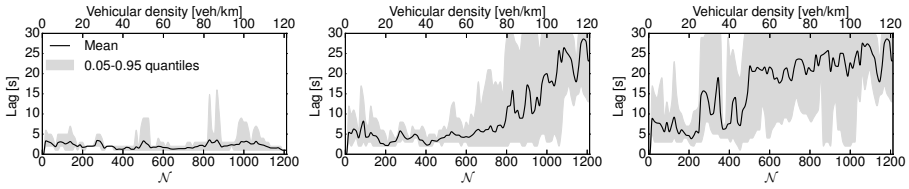


Figure 13: Correlogram heatmap for $\mathcal{S}_{max}$. Figure best viewed in color.

The same probability grows to 80% by considering a slightly denser network with $\mathcal{N} = 500$. These numbers imply that the network is unavailable 70% of the time in the first case, and 20% in the second.

In addition, a comparative analysis of the plots in Fig. 12 yields the following remarks. First, the results confirm the dramatic impact of R. If vehicles that are 200 m apart can communicate, 1.0-availability (i.e., full network connectivity) is around as probable as 0.1-availability with $R = 50$ m, and 0.25-availability with $R = 100$ m. Conversely, the network is never 1.0-available with a probability higher than 80% if $R = 50$ m. Second, most curves are quite steep as a function of $\mathcal{N}$, indicating that *percolation thresholds in $\mathcal{N}$ often characterize the network availability*: if the system operates around the threshold, small variations of vehicular density (in the order of a a few vehicles/km) can drastically change the probability that the network is γ-available, for a given γ. A notable, persistent exception to the percolation behavior is visible in the longer tail of high-availability curves (i.e., $\gamma \geq 0.75$): this implies that *ensuring with certainty that the vehicular network is highly available demands a significant additional effort, for any R.*

Network stability. The notion of *stability* concerns the amount of time for which the vehicular network maintains the same connectivity properties. We investigate stability by focusing on the largest network component, as it represents the portion of the network that can best support practical services based on multi-hop vehicle-to-vehicle communication. More precisely, we map the



(a) $R = 50$ m (b) $R = 100$ m. (c) $R = 200$ m.

Figure 14: Vehicular network average temporal stability versus the number of nodes $\mathcal{N}$, for different values of R.

stability of such a component to the temporal autocorrelation of its size, $\mathcal{S}_{max}$. The rationale is that if $\mathcal{S}_{max}$ is strongly autocorrelated over long time periods, then we can expect that the most significant portion of the network conserves stable topological properties.

An intuitive explanation of the analysis we carry out is provided in Fig. 13. There, we show heatmaps of the correlograms of time series of $\mathcal{S}_{max}$. To derive the plots, time series are divided into 10-minute windows, and, for each window, the temporal autocorrelation at different lags is calculated. Values in the heatmap hence represent the autocorrelation value for each window (along the horizontal axis) and lag (along the vertical axis) pair. The heatmaps provide complete information on the level of stability of the vehicular network over time. In particular, it proves how stability can be highly time-varying: as an example, we can remark that, for $R = 100$ m, a strong $\mathcal{S}_{max}$ autocorrelation peak, denoting a network much more stable than usual, appears just before 8 a.m.

The heatmap representation allows introducing a more formal definition of stability: we say the vehicular network to be stable if the size of its largest component yields a temporal autocorrelation higher than 0.7 [41]. Fixing this autocorrelation threshold allows pinpointing a precise lag time at each instant in time and for each R, i.e., for each point in the $(R, \mathcal{N})$ space. The result is portrayed in Fig. 14, which provides a neat representation of the stability one can expect from the vehicular networks in the highway scenarios we consider.

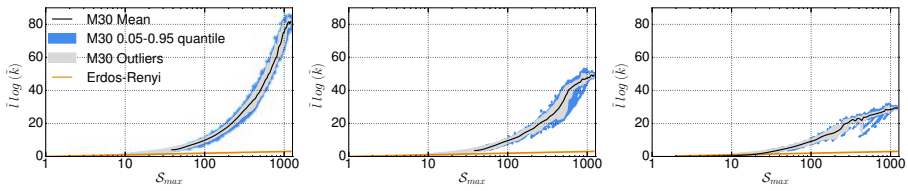
We remark that, when $R = 50$ m, in Fig. 14a, the network is very unstable, as the low threshold lag implies that the largest component undergoes significant size variations every 2-3 seconds on average. This behavior is independent of $\mathcal{N}$. As the communication range grows to 100 m, the stability only slightly improves over the $R = 50$ m case, raising to 3-5 s. This time, however, $\mathcal{N}$ starts having some impact, even if only at rather high vehicular densities around 80 vehicles/km that already denote road traffic at the boundary between free flow and congestion. In such traffic conditions, increasing $\mathcal{N}$ favors stability, and large components that persist over intervals of 10-25 s can be observed. For $R = 200$ m, as soon as $\mathcal{N}$ grows beyond sparse overnight traffic, at around 50 vehicles/km, a more stable behavior emerges, with large components that typically endure 20-25 s. By looking at the absolute values of the network stability that we identify, we note that, under all system parametrizations, *the stability of the vehicular network is in the order of a few tens of seconds at most.*

5.4. Internal structure

Having assessed that the spontaneous vehicular network yields poor availability and stability, we study its level of *navigability*, i.e., its predisposition to support multi-hop communication [40]. To that end, we analyze internal structural properties of the largest network component, where multi-hop V2V transfers can actually occur. Also relevant to the network navigability is the duration of V2V contacts: indeed, it determines the amount of time during which two vehicles can communicate, and thus the usability of contact opportunities.

Small-world property. A network is said to be a *small-world* if the distance among its vertices stays small as the network size grows. More rigorously, a typical example of small-world network is the Erdös-Rényi random graph, whose average shortest path length, i.e., $\tilde{l}$, scales logarithmically in the number of vertices $\mathcal{N}$. In fact, in a Erdös-Rényi network, $\tilde{l} \sim log(\mathcal{N})/log(\tilde{k})$, where $\tilde{k} = \sum_{i=1}^{\mathcal{N}} k_i/\mathcal{N}$ is the average vertex degree.

We compare the instantaneous vehicular networks in our reference highway scenarios to the Erdös-Rényi random graph, in Fig. 15. The plots show the



(a) $R = 50$ m (b) $R = 100$ m (c) $R = 200$ m

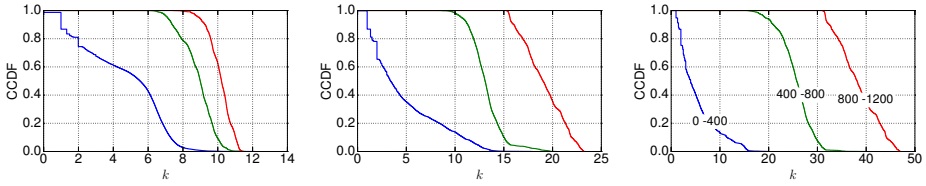
Figure 15: Scaling properties of the average shortest path $\tilde{l}$ measured in the largest component C_{max} of the vehicular network, for different values of R. The result is compared to an Erdös-Rényi random graph of equivalent size $\mathcal{S}_{max}$.

average shortest path times the logarithm of the average node degree, versus $\mathcal{S}_{max}$, along log-linear axes. Therefore, the Erdös-Rényi model portrays as a line of unit slope.

By observing how the same measure scales with the component size in our case study, we conclude that *instantaneous vehicular networks in the highway scenarios we consider are not small-world*: the empirical curves lay well above the logarithmic scaling of a typical small-world graph. The effect of R is again evident, as the average multi-hop distance among vehicles is reduced threefold for significant component sizes, i.e., $\mathcal{S}_{max} \geq 100$, when R grows from 50 m to 200 m. However, the super-logarithmic trend of the mean for any R implies that adding nodes to the network pushes the largest component farther away from a small-world behavior, making it harder to navigate. From a networking perspective, this implies that *increasing the vehicular density leads to larger components where multi-hop communication among node pairs becomes much more challenging and delay-prone.*

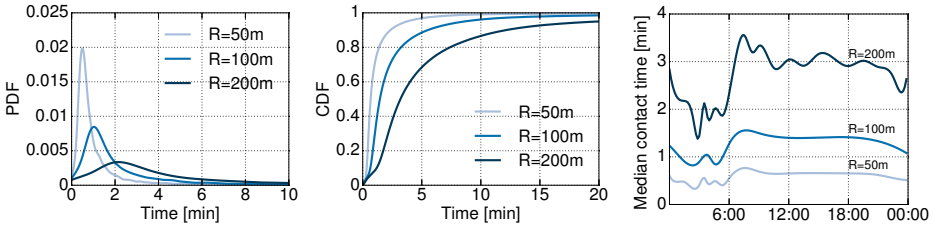
Scale-free property. A *scale-free* network retains the same functional form of its vertex degree distribution at all scales. In other words, the probability distribution of the degree obeys a power law $P(k) \sim k^{-\alpha}$, with the exponent α typically lying between 2 and 3 [40]. This property is known to result in an easily navigable network, with a backbone of high-degree hub nodes that interconnect that majority of low-degree leaf nodes.

This is, however, not the case in the vehicular networks we consider, as shown



(a) $R = 50$ m (b) $R = 100$ m. (c) $R = 200$ m.

Figure 16: CCDFs of the node degree k, for different values of R, separated according to $\mathcal{S}_{max}$ ranges.



(a) PDF (b) CDF (c) Time series

Figure 17: Vehicle-to-vehicle contact duration on M30, under different values of the communication range R. (a) Time-aggregate PDF. (b) Time-aggregate CDF. (c) Time series, over 24 hours.

in Fig. 16. There, Complementary Cumulative Distribution Functions (CCDFs) of the vertex degree, separated for different $\mathcal{S}_{max}$ ranges (0-400, 400-800, and 800-1200, respectively) are plotted for each value of R. It is evident that the distributions are not power laws, and thus *the highway vehicular networks we consider in our study are not scale-free*. Instead, they are characterized by a remarkably small range over the node degree k: *vehicles traveling on highways have one-hop communication neighborhoods of rather constant size over time*, with a variability in the order of a few units at most. We remark that this is very different from what is observed in urban scenarios [34].

Contact duration distributions. As anticipated, the duration of communication links established by vehicles is an important metric that characterizes how easy (or difficult) it is to exploit the data transfer opportunities created

by vehicular mobility. Indeed, experimental works showed that, in presence of short-lived V2V contacts, even simple signalization procedures induce a significant overhead and waste precious communication time [36, 37].

Fig. 17a and Fig. 17b show the PDF and CDF of the V2V contact duration. Each curve refers to a different value of the communication range R. We observe that also in our large-scale scenarios, contacts typically last from a few tens of seconds to a few minutes, and are thus quite short, as one could expect in a highly dynamic environment such as the vehicular one.

The communication range has a significant impact on the contact duration, which can be expected again. In order to highlight the dramatic effect of R, we can, e.g., underscore that 10% of contacts last more than two minutes when $R = 50\,\text{m}$, while the same percentage grows to 90% when $R = 200\,\text{m}$. Similarly, the median contact duration grows from 30 seconds to 3 minutes when R increases from 50 m to 200 m.

These results are aggregated over a full day of measurements on M30. An interesting question is then if the different road traffic conditions we observed to occur on the highway throughout the day affect the duration of V2V contacts. Fig. 17c portrays time series of the median contact duration over 24 hours, under different communication ranges. The results highlight once more the critical importance of R, but also the minor variability of contact durations throughout the day. Except for slightly shorter contacts at night, between midnight and 6 a.m., and slightly longer contacts during the morning traffic peak, contacts tend to have the same duration. The reaction to different road traffic conditions is intuitive, since the higher (respectively, lower) speed recorded at night (respectively, during congestion) leads to shorter (respectively, longer) lived V2V communication links. However, even the maximum variability due to different traffic conditions is not dramatic, especially when compared to that induced by different values of R.

Overall, the results above underscore that vehicular network components are not small-world nor scale-free, and that they are in fact the result of fairly short-lived V2V contacts, in the order of a few tens of seconds at most. All

Table 3: Impact of system parameters on vehicular connectivity properties, and implications for the design of networking solutions.

	Connectivity properties	Networking insights
Network-wide	**Three-phase network connectivity model** in Fig. 10. The model is a function of $\mathcal{N}$ (determines the overall shape) and R (introduces a scaling factor).	• $R \geq 200$ m induces a fully connected network, but connectivity rapidly deteriorates for lower R, until complete fragmentation is reached for $R \leq 50$ m. • Vehicular connectivity is easily modeled and predicted given $\mathcal{N}, R$. • Employ store-carry-and-forward for delay-tolerant data dissemination. • Resort to cellular networks for long-range or QoS data transfers.
Components	• **Large connected components are typically unavailable**, and 100% availability is extremely difficult to obtain under any system settings. • Large connected components remain **stable for a few tens of seconds** at most.	• Reference chart of component availability as a function of $\mathcal{N}, R$ in Fig. 12. • Network- and transport-layer protocols need to be highly reactive to topology changes.
Within components	• **No small-world** property observed in highway vehicular networks. • **No scale-free** property observed in highway vehicular networks. • **Short-lived links** established among vehicles traveling on highways. • **Stable one-hop neighborhood** as far as size is concerned.	• Use effective (geographical) routing. • Need for extremely rapid V2V link establishment (across the whole protocol stack). • MAC-layer solution design: – Data rate adaptation must operate on fast (order of 100 ms) dynamics, while channel contention and power control can be less reactive. – Different MAC-layer algorithm are needed for highway and urban environments. – Reference chart of expected contention as a function of $\mathcal{N}, R$ in Fig. 16.

these aspects together let us conclude that *vehicular networks in highway environments have poor navigability properties.* We remark that, in this regard, highway vehicular networks are comparable to urban ones [34].

5.5. Discussion and networking insights

The results presented in Sec. 5.1–5.4 have significant implications in terms of viability of communication paradigms and design of network architectures and protocols in vehicular environments. Below, we summarize our findings and discuss how they are useful to the networking community. Tab. 3 provides a useful reference in that sense.

The limitations of the network connectivity may be even more severe than expected. The positive impact of factors such as R and the vehicular density on the instantaneous connectivity of vehicles is a quite intuitive result

that has already been observed in the past, as also indicated in Sec. 6. However, in addition to confirming the findings of previous works, our study allows unveiling for the first time the exact proportions of the phenomenon on a fairly large set of realistic highway mobility traces. The results we obtain indicate that *a communication range above 200 m guarantees a well-connected network independently of the traffic conditions, but reducing that value causes the topology to break apart dramatically fast.*

This is a troubling observation at the light of experimental studies that found 50 m to be a credible value of R in highway scenarios [38, 39]. With such a communication range, the network is normally so fragmented that it is barely exploitable, and traffic jams represent the only hope for V2V connectivity. On the one hand, this lets us *advocate in favor of store-carry-and-forward approaches to data dissemination in spontaneous highway vehicular networks.* On the other hand, a more controversial conclusion is that, given the coverage of the diverse radio interfaces envisioned to be embedded in cars, *vehicle-to-vehicle communication may just be unfit to long-range (e.g., order of km) delay-bounded (e.g., order of seconds) transfers in highway environments, and, in such cases, vehicles may have to resort to cellular transfers for reliable and time-bounded data delivery.* In other words, the vehicular network may not support some services it is envisioned to enable, such as those based on the decentralized floating car data paradigm supported by ETSI [42].

Network-wide vehicular connectivity is easily predictable. We unveil the three-phase relationship that drives the network-wide instantaneous connectivity of a spontaneous highway vehicular network. This relationship captures well the full diversity in connectivity dynamics, and relies on two factors only: (i) the communication range, R, and (ii) the vehicular density, $\mathcal{N}$. All other settings have small impact on the network topology, and one can safely neglect information on the daytime, day of the week, number of lanes, and speed limits when estimating the level of connectivity of the network. Similarly, the fact that we find consistent dynamics throughout a variety of highway scenarios (M30, M40, A6) and road traffic conditions (sparse overnight traffic, daytime freeflow

traffic, congested traffic during rush hours) is a promising result with respect to the generality of our study. Indeed, it suggests that our conclusions may hold for a vast range of highways, different than those modeled by our road traffic datasets.

Overall, these considerations imply that *network-wide vehicular connectivity is especially simple to model and anticipate, as the knowledge of two parameters is sufficient to comprehensively describe the system.*

Vehicular multi-hop clusters are not stable. The communication range R has a paramount importance to both the availability and stability of connected components in the vehicular network. Indeed, a slightly larger R makes such a well connected network emerge much more frequently and sustain for a longer timespan. Our evaluation suggests that *when the communication range shifts from 50 to 200 m, the network becomes roughly 10 times more available and stable.*

Still, the stability of the vehicular network never exceeds a few tens of seconds, which imposes *strict requirements on protocols operating at the network layer and above, in terms of reactivity to very frequent topology changes.*

Finally, our analysis includes figures that pinpoint the level of availability of spontaneous highway vehicular networks in the $(R, \mathcal{N})$ space. In this sense, *Fig. 12 represents a useful reference chart for networking practitioners to understand the network availability they can expect, given their specific R and $\mathcal{N}$ settings.*

MAC-layer requirements are heterogeneous. Unlike multi-hop clusters, we observed one-hop neighborhoods to be relatively stable: at least the size of the neighborhood of a vehicle tends to remain the same for fairly long time periods. This means that, from a MAC-layer protocol perspective, *the requirements, in terms of reactivity, of wireless channel contention and power control algorithms are not especially stringent as long as vehicles stay on highways.*

Since this result is very different from what happens in urban scenarios [34], *diverse, dedicated MAC solutions shall be adopted for highway and urban environments, for optimal operation.* In the highway case, *Fig. 16 can be leveraged as*

a reference chart to estimate MAC-layer channel contention and power control settings, based on the current operational point in the $(R, \mathcal{N})$ *space.*

However, we also found that pairwise links in the network are short-lived no matter the traffic conditions. This corroborates the results obtained in small-scale field tests [36, 37], and confirms that *MAC- and network-layer protocols have to rapidly establish V2V links, so that the time available for data transfer is maximized*. The latter constraint also applies to MAC-layer solutions, stressing how *vehicular networks require effective and highly adaptive data rate adaptation algorithms.*

The limited duration of V2V links adds to the fact that the vehicular network is not small-world nor scale-free: all these undesirable features determine the poor navigability of the network. From this viewpoint, and despite their simpler quasi-unidimensional road layout, highway vehicular networks resemble urban ones [34]: thus, similar considerations apply, i.e., *effective geographical routing techniques are highly recommended to move data throughout the intrinsically complex vehicular network topology.*

6. Related Work

Our work relates to two main research directions in vehicular networking, i.e., mobility modeling and connectivity analysis. Below, we separately discuss the relevant literature, and how our study compares to it.

6.1. Vehicular mobility modeling

The impact of realistic mobility modeling in the simulation of communication protocols tailored for vehicular networks has been emphasized in many works [6, 14, 16, 17]. As a result, in the last decade, the research community has devoted significant effort to the quest for ever-increasing realism of road traffic traces used in network simulators. A first approach consists in directly recording real-world mobility traces, by logging the position of vehicles during their movements. Unfortunately, these traces are currently limited to subsets

of the overall traffic, i.e., fleets of specific vehicles such as buses [8] or taxis [7], which prevents the analysis of full-fledged vehicular networks; moreover, none of such datasets is specific to the highway environment we target.

Other works have focused on the generation of synthetic vehicular traces by feeding real-world road topologies of different cities to microscopic traffic simulators such as SUMO [9] or VanetMobiSim [10]. In order to characterize the number, origin, destination and time of trips, these works usually make use of macroscopic data (e.g., origin-destination matrices) collected from user surveys [13, 14] or from roadside detectors [15]. However, all the works above deal with synthetic traces of road traffic in cities like Zurich [13], Cologne [14] or Luxembourg [15]. Yet, the dynamics of traffic over urban regions are not comparable to those of highways: the former are characterized by vehicles traveling at low or medium speed, and often crossing intersections regulated by traffic lights or roundabouts; the latter feature instead high speeds and frequent overtaking. Moreover, none of the aforementioned works considers fine-tuning of microscopic mobility models, as we do in this study.

The work in [17] is closer to our approach, as it uses two empirical datasets are used to generate synthetic highway mobility traces. The first dataset was collected on the I-80 highway near Berkeley, CA, USA, using dual-loop detectors that log information on individual vehicles; the second dataset contains 20-second aggregated traffic on the Gardiner expressway, near Toronto, Canada, recorded using metal detectors. The authors assume vehicle inter-spacing and car speed to be exponential and Gaussian random variables, respectively, and use the empirical data to derive the distribution parameters. Then, a mobility generator implementing these probabilistic models is used to create synthetic traces of road traffic. Our study improves that in [17] from several viewpoints: (i) the traffic-count datasets we employ are more detailed and heterogeneous, and do not accommodate the exponential inter-arrival assumption, as detailed in Sec 2.3. (ii) we use validated microscopic car-following and lane-changing models to describe the behavior of drivers, instead of simple stochastic representations; (iii) the synthetic mobility traces we generated are publicly available.

Table 4: Highway road traffic datasets in the vehicular networking literature.

Study	Macroscopic features			Microscopic features		Measurements	Availability
	Stationarity	Road heterogeneity	Traffic heterogeneity	Speed adjustment	Overtaking		
[22]	perfect	no	no	no	no	no	–
[23]	perfect	no	no	no	no	no	–
[24]	perfect	no	3×2 hours	no	no	high-detail	no
[25]	perfect	no	no	no	no	no	–
[26]	perfect	no	no	no	no	no	–
[43]	perfect	no	no	no	no	no	–
[44]	perfect	no	no	no	no	no	–
[45]	perfect	no	no	no	no	no	–
[46]	perfect	no	no	no	no	no	–
[47]	perfect	no	no	no	no	no	–
[17]	perfect	two US highways	48 hours	no	no	high-detail	no
[27]	quasi	no	no	no	Nagel-Schreckenberg	no	no
[18]	non	two US highways	14×30 minutes	Krauss	Krajzewicz	low-detail	yes
Ours	quasi	three Spanish highways	1×24 hours 16×30 minutes	IDM	MOBIL	high-detail	yes

Highway scenarios are also considered in [18], where empirical aggregated data from the Freeway Performance Measurement System (PeMS) is fed to the SUMO simulator to generate synthetic highway traces. More precisely, the real-world data, from road sensors on the I5 and I880 highways, CA, USA, is used to determine the assignment of the vehicular traffic flow and the average speed values over the road. However, the traffic count dataset features a coarse time granularity, with a sampling interval of the flow and speed from 30 seconds to 5 minutes.

Although those in [17, 18] are the only previous works that employ real-world traffic count data, other attempts at modeling highway traffic have been also made. We summarize in Tab. 4 the features of the mobility traces considered in a representative set of aforementioned works that study vehicular networks in highway environments. In the table, columns are read as follows.

- *Stationarity* refers to whether the considered vehicular mobility is quasi-stationary (i.e., macroscopic traffic conditions are comparable in between the in-flow and out-flow boundaries of the simulated road segments, but microscopic dynamics are modeled), or in perfect stationarity (i.e., not only macroscopic traffic conditions are uniform, but vehicles all travel

at constant speed and maintain a constant inter-vehicle distance, e.g., distributed according to some random variable). Non-stationarity refers to the presence of in- and out-ramps that induce variations in the vehicular density along the simulated road segment.

- *Macroscopic heterogeneity* is further expressed in terms of whether different highways and traffic conditions (i.e., time periods featuring diverse traffic levels) are considered.

- *Speed adjustment* and *overtaking* are at the base of microscopic dynamics of road traffic. The former allows vehicles to accelerate and decelerate depending on the surrounding traffic, while the latter enables lane changes towards the left and right lanes.

- *Measurements* can be at the origin of the mobility traces. If so, we tell apart high-detail measurement data, which records information on each vehicle separately, and low-detail data, which provides aggregate information at some periodicity (typically in the order of tens of seconds).

- *Availability* refers to whether the datasets are publicly available.

A vast majority of works consider road traffic in perfect stationarity, thus neglecting microscopic-level dynamics determined by individual drivers' behavior. In fact, most of such works propose analytical models of highway traffic, which build on simplifying assumptions that make problems analytically tractable. Common assumptions include randomly distributed speeds and inter-arrivals of vehicles, which are then kept constant: thus most of the works dealing with perfect stationary conditions also consider one single road type, and no microscopic-level models of acceleration/deceleration or lane changing. A few works dealing with perfect-stationary mobility build on measurement data, and account for heterogeneous real-world traffic conditions, at different time periods. However, the relevant datasets do not consider detailed microscopic modeling, nor they are publicly available.

Only a pair of previous works in Tab. 4 consider quasi- or non- stationarity, i.e., account for the microscopic dynamics of highway traffic. However, the work in [27] does not build on real- world data, but on assumptions about stochastic features of traffic. The work in [18] is that closest to ours, however it employs coarse-grained measurement data that does not allow reproducing the arrivals and velocity of vehicles with the same level of detail as in our dataset. In addition, the vehicular mobility traces in [18] are representative of 30-minute time intervals, whereas our M30 dataset covers one full day and thus enables a larger variety of networking studies (e.g., those targeting scalability, adaptability and reactiveness of network solutions to temporal variations of road traffic conditions).

In the light of the considerations above, we summarize the advantages of our proposed methodology for the generation of vehicular mobility (detailed in Sec. 4) as follows.

- With respect to other attempts at generating synthetic mobility traces in quasi- or non-stationary conditions (i.e., through vehicular mobility simulators that capture microscopic dynamics), ours is the first work that employs fine-grained traffic counts (i.e., containing per-vehicle statistics) collected through real-world measurements. Accounting for the actual inter-arrivals yields a higher accuracy than considering deterministic or random inter-arrivals, derived from measurements of road traffic flow with order-of-minute precision. We underscore that integrating such fine-grained traffic counts in a microscopic mobility generator is not a trivial task, and requires an original, dedicated parametrization as that presented in Sec. 3.2 of the main document.

- When considering perfect-stationary mobility (employed in semi-analytical or analytical models) the works in [17, 24] are the only using fine-grained traffic counts comparable to those we employ. However, synthetic traces are, by their own nature, more accurate than measurement-based semi-analytical or analytical models. Specifically, synthetic traces such as those

we generate are based on validated representations of drivers' acceleration, deceleration, and lane change behaviors (see Sec. 3.1 of the main document). They thus convey a richness of microscopic dynamics in vehicular movement (e.g., different drivers' target speeds and safe time headway, left- and right-lane movements, overtakes, etc.) that mathematical representations of highway traffic proposed in the networking literature (based on, e.g., constant speed, fixed vehicle inter-distance, no lane changes, etc.) cannot capture.

We also point out that the methodology proposed in this paper advances that appeared in an earlier version [48]. Specifically, the conference version of the work only considered the short 30-minute traces, while we base the analysis in this manuscript on the 24-hour M30 dataset, Moreover, the calibration of the microscopic mobility models in [48] operated on a per-lane basis in the conference version – an appoach that could not accommodate the more demanding road traffic conditions present in the new traffic dataset. Thus, the calibration proposed in this paper is different for each vehicle, and much more flexible.

6.2. Vehicular network connectivity

As far as vehicular network connectivity studies are concerned, some seminal works have considered urban areas [34, 49]. However, their findings do not necessarily apply to the highway scenarios we are interested in, due to the significant differences between urban and highway road traffic. Concerning the latter environment, a large number of studies have addressed the problem from an analytical perspective [22–24], characterizing features such as the mean component size [25], the probability of attaining a single connected component [26], or the impact of a dedicated roadside infrastructure [43].

Far fewer analyses have instead employed realistic traces to investigate the instantaneous connectivity of highway vehicular networks. Pioneering results on the connectivity of free flow highway traffic are provided in [27]: the authors use synthetic data generated by a simple microscopic simulator to prove that higher vehicular densities help connectivity. Subsequent studies confirmed this

conclusion, observing that the communication range is another primary factor affecting the network connectivity [23, 25]. However, these works are based on less detailed mobility traces, and only provide a basic assessment of the structural properties of the network topology.

More recently, the focus has shifted towards the internal structure of highway vehicular networks. In [18], the authors characterize distributions of the centrality, clustering coefficient, and vertex degree in the Alameda County road traffic traces presented above. In [28], the aforementioned I-80 mobility trace is leveraged to investigate the small-word and scale-free properties of vehicular networks. Our study confirms the findings of these statistical analyses on more detailed and comprehensive mobility traces. In addition, we take a step forward in the topological analysis, and unveil previously unknown properties, such as the invariant three-phase dependence of the connectivity on the network size, or the actual availability and stability of highway vehicular networks. These findings are new even with respect to those in the earlier version of the work in [48].

7. Conclusions and open issues

In this paper, we employed fine-grained road traffic counts collected on real-world highways in proximity of Madrid, Spain, to generate synthetic traces of vehicular mobility along those road segments. An original approach to the parameterization of well-known microscopic vehicular mobility models allowed us to obtain realistic descriptions of quasi-stationary unidirectional traffic in heterogeneous conditions, including different highways, weekdays and measurement hours. These traces are publicly available and, to the best of our knowledge, represent the current state of the art in highway traffic datasets for networking studies.

We carried out a comprehensive topological analysis on the mobility traces, confirming that: (i) the communication range and the vehicular density are the factors that primarily control the connectivity of highway vehicular networks;

(ii) vehicular networks are not small-world or scale-free in nature. In addition, we unveiled the three-phase dependence of connectivity on network size, and its potential general validity across highway scenarios. We also quantified for the first time the actual availability and stability of the system.

Our study also has limitations that open the way for future research activities. First, our analysis is based on data collected on three highways, and all results are thus specific to those scenarios. Some promising results on the potential generality of our conclusions come from the invariance of the connectivity dynamics in all such different datasets, in Sec. 5. Still, a much larger set of measurements is required to generalize our findings.

Second, our connectivity analysis builds on a unit-disc representation of the radio signal propagation. Considering some model of signal fading would add the rapid variability induced by radio signal fluctuations on top of the mobility-dependent dynamics we observe in our study. Ultimately, that would lead to an even finer-grained description of the vehicular connectivity.

Third, in this paper we only investigate the instantaneous connectivity of highway vehicular networks. The temporal analysis of vehicular connectivity, aimed at the characterization of delay-tolerant network properties, would require a completely different approach (based, e.g., on time-expanded representations), and it is an interesting extension of our work.

References

[1] Wireless LAN medium access control (MAC) and physical layer (PHY) specifications. Institute of Electrical and Electronics Engineers (IEEE). doi: 10.1109/ieeestd.2012.6178212. URL http://dx.doi.org/10.1109/IEEESTD.2012.6178212.

[2] IEEE 1609 - family of standards for wireless access in vehicular environments (WAVE). U.S. Department of Transportation, 2006.

[3] Communications access for land mobiles (ITS-CALM-M5). OSI 21215 Standard.

[4] Intelligent transportation systems (ITS). ETSI Standard EN 302 665, .

[5] Obama backs highway fund fix, touts 'talking' cars. http://nyti.ms/1nssF8J. Accessed: 2015-08-05.

[6] Marco Fiore and Jérôme Härri. The networking shape of vehicular mobility. In *Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing - MobiHoc '08*. Association for Computing Machinery (ACM), 2008. doi: 10.1145/1374618.1374654. URL http://dx.doi.org/10.1145/1374618.1374654.

[7] Hong-Yu Huang, Pei-En Luo, Minglu Li, Da Li, Xu Li, Wei Shu, and Min-You Wu. Performance evaluation of SUVnet with real-time traffic data. *IEEE Trans. Veh. Technol.*, 56(6):3381–3396, nov 2007. doi: 10.1109/tvt.2007.907273. URL http://dx.doi.org/10.1109/TVT.2007.907273.

[8] Michael Doering, Tobias Pögel, Wolf-Bastian Pöttner, and Lars Wolf. A new mobility trace for realistic large-scale simulation of bus-based DTNs. In *Proceedings of the 5th ACM workshop on Challenged networks - CHANTS '10*. Association for Computing Machinery (ACM), 2010. doi: 10.1145/1859934.1859950. URL http://dx.doi.org/10.1145/1859934.1859950.

[9] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker. Recent development and applications of sumo - simulation of urban mobility. *International Journal On Advances in Systems and Measurements*, 5(3), 2012.

[10] J. Harri, M. Fiore, F. Filali, and C. Bonnet. Vehicular mobility simulation with VanetMobiSim. *SIMULATION*, 87(4): 275–300, sep 2009. doi: 10.1177/0037549709345997. URL http://dx.doi.org/10.1177/0037549709345997.

[11] C Sommer, R German, and F Dressler. Bidirectionally coupled network and road traffic simulation for improved IVC analysis. *IEEE Transactions on Mobile Computing*, 10(1):3–15, jan 2011. doi: 10.1109/tmc.2010.133. URL http://dx.doi.org/10.1109/TMC.2010.133.

[12] EU FP7 iTetris. http://ict-itetris.eu. Accessed: 2015-08-05.

[13] Bryan Raney, Nurhan Cetin, Andreas Völlmy, Milenko Vrtic, Kay Axhausen, and Kai Nagel. *Networks and Spatial Economics*, 3(1):23–41, 2003. doi: 10.1023/a:1022096916806. URL http://dx.doi.org/10.1023/A:1022096916806.

[14] Sandesh Uppoor, Oscar Trullols-Cruces, Marco Fiore, and Jose M. Barcelo-Ordinas. Generation and analysis of a large-scale urban vehicular mobility dataset. *IEEE Transactions on Mobile Computing*, 13(5):1061–1075, may 2014. doi: 10.1109/tmc.2013.27. URL http://dx.doi.org/10.1109/TMC.2013.27.

[15] Yoann Pigné, Grégoire Danoy, and Pascal Bouvry. A vehicular mobility model based on real traffic counting data. In *Lecture Notes in Computer Science*, pages 131–142. Springer Science + Business Media, 2011. doi: 10.1007/978-3-642-19786-4_12. URL http://dx.doi.org/10.1007/978-3-642-19786-4_12.

[16] Stefan Joerer, Christoph Sommer, and Falko Dressler. Toward reproducibility and comparability of IVC simulation studies: a literature survey. *IEEE Commun. Mag.*, 50(10):82–88, oct 2012. doi: 10.1109/mcom.2012.6316780. URL http://dx.doi.org/10.1109/MCOM.2012.6316780.

[17] Fan Bai and Bhaskar Krishnamachari. Spatio-temporal variations of vehicle traffic in VANETs. In *Proceedings of the sixth ACM international workshop on VehiculAr InterNETworking - VANET '09*. Association for Computing Machinery (ACM), 2009. doi: 10.1145/1614269.1614278. URL http://dx.doi.org/10.1145/1614269.1614278.

[18] Nabeel Akhtar, Sinem Coleri Ergen, and Oznur Ozkasap. Vehicle mobility and communication channel models for realistic and efficient highway VANET simulation. *IEEE Trans. Veh. Technol.*, 64(1):248–262, jan 2015. doi: 10.1109/tvt.2014.2319107. URL http://dx.doi.org/10.1109/TVT.2014.2319107.

[19] Boris S. Kerner. *The Physics of Traffic*. Springer Berlin Heidelberg, 2004. doi: 10.1007/978-3-540-40986-1. URL http://dx.doi.org/10.1007/978-3-540-40986-1.

[20] Marco Gramaglia, Pablo Serrano, Jose Alberto Hernandez, Maria Calderon, and Carlos J. Bernardos. New insights from the analysis of free flow vehicular traffic in highways. In *2011 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks*. Institute of Electrical and Electronics Engineers (IEEE), jun 2011. doi: 10.1109/wowmom.2011.5986384. URL http://dx.doi.org/10.1109/WoWMoM.2011.5986384.

[21] Marco Gramaglia, Marco Fiore, and Maria Calderon. Measurement-based modeling of interarrivals for the simulation of highway vehicular networks. *IEEE Communications Letters*, 18(12): 2181–2184, dec 2014. doi: 10.1109/lcomm.2014.2363114. URL http://dx.doi.org/10.1109/LCOMM.2014.2363114.

[22] Marco Gramaglia, Ignacio Soto, Carlos J. Bernardos, and Maria Calderon. Overhearing-assisted optimization of address autoconfiguration in position-aware VANETs. *IEEE Trans. Veh. Technol.*, 60 (7):3332–3349, sep 2011. doi: 10.1109/tvt.2011.2158597. URL http://dx.doi.org/10.1109/TVT.2011.2158597.

[23] S. Yousefi, E. Altman, R. El-Azouzi, and M. Fathy. Analytical model for connectivity in vehicular ad hoc networks. *IEEE Trans. Veh. Technol.*, 57(6):3341–3356, nov 2008. doi: 10.1109/tvt.2008.2002957. URL http://dx.doi.org/10.1109/TVT.2008.2002957.

[24] Andre Cardote, Susana Sargento, and Peter Steenkiste. On the connection availability between relay nodes in a VANET. In *2010 IEEE Globecom Workshops*. Institute of Electrical and Electronics Engineers (IEEE), dec 2010. doi: 10.1109/glocomw.2010.5700255. URL http://dx.doi.org/10.1109/GLOCOMW.2010.5700255.

[25] M. Khabazian and M. Ali. A performance modeling of connectivity in vehicular ad hoc networks. *IEEE Trans. Veh. Technol.*, 57(4):2440–2450, jul 2008. doi: 10.1109/tvt.2007.912161. URL http://dx.doi.org/10.1109/TVT.2007.912161.

[26] Salman Durrani, Xiangyun Zhou, and Abhas Chandra. Effect of vehicle mobility on connectivity of vehicular ad hoc networks. In *2010 IEEE 72nd Vehicular Technology Conference - Fall*. Institute of Electrical and Electronics Engineers (IEEE), sep 2010. doi: 10.1109/vetecf.2010.5594505. URL http://dx.doi.org/10.1109/VETECF.2010.5594505.

[27] M.M. Artimy, W. Robertson, and W.J. Phillips. Connectivity in inter-vehicle ad hoc networks. In *Canadian Conference on Electrical and Computer Engineering 2004 (IEEE Cat. No.04CH37513)*. Institute of Electrical and Electronics Engineers (IEEE), 2004. doi: 10.1109/ccece.2004.1345014. URL http://dx.doi.org/10.1109/CCECE.2004.1345014.

[28] Romeu Monteiro, Susana Sargento, Wantanee Viriyasitavat, and Ozan K. Tonguz. Improving VANET protocols via network science. In *2012 IEEE Vehicular Networking Conference (VNC)*. Institute of Electrical and Electronics Engineers (IEEE), nov 2012. doi: 10.1109/vnc.2012.6407428. URL http://dx.doi.org/10.1109/VNC.2012.6407428.

[29] Martin Treiber, Ansgar Hennecke, and Dirk Helbing. Congested traffic states in empirical observations and microscopic simulations. *Physical Review E*, 62(2):1805–1824, aug 2000. doi: 10.1103/physreve.62.1805. URL http://dx.doi.org/10.1103/PhysRevE.62.1805.

[30] Martin Treiber and Dirk Helbing. Realistische mikrosimulation von strassenverkehr mit einem einfachen modell. In *16th Symposium Simulationstechnik ASIM*, volume 2002, page 80, 2002.

[31] Björn Filzek and Bert Breuer. Distance behaviour on motorways with regard to active safety: A comparison between adaptive-cruise-control (acc)

and driver. In *Proceedings of the International Technical Conference on Enhanced Safety of Vehicles*, volume 201, pages 1–8, 2001.

[32] Jesse White. 2014 rules of the road. http://www.cyberdriveillinois.com/publications/pdf_publications/dsd_a112.pdf. Accessed: 2015-08-05.

[33] Seokheon Cho, Rene Cruz, Ramesh Rao, and Anush Badii. Time-gap based traffic model for vehicular traffic flow. In *2014 IEEE 79th Vehicular Technology Conference (VTC Spring)*. Institute of Electrical and Electronics Engineers (IEEE), may 2014. doi: 10.1109/vtcspring.2014.7023125. URL http://dx.doi.org/10.1109/VTCSpring.2014.7023125.

[34] Diala Naboulsi and Marco Fiore. On the instantaneous topology of a large-scale urban vehicular network. In *Proceedings of the fourteenth ACM international symposium on Mobile ad hoc networking and computing - MobiHoc '13*. Association for Computing Machinery (ACM), 2013. doi: 10.1145/2491288.2491312. URL http://dx.doi.org/10.1145/2491288.2491312.

[35] Code of federal regulations, title 47, part 90, private land mobile radio services.

[36] David Hadaller, Srinivasan Keshav, Tim Brecht, and Shubham Agarwal. Vehicular opportunistic communication under the microscope. In *Proceedings of the 5th international conference on Mobile systems, applications and services - MobiSys '07*. Association for Computing Machinery (ACM), 2007. doi: 10.1145/1247660.1247685. URL http://dx.doi.org/10.1145/1247660.1247685.

[37] Fan Bai, Daniel D. Stancil, and Hariharan Krishnan. Toward understanding characteristics of dedicated short range communications (DSRC) from a perspective of vehicular network engineers. In *Proceedings of the sixteenth annual international conference on Mobile*

computing and networking - MobiCom '10. Association for Computing Machinery (ACM), 2010. doi: 10.1145/1859995.1860033. URL http://dx.doi.org/10.1145/1859995.1860033.

[38] Francesca Martelli, M. Elena Renda, Giovanni Resta, and Paolo Santi. A measurement-based study of beaconing performance in IEEE 802.11p vehicular networks. In *2012 Proceedings IEEE INFOCOM*. Institute of Electrical and Electronics Engineers (IEEE), mar 2012. doi: 10.1109/infcom.2012.6195517. URL http://dx.doi.org/10.1109/INFCOM.2012.6195517.

[39] Paul Alexander, David Haley, and Alex Grant. Cooperative intelligent transport systems: 5.9-GHz field trials. *Proceedings of the IEEE*, 99(7):1213–1235, jul 2011. doi: 10.1109/jproc.2011.2105230. URL http://dx.doi.org/10.1109/JPROC.2011.2105230.

[40] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74 (1):47–97, jan 2002. doi: 10.1103/revmodphys.74.47. URL http://dx.doi.org/10.1103/RevModPhys.74.47.

[41] Martin Pinzger, Nachiappan Nagappan, and Brendan Murphy. Can developer-module networks predict failures? In *Proceedings of the 16th ACM SIGSOFT International Symposium on Foundations of software engineering - SIGSOFT '08/FSE-16*. Association for Computing Machinery (ACM), 2008. doi: 10.1145/1453101.1453105. URL http://dx.doi.org/10.1145/1453101.1453105.

[42] Intelligent transportation systems (ITS). ETSI Technical Report TR 102 638, .

[43] Seh Chun Ng, Wuxiong Zhang, Yu Zhang, Yang Yang, and Guoqiang Mao. Analysis of access and connectivity probabilities in vehicular relay networks. *IEEE J. Select. Areas Commun.*, 29 (1):140–150, jan 2011. doi: 10.1109/jsac.2011.110114. URL http://dx.doi.org/10.1109/JSAC.2011.110114.

[44] Ashish Agarwal, David Starobinski, and Thomas D. C. Little. Analytical model for message propagation in delay tolerant vehicular ad hoc networks. In *VTC Spring 2008 - IEEE Vehicular Technology Conference*. Institute of Electrical and Electronics Engineers (IEEE), may 2008. doi: 10.1109/vetecs.2008.333. URL http://dx.doi.org/10.1109/VETECS.2008.333.

[45] Hao Wu, R.M. Fujimoto, G.F. Riley, and M. Hunter. Spatial propagation of information in vehicular networks. *IEEE Trans. Veh. Technol.*, 58(1):420–431, jan 2009. doi: 10.1109/tvt.2008.923689. URL http://dx.doi.org/10.1109/TVT.2008.923689.

[46] Zijie Zhang, Guoqiang Mao, and Brian D. O. Anderson. On the information propagation process in mobile vehicular ad hoc networks. *IEEE Trans. Veh. Technol.*, 60(5):2314–2325, 2011. doi: 10.1109/tvt.2011.2145012. URL http://dx.doi.org/10.1109/TVT.2011.2145012.

[47] Emmanuel Baccelli, Philippe Jacquet, Bernard Mans, and Georgios Rodolakis. Information propagation speed in bidirectional vehicular delay tolerant networks. In *2011 Proceedings IEEE INFOCOM*. Institute of Electrical and Electronics Engineers (IEEE), apr 2011. doi: 10.1109/infcom.2011.5935199. URL http://dx.doi.org/10.1109/INFCOM.2011.5935199.

[48] Marco Gramaglia, Oscar Trullols-Cruces, Diala Naboulsi, Marco Fiore, and Maria Calderon. Vehicular networks on two madrid highways. In *2014 Eleventh Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. Institute of Electrical and Electronics Engineers (IEEE), jun 2014. doi: 10.1109/sahcn.2014.6990380. URL http://dx.doi.org/10.1109/SAHCN.2014.6990380.

[49] Wantanee Viriyasitavat, Fan Bai, and Ozan K. Tonguz. Dynamics of network connectivity in urban vehicular networks. *IEEE J. Select. Areas Commun.*, 29(3):515–533, mar 2011. doi: 10.1109/jsac.2011.110303. URL http://dx.doi.org/10.1109/JSAC.2011.110303.