

# On the Structure and Evolution of Vehicular Networks

George Pallis\*, Dimitrios Katsaros†, Marios D. Dikaiakos\*, Nicholas Loulloudes\*, Leandros Tassioulas†

\*Department of Computer Science

University of Cyprus

Nicosia, Cyprus

†Computer & Communication Engineering Department

University of Thessaly

Volos, Hellas

**Abstract**—Vehicular ad hoc networks have emerged recently as a platform to support intelligent inter-vehicle communication and improve traffic safety and performance. The road-constrained and high mobility of the vehicles, their unbounded power source, and the emergence of roadside wireless infrastructures make VANETs a challenging research topic. A key to the development of protocols for intervehicle communication and services lies in the knowledge of the topological characteristics of the VANET communication graph. This article provides answers to the general question: *how does a VANET communication graph look like over time and space?* This study is the first one that examines a very large-scale VANET graph and conducts a thorough investigation of its topological characteristics using several metrics, not examined in previous studies. Our work characterizes a VANET graph at the connectivity (link) level, quantifies the notion of “qualitative” nodes as required by routing and dissemination protocols, and examines the existence and evolution of communities (dense clusters of vehicles) in the VANET. Several latent facts about the VANET graph are revealed and incentives for their exploitation in protocol design are examined.

## I. INTRODUCTION

Inter-vehicle communication (IVC) has emerged as a promising field of research, where advances in wireless and mobile ad-hoc networks can be applied to real-life problems (traffic jams, fuel consumption, pollutant emissions, and road accidents) and lead to a great market potential. Already, several major automobile manufacturers and research centers are investigating the development of IVC protocols, systems (e.g., DSRC, 802.11p) and the use of inter-vehicle communication for the establishment of Vehicular Ad-Hoc NETWORKS (VANETs).

A vehicular network is a challenging environment since it combines a fixed infrastructure (roadside units, e.g., proxies), and ad hoc communications among vehicles. Despite the fact that it presents similarities with the traditional Mobile Ad hoc NETWORKS (MANETs), the mobile nodes in a VANET (i.e., the vehicles) are not energy-starving, are highly mobile, and their mobility is constrained by the underlying road network topology. Moreover, the existence of roadside infrastructure creates more opportunities for optimized communications. Apart from the networking aspects, the applications which are expected to run over a VANET make it also a unique environment: safety

applications [25], [28] (accident avoidance near intersections, speed “regulation” for road congestion avoidance), peer-to-peer music sharing [16], Internet access [23], they all pose interesting questions related to protocol design and network deployment.

During the process of designing and deploying a VANET, various questions must be answered that pertain to protocol performance and usefulness. For instance, when deciding the placement of roadside proxies [20], in order to reduce the average path length between the vehicles and the access points, we need to know the distribution of the position of vehicles; when performing message routing, the corner-stone question is “which are the highest-quality nodes (vehicles)?” [9] to carry out the forwarding process; when performing geocasting, the question is how we can spread the emergency messages with the minimal number of rebroadcasts so as to reduce collisions and latency; when designing mobility models [21], we need to know the distribution of “synapses” per node, i.e., whether there are any clusters (communities); when the network is disconnected, a significant question concerns the identification of bridge nodes [4] which are encharged with the delivery/ferrying of the messages.

All these questions and many more require knowledge of the topological characteristics of the VANET communication graph, where vehicles correspond to vertices and communication links to edges. Despite the fact that such knowledge is of paramount importance, the relevant literature is relatively poor w.r.t. the study of the characteristics of a VANET communication graph (for a detailed presentation of the relevant work see Section II). Notable exceptions are the works reported in [10] and [27], which study quantities like wireless link lifetime, network diameter, node degree, number, size of groups (clusters) of vehicles and the intra-group connectivity strength. In spite of the usefulness of these metrics, they are not capable of revealing a deep image of the VANET graph.

The objective of this work is to go one step further and present “higher order” knowledge of the time-evolving topological characteristics of a VANET communication graph, as compared to the “first-order” knowledge provided by the studies reported in [10], [27]. Most of real-world networks have

been proved to follow some topological statistical features (i.e., features of scale-free networks, small-world properties, power-law degree distribution etc.) [19]. Considering that VANETs are not static but evolve over time by additions and deletions of nodes, it is important to examine the network properties and topological statistical features that characterize the structure and behavior of vehicular networks. This study is the first one that examines a very large scale VANET graph and conducts a thorough investigation of its topological characteristics using numerous metrics not examined in any previous study, existence and evolution of communities (dense clusters of vehicles) in the VANET. In particular, the paper’s major contributions are the following:

- A thorough study of the visible and “latent” structure of a VANET communication graph, including metrics used in earlier studies [10], [27], as well as several other metrics traditionally used in the field of social network analysis, i.e., centrality measures.
- A detailed study of clusters and dense subgraphs established inside a VANET communication graph including their dynamic properties.
- A study of the connectivity properties of the road network. The results obtained provide a more definite understanding of the impact of the road network on the properties of the VANET graph.
- Discussion of the implications of the findings upon the design of protocols for the MAC, Network, and Application layers of a VANET.

The rest of the article is organized as follows: Section II briefly surveys the relevant work; Section III describes the metrics used in the present study to characterize the evolution of VANET communication graph; Section IV describes the source of the data studied here. Section V records the findings of the study, and Section VI examines the implications of the these findings in the design of protocols. Finally, Section VII concludes the article.

## II. RELEVANT WORK

Network graph analysis has been conducted both for MANETs and VANETs. In [2], the authors study the temporal evolution of the diameter of opportunistic mobile networks, which follow the random graph model. Results showed that the diameter increases slowly with the network size. Härrilä et al. in [13] introduced the concept of kinetic graphs to capture the dynamics of mobile graph structures so as to efficiently support network-wide operations, e.g., broadcasting. A kinetic graph comprises a generalization of the static network graph able to model the trajectories of the mobile nodes and supports the notion of the “probabilistic existence” of graph edges.

Well-known concepts from social network analysis have been used as primitives to design advanced protocols for routing and caching in delay-tolerant networks (DTNs) and in ad hoc sensor networks. In [4], the betweenness centrality index and its combination with a similarity metric (comprising both the SimBet metric) have been used to select forwarding nodes to support routing in DTNs. Results showed that data dissemination is improved if the messages are delivered through

nodes which have high SimBet utility values. The betweenness centrality has also been used in [7] to design a cooperative caching protocol for wireless multimedia sensor networks. This protocol selects the mediator nodes that coordinate the caching decisions based on their “significant” position in the network. Yoneki studied the impact of connective information (clustering, network transitivity, and strong community structure) on epidemic routing in a series of works [30]. The value of the connectivity analysis of ad hoc networks is so fundamental that recently a competition-experiment has been started — the MANIAC experiment [27]— to study network connectivity, diameter, node degree distribution, clustering, frequency of topology changes, route length distribution, route asymmetry, frequency of route changes, and packet delivery ratio. The obtained results show a high degree of topology and route changes, even when mobility is low, and a prevalence of asymmetric routes, both of which contradict assumptions commonly made in MANET simulation studies.

In the context of vehicular networking, there has been relatively little work on exploring the properties of the time evolution of VANET graphs. The authors of [3] present a preliminary characterization of the connectivity of a VANET operating in an urban environment. They transform the vehicular network into a transitive closure graph. Then, the graph temporal evolution of the average node degree is presented. However, the authors do not make a deep analysis of the networking shape of vehicular mobility. They study only the average node degree for a small time interval. In [26], the authors set up a real-world experiment consisting of 10 vehicles making loops in a 5-mile segment of a freeway. They focus on the connectivity issues without investigating the topological properties of the VANET graph. Authors in [10] study the node degree distribution, link duration, clustering coefficient and number of clusters for VANET graphs under various vehicular mobility models. The objective of [10] focuses on studying the topological properties of different mobility models and explaining why different models lead to dissimilar network protocol performance. Finally, the authors of [15] provide an analysis of the connectivity of vehicular networks by leveraging on well-known results of percolation theory. Using a simulation model, they study the influence of vehicle density, the proportion of equipped vehicles, transmission range, traffic lights and roadside units.

The paper at hand goes one step further from the previous studies in [10], [15], [26], [27] and provides a thorough study of the visible and “latent” structure of a VANET communication graph, including metrics not examined before.

## III. GRAPH METRICS EXAMINED

This section contains the definitions of the metrics used in the study. We categorize the examined metrics as *network-wide* (require knowledge of the complete VANET graph), *localized*, and *community-oriented*. All node IDs mentioned in this section refer to the sample graph of Figure 1. For the sequel, we will consider  $G(t)$  to be an undirected graph of VANET at time  $t$ , where vehicles correspond to the set of

vertices  $V(t) = \{u_i\}$  and communication links to the set of edges  $E(t) = \{e_{ij}\}$ . An edge  $e_{ij}(t)$  exists, if  $u_i$  can communicate directly with  $u_j$  at time  $t$ , with  $i \neq j$ .

#### A. Localized metrics

- **Node degree.** The number of vehicles within the transmission range of a node. Formally, the degree of  $u_i$  at time  $t$  is defined as  $D_i(t) = \|\{u_j \mid \exists e_{ij}(t)\}\|$ .
- **Lobby Index [18].** The lobby index of a given vehicle  $u_i$  at time  $t$ , denoted as  $L_i(t)$ , is the largest integer  $k$  such that the number of one-hop neighbors of  $u_i$  in graph  $G(t)$  with degree at least  $k$  equals  $k$ . This metric can be seen as a generalization of  $D_i(t)$ , conveying information about the neighbors of the node as well (e.g., node with ID 8 has lobby index 2).
- **Link duration.** The time between the instance at which a vehicle enters within transmission range of another vehicle, and the instance at which the physical connection is lost. Formally, the duration  $l_{ij}(t)$  of the link from  $u_i$  to  $u_j$  at time  $t$  is defined as  $l_{ij}(t) = t_c - t_o$ , if  $\exists e_{ij}(t)$ , where  $t \in [t_o, t_c]$ .

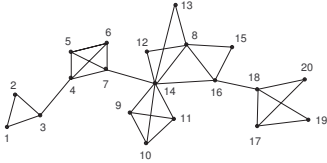


Fig. 1. Snapshot of a sample VANET graph.

#### B. Network-wide metrics

- **Diameter.** It is the longest distance between any two nodes in the network, where the distance is defined as the shortest path between the nodes.
- **Closeness Centrality [29].** It is defined as the inverse of the sum of the distances between a given node and all other nodes in the network. The closeness centrality of a vehicle  $u_i$  at time  $t$  is:

$$C_i(t) = \frac{1}{\sum_{j \neq i} \text{distance}(u_i, u_j)} \quad (1)$$

where  $\text{distance}(u_i, u_j)$  is the distance between  $u_i$  and  $u_j$ . Closeness centrality measures how long it will take information to spread from a given vehicle to other vehicles in the network. A sample node with low closeness centrality is the node with ID 14 ( $C_{14} = 0.026$ ).

- **Betweenness Centrality [29].** It is defined as the fraction of the shortest paths between any pair of nodes that pass through a node. The betweenness centrality of a vehicle  $u_i$  at time  $t$  is:

$$BC_i(t) = \sum_{j \neq k} \frac{sp_{j,k}(u_i, t)}{sp_{jk}(t)} \quad (2)$$

where  $sp_{jk}$  is the number of shortest paths linking vertices  $j$  and  $k$  at time  $t$  and  $sp_{j,k}(u_i, t)$  is the number of shortest paths linking vertices  $j$  and  $k$  that pass through  $u_i$  at time  $t$ . Betweenness centrality is a measure of the extent to which a vehicle has control over information flowing between others (e.g.,  $BC_{14} = 0.668$ ).

- **Bridging Centrality [14].** It is computed by multiplying the betweenness centrality by a bridging coefficient. The bridging centrality of  $u_i$  at time  $t$  is:

$$BR_i(t) = BC_{u_i}(t) \cdot \beta(u_i), \quad (3)$$

where  $\beta(u_i)$  is the bridging coefficient of  $u_i$ . The bridging coefficient is the ratio of the inverse of a node degree to the sum of the inverses of all its neighbor degrees. The bridging centrality metric attempts to find nodes that are central to the graph, but also they have a low number of direct connections relative to their neighbor connections (e.g.,  $BR_7 = 3.345$ ).

At this point we emphasize that localized versions of these centrality metrics can be devised as well, taking into account only  $k$ -hop neighborhoods.

#### C. Community metrics

- **Number of Clusters.** The number of co-existing, non-connected clusters of nodes at a given instant. We define as cluster a connected group of vehicles. A connected group is a subgraph of the network such that there is a path between any pair of nodes.
- **Clustering Coefficient.** It measures the cliquishness of a network. The clustering coefficient  $p_k(t)$  of a cluster  $k$  at time  $t$  (as defined in [10]) is:

$$p_k(t) = \frac{2|E_k(t)|}{|N^k(t)|(|N^k(t)| - 1)}, \quad (4)$$

where  $|E_k(t)|$  is the number of existing links in cluster  $k$  at time  $t$  and  $|N^k(t)|$  is the number of nodes in cluster  $k$  at time  $t$ . The clustering coefficient has a maximum value 1 if the cluster is a clique.

- **Localized Clustering Coefficient.** We define a localized version of the clustering coefficient as follows: For a vehicle  $i$ , which has  $D_i(t)$  neighboring vehicles at time  $t$ , and there are  $z_i(t)$  edges between its neighbors, then the localized clustering coefficient  $lp_i(t)$  of vehicle  $i$  is  $lp_i(t) = D_i(t)/z_i(t)$  if  $z_i(t) > 1$ , else  $lp_i(t) = 0$ , if  $z_i(t) = 0$  or  $z_i(t) = 1$ .
- **Number of Communities [17].** The number of existing communities at a given instant. A community is defined as a dense sub-graph where the number of intra-community edges is larger than the number of inter-community edges. In order to identify communities, we transform  $G(t)$  to directed graph so as  $D_i^{in}(t) = D_i^{out}(t) = D_i(t)$ , where  $D_i^{in}(t)$ ,  $D_i^{out}(t)$  is the in-degree and out-degree of node  $u_i$  at time  $t$ . Formally, a subgraph  $U(t)$  of a VANET graph  $G(t)$  at time  $t$  constitutes a community, if it satisfies:

$$\sum_{u_i \in U(t)} (D_i^{in}(t))(U(t)) > \sum_{u_i \in U(t)} (D_i^{out}(t))(U(t)), \quad (5)$$

i.e., the sum of all degrees within the community  $U(t)$  is larger than the sum of all degrees toward the rest of graph the  $G(t)$ <sup>1</sup>.

<sup>1</sup>Other notions of communities can be defined as well, but we use that defined in [17] since it allows for overlap.

#### IV. TRAFFIC DATA STUDIED

Apparently, the usefulness of the findings of any study depends on the realism of the data upon which the study operates. For the first time at such a large scale, we study the structure and evolution of a VANET communication graph using realistic vehicular traces<sup>2</sup> from the city of Zurich. These traces are obtained from a multi-agent microscopic traffic simulator (MMTS) [22]. MMTS is capable of simulating public and private traffic over real regional road maps of Zurich with a high level of realism. Each vehicle makes its plans and MMTS executes all those plans simultaneously. The route choice of each vehicle is dynamic in order to react adequately to time-dependent congestion effects. We skip a re-proof of the realism of these traces, since this has been done in [24]. We have extracted a rectangular street area of size  $5km \times 5km$ , which covers the centre of Zurich and contains around 200,000 distinct vehicle trajectories during a 3 hours interval in morning rush hour. We have also considered 1 hour warm-up period (5:00-6:00). During this period, vehicles have evenly distributed throughout the map. Assuming that all vehicles are equipped with vehicular communication hardware and software, we study the networking shape evolution of VANET, by observing snapshots of this network taken at regularly spaced time instances. Future work is to study the impact of market penetration of the communication equipment as well as the resultant background traffic on connectivity.

We also assume that a vehicle can communicate effectively with neighboring vehicles and stationary access points that are within a range of at most 50 – 100 meters from it. This assumption is established upon recent experimental results that investigated the viability of IEEE 802.11b for vehicle-to-vehicle communication in urban environments [12]. Consequently, in our study, we examine network graphs that capture wireless network connections established between vehicles separated by a distance of at most 100 meters. The datasets examined in this article comprise two VANET communication graphs: one corresponds to transmission range equal to  $T = 50m$  (urban scenario with non-line of sight transmission distance for bandwidth  $> 4Mbps$  with 20 – 30% lost of packets), and the other to a transmission range of  $T = 100m$  (urban scenario with line of sight transmission distance for bandwidth  $> 4Mbps$  with 20 – 30% lost of packets).

#### V. THE STRUCTURAL PROPERTIES OF A VANET COMMUNICATION GRAPH

This section presents the findings of the study related to laws governing the nodes, edges, diameter of the VANET network (§ V-A), node centralities (§ V-B), the characteristics of the network concerning the link duration (§ V-C), the community-level analysis of the graph (§ V-D), and the network robustness (§ V-E).

##### A. Network analysis

Typically, large real world networks evolve over time. Lescovec et al. in [19] studied the temporal evolution of

<sup>2</sup>The traces are publicly available from <http://www.lst.inf.ethz.ch/research>

several real graphs arising in a wide range of domains (i.e., autonomous systems, e-mail networks, citations) and made the following empirical observations: i) the average degree increases as the network grows, with the number of edges growing super-linearly in the number of nodes, ii) the diameter is decreasing as the network grows, iii) real networks have relatively small average node degrees and diameters.

*QUESTION 1: What are the laws that govern the temporal evolution of VANET-graph properties?*

For each graph dataset, we have several time snapshots for which we study the number of nodes  $V(t)$  and edges  $E(t)$ . Figures 2 illustrates the number of nodes and edges over time, respectively. As expected, the VANET graph grows with the number of vehicles injected into the map and the transmission range of their wireless antennae. We investigate further the relation between edges and nodes: Figure 3 depicts that the *VANET graphs obey a power-law with a consistently good fit*. Specifically, we find that they follow the relation  $E(t) \propto V(t)^\alpha$ , where  $\alpha \simeq 1.77$ ; also, that this relation is independent of transmission range. Specifically, we examined this relation for several communication ranges (e.g.,  $T=50m$ ,  $T=200m$ ) and we observed that it holds for any  $T$ ; due to the interest of space we do not present the figures. This means that the *VANET graph is dense*; note that  $\alpha = 2$  corresponds to an extremely dense graph where each node has, on average, edges to a constant fraction of all nodes. Practically, this observation is important since we can have estimate the number of communication links in the network.

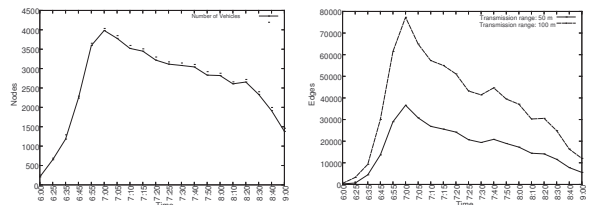


Fig. 2. (Left) Nodes over time. (Right) Edges over time.

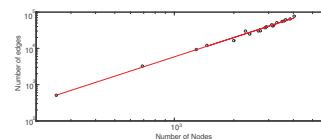


Fig. 3. Number of edges vs. number of nodes in log-log scale. Slope  $\alpha = 1.77$  ( $T=100m$ ).

Figure 4 illustrates the average degree of separation (i.e., average length of shortest path between pairs of nodes) over time. Our findings show that the VANET graphs *do not exhibit small world properties*. That is important as this metric provides an indication about how long a vehicle has to wait on average before obtaining a desired piece of information. For instance, at 7:00 am the average degree of separation is 60 hops. We also observe that the average number of hops between two vehicles for the  $T=50m$  case is large and exhibits wide variability as the network grows in size (right chart of Figure 4). On the



other hand, the degree of separation for  $T=100m$  changes more smoothly and is less variant. Note, that the small degrees of separation that we observe in some timestamps (e.g., 6:00, 6:35, 9:00) are side-effects of the shuttering of the whole VANET into smaller clusters. A thorough study for the clusters of the VANET communication graph is presented in § V-D.

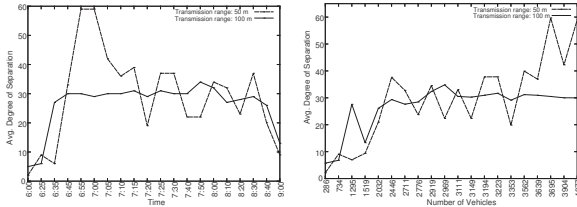


Fig. 4. (Left) Avg. degree of separation over time. (Right) Average degree of separation over number of vehicles.

Next, we examine the evolution of the graphs’ diameter over time. Our analysis shows that graph diameter follows the average degree of separation over time, and gets quite large values<sup>3</sup>.

Finally, we estimate the geographic distance between the two most distant vehicles on the map, in terms of hops, with each hop being equal to the transmission range of the wireless medium. We call this measure the “*geographic diameter*”. The geographic diameter gives us an estimate of the overall span of our vehicles on the area of our map and a worst-case bound about the end-to-end delay to deliver a message from a source to a destination or the total delay to perform network-wide broadcasting. As we can see from Figure 5, *the geographic diameter is practically constant and independent of the network’s size*. The geographic diameter of the network is a useful metric for taking decisions about base stations placement in a vehicular network [1].

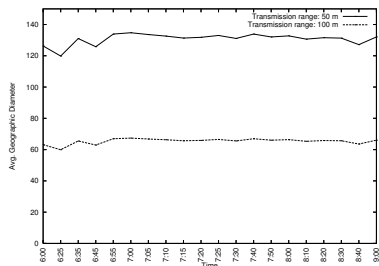


Fig. 5. Avg. geographic diameter over time.

Figure 6 shows the average node degree over time. We observe that *the average node degree increases with the number of nodes in the vehicular network*, and has a pattern similar to the number of edges (Figure 2); both the number of edges and the average node degree for  $T = 100m$  are two times larger than  $T = 50m$ . We also observe that the variance of the average node-degree values is quite large, but that their distribution is rather uniform. This means that a vehicular network includes a significant percentage of nodes

with high and low degree (i.e., vehicles which are in junctions and vehicles which are in places with low traffic, respectively).

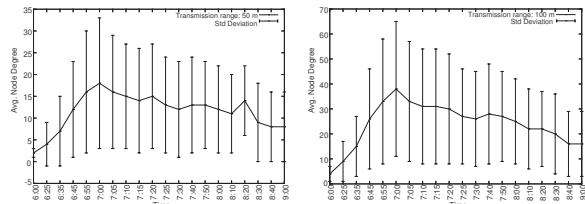


Fig. 6. Avg. node degree over time: (Left)  $T=50m$ . (Right)  $T=100m$ .

To sum up, our network analysis concluded to the following empirical observations:

- The VANET is a dense graph; the number of edges vs. the number of nodes follows power-law distribution:  $E(t) \propto V(t)^\alpha$ , where  $\alpha \simeq 1.77$ .
- The diameter and the average node degree are, in most cases, increasing as the vehicular network grows in size.
- The average node degrees present large standard deviation values.
- The VANET does not exhibit small world properties; high values of diameter and degrees of separation exist.
- The geographic diameter (for any  $T$ ) and the average degree of separation (for  $T \geq 100m$ ) are, in most cases, constant and smooth.

## B. Centrality metrics

Looking deeper into the VANET communication graph, we examined the average value of the centrality measures as a function of time and communication range<sup>4</sup>. The results are depicted in Figures 7 and 8. The key question we investigate is:

*QUESTION 2: Do the centrality metrics identify “quality” nodes and what is the spatial distribution thereof?*

The general observation is that the distribution of “central” nodes is not affected by the communication range; the distributions have similar shapes for both  $T = 50m$  and  $T = 100m$ . The centrality metrics reflect quite reliably the variation in traffic conditions, i.e., density and relative positions of the vehicles. Therefore, *centrality is not an artefact of the communication range but an indication of the latent “behavior” of the vehicles*, i.e., road network and drivers’ intentions, which ultimately define the network position of the vehicles.

A very interesting question is whether all these centrality metrics reveal different patterns of the graph, and what is their relation to the node degree. From Figure 7, we can conclude that the betweenness, closeness, and bridging centrality indexes follow more or less similar distributions. For the case where the transmission range is equal to  $50m$ , closeness centrality is practically constant over time. This is due to the fact that the resulting network is quite sparse without large dense components, which in turns means that the (average) distance of each vehicle to the rest of the network’ vehicles is practically the same.

<sup>4</sup>We have also examined the case of Spectral Centrality metrics (e.g., PageRank), but it did not show any significant results, and thus we did not consider presenting them into this article.

<sup>3</sup>Avg. diameter=90.8 ( $T=50m$ ) | 76.2 ( $T=100m$ )

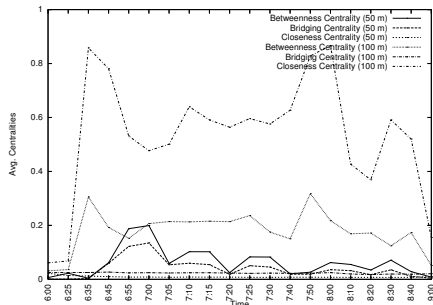


Fig. 7. Avg. centralities over time and range.

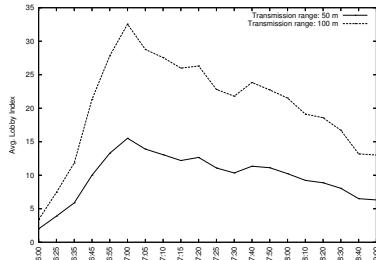


Fig. 8. Avg. lobby centrality over time and range.

The case of lobby index (Figure 8) is quite interesting and deviates from the behavior of the aforementioned centrality metrics. Recall that its definition implies that it is a generalization of node degree and some sort of “simplification” of betweenness centrality. Indeed, it follows the general pattern of betweenness, but without the abrupt changes. It is worth noting that we have identified a significant number of vehicles with the same (quite large) value for the lobby index. This was not the case for betweenness centrality, where only (relatively) few vehicles had quite large value for the betweenness centrality. This observation is quite significant, because, during some protocol design we need to identify quite a large number of “quality” nodes to assign to them special roles, whereas in some other case we need to identify only the highest “quality” vehicle. Therefore, we can conclude that *the betweenness centrality and lobby centrality indexes are sufficient and appropriate for capturing the structural properties of a VANET communication graph, without any of them superseding the other.*

Since prior studies have investigated the degree distribution, it is interesting to examine whether the centrality metrics studied in this article are correlated to the degree, i.e., whether the high-degree nodes are also high-quality nodes. To answer this, we computed the Pearson correlation coefficient for all nodes at a specific<sup>5</sup> time (7:00 am,  $T = 100m$ ). In general, the Pearson correlation coefficient ranges from  $-1$  to  $1$ , where  $-1$  or  $1$  indicates a “perfect” relationship. The further the coefficient is from  $0$ , regardless of whether it is positive or negative, the stronger the relationship is between the two variables. It is obvious from Table I that the high-degree nodes are not correlated with betweenness centrality

<sup>5</sup>We examined this for many time instances, but due to lack of space we only present one representative case.

and bridging centrality metrics. On the other hand, there is a low positive correlation with closeness centrality. Therefore, *the node degree is not able to identify “quality” nodes in a VANET; betweenness and lobby index do it better.*

	Betweenness	Bridging	Closeness	Lobby
Degree	0.044	-0.008	0.36	0.106

TABLE I  
PEARSON CORRELATION COEFFICIENT (CORRELATION IS SIGNIFICANT AT THE 0.01 LEVEL).

Examining in greater depth the variation of one of the prevalent centrality metrics, i.e., betweenness, we plotted its actual values (instead of averages) as a function of time and geographic location. Due to space limitations, we present the betweenness centrality values at 08 : 00. The results are illustrated in the diagrams of Figure 9<sup>6</sup>. This diagram reflects the BC values of vehicles: each vehicle is colored according to its BC value. Clearly, the road topology is not the decisive parameter for the betweenness, even though it affects it (observe that many of the vehicles with high BC value are moving along the same roads). Nevertheless, vehicles with high centrality values appear at any geographic location independently of the road-network’s structure. Therefore, *the road network alone, e.g. position of junctions, does not determine the positions of possible “significant” nodes.*

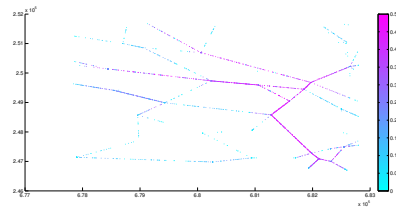


Fig. 9. Betweenness centrality over geographic location.

Finally, we investigated the correlation of centrality metrics (degree, betweenness, closeness, bridging, lobby) to the vehicles’ density, and also among them. The results are illustrated in Figure 10 and Table II. The color of each junction in Figure 10 represents the average quality value of the vehicles which are close enough ( $\leq 50m$ ) to the junction. For the interest of space, in Figure 10 we illustrate graphically the true values for the junctions only for the vehicles density and lobby index. Although not identical, we are able to identify a significant correlation of the two metrics in the junctions; indeed this is also the case for the rest of the metrics (see Table II).

The metrics are highly correlated with each other regarding their value in the junctions; they are also correlated to the vehicles’ density. To confirm our observation, we calculated the Pearson correlation coefficient and the results are presented

<sup>6</sup>The use of colored figures here and in the sequel, is necessary to help the reader gain a better image of the results.

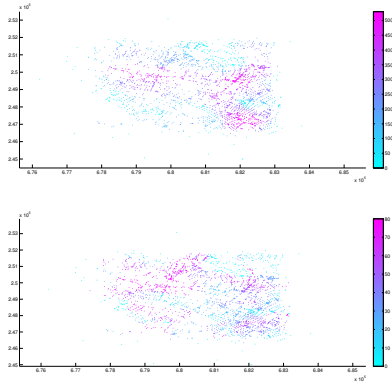


Fig. 10. Geographic distribution of: (top) vehicles density, (bottom) lobby index.

in Table II. We see that *the localized metrics are positively correlated to a great degree with the network-wide metrics in the case of junctions*. We view this finding as quite significant, since the localized metrics are easier to compute than the network-wide ones, and we can exploit them easier in the road junctions.

	Density	Degree	Betweenness	Bridging	Closeness	Lobby
Density	1	0.711	0.299	0.231	0.643	0.711
Degree	0.711	1	0.508	0.396	0.816	0.681
Betweenness	0.299	0.508	1	0.829	0.684	0.482
Bridging	0.231	0.396	0.829	1	0.606	0.472
Closeness	0.643	0.816	0.684	0.606	1	0.801
Lobby	0.576	0.681	0.482	0.472	0.801	1

TABLE II  
PEARSON CORRELATION COEFFICIENT IN JUNCTIONS  
(CORRELATION IS SIGNIFICANT AT THE 0.01 LEVEL).

In summary, we made the following observations:

- Centrality metrics are an indication of the latent “behavior” of the vehicles without being affected by the communication range.
- Betweenness centrality and lobby centrality indexes are sufficient and appropriate for identifying the “quality” (more central nodes) of vehicles. On the other hand, the node degree is not able to identify “quality” nodes in a VANET.
- The road network alone is not sufficient information to identify the positions of possible “significant” nodes in a VANET.
- The localized metrics of vehicles are highly correlated with their network-wide metrics regarding their values in roads junctions.

### C. Link duration analysis

The analysis of link duration contributes to the prediction of network-link lifetime. Link duration is influenced by driving situations and vehicle speed. For instance, vehicles establish short-lived links when travelling fast on the opposite direction. A recent study showed that the link duration is high when vehicles are moving in a highway, with most connections formed in the freeway lasting between 15 and 30 secs, with a median of 23 secs [26].

**QUESTION 3:** Which are the link duration statistics in VANET when the vehicles are moving in urban areas?

Table III presents the link duration statistics for our data sets. As expected, for larger  $T$ , the link duration is longer. Specifically, the link duration is almost doubled when  $T$  is doubled. As depicted in Figure 11, *most vehicles have low link duration times*. However, the link durations measured, can accommodate the time required for service interactions over the VANET [6]. Prior simulations have shown that the average time of a successful transaction can be less than 0.1 sec. From a comparison of the mean and median values of link duration, we can conclude that there is a high variability in link-duration values.

An interesting question is to identify which vehicles have long link duration times. Our experiments showed the following trend: *the vehicles with high degree values usually have longer link duration time than the ones with low-degree values*. This means that vehicles which have a large number of communication links are better and longer connected than the ones with a small number of links.

Transmission range	50 m	100 m
Time	6:00 – 9:00	6:00 – 9:00
Total links	21922350	23705232
Min	1 sec	1 sec
Max	978 sec	1105 sec
Mean	6.7531 sec	13.2038 sec
Median	3 sec	7 sec
Standard deviation	21.2401 sec	34.2413 sec

TABLE III  
LINK DURATION STATISTICS.

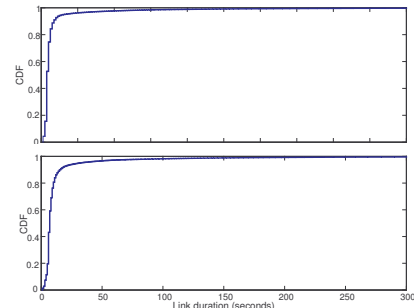


Fig. 11. Link duration CDF (top: T=50m, bottom: T=100m).

### D. Cluster analysis

For services requiring dissemination of messages across the whole VANET, it is mandatory to know whether the network is connected or not. Additionally, for services that require the delivery of messages inside a specific geographic region (geocasting), it would be useful to have an estimation of the density of communication links among vehicles inside this region so as to exploit or avoid the use of the flooding primitive. Therefore, it is useful to investigate the following: **QUESTION 4:** Does the VANET consist of a single connected component? Are there any dense subgraphs inside the VANET?

We performed a cluster analysis of the VANET graph, and in Figure 12 we present the number of clusters and the clustering coefficient over time for our networks. We observe that the number of clusters is mainly affected by transmission range. Specifically, *the number of clusters when  $T = 50m$  is 3 times larger than that for  $T = 100m$* . On the other hand, *the clustering coefficient is stable (about 0.73) without influencing by vehicle density and transmission range*. The existence of clusters means that the vehicular network graph is not connected. Another interesting observation is that most vehicles belong to the same cluster. A snapshot of the existing clusters in a specific time is illustrated in Figure 13, where the vehicles that belong to the same cluster have the same color. Figure 14 shows that about 80% and about 50% of vehicles belong to the same cluster for  $T = 100m$  and  $T = 50m$ , respectively.

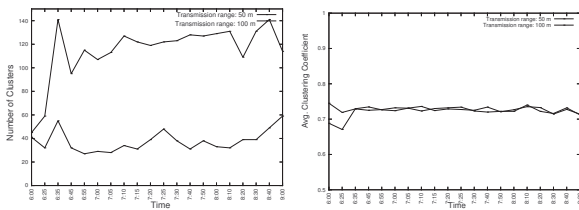


Fig. 12. (Left) Number of clusters over time. (Right) Clustering coefficient over time (all clusters).

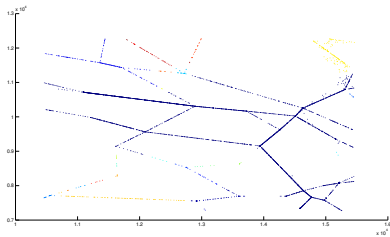


Fig. 13. The clusters (color indicates membership) at 7.00 am ( $T=100$ ).

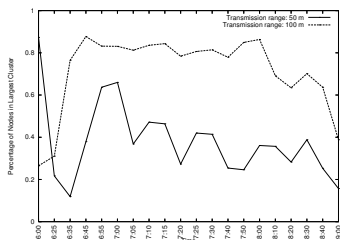


Fig. 14. Size of the largest cluster over time.

The right chart of Figure 12 tells us that the *connectivity within a cluster remains stable over time*. Despite the usefulness of such observation, the complementary question of what is the relation of node degree and connectivity, which can be used to quantify a purely localized connectivity behavior, seems much more interesting. Figure 15 investigates this question. It shows that *dense clusters can contain nodes*

*with both small and large degree*. Beyond a certain, very large degree (e.g., 100) though, the localized clustering coefficient stabilizes at the value of 0.5 implying that the neighborhoods are “half” dense and this is expected since there can not be too many links between nodes of a neighborhood which is comprised by more than 100 nodes. Nevertheless, the strong practical meaning is that *in a VANET we can easily find “almost-cliques” comprised by many vehicles, and not simply triangles*. This is particularly appealing, because in a dissemination protocol for instance, a single broadcast will reach a lot (depending on the cluster size) of nodes.

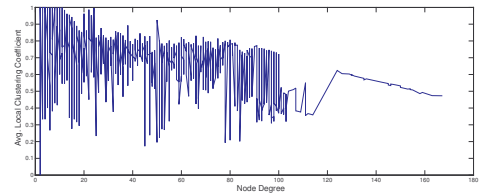


Fig. 15. Avg. localized clustering coefficient vs. node degree.

To better understand the properties of the largest cluster, we further study its change over time. For each second we identify how many vehicles are inserted, left and remained in the largest cluster. Figure 16 depicts the Q-Q plots of arrival/departure processes of vehicles and we observe that they follow the Pareto distribution. Specifically, the Pareto distribution has two parameters associated with it: the shape parameter  $\alpha > 0$  and the scale parameter  $\kappa > 0$ . The cumulative distribution function of inter-arrival and inter-departure time durations is  $F(x) = 1 - (\frac{\kappa}{x+\kappa})^\alpha$ . This distribution is heavy-tailed with unbounded variance when  $\alpha < 2$ . In our datasets,  $\alpha = 0.518$  for inter-arrival of vehicles and  $\alpha = 0.489$  for inter-departure of vehicles. The scale parameter takes the value  $\kappa = 4$ . This means that *vehicles inter-arrivals and inter-departures from the largest cluster exhibit burstiness on several time scales*. The Q-Q plot in the bottom of Figure 16 illustrates that *the number of vehicles that remain in the same cluster follow Normal distribution*. We view these findings as particularly important since we can predict the evolution of clusters over time.

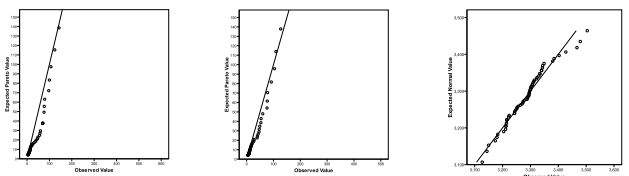


Fig. 16. (Left) Distribution of arrival nodes in the largest cluster ( $\alpha = 0.518$ ,  $\kappa = 4$ ). (Middle) Distribution of departure nodes in the largest cluster ( $\alpha = 0.489$ ,  $\kappa = 4$ ). (Right) Distribution of static nodes in the largest cluster.

Finally, we used the state-of-the-art, CiBC algorithm [17], in order to identify the communities in VANET. CiBC identifies overlapping communities in a graph without requiring to preset their number. Figure 17 depicts the identified communities



over time. Results have shown that a significant number of communities is identified in vehicular networks. As expected, the number of communities are mainly influenced by the transmission range. More edges to the network leads to fewer and longer communities.

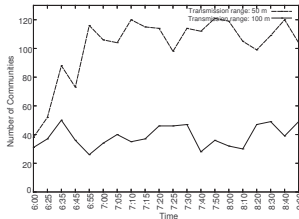


Fig. 17. Number of communities over time.

We further sharpen the quantitative aspects of these findings. In particular, we observe that *tight communities scale at very small sizes (up to 150 nodes)*. Communities that scale at very large sizes are not tight and thus become less “community-like”. The existence communities in vehicular networks is important since this means that there are groups of vehicles in VANET which are strongly connected with each other. In other words, VANET contains groups of vehicles that interact more strongly amongst themselves than with the outside world.

To sum up, we have made the following observations:

- The VANET graph includes a giant cluster.
- Vehicles’ inter-arrivals and inter-departures from the largest cluster exhibit burstiness on several time scales.
- Clusters’ connectivity remains stable over time.
- Dense clusters contain nodes with both small and large degree.
- The VANET includes overlapping communities.
- Tight communities scale at very small sizes.

#### E. Network resilience

The notion of network resilience to the removal of vertices is a very significant property of any VANET, since it directly impacts its cohesion (thus, the possibility of disconnection), and also the immunization of the network against malicious attacks to vehicles concerning their communication. Thus, the following question is particularly important:

*QUESTION 5: How robust is a VANET?*

If vertices are removed from the network, the typical length between pairs of vertices will increase. Thus, even though there are various ways to address the above question, in this article we adopt a simple, though powerful metric to quantify the network resilience [5]. We investigate how the diameter of the network changes when we remove nodes with the highest betweenness centrality values, and how the number of clusters change after these removals. Figure 18 illustrates the obtained results.

We investigated the increase/decrease of the network diameter when removing 10% of nodes with the highest betweenness centrality index. Figure 18 illustrates that the diameter change distribution follows a power-law with the relation  $N_i \propto diameter_i^\alpha$ , where  $N_i$  is the number of nodes remaining in the graph at the  $i$ -th removal of nodes and  $diameter_i$

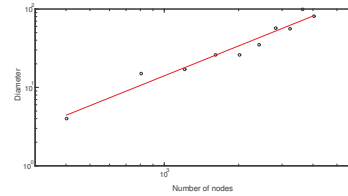


Fig. 18. Diameter vs. number of nodes in log-log scale. Slope  $\alpha = 1.26$  ( $T=100m$ ).

is the diameter of the network which has  $N_i$  nodes. The value of  $\alpha$  determines how robust the VANET is. If  $\alpha$  is close to 1 then the network can be considered as robust. According to our findings, *the VANET communication graph is not robust* since a significant loss of high quality nodes affects the characteristics of graph.

## VI. IMPLICATIONS ON PROTOCOL DESIGN

In the previous section we conducted a thorough analysis of the topological characteristics of a large scale VANET graph and gained a deep understanding of its shape. The question that remains to be answered is whether this information is useful from an engineering perspective. The astute reader will have already recognized many uses for our results. In the interest of space, we provide only some generic directions where our results can be useful.

One of the most sound observations of our study is that a VANET graph is quite dense. Therefore, protocols based on flooding will choke the network; thus, topology control algorithms are necessary. Though, the high mobility of the vehicles, make graph-based topology-control methods (spanning trees, Gabriel/Yao graph, RNG) not appealing; clustering is preferable. Which could be the clusterheads? Not necessarily high-degree nodes, but those with large betweenness centrality (if we need a few clusters), or those with high lobby centrality (if we need a lot of clusters). Similarly, the network’s density makes the use of power transmission adjustment mandatory. The difference in a VANET setting is that this procedure must be *continuous*, and can not be decided once in advance. Such continuous power adjustment is not easy to achieve, unless it is done on a per-road-segment basis, i.e., decide power transmission levels for segments of the roads, based on observed vehicle density (i.e., in places where vehicles’ communities exist). How to segment the roads, is a matter of investigation.

Article [9] asked the question: which nodes will be the forwarders in routing? Our study is able to provide an answer to this: we can draw such nodes among those with high centrality value. These nodes are also perfect candidates for message ferrying in case of network partitionings. Similarly, for geocasting applications, nodes with high lobby index are ideal for carrying out the rebroadcasts so as to spread the message with as few rebroadcasts as possible.

The very short lifetime of communication links makes the exchange of time-critical info problematic, thus the use of cooperative caching [11], [8] would be beneficial. Moreover,

for applications requiring awareness of the positions of other vehicles through periodic beacons, or the distribution of traffic related data through periodic beacons, the exploitation of the more “central” vehicles for these tasks could relieve the network from redundant broadcasts and reduce the collisions. Also, the roadside infrastructure is suggested. But, where to place these roadside units? In points where borders of clusters exist, or in places where vehicle communities exist. Similarly, installation of roadside units is suggested in places where the nodes have low localized clustering coefficient (sparse network), and thus the delivery of messages would require a significant amount of time without the infrastructure. Additionally, since the link lifetime is very short, it arises the need to design MAC protocols which, based on the prediction of link lifetime using information like the direction and velocity of vehicles, will prioritize broadcasts based on estimated link duration, or design appropriate handover techniques.

Finally, the existence of communities implies that mobility models like the Random Way Point, which are based on types of random walks should be abandoned, because they do not produce clusterings of the vehicles and additionally they do not support the existence of “hub” vehicles that explain the distributions of the centrality metrics. Therefore, research towards richer models (e.g., [21]) must be conducted. Furthermore, the existence of communities implies also that leader election algorithms will work successfully, especially if we incorporate the centrality metrics in their selection.

## VII. CONCLUSIONS

This paper provides a thorough study of the topological characteristics and statistical features of a VANET communication graph. Specifically, our work provides answers to some critical questions: How do VANET graphs evolve over time and space? Do the centrality metrics identify “quality” (more central) nodes, and what is the spatial distribution of these nodes? Which are the link duration statistics in VANET when the vehicles are moving in urban areas? Does the VANET consist of a single connected component? Are there any dense subgraphs inside the VANET? How robust is a VANET? We view our findings as particularly important since the obtained results have a wide range of implications upon the development of high-performance, reliable, scalable, secure, and privacy-preserving vehicular technologies.

**Acknowledgement:** This work was supported in part by the European Commission under the Seventh Framework Programme through the SEARCHiN project (Marie Curie Action, contract number FP6-042467) and the project “Control for Coordination of Distributed Systems”, funded by the EU.ICT program, Challenge ICT-2007.3.7.

## REFERENCES

- [1] N. Banerjee, M. D. Corner, D. Towsley, and B. N. Levine. Relays, base stations, and meshes: Enhancing mobile networks with infrastructure. In *Proceedings ACM MobiCom*, pages 81–91, 2008.
- [2] A. Chaintreau, A. Mtibaa, L. Massoulie, and C. Diot. The diameter of opportunistic mobile networks. In *Proceedings ACM CoNEXT*, 2007.
- [3] H. Conceição, M. Ferreira, and J. Barros. On the urban connectivity of vehicular sensor networks. In *Proceedings DCOSS*, volume 5067 of *LNCS*, pages 112–125, 2008.
- [4] E. M. Daly and M. M. Haahr. Social network analysis for routing in disconnected delay-tolerant MANETs. In *Proceedings ACM MobiHoc*, pages 32–40, 2007.
- [5] A. H. Dekker and B. D. Colbert. Network robustness and graph topology. In *Proceedings ACCS*, pages 359–368, 2004.
- [6] M. D. Dikaiakos, A. Florides, T. Nadeem, and L. Iftode. Location-aware services over vehicular ad-hoc networks using car-to-car communication. *IEEE JSAC*, 25(8):1590–1602, 2007.
- [7] N. Dimokas, D. Katsaros, and Y. Manolopoulos. Cooperative caching in wireless multimedia sensor networks. *ACM MONET*, 13(3–4):337–356, 2008.
- [8] N. Dimokas, D. Katsaros, L. Tassioulas, and Y. Manolopoulos. High performance, low overhead cooperative caching for wireless sensor networks. In *Proceedings of IEEE WoWMoM*, 2009.
- [9] V. Erramilli, M. Crovella, A. Chaintreau, and C. Diot. Delegation forwarding. In *Proceedings ACM MobiHoc*, pages 251–259, 2008.
- [10] M. Fiore and J. Härrri. The networking shape of vehicular mobility. In *Proceedings ACM MobiHoc*, pages 261–272, 2008.
- [11] M. Fiore, F. Mininni, C. Casetti, and C. F. Chiasserini. To cache or not to cache? In *Proceedings of IEEE INFOCOM*, 2009.
- [12] V. González, A. Alberto, Los Santos, C. Pinart, and F. Milagro. Experimental demonstration of the viability of IEEE 802.11b based inter-vehicle communications. In *Proceedings ICST TRIDENTCOM*, 2008.
- [13] J. Härrri, C. Bonnet, and F. Filali. Kinetic graphs: A framework for capturing the dynamics of mobile structures in MANET. In *Proceedings ACM PM2HW2N*, pages 88–91, 2007.
- [14] W. Hwang, T. Kim, M. Ramanathan, and A. Zhang. Bridging centrality: Graph mining from element level to group level. In *Proceedings ACM SIGKDD*, pages 336–344, 2008.
- [15] M. Kafsi, P. Papadimitratos, O. Dousse, T. Alpcan, and J.-P. Hubaux. VANET Connectivity Analysis. In *Proceedings of the IEEE Workshop on Autonet*, 2008.
- [16] Y. Karydis, A. Nanopoulos, A. Papadopoulos, D. Katsaros, and Y. Manolopoulos. Music retrieval over wireless ad-hoc networks. *IEEE TASLP*, 16(6):152–162, 2008.
- [17] D. Katsaros, G. Pallis, K. Stamos, A. Vakali, A. Sidiropoulos, and Y. Manolopoulos. CDNs content outsourcing via generalized communities. *IEEE TKDE*, 21(1):137–151, 2009.
- [18] A. Korn, A. Schubert, and A. Telcs. Lobby index in networks. *Physica A: Statistical Mechanics and its Applications*, 388(11):2221 – 2226, 2009.
- [19] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM TKDD*, 1(1), 2007.
- [20] P. Li, X. Huang, Y. Fang, and P. Lin. Optimal placement of gateways in vehicular networks. *IEEE TVT*, 56(6):3421–3430, 2008.
- [21] M. Musolesi and C. Mascolo. Designing mobility models based on social network theory. *ACM SIGMOBILE Mobile Computing and Communications Review*, 11(3):59–70, 2007.
- [22] V. Naumov, R. Baumann, and T. Gross. An evaluation of inter-vehicle ad hoc networks based on realistic vehicular traces. In *Proceedings ACM MobiHoc*, pages 108–119, 2006.
- [23] S. Pack, H. Rutagemwa, X. Shen, J. W. Mark, and K. Park. Proxy-based wireless data access algorithms in mobile hotspots. *IEEE TVT*, 57(5):3165–3177, 2008.
- [24] B. Raney, A. Voellmy, N. Cetin, M. Vrtic, and K. Nagel. Towards a microscopic traffic simulation of all of switzerland. In *Proceedings of the International Conference on Computational Science*, pages 371–380, 2002.
- [25] M. Raya, P. Papadimitratos, I. Aad, D. Jungels, and J.-P. Hubaux. Eviction of misbehaving and faulty nodes in vehicular networks. *IEEE JSAC*, 25(8):1557–1568, 2007.
- [26] K. Seada. Insights from a freeway car-to-car real-world experiment. In *Proceedings ACM WiNTECH*, pages 49–55, 2008.
- [27] V. Srivastava, A. B. Hilal, M. S. Thompson, J. N. Chattha, A. B. MacKenzie, and L. A. DaSilva. Characterizing mobile ad hoc networks: The MANIAC challenge experiment. In *Proceedings ACM WiNTECH*, pages 65–72, 2008.
- [28] Y. Toor, P. Mühlenthaler, A. Laouiti, and A. De La Fortelle. Vehicle ad hoc networks: Applications and related technical issues. *IEEE Communications Surveys & Tutorials*, 10(3):74–88, 2008.
- [29] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.
- [30] E. Yoneki. Visualizing communities and centralities from encounter traces. In *Proceedings ACM CHANTS*, pages 129–132, 2008.