Car Rank: An Information-Centric Identification of Important Smart Vehicles for Urban Sensing

Junaid Ahmed Khan, Yacine Ghamri-Doudane

CarRank: An Information-Centric Identification of Important Smart Vehicles for Urban Sensing

Junaid Ahmed Khan*† and Yacine Ghamri-Doudane†
*University Paris-Est, LIGM Lab, Marne-la-Vallée, France
†L3i Lab, University of La Rochelle, France
khan@univ-mlv.fr, yacine-ghamri@univ-lr.fr

Abstract—**Future cars are becoming powerful sensor platforms capable to collect, store and share large amount of sensory data by constant monitoring of urban streets. It is quite challenging to upload such data from all vehicles to the infrastructure due to limited bandwidth resources and high upload cost. This invoke the need to identify the appropriate vehicles within the Vehicular Ad-hoc Network, that are important for different urban sensing tasks based on their natural mobility and availability. This paper address this problem leveraging the self-decision making ability of a "Smart Vehicle" regarding its importance in the network. To do so, we present CarRank, an Information-Centric algorithm for a vehicle to first rank different location-aware information. It then uses the information importance, its spatio-temporal availability and neighborhood topology to analytically find its relative importance in the network. CarRank is the first step towards identifying the best set of information hubs to be used in the network for the efficient collection, storage and distribution of urban sensory information. We evaluate CarRank under a scalable simulation environment using realistic vehicular mobility traces. Results show that CarRank is an efficient ranking algorithm to identify socially important vehicles in comparison to other ranking metrics used in the literature.**

Keywords—*Information-Centric Vehicular Networking, Urban Sensing and monitoring, Vehicular Ad-hoc Networks*

I. INTRODUCTION

Nowadays vehicles are equipped with a lots of electronic components including sensors, cameras and communication devices to facilitate towards our utmost travel comfort and safety. Such "Smart Vehicles" can be considered as an instance of Internet of Things (IoT) aimed to harvest and share different sensory and multimedia data from urban streets supporting various Intelligent Transportation System (ITS) applications such as efficient traffic management and urban environment sensing. Urban Sensing and vicinity monitoring using vehicles has attracted lots of researchers in the past few years and several schemes are proposed [1] [2] [3], where sensor-equipped vehicles sense and share data in a vehicular network.

At the same time, the issue is that each vehicle is constantly generating a tremendous amount of data that cannot be uploaded to the cloud or Internet due to its large volume. Moreover, most of the generated content is of "local relevance" as the intended users lies within the vehicular network. Relying on the infrastructure network for the collection, storage and distribution of such heterogeneous *Big-Data* from vehicles can thus prove costly and inadequate to its usage. Pre-advertising or broadcasting all the sensing data from each vehicle would result in a massive advertising overhead and a redundant information storm within the network. The major problem is

to efficiently locate and collect the user relevant data from the fleet of vehicles with the underlying challenge of intermittent connectivity and mobility in a Vehicular Ad-hoc Network (VANET).

This motivates the need to identify important vehicles to be recruited as distributed information hubs based on their daily commute and their social importance with respect to the frequently visited neighborhoods. To identify important nodes, network analysis typically rely on different variants of centrality measures such as Degree, Closeness, Betweenness, Eigenvector centrality. However, such schemes are difficult to use in the sporadic vehicular network topology. Therefore, the challenge is to find the right vehicle available at the right time and place for efficient data collection, storage and distribution through low-cost inter-vehicle communications.

To address this problem, for the first time in vehicular networks, we propose a new concept of finding important vehicles, allowing a smart vehicle to rank itself based on its popularity with respect to the user interests, spatio-temporal availability and its neighborhood in an urban scenario. We envision such vehicles as buses, taxis, commuters available to address user interests in the network. Therefore, the target of this paper is to introduce an innovative vehicle ranking algorithm "CarRank" for the identification of Information Facilitator Vehicles (IFVs), responsible for the efficient gathering, storing and publishing of urban sensing data. The vehicle first ranks the information associated to it taking into consideration the relevance to the users interest. It then considers the associated location-aware information popularity to find its relative importance in the network using CarRank algorithm as its *vehicle centrality*.

Recently, we observe a shift towards Information-Centric Networking (ICN) [4] in [5] and [6] as the underlying routing protocol for vehicular networks. ICN is a content-centric networking architecture proposed to replace the current IP based Internet. In ICN, a user broadcasts an interest for a content by its name, any corresponding host in the network replies back with the desired content. Additionally, it offers In-Network caching at intermediate nodes while forwarding and responding to subsequent user interests. Therefore, CarRank considers the importance of the location-aware information associated to vehicles as an information-centric approach instead of relying on physical hosts in the ephemeral vehicular network topology. The major contributions to this paper can be summarized as follows:

- We propose a novel algorithm enabling a vehicle to

rank important location-aware information associated to it based on the satisfied user interests, without relying on any infrastructure network.

- An innovative vehicle ranking algorithm, "CarRank", is proposed, where each vehicle can find its importance in the network. This importance is linked to the importance of the associated information, vehicle spatio-temporal availability and the neighborhood topology.

- We validate the scalability as well as ICN compliance for both algorithms by performing extensive simulations comprising around three thousand vehicles using realistic mobility traces to identify popular IFVs in time evolving VANETs.

The obtained results show that the proposed algorithms are well suited to help in the efficient identification of the best IFVs in the network using information-centric vehicular networking. The rest of the paper is organized as follows. The next Section highlights the major related work. In Section III, we present the information-centric CarRank algorithm to analytically compute the vehicle centrality. Section IV explains the performance evaluation discussing the simulations. In Section V we conclude the paper along some insight for future research.

II. RELATED WORK

Identification of influential information hubs for publishing/spreading information is required in applications such as social networks. Another interesting application is found in medical sciences to find epidemic disease spreaders [7]. Similarly, Google's PageRank [8] algorithm ranks the importance of a web-page in an Internet search based on the number of web links directed towards it. The above applications usually rely on well known network centrality schemes such as Degree, Closeness, Betweenness and Eigenvector centrality.

Degree centrality considers the number of direct (one hop) neighbors of a node, where Closeness centrality is the inverse of the sum of the lengths of the shortest paths from a node to the rest of the nodes in the network. Betweenness centrality is the fraction of all pairs of shortest paths passing through a node, where Eigenvector centrality is the node's influence measure in the network [9]. By tweaking these centrality measures, algorithms such as BubbleRap [10] and ML-SOR [11] are proposed, where nodes with high centrality score are preferred for data dissemination and routing in Opportunistic Mobile Social Networks.

Unlike the above mentioned applications, it is unfeasible to use centrality-based popularity schemes to find the best information hubs in VANETs for multiple reasons; First, The rapid topological changes due to the high mobility of vehicles requires a continuous time varying analysis of the VANETs which is unfeasible by a practical scheme. Indeed, typical schemes assume a static graph topology with respect to time where the temporal network characteristics of VANETs would be ignored. Second, centrality measures such as Betweenness, Closeness and Eigenvector centrality computation requires network wide parameters, while in VANETs, a vehicle cannot obtain such information to make run-time decisions. Third,

existing schemes consider shortest path metric to compute a node's importance, while the highly dynamic VANET topologies does not ensure the existence of a stable path between nodes. Therefore, a new vehicle ranking algorithm adapted to VANETs and enabling vehicles to decide their relative importance in the network by overcoming the above mentioned constraints need to be thought about.

III. CARRANK: A FULLY DISTRIBUTED APPROACH

We present CarRank as the vehicle centrality measure enabling each vehicle to autonomously find its importance in the network. It is difficult to use the vehicle contact frequency and duration to decide its importance in the network due to the rapid changes in the time evolving vehicular network topology. To overcome this, CarRank simultaneously considers three novel albeit essential parameters, the information importance, the vehicle spatio-temporal availability and its network connectivity. Additionally, the user's interest satisfaction for a content is also considered as a key metric for a vehicle's importance as it regularly responds to user interests. The interests are assumed to be generated and received from the neighboring vehicles using multi-hop interest forwarding. We consider the following local parameters known to the vehicle for analytically finding its importance:

Information Importance: Information importance measures vehicle relevance to users for a particular content, i.e. The interest-response frequency is a vital factor to classify a content's importance. A vehicle associated to contents related to popular locations is considered as an important information hub in the network.

Spatio-Temporal availability: It reflects the social-behavior based on the vehicle's habitual routes as a factor of the daily commute. Spatial availability reflects the vehicle's recursive presence in an area, while temporal availability refers to its relevance in time for a location.

Neighborhood Importance: Neighborhood importance shows vehicle topological connectivity in order to be capable perform various sensing tasks. An easily reachable and well connected vehicle in a network topology can act as an efficient IFV.

The following section defines the network model enabling the vehicle to compute its respective CarRank score.

A. Network Model

We consider a time varying VANET modeled as an undirected vehicular graph $G(\mathbb{V}(t), \mathbb{E}^v(t))$, where $\mathbb{V}(t) = \{v\}$ is a set of vertices v, each representing a vehicle on the road at time t. $\mathbb{E}^v(t) = \{e_{jk}(t) \mid v_j, v_k \in \mathbb{V}, j \neq k\}$ is the set of edges $e_{jk}(t)$ modeling the existence of a direct communication link between vehicles v_j and v_k at time t. The number of edges $\mathbb{E}^v(t)$ depends on the transmission range of each vehicle as shown in Figure 1a. We assume it as a simple unit disk model bounded by its communication range. The city map is represented by the undirected graph $G(\mathbb{X}, \mathbb{E}^x)$ as in Figure 1b, the set of vertices $\mathbb{X} = \{x\}$ contains different urban zones x and the set of edges $\mathbb{E}^x = \{e_{pq} \mid x_p, x_q \in \mathbb{X}, p \neq q\}$ are their respective boundaries that connects different zones through the underlying road network.
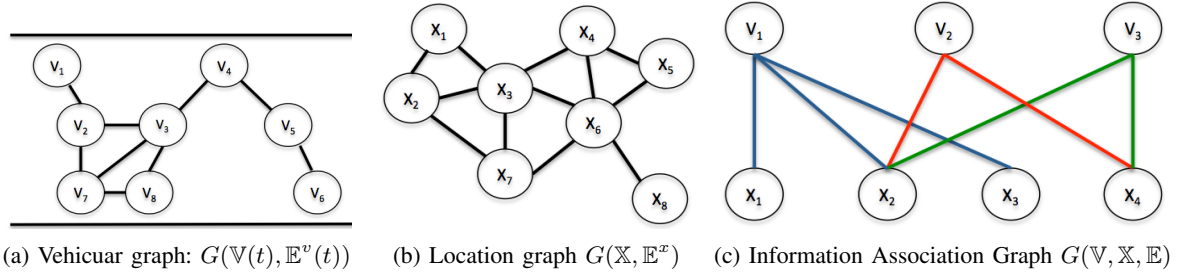
(a) Vehicuar graph: $G(\mathbb{V}(t), \mathbb{E}^v(t))$ (b) Location graph $G(\mathbb{X}, \mathbb{E}^x)$ (c) Information Association Graph $G(\mathbb{V}, \mathbb{X}, \mathbb{E})$

Figure 1: Network Model

Information Association: Information association is defined as a bipartite graph $G(\mathbb{V}, \mathbb{X}, \mathbb{E})$, where $\mathbb{V}$ is the set the vertices in the vehicular graph $G(\mathbb{V}(t), \mathbb{E}^v(t))$ and $\mathbb{X}$ is the set of locations in the city map $G(\mathbb{X}, \mathbb{E}^x)$ as shown in Figure 1c. The edge $\mathbb{E} = \{e_{ij} \,|\, v_i \in \mathbb{V}, x_j \in \mathbb{X}\}$ associates each vehicle to a set of regions $X_v \subset \mathbb{X}$ with respect to the user relevant content.

The associated information is classified by clustering the regions using ICN hierarchical naming convention *"/region/road-section/information-type"*. Information type comprises different Intelligent ITS applications (Safety warnings, Road congestion information, Infotainment...) with varying content popularity and priority. The regions are clustered using voronoi tessellation [12] where the vehicles concentrated in an area are considered in a single voronoi region $x \in G(\mathbb{X}, \mathbb{E}^x)$. The advantage of voronoi zones is that it retain the realism of city road structure, while covering the entire map. Additionally, it provides us an abstraction for the user relevant information regarding different neighborhoods, while not relying on individual road-sections.

For temporal VANET analysis, we divide the time $T = (\overline{t_1}, \overline{t_2}, ...)$ into a sequence of regular time-slots, where the kth time-slot is $\overline{t_k} = [t_k, t_{k+1})$. Each vehicle finds its centrality at the time instant t_{k+1} from the known information in the current time-slot, where t_k is the time instant at the beginning of the time-slot $\overline{t_k}$. We will refer to content/information or location/areas/zones interchangeably in the text since contents are associated to locations in the urban map.

The content distance $d(x, x_k)$ is the Euclidean distance between the content location x and x_k, the vehicles current position at time instant t_{k+1}, where $x, x_k \in X_v$. We assume each vehicle knows the city map, i.e. the graph $G(\mathbb{X}, \mathbb{E}^x)$. However, its knowledge is limited to the information regarding the locations relevant to itself. This is due to each vehicle's limited storage and coverage scope based on the driver's daily commute as it is unlikely for instance that it visited, and thus stored data, of all the traversable roads in the city.

B. Information Importance

We assume vehicles in a distributed VANET encountering each other constantly receiving interests from neighboring vehicles for different location-dependent information. Some of such information can be of more importance to the intended users in the network which can be easily identified by the vehicle by the amount of user interests received for it. We assume that it is capable of recording the time and position each time it responds as the content provider to a user interest.

Table I: List of Notations

Notation	Description
$\mathbb{V}$	Set of vehicles
$\mathbb{X}$	Set of locations/regions
$\mathbb{E}^x/\mathbb{E}^v$	Set of edges between locations/vehicles
$\mathbb{E}$	Edge between vehicles and locations
$\overline{t}_k$	Time-slot k for CarRank computation
t_k/t_{k+1}	Current time instant/next time instant
X_v	Set of locations associated to vehicle v
$d(x, x_k)$	Distance from current location x_k to x
I_x^v	Interests satisfaction frequency for x
r_x	Number of successful responds for x in the previous slot
R_x	Total successful responds for x
R_T	Vehicle responds count for all contents
t_x^f	Last successful respond time for x
$\overline{t}_d$	Average interest deadline
n	Total received interests in the previous slot
t_x	Interest validity deadline for content x
τ	Information timeliness
δ	Tuning parameter for information validity
C_x^v	Content x importance for vehicle v
λ	Tune importance based on distance from x
s_x^v	vehicle reliability as content source for x
w_x	Information x weight with respect to vehicle
f_I^v	Information importance function
$p_x^v(t_k, x_k)$	Probability of satisfying interests for location x at current time t_k and position x_k
$R_x^v(t_k, x_k)$	Interests satisfied for content x at current time t_k and position x_k
$I_x^v(t_k; x_k)$	Mutual information shared between the current time and location for content x
$p_x^v(t_k)$	Marginal probability of interest responds at current time
$p_x^v(x_k)$	Marginal probability of interest responds at current location
$f_{T,X}^v$	Vehicle spatio-temporal availability function
k_v	Vehicle degree (number of neighbors)
k_Γ^v	Vehicle average neighbor degree
Γ_v	Set of neighbors for vehicle v
C_Γ^v	Neighbor vehicle centrality
f_Γ^v	Vehicle neighborhood importance function
C_v	Vehicle centrality
$\alpha/\beta/\gamma$	Tuning parameters for each function
θ	Smoothing factor for vehicle centrality

Therefore, a vehicle considers an information as popular if it observes an increase in the number of user interests for it. For this reason, information importance takes into account the vehicle latent ability to satisfy more user interests with its natural mobility pattern.

Definition 1: (Interest Satisfaction Frequency) We define $I_x^v(\overline{t}_k) = \frac{r_x(\overline{t}_k)}{R_x}$ as the frequency of user interests satisfied in the time-slot $\overline{t}_k$, where $r_x(\overline{t}_k)$ are the number of successful responds in the previous slot and R_x are the total successful responds for the content $x \in X_v$ associated to vehicle $v \in \mathbb{V}$.

The importance of each location-aware content is periodically updated based on the interest satisfaction frequency by the vehicle. Interest for each content specify a temporal scope for the information validity, For instance, road congestion information is only valid during congestion. Therefore, in order to ensure the information importance is not substantially increased after the desired deadline, let t_x^f be the last successful respond time for the content x and the average interest deadline as $\bar{t}_d = \frac{1}{n} \sum_n t_x$ associated with each content, where n are the total number of interests in the previous time-slot and t_x is the deadline of each interest for the content x.

Definition 2: (Information Timeliness) The information timeliness defined as $\tau(t_{k+1}) = \begin{cases} 1 & t_{k+1} \leq t_x^f + \bar{t}_d \\ e^{-\delta \bar{t}_d} & t_{k+1} > t_x^f + \bar{t}_d \end{cases}$ is the measure of the temporal information validity scope where $\delta \in [0,1]$ is the tuning parameter depending on the application needs (E.g. maximum 1 hour for accident information validity).

For each information type, the information timeliness parameter τ considers its validity at the importance computation time instant t_{k+1}. If there are no active interests in the previous slot and the average interest validity deadline has passed, the information importance follows an exponential decay since the information is of less importance in the network. On the other hand, τ is set to unity for the information required to be always available in the network.

The corresponding information importance at the next time instant t_{k+1} is updated as follows:

$$C_x^v(t_{k+1}) = C_x^v(t_k) + \tau(t_{k+1})I_x^v(\bar{t}_k)(1 + d(x,x_k))^{-\lambda} \\ + s_x^v(t_{k+1}) \quad (1)$$

The information importance depends on the its value $C_x^v(t_k)$ at the beginning of the time-slot (time instant t_k). If a content is not responded in the previous slot, then $I_x^v(\bar{t}_k) = 0$ avoids unnecessary increase in the information importance. The term $s_x^v \in [0,1]$ represents the percentage of time the vehicle itself acted as the original source for the content x. It is updated regularly to ensure the content relevant to the vehicle retain its association in case it does not respond in the previous slot. Thus, the interests for a particular content later in time could finally route to the original source vehicle in the network. The tuning parameter $\lambda \in [0,1]$ decides the effect of the vehicle distance from the associated content location.

Algorithm 1 shows the steps allowing a vehicle to find the respective information importance. The vehicle considers its distance from the information location, its reliability as the content source and the information timeliness to find each location-aware content importance. Additionally, It computes the interest satisfaction ratio as the key metric, based on which it decides to rank each associated content. The information importance is then updated for the next time-slot at time instant t_{k+1} using (1). The corresponding information importance of all locations is then used to find the vehicle information importance function. It measures the vehicles ability to respond to interests taking into account the importance of each location. The vehicle finds the information-centric centrality function using the relation below:

Algorithm 1 Information Importance

INPUT: $G(\mathbb{V}, \mathbb{X}, \mathbb{E})$: information association graph
OUTPUT: Updated information importance for the next time-slot at time-instant t_{k+1}
for each vehicle $v \in \mathbb{V}$ **do**
  for each content $x \in X_v$ in cache **do**
    Find $d(x, x_k), \tau(t_{k+1}), s_x^v(t_{k+1}), w_x$
    Compute $I_x^v(\bar{t}_k) \leftarrow \frac{r_x(\bar{t}_k)}{R_x}$
    if $I_x^v(\bar{t}) \neq 0$ **then**
      Update $C_x^v(t_{k+1})$ using (1)
    else
      $C_x^v(t_{k+1}) = C_x^v(t_k) + s_x^v(t_{k+1}),$
    end if
  end for
end for
return $C_x^v(t_{k+1})$

$$f_I^v(t_{k+1}) = \frac{1}{|X_v|} \sum_{x \in X_v} C_x^v(t_{k+1}) \cdot w_x \quad (2)$$

For all content $x \in X_v$ associated to v, $C_x^v(t_{k+1})$ is the respective information importance at time instant t_{k+1}, where $w_x = \frac{R_x}{R_T}$ is the weight of interest satisfied for the content x among all the content in cache. R_x represents the number of responds for x and R_T is the number of responds for all content associated to vehicle v. The term $|X_v|$ is the cardinality of the sub-graph $X_v \subset \mathbb{X}$, i.e. set of all regions associated to the vehicle v.

C. Spatio-temporal Availability

Spatio-temporal availability of the vehicle reflects the driver social behavior. It considers the vehicle physical availability in an area while taking into account different times of the day. For example, we drive the same route around the same time of the day to go to places we visit habitually such as our work place or the gym. Users are likely to be located in the same city neighborhood which is related to their daily routine. The challenge lies in the fact that each user natural mobility scope is bounded by the geographical regions that are only relevant to its daily commute, thus making it difficult to derive a distributed method to find its importance without relying on the complete network topology.

However, to incorporate such social behavior, we borrow tools from information theory to find to what extent the current time and location contribute to the vehicle's importance. Since the vehicle does not have network-wide information to find its relevance, we continue with our proposed interest satisfaction ratio based assumption. The probability of the vehicle v satisfying interests for content for location x at the current time t_k and position x_k is $p_x^v(t_k, x_k) = \frac{R_x^v(t_k, x_k)}{R_T}$, where $R_x^v(t_k, x_k)$ are the interests satisfied for content x at the current time and location in the past and R_T are the total interests satisfied by the vehicle v. The current time in the past refers to the time-slot around the same time in the day for all the days before the present day Y with respect to each content as shown in Figure 2. For example, for finding the spatio-temporal availability between 7 AM and 8 AM, it compares the interest satisfied
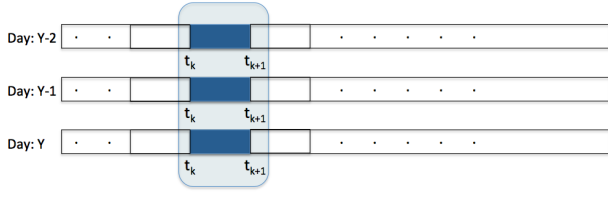
Figure 2: Spatio-temporal availaibility in the same time-slot

Figure 3: Neighborhood Centrality Exchange

in the same area around 7 AM and 8 AM in the past for all content in cache.

For content associated to location x, the mutual information shared between all the correlating time-slots and the locations is:

$$I_x^v(t_k; x_k) = \sum_{\forall t_k \in T} \sum_{\forall x \in X_v} p_x^v(t_k, x_k) \log \left(\frac{p_x^v(t_k, x_k)}{p_x^v(t_k) p_x^v(x_k)} \right),$$
(3)

where $p_x^v(t_k)$ and $p_x^v(x_k)$ are the marginal probabilities of the content responds in the current time and the current location, respectively. Now, the vehicle finds its spatio-temporal availability function for all locations:

$$f_{T,X}^v(t_{k+1}) = \frac{1}{|X_v|} \sum_{x \in X_v} I_x^v(t_k; x_k) . w_x$$
(4)

The function in (4) indicates a vehicle's importance at the time and position of CarRank computation. If it correlates more to the associated contents at the current time and location, it counts more towards computing its respective score at the same hour of the day and the same location.

D. Neighborhood Importance

The neighborhood of the vehicle in a distributed system is important for efficient content distribution and storage. We incorporate the neighborhood information by letting the vehicles in transmission range share their respective importance as well as their connectivity information. The idea is to consider better connected vehicles with better spreading capabilities. This instantiates the use of the vehicles physical topological information. For this purpose, we consider the vehicle's assortativity as its average neighbor degree k_Γ^v. Besides topological connectivity, each neighbor centrality C_Γ^v within communication range at time t_k is also taken into account. The neighborhood importance function for the time-slot t_{k+1} is expressed as:

$$f_\Gamma^v(t_{k+1}) = \frac{1}{k_v} \sum_{\Gamma_v \in \mathbb{V}} C_\Gamma^v(t_k) \cdot k_\Gamma^v$$
(5)

where k_v is the vehicle degree at time t in the graph $G(\mathbb{V}(t), \mathbb{E}^v(t))$. Since it is impossible to use any network-wide centrality measure unknown to the vehicle at the time of importance computation. Therefore, the function $f_\Gamma^v(t_{k+1})$ in (5) considers more information than just the degree of the vehicle while maintaining a local scope, thus, relying only on local information within the vehicle range as shown for the node V_3 in Figure 3.
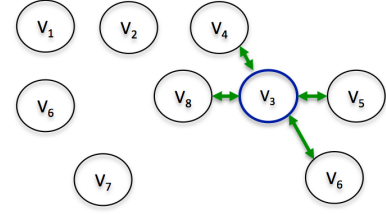
E. Vehicle Centrality

The vehicle centrality for the next time instant t_{k+1} is updated as the Exponential Weighted Moving Average (EWMA) function of the current and previous vehicle centrality, where (2), (4) and (5) contributes to the overall CarRank computation:

$$\begin{aligned} C_v(t_{k+1}) = \theta C_v(t_k) + (1-\theta)[\alpha f_I^v(t_{k+1}) \\ + \beta f_{T,X}^v(t_{k+1}) + \gamma f_\Gamma^v(t_{k+1})] \end{aligned}$$
(6)

Algorithm 2 CarRank

INPUT: Information association graph $G(\mathbb{V}, \mathbb{X}, \mathbb{E})$:
OUTPUT: $C_v(t_{k+1})$: Updated CarRank for the next time-slot t_{k+1}
for each vehicle $v \in \mathbb{V}$ **do**
  for each associated content $x \in X_v$ **do**
    Compute associated information importance using (2)
    Compute mutual information with respect to the content from (3)
  end for
  for each neighbor vehicle $\Gamma_v \in \mathbb{V}$ **do**
    $k_\Gamma^v \leftarrow$ average neighbor degree
    $C_\Gamma^v(t_k) \leftarrow$ neighbor centrality
  end for
  Find spatio-temporal availability using (4)
  Compute neighborhood importance using (5)
  Update vehicle centrality from (6)
end for
return $C_v(t_{k+1})$

Each function's contribution is normalized by the terms α, β and γ, where $\alpha + \beta + \gamma = 1$, where $\theta \in [0, 1]$ allows the vehicle to increase its importance with respect to the previous time-slot. The impact of each parameter differs with respect to different applications. For example, if the vehicle is located in a better connected neighborhood, it can easily spread information. Therefore, the corresponding vehicle weights the information importance along the neighborhood more than the spatio-temporal availability.

The different steps required for a vehicle to find its Car-Rank score are described in Algorithm 2. Using the importance of all the associated contents, the vehicle finds its CarRank score from the respective information importance function obtained using (2). We also need the vehicle spatial and temporal availability by finding the mutual information shared between the current time and location for all the associated contents. The vehicle finds its topological connectivity measure from the neighborhood. It exchanges the average neighbor degree along the centrality score with the neighboring vehicles

Table II: Simulation Parameters

Parameter	Value
Simulation platform	NS-3
Number of nodes	2986
Mobility trace	Cologne, Germany
Area	$6X6km^2$ city center
Duration	1 hour
Communication range	100m
Packet size	1024 bytes
Time granularity	1 sec
Simulation Runs	5

in range. The information importance function, spatio-temporal availability and neighborhood all-together contribute to the vehicle centrality score for the next time-slot.

CarRank metrics described above are defined as the local scope of the information relevance with respect to a particular location in time and space. Regular visits to popular locations at well interesting time of the day will increase the vehicle's importance in the network. However, the vehicle global mobility pattern in a city is bounded by the regions only known to the vehicle (visited before). Thus, the more the number of popular locations visited by the vehicle, the more it increases its centrality. Moreover, Stale information is automatically deleted from the cache after some time due to the limited size storage buffer at vehicles. One should note that different cache management schemes are developed in ICN which are not the scope of our work [13].

IV. PERFORMANCE EVALUATION

One of the basic requirement for evaluating the efficiency of CarRank is scalability. Therefore, we use Network Simulator-3 (NS-3) as a scalable simulation platform for upto three thousand vehicles. The performance of CarRank is validated by a set of simulation runs under a realistic mobility scenario using traces from Cologne, Germany. To the best of our knowledge, it is considered as the most accurate mobility trace available for Vehicular Networks [14]. The vehicle availability as well as its mobility pattern is extracted using this trace. The simulation parameters are summarized in Table II, followed by a description of the simulation scenarios implemented for the performance evaluation.

A. Simulation Scenario

We simulate a VANET urban sensing scenario using the ndnSIM [15] module available for NS-3. ndnSIM integrates the Named Data Networking (NDN) communication model where the name based architecture replaces the traditional IPv4/v6 NS-3 network-layer modules. The simulation scenario implements the following two applications:

Producer: We define a producer vehicle to be the content source in the network. The areas visited by a vehicle in a time-slot before the CarRank computation time are considered as locations associated with the producer.

Consumer: Consumer vehicles are the potential user nodes planning to visit an area. Each consumer vehicle generates an interest for a content associated to a location in the city, which is routed to producer vehicles.
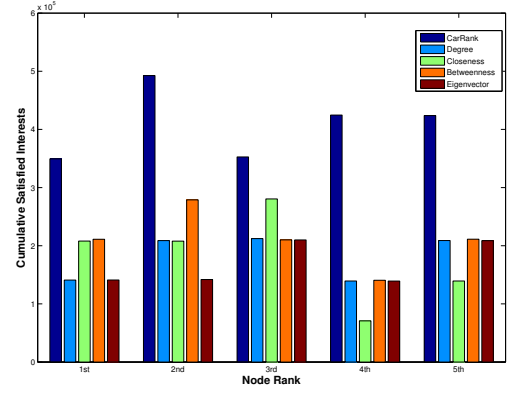

Figure 4: Cumulative Satisfied Interests by top identified nodes using each scheme

Table III: CarRank in different set of Simulations

Simulation	1		2		3		4		5		
Car Rank	ID	Score	ID	Score	ID	Score	ID	Score	ID	Score	Mean
1	764	1	210	1	1179	1	271	1	36	1	1
2	1356	0.8182	178	0.6978	907	0.9566	295	0.9657	395	0.6212	0.8119
3	294	0.8177	298	0.6770	444	0.8568	595	0.8329	1902	0.5166	0.7402
4	46	0.7831	424	0.6012	2276	0.8511	1179	0.8116	1926	0.4410	0.6976
5	1454	0.7361	428	0.5701	682	0.8384	1926	0.7878	1147	0.4166	0.6698
6	169	0.7289	132	0.5420	2325	0.8285	2300	0.7709	46	0.4127	0.6566
7	969	0.7287	444	0.5361	653	0.8255	2064	0.7642	1384	0.3991	0.6523
8	1384	0.7185	270	0.5149	2527	0.8210	2436	0.7613	895	0.3903	0.6412
9	949	0.7174	39	0.4952	1581	0.7907	46	0.7375	2251	0.3559	0.6193
10	1115	0.7157	169	0.4934	399	0.7532	1386	0.6734	477	0.3367	0.5945

We assume the interests follow a Zipf distribution, where interests for popular contents are more frequent. The city map is divided into zones/areas as voronoi tessellation where vehicles in proximity of each other by average values of their coordinates are co-located within the same voronoi region at the current time-slots. Any producer acting as source for an area upon receiving the interest responds with the desired content. For each vehicle, the vehicle centrality is computed at regular instants using Algorithm 2. We perform each simulation upto five times by analyzing different set of nodes as information producers and consumers. The tuning parameters α, β and γ are set to 0.33 to maintain generality since the significance of each parameter depends on the application requirements. For each simulation scenario, we rank the top information facilitator vehicles in the network by comparing their CarRank score with the respective Degree, Closeness, Betweenness and Eigenvector centrality score.

B. Simulation Results

The objective of our simulation study is to find answers to the fundamental question: *How well can CarRank identify the top IFVs?* We highlight the top 10 IFVs after each simulation run with their respective CarRank score in Table III. The last two column shows the average score at each rank and the confidence interval for a confidence level of 95%. Trace results for all the vehicles are obtained from five set of simulation runs but we are interested to analyze only the top IFVs. The centrality score is normalized with respect to the top identified node, i.e. the top IFV will have unity score followed by the relative score of other vehicles. We will use the same convention to interpret results in the later sections. In the first simulation run, the vehicle 764 is identified to have the top

CarRank score among all the vehicles in the network. One reason is that it responded more frequently to the incoming interests throughout the simulation for the respective associated locations. At the same time, its spatio-temporal availability with respect to the associated content and the neighborhood connectivity also contribute towards its score.

For better analysis of the performance of CarRank in different simulation scenarios, we consider the following performance metrics in comparison with the state of the art centrality schemes (Degree, Closeness, Betweenness and Eigenvector centrality):

- Cumulative Satisfied Interests (CSI) for the top identified nodes by each scheme

- Comparison of top nodes identified by each scheme with their respective centrality scores

- Average aggregated throughput of the identified top ranked nodes by each scheme

- Cache hit rate for the top nodes by each scheme to evaluate CarRank along ICN in VANET mobility scenarios

1) Cumulative Satisfied Interests: Cumulative Satisfied Interests refers to the total number of user interests satisfied during simulation duration. We compare the CSI score of CarRank with the state of the art centrality schemes. Figure 4 shows the CSI score of the top five nodes identified by all these schemes in an average of five set of simulations. Typical ranking schemes only takes into account physical topology towards computing a node importance in the network, ignoring the satisfied user interests. Nevertheless, the tops IFVs identified by CarRank satisfied more user interests than other schemes in all the five set of simulations due to the consideration of user interest satisfaction for important information as a key factor towards vehicle centrality in the network.

2) Temporal behavior analysis of top nodes: The time varying behavior of the relative score of the top five nodes identified by all schemes are shown by periodic network snapshots after each 15 minutes interval in Figure 5. We consider the top node identified by each scheme as benchmark by assigning it a unity score. We identified an outlier node 764, persistently ranked as top IFVs by CarRank, thought the other schemes underrated it. This is because we consider relatively stable factors such as the importance of associated information and the vehicle spatio-temporal availability besides the topological information. Vehicles also change places along the ranking order. For example, the node 294, ranked 3rd in the first half hour swap place with the node 46 around 45 minutes and finally retake the 3rd place.

An interesting results was observed in the first half hour: Only one node yields a high Eigenvector centrality score followed by other nodes with a negligible Eigenvector centrality score. We investigate this effect and found that the principle eigenvalue yields the top nodes where the eignvector is shifted towards the principle component. Thus, providing one major central node. This shows that the famous Eigenvector centrality fails to assign significant score to a large fraction of nodes in a large network, while CarRank does not present such behavior. Other centrality schemes result in different set of top nodes at every snapshot. It is because such schemes only consider the instantaneous shortest paths towards ranking the vehicles at a particular time instant which require the complete topological information. However, such complete network information is not available to an individual vehicle in highly unstable VANETs. CarRank ensures more stable set of top IFVs as it is clear from the time varying VANET analysis that it is not affected by the network dynamics since we are able to rank each vehicle considering relatively stable metrics, which is not the case for other schemes.

3) Aggregated Per Node Throughput: We also evaluate the ranking scheme by analyzing the throughput at important nodes in the network. Figure 6 shows the aggregated per node throughput of the top nodes identified by each scheme. The average aggregated throughput (Kbps) is computed over the entire simulation duration for five set of simulations. The top nodes identified by CarRank yields more throughput compared to other schemes. We also observe that the throughput of the fourth node is relatively higher, thus inferring a variation between different ranks. Similar variations are seen for Degree, Betweenness and Eigenvector centrality. However, Closeness centrality follow a decrease along the vehicles ranking order. CarRank outperformed all schemes as it incorporates additional factors towards vehicle importance computation such as the information importance and the spatio-temporal availability, while other schemes rely only on topological measures (node degree or shortest paths) towards vehicle importance computation.

4) ICN Evaluation - In-Network Caching: We evaluate the top IFVs for the ICN built-in feature of In-Network caching at nodes. For this purpose, we computed the cache hit rate at the top five nodes identified by each scheme. A second successful response by a node for the same content is considered a cache hit. The cumulative cache hit rate is computed for the entire simulation duration for five set of simulations. Figure 7 shows the cache hit rate for top nodes identified by each scheme. The top nodes identified by CarRank yield a higher hit rate than all the other schemes in all the five simulations. This is because CarRank considers content popularity as a key factor, thus, the vehicle containing important information responds and subsequently cache more frequently compared to other vehicles. Moreover, the top node identified by CarRank cached more important content due to their better neighborhood and spatio-temporal availability. This proves that In-Network caching offered by ICN in CarRank implementation overcomes the mobility and intermittent connectivity constraints in VANETs for efficient content access.

Finally, we are able to comment on the question we posed in the beginning of this section: *How well can it identify the top IFVs?* From the simulation results, It is evident that a relatively stable set of top IFVs are identified by CarRank compared to the other schemes. It is also clear that CarRank identify nodes which satisfied more user interests with higher aggregated per node throughput and more cache hit rate compared to the other schemes. Thus, the overall comparative analysis of CarRank with different network ranking schemes in the literature suggests it as an efficient vehicle ranking algorithm.
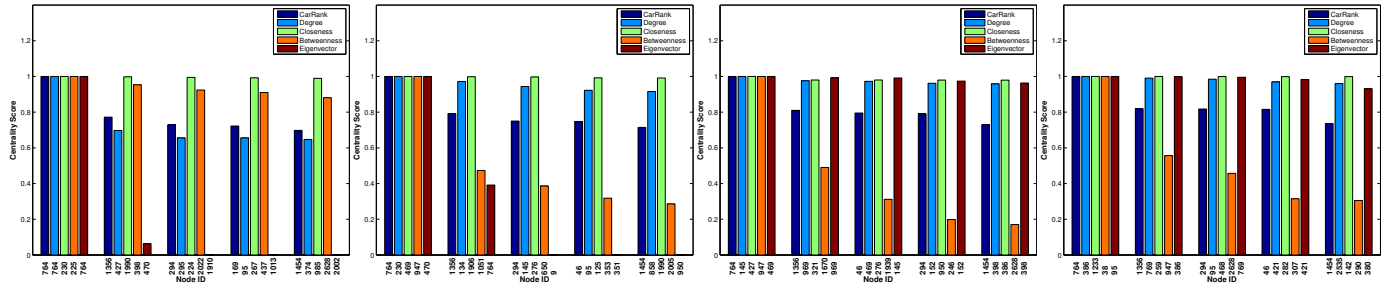
Figure 5: Temporal Snapshots of comparing top nodes identified by all schemes (CarRank, Degree, Closeness, Betweenness, Eigenvector centrality)
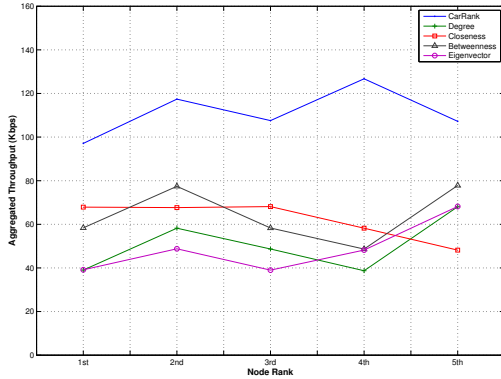

Figure 6: Average aggregated throughput by the top identified nodes using each scheme in five simulations
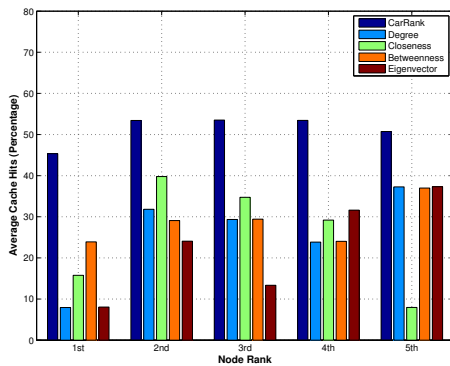

Figure 7: Average cumulative cache hit rate by the top identified nodes using each scheme in five simulations

V. CONCLUSIONS AND FUTURE DIRECTIONS

This paper introduced an innovative vehicle ranking algorithm "CarRank", allowing smart vehicles to rank themselves in a fully distributed VANET based on user relative importance. The vehicle considers the importance of location-aware information, its neighborhood topology along its spatio-temporal availability to find its importance in the network. The best ranked vehicles for urban sensing are then selected to perform different city-wide urban sensing tasks. Results by comparing with state of the art centrality schemes revealed that CarRank is best suited to efficiently identify important information facilitator vehicles in VANETs compared to other centrality schemes. Identification of important vehicles can be helpful in different smart city applications such as urban sensing mentioned in the paper. Popular IFVs can be used for efficient data collection, storage and distribution in VANETs. Designing such efficient schemes will be the subject of our future research.

REFERENCES

[1] U. Lee, B. Zhou, M. Gerla, E. Magistretti, P. Bellavista, and A. Corradi, "Mobeyes: smart mobs for urban monitoring with a vehicular sensor network," *Wireless Communications, IEEE*, vol. 13, no. 5, pp. 52–57, 2006.

[2] H. Wang, Y. Zhu, and Q. Zhang, "Compressive sensing based monitoring with vehicular networks," in *INFOCOM, 2013 Proceedings IEEE*. IEEE, 2013, pp. 2823–2831.

[3] R. Bruno and M. Nurchis, "Efficient data collection in multimedia vehicular sensing platforms," *Pervasive and Mobile Computing*, 2014.

[4] B. Ahlgren, C. Dannewitz, C. Imbrenda, D. Kutscher, and B. Ohlman, "A survey of information-centric networking," *Communications Magazine, IEEE*, vol. 50, no. 7, pp. 26–36, 2012.

[5] M. Amadeo, C. Campolo, A. Molinaro, and G. Ruggeri, "Content-centric wireless networking: A survey," *Computer Networks*, vol. 72, pp. 1–13, 2014.

[6] M. Gerla, E.-K. Lee, G. Pau, and U. Lee, "Internet of vehicles: From intelligent grid to autonomous cars and vehicular clouds," in *Internet of Things (WF-IoT), World Forum on*. IEEE, 2014, pp. 241–246.

[7] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, "Identification of influential spreaders in complex networks," *Nature Physics*, vol. 6, no. 11, pp. 888–893, 2010.

[8] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web." 1999.

[9] S. P. Borgatti, "Centrality and network flow," *Social networks*, vol. 27, no. 1, pp. 55–71, 2005.

[10] P. Hui, J. Crowcroft, and E. Yoneki, "Bubble rap: Social-based forwarding in delay-tolerant networks," *Mobile Computing, IEEE Transactions on*, vol. 10, no. 11, pp. 1576–1589, 2011.

[11] A. Socievole, E. Yoneki, F. De Rango, and J. Crowcroft, "Ml-sor: Message routing using multi-layer social networks in opportunistic communications," *Computer Networks*, vol. 81, pp. 201–219, 2015.

[12] A. Okabe, B. Boots, K. Sugihara, and S. N. Chiu, *Spatial tessellations: concepts and applications of Voronoi diagrams*. John Wiley & Sons, 2009, vol. 501.

[13] I. Psaras, W. K. Chai, and G. Pavlou, "In-network cache management and resource allocation for information-centric networks," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 25, no. 11, pp. 2920–2931, 2014.

[14] S. Uppoor, O. Trullols-Cruces, M. Fiore, and J. M. Barcelo-Ordinas, "Generation and analysis of a large-scale urban vehicular mobility dataset," *Mobile Computing, IEEE Transactions on*, vol. 13, no. 5, pp. 1061–1075, 2014.

[15] A. Afanasyev, I. Moiseenko, and L. Zhang, "ndnSIM: NDN simulator for NS-3," NDN, Technical Report NDN-0005, October 2012. [Online]. Available: http://named-data.net/techreports.html