# GRank - An Information-Centric Autonomous and Distributed Ranking of Popular Smart Vehicles

Junaid Ahmed Khan*[†] Yacine Ghamri-Doudane[†] and Dmitri Botvich*
*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, dbotvich@gmail.com

*Abstract*—**Modern cars are transforming towards autonomous cars capable to make intelligent decisions to facilitate our travel comfort and safety. Such "Smart Vehicles" are equipped with various sensor platforms and cameras that are capable to constantly sense tremendous amount of heterogeneous data from urban streets. This paper aims to identify the appropriate vehicles, important to be selected as information hubs for the efficient collection, storage and distribution of such massive data. Therefore, we propose an Information-Centric algorithm, "GRank" for vehicles to autonomously find their importance based on their reachability for different location-aware information in a collaborative manner, without relying on any infrastructure network. GRank is the first step to identify socially important information hubs to be used in the network. Results from scalable simulations using realistic vehicular mobility traces show that GRank is an efficient ranking algorithm to find important vehicles in comparison to other ranking metrics in the literature.**

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

## I. INTRODUCTION

The era of autonomous cars is around the corner as vehicles today are equipped with myriad of electronic components, sensors, cameras and communication devices to assist drivers regarding the travel safety and comfort. In an urban environment, "Smart Vehicles" can now be considered as an instance of the Internet of Things (IoT) aimed to collect and share sensory and multimedia data from different city locations assisting various Intelligent Transportation System (ITS) applications. These include efficient traffic management, urban environment sensing and vicinity monitoring [1] [2].

At the same time, these vehicles are constantly generating a tremendous amount of sensory data that cannot be uploaded to the 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 challenge is to efficiently locate, collect and store the user relevant data from the fleet of vehicles with the challenge of intermittent connectivity and vehicle mobility in a Vehicular Ad-hoc Network (VANET).

This evoke the need to find the appropriate set of vehicles that are important to be selected as information hubs forming a distributed social network of vehicles based on their

reachability within VANET. To identify important nodes in a network, social network analysis typically rely on different variants of centrality measures such as Degree, Closeness, Betweenness, Eigenvector centrality and communicability [3]. However, such schemes are difficult to use in the sporadic vehicular network topology. Therefore, we propose a new concept of finding important vehicles by leveraging the vehicle self-decision making, allowing it to rank itself based on the daily commute and the popularity of the neighborhoods it visits frequently.

In this paper, we present an innovative vehicle ranking algorithm, GRank, allowing a vehicle to use a new stable metric named "Information communicability" to rank different locations in the city and rank itself accordingly. Using GRank, the vehicle finds each location reachability and popularity taking into consideration the user interest satisfaction related to the location. It also considers its mobility pattern between different locations in the city along its availability in each location. Vehicles available in popular locations in the city qualify as important Information Facilitator Vehicles (IFVs) with higher *vehicle centrality* score in the network.

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 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, GRank 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:

- A novel distributed approach enabling a vehicle to find popular locations in an urban environment based on the satisfied user interests.
- An innovative vehicle ranking algorithm, "GRank", where vehicle can find its importance in the network, without relying on any infrastructure network.
- Scalability as well as ICN compliance validation through extensive simulations using around three thousand vehicles utilizing realistic mobility traces.

The obtained results show that the proposed algorithm is well suited to help in the efficient identification of the top IFVs in the network as the vehicles identified by GRank satisfied

three times more interests compared to other schemes. The rest of the paper is organized as follows. The next Section highlights the major related work. In Section III, we propose GRank followed by the performance evaluation discussing the simulations in Section IV. Section V concludes the paper along with 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 [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. GRANK: A FULLY DISTRIBUTED APPROACH

We present GRank as a global centrality measure enabling each IFV to autonomously find its importance independently from a centralized infrastructure. 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. Similarly, it is not always necessary for two nodes to communicate through shortest paths but they can possibly follow non-shortest paths in a dynamic VANET topology. Therefore, inspiring from the concept of communicability in complex networks [3], we introduce a new metric called

"Information communicability", where each IFV periodically finds its importance in the network based on the reachablity of the associated information with respect to the satisfied user interests. The interests are assumed to be generated and received from the neighboring vehicles using multi-hop interest forwarding.

### A. Network Model

We model a time varying VANET as an undirected vehicular graph $G(\mathbb{V}(t), \mathbb{E}^v(t))$, where $\mathbb{V} = \{v\}$ 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 $j$ and $k$ at time $t$. The number of edges $\mathbb{E}^v(t)$ depends on the transmission range of each vehicle. We assume it as a simple unit disk model bounded by the communication range. The city map is represented by the undirected graph $G(\mathbb{X}, \mathbb{E}^x)$, 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 a road network.

Table I: List of Notations

| Notation | Description |
|---|---|
| $\mathbb{V}$ | Set of vehicles |
| $\mathbb{X}$ | Set of locations/regions |
| $\mathbb{E}^x / \mathbb{E}^v(t)$ | Edges set between locations/vehicles at time $t$ |
| $\mathbb{E}$ | Edges set between vehicles and locations |
| $\bar{t}_k$ | $k^{th}$ time-slot for GRank computation |
| $N$ | Total number of locations |
| $X_v$ | Set of locations associated to vehicle $v$ |
| $P_x$ | Information association profile for a vehicle |
| $p_{x_i}$ | Probability to satisfy interests for location $x_i$ |
| $k$ | number of regions between two regions |
| $W$ | Vehicle adjacency matrix of information association profile for all locations |
| $C_{x_i x_j}^v$ | Information communicability between locations $x_i$ and $x_j$ |
| $\Gamma_v / \Gamma_x$ | Set of neighbors for vehicle $v$ / locations $x$ |
| $f_{x_i}^{\Gamma_v}$ | Neighbors communicability function |
| $f_{x_i}^v$ | Information centrality function |
| $P_v$ | Vehicle mobility transition probability matrix |
| $p_{x_j, x_i}$ | State transition probability between regions |
| $\pi_x$ | Steady state probability of being in region $x$ |
| $\rho_{x_i}^v$ | Location importance parameter for location $x_i$ |
| $G_{x_i}^v$ | Information global centrality for each location |
| $\alpha$ | Tuning parameter for neighbors contribution |
| $\beta$ | Tuning between the recent location popularity and overall location popularity |
| $\gamma$ | Tune past GRank score with the current score |
| $f_v$ | Vehicle centrality function |
| $C_v$ | Vehicle global centrality |

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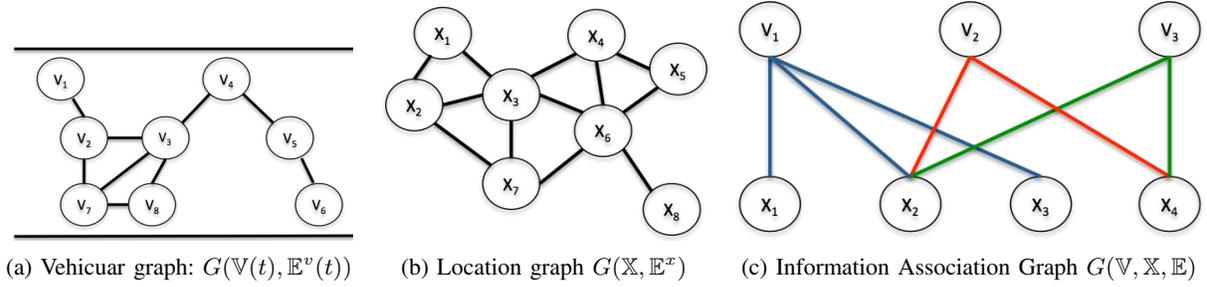
(a) Vehicuar graph: $G(\mathbb{V}(t), \mathbb{E}^v(t))$    (b) Location graph $G(\mathbb{X}, \mathbb{E}^x)$    (c) Information Association Graph $G(\mathbb{V}, \mathbb{X}, \mathbb{E})$

Figure 1: Network Model

Figure 1 shows an instance of the respective vehicular, location and the corresponding information association graph.

***Definition 1****:* (Information Association Profile) The information association profile for vehicle $v$ from the sub-graph $X_v \subset \mathbb{X}$ is the $|N| \times 1$ probability vector $P_x = [p_{x_1}, ... p_{x_N}]^T$, for $N$ locations, where $(\cdot)^T$ is the matrix transpose and $p_{x_i}$ is the probability of satisfying user interests for information associated to location $x_i \in X_v$, where $\sum_{x_i \in X_v} p_{x_i} = 1$.

We further classify the associated information by clustering the regions using ICN hierarchical naming convention *"/region/road-section/information-type"*. Information type comprises different Intelligent Transportation Systems (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 associated to the set of roads in a single Voronoi region $x \in \mathbb{X}$.

For temporal VANET analysis, we divide the time $T = (\overline{t_1}, \overline{t_2}, ...)$ into a sequence of regular time-slots, where the $k^{th}$ time-slot is represented as $\overline{t_k} = [t_k, t_{k+1})$. Each vehicle finds its centrality at the time instant $t_{k+1}$ based on 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 interchangeably in the text since we associate content to locations in the urban map.

### B. Information Global Centrality

In this section, we define the parameters each vehicle use to analytically find the information popularity for each location in the city. It periodically compute and exchange its own knowledge about the information popularity for each city location. It also uses the corresponding information popularity received from neighboring IFVs to rank all locations.

***Definition 2****:* (Information Walk) We define an information walk of length $k$ as a sequence of locations $x_0, x_1, ..., x_{k-1}, x_k$, such that, for each $i = 1, 2, ..., k$, there is a link from $x_i$ to $x_{i+1}$ in $X_v$.

It reflects the vehicle commute between the two regions taking into consideration all the possible regions it can visit in the way. For example, the set of intermediate regions with all possible set of information walks between location $x_A$ and $x_B$ are shown in Figure 2. The idea of using walk instead of path allows revisiting intermediate regions several times since it is sometimes inevitable to avoid traversing an intermediate region due to the static urban road structure.
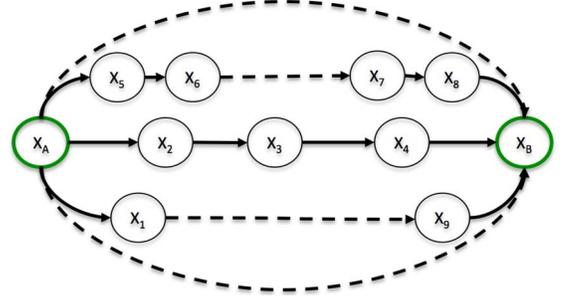


Figure 2: Information Walk

***Definition 3****:* (Information Communicability) For each vehicle $v$, the information communicability for associated location $x_i$ is the information reachability through the weighted sum of all possible information walks starting at location $x_i$ and ending at $x_j$, $\forall x_i, x_j \in \mathbb{X}$:

$$C_{x_i x_j}^v(t_{k+1}) = \sum_{k=0}^{\infty} \frac{(W^k)_{x_i x_j}}{k!} = (e^W)_{x_i x_j}, \qquad (1)$$

The term $W$ is the weighted adjacency matrix with entries as area weight $w_{ij} = p_{x_i} p_{x_j}$, where $p_{x_i}$ and $p_{x_j}$ reflects the vehicle information association profile for location $x_i$ and $x_j$ considering the vehicle association to both regions. $(W^k)_{x_i x_j}$ uses the $k^{th}$ power of the adjacency matrix in order to take into account the number of information walks of length $k$ starting at location $x_i$ and ending at location $x_j$.

The matrix function $e^W$ can be defined using the Taylor series expansion: $e^W = I + W + \frac{W^2}{2!} + \cdots + \frac{W^k}{k!} + \cdots$. The identity matrix in the expansion does not affect the communicability as it only adds a constant to every value of the communicability measures. As a special case, the vehicle can find the reachability of its home location using self-communicability as the sub-graph centrality by replacing $x_j$ in (1) by the home location $x_i$.

The information communicability for a location considers all the possible routes to it, however information walks with more number of intermediate regions should be given lower priority. To implement such penalty, taking $k!$ in the denominator ensures that longer routes are not weighted more than the shorter ones. The advantage is to prefer possible shorter routes, not necessarily the shortest path, for communicating between two regions.

*Neighbor Communicability:* The vehicle receives information communicability for each location from the neighboring

vehicles which is used to find information centrality function. The neighbors communicability function is defined as;

$$f_{x_i}^{\Gamma_v}(t_{k+1}) = \frac{1}{|\Gamma_v| \cdot |X_v|} \sum_{u \in \Gamma_v} \sum_{x_j \in X_v} C_{x_i x_j}^u(t_{k+1}) \cdot C_u(t_{k+1}),$$

(2)

The neighbors belong to set of vehicles $\Gamma_v \subset \mathbb{V}(t_{k+1})$ at time instant $t_{k+1}$, where $C_{x_i x_j}^u(t_{k+1})$ is the neighbor information communicability for location $x_i$ and $C_u(t_{k+1})$ is the respective neighbor vehicle centrality. A well popular neighborhood can easily be identified as the location where vehicles with high centrality exchange information more frequently. The information centrality function at the vehicle is given as:

$$f_{x_i}^v(t_{k+1}) = \frac{\alpha}{|X_v|} \sum_{x_j \in X_v} C_{x_i x_j}^v(t_{k+1}) + (1 - \alpha) \cdot f_{x_i}^{\Gamma_v}(t_{k+1}),$$

(3)

The vehicle uses the neighbor communicability function in (2) to find each location information centrality function, where $\alpha \in [0, 1]$ is the tuning parameter to adjust the neighbors contribution.

*Vehicle mobility pattern:* The road sections in each region directly depends on the corresponding set of road sections in the adjacent regions. Therefore, the vehicle mobility pattern is modeled as a Markov chain where each region $x \in \mathbb{X}$ is represented as a state. The entries $p_{x_i, x_j}$ of the vehicle state transition probability matrix $P_v$ represents the transition probability between region $x_i$ and the neighboring region $x_j \in \Gamma_{x_i}$ where $\Gamma_{x_i}$ is the set of neighboring regions for $x_i$. For each vehicle, the transition probability matrix for the mobility between $N$ regions is:

$$P_v = \begin{pmatrix} p_{x_1, x_1} p_{x_1, x_2} & \cdots & p_{x_1, x_N} \\ \vdots & \ddots & \vdots \\ p_{x_N, x_1} p_{x_N, x_2} & \cdots & p_{x_N, x_N} \end{pmatrix}$$

The vehicle steady state probability $\pi_{x_i}$ to be in region $x_i$ as a factor of the neighboring regions given by the relation $\pi_{x_i} = \sum_{x_j \in \Gamma_{x_i}} p_{x_j, x_i} \pi_{x_j}$. The steady state probability $\pi_{x_j}$ is the probability of the vehicle availability in the $j^{th}$ neighboring region, where $\sum_{x \in \mathbb{X}} \pi_x = 1$, and $\sum_{x_j \in \mathbb{X}} p_{x_j, x_i} = 1$.

The vehicle mobility pattern described above associates the vehicle availability in the region which helps in finding each region importance with respect to the vehicle. For instance, higher steady state probabilities of being in a region yields more association. It is also noteworthy that a vehicle might be unable to visit a neighboring region in case of pre-defined lane rules defined by the city administration. Therefore, by modeling the vehicle mobility pattern as a markov chain, we take into account the regions where it will subsequently visit being in the current region, independent of the previous travel pattern and the underlying road structure.

The importance of a location $x_i$ with respect to the vehicle availability pattern as well as the associated information reachability is given as $\rho_{x_i}^v(t_{k+1}) = \pi_{x_i} \cdot f_{x_i}^v(t_{k+1})$, where $\pi_{x_i}$ yields the probability of the vehicle availability in the region and $f_{x_i}^v$ considers the respective information centrality function using (3). The information global centrality for each location can now be defined as:

$$G_{x_i}^v(t_{k+1}) = \beta \cdot \rho_{x_i}^v(t_{k+1}) + (1 - \beta)\rho_{x_i}^v(t_k) \quad (4)$$

where $\rho_{x_i}^v(t_k)$ is maintained by each vehicle as the location importance value at the time $t_k$. The parameter $\beta \in [0, 1]$ is used to tune between the recent location popularity and overall location popularity. Thus, we can identify popular locations at time instant $t_{k+1}$ in the city with the maximum global centrality $\arg\max_{v \in \mathbb{V}, x_i \in \mathbb{X}} G_{x_i}^v(t_{k+1})$ with respect to all IFVs. However, popularity of locations depends on several factors such as the information-type depending on the application requirements as well as time of the day. Similarly, we can use the maximum location importance $\max_{v \in \mathbb{V}, x \in \mathbb{X}} \rho_{x_i}^v$ to identify popular neighborhoods for a longer time span.

## C. Vehicle Centrality

The vehicle considers its importance with respect to all locations in the city in order to measure its influence in the network. Each vehicle information association profile plays an important role in deciding its importance in the network. Thus, it should receive and subsequently satisfy more frequently the interests for information regarding the popular regions in the city. The vehicle centrality function at the time instant $t_{k+1}$ is given as the average information global centrality for all associated locations:

$$f_v(t_{k+1}) = \frac{1}{|X_v|} \sum_{x_i \in X_v} G_{x_i}^v(t_{k+1}),$$

where the vehicle regularly updates its global centrality as the Exponential Weighted Moving Average (EWMA) of the vehicle centrality:

$$C_v(t_{k+1}) = (1 - \gamma)C_v(t_k) + \gamma f_v(t_{k+1}),$$

where $\gamma \in [0, 1]$ is a tuning parameter to tune between the past vehicle centrality score and the score in the current time-slot.

## D. Summary

The steps required for a vehicle in order to find its global importance at regular time instants are summarized in Algorithm 1. For each associated location, the vehicle first finds the information communicability as the information reachability with respect to the location. Similarly, it exchanges the information communicability for each location with the neighboring IFVs in range along its respective local vehicle centrality. The vehicle uses its own as well as the neighbors information communicability to find the information centrality function for each location using (3). It then finds each location's importance taking into account its availability in that location. Each location's importance is then used to find the information global centrality. Each vehicle regularly updates the global information centrality in order to find its respective global centrality in the network. The vehicle global centrality is the average global information centrality with respect to all locations in the city. Thus, highly reachable and available vehicles with respect to all locations are considered as the most central vehicles in the city.

**Algorithm 1** GRank

---

**INPUT:** $G(\mathbb{V}, \mathbb{X}, \mathbb{E})$ : information association graph , Information global centrality, GRank in previous time-slot

**OUTPUT:** Updated GRank for the next time-slot at time-instant $t_{k+1}$

**for** each vehicle $v \in \mathbb{V}$ **do**
    **for** each associated location $x_i \in X_v$ **do**
        find information communicability $C_{x_i x_j}^v, \forall x_i x_j \in \mathbb{X}$ using (1)
        **for** each vehicle neighbor $\Gamma_v \in \mathbb{V}$ in range **do**
            receive neighbor communicability $C_{x_i x_j}^{\Gamma_v}$, neighbor centrality $C_{\Gamma_v}$,
        **end for**
        compute neighbors communicability function $f_{x_i}^{\Gamma_v}$ using (2)
        find information centrality function $f_{x_i}^v$ using (3), then location importance $\rho_{x_i}^v$
        compute information global centrality $G_{x_i}^v$ using (4)
    **end for**
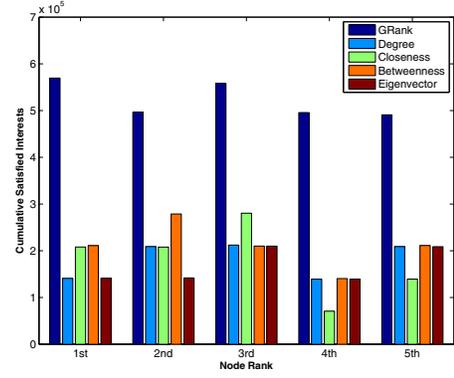    compute $f_v$, $C_v$
**end for**
**return** $C_v(t_{k+1})$

---



Figure 3: Cumulative Satisfied Interests by top identified nodes using each scheme over an average of five different simulations

## 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 |

## IV. PERFORMANCE EVALUATION

One of the basic requirement for evaluating the efficiency of GRank is scalability. Therefore, we use Network Simulator-3 (NS-3) as a scalable simulation platform for upto three thousand vehicles. The performance of GRank 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 [13]. The simulation parameters are summarized in Table II, followed by a description of the simulation scenarios used for the performance evaluation.

## Table III: GRank in different set of Simulations

| Simulation | 1 | | 2 | | 3 | | 4 | | 5 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| GRank | ID | Score | ID | Score | ID | Score | ID | Score | ID | Score | Mean |
| 1 | 2687 | 1 | 483 | 1 | 1999 | 1 | 238 | 1 | 1239 | 1 | 1 |
| 2 | 259 | 0.9988 | 957 | 0.9906 | 9 | 0.9996 | 1105 | 0.9893 | 1407 | 0.9988 | 0.9954 |
| 3 | 768 | 0.9982 | 1932 | 0.9898 | 1215 | 0.9993 | 925 | 0.9889 | 34 | 0.9971 | 0.9946 |
| 4 | 1301 | 0.9980 | 103 | 0.9896 | 2471 | 0.9991 | 1867 | 0.9889 | 401 | 0.9965 | 0.9942 |
| 5 | 797 | 0.9971 | 950 | 0.9896 | 2204 | 0.9990 | 1902 | 0.9888 | 1195 | 0.9964 | 0.9941 |
| 6 | 216 | 0.9971 | 318 | 0.9788 | 57 | 0.9988 | 1349 | 0.9877 | 653 | 0.9963 | 0.9917 |
| 7 | 92 | 0.9969 | 2728 | 0.9770 | 808 | 0.9983 | 313 | 0.9866 | 444 | 0.9963 | 0.9910 |
| 8 | 76 | 0.9967 | 268 | 0.9738 | 849 | 0.9982 | 46 | 0.9862 | 508 | 0.9959 | 0.9901 |
| 9 | 658 | 0.9966 | 54 | 0.9735 | 1940 | 0.9968 | 1682 | 0.9860 | 108 | 0.9958 | 0.9897 |
| 10 | 1603 | 0.9965 | 150 | 0.9729 | 771 | 0.9966 | 158 | 0.9860 | 1904 | 0.9957 | 0.9895 |

### A. Simulation Scenario

We simulate a VANET urban sensing scenario using the ndnSIM [14] 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 information source in the network. The areas visited by a vehicle in a time-slot before the GRank computation time are considered as content 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.

We assume interests follow a Zipf distribution, where interests for popular contents are more frequent, thus, results in a high interest probability for popular city locations. Any producer acting as source for an area upon receiving the interest responds with the desired content. 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 region at the current time-slots.

For each vehicle, the vehicle centrality is computed using the information global centrality for each region. We perform each simulation upto five times by analyzing different set of nodes as information producers and consumers in order to compute upto 95% confidence intervals. The tuning parameters $\alpha$, $\beta$ and $\gamma$ are set to $0.5$ in order to maintain generality in our evaluation, however, 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 GRank score with the respective Degree, Closeness, Betweenness and Eigenvector centrality.

### B. Simulation Results

The objective of our simulation study is to find answers to the question: *How well can it identify the top IFVs?* GRank score for the top 10 IFVs from five simulation runs are shown in Table III. For each rank, The average score lies within a confidence interval of $0.01$ for a confidence level of $95\%$. We rank all the vehicles in the simulation scenario, however, we
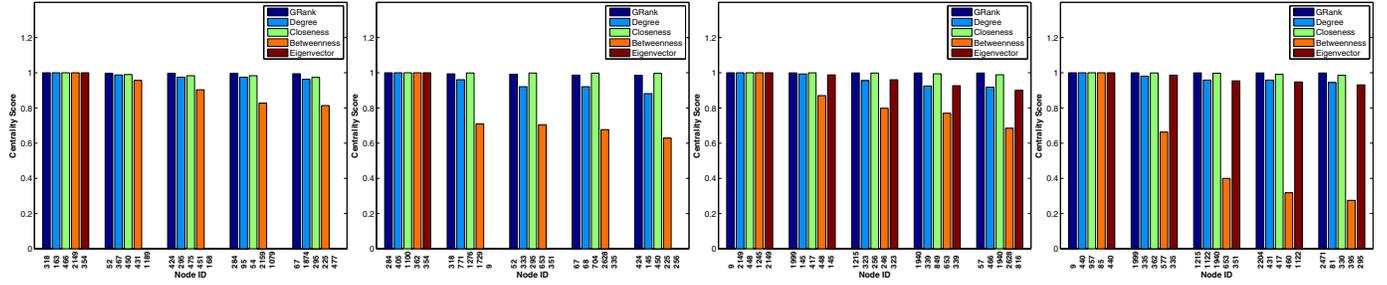
Figure 4: Temporal Snapshots after each 15 minutes comparing top identified nodes by each schemes

are interested only to identify the top IFVs in each simulation. Therefore, GRank score is normalized with respect to the top identified node, i.e. the top node 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, due to its natural mobility pattern, the vehicle 2687 is identified to have the top GRank score among all vehicles in the network as it is able to satisfy more frequently the incoming interests throughout the simulation. It is also noteworthy that GRank will result in different set of top nodes by tuning the parameters $\alpha$, $\beta$ and $\gamma$ depending on the application requirement.

For better analysis of GRank in different simulation scenarios, we consider the following performance metrics in comparison with the state of the art importance computation 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 GRank along ICN in VANETs

*1) Cumulative Satisfied Interests:* Cumulative Satisfied Interests refers to the total number of user interests satisfied during the simulation duration. Figure 3 shows the CSI score of the top five nodes identified by all these schemes in an average of five set of simulations with random set of consumers and producers. Typical ranking schemes only takes into account physical topology towards computing a node importance in the network, ignoring the satisfied user interests. Nevertheless, GRank satisfied around 50% more user interests than other schemes in all the five set of simulations due to the consideration of vehicle interest satisfaction profile along with its reachability and availability as key factors towards the vehicle importance.

*2) Temporal behavior analysis of top nodes:* It is important to efficiently analyze the time varying behavior of our algorithm under the dynamic VANET topology. 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 4. We consider the top node

identified by each scheme as benchmark by assigning it a unity score followed by the other IFVs in the ranking order.

At the beginning, vehicle 318 is ranked as the top IFV by GRank, thought the other schemes underrated it. This is because we consider stable metric such as fixed locations associated to the vehicle instead of using ephemeral topological information. Vehicles also change places along the ranking order. For example, around 15 to 30 minutes in the simulation, the top ranked vehicle 318 is replaced by 284, which is then replaced by the vehicle 9 at around 30 minutes till the end of the simulation duration.

An interesting results was observed in the initial 30 minutes (Figure 4 a,b): Only one node yields a high Eigenvector centrality score followed by other nodes with a negligible Eigenvector centrality score. We investigate this behavior and found that the principle eigenvalue yields the top nodes where the eignvector is shifted towards the principle component. Thus, resulting in a single vehicle as the top IFV. This proves that the famous Eigenvector centrality fails to assign significant score to a large fraction of nodes in a larger network, while GRank does not show such behavior. Typical centrality schemes result in different set of top IFVs at each snapshot. It is because such schemes only consider the instantaneous shortest paths between vehicles towards ranking the vehicles at a particular time instant as well as require the complete network topological information. However, such complete network information is not available to an individual vehicle in highly unstable VANETs. GRank considers the information walks between locations in the city ensuring more stable set of top nodes, thus, not affected by the network dynamics which is not the case for other schemes.

*3) Aggregated Per Node Throughput:* We evaluate the proposed ranking scheme by analyzing the throughput at important nodes in the network. Figure 5 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 GRank yields three times more throughput compared to centrality schemes since they satisfied more frequently interests regarding popular locations. For GRank, the per node throughput of the top ranked IFVs shows a decrease between the ranking order while a variation is seen for Degree, Betweenness and Eigenvector centrality. It is clear that GRank outperformed all schemes as it incorporates information importance in order to rank vehicle, while other schemes entirely rely on topological measures such as the node
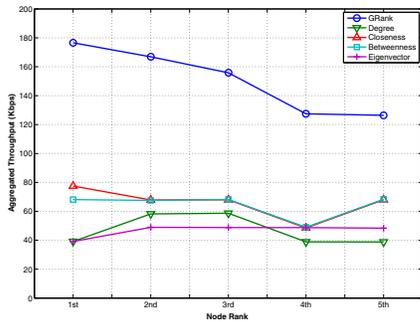
Figure 5: Average aggregated throughput by the top identified nodes using each scheme in five simulations
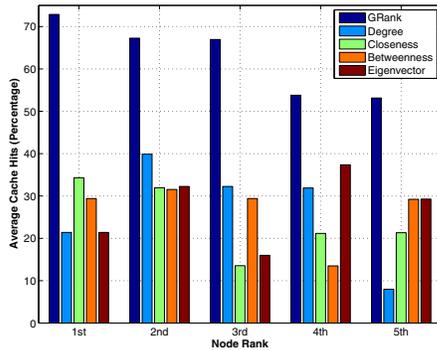


Figure 6: Average cumulative cache hit rate by the top identified nodes using each scheme in five simulations

degree or shortest paths towards finding vehicle importance. Moreover, higher per nod throughput also shows that efficiency of top ranked IFVs in various network operations.

*4) ICN Evaluation - In-Network Caching:* Let us evaluate now GRank for the ICN built-in feature of In-Network caching at the intermediate 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 vehicle for the same content is considered a cache hit. Figure 6 shows the cumulative cache hit rate for the top nodes identified by each scheme for the entire simulation duration for five set of simulations. The top vehicles identified by GRank yield a higher hit rate than those identified by the other schemes as vehicle containing important information relevant to the users responds and subsequently cache more frequently leveraging their mobility pattern compared to other vehicles. This proves that In-Network caching offered by ICN in GRank implementation overcomes the mobility and intermittent connectivity constraints of VANETs.

### C. Summary

Responding to the question posed, *How well can it identify the top IFVs?* From the simulation results, it is clear that a relatively stable set of top IFVs are identified by GRank compared to the other schemes in dynamic VANETs since we consider location-aware information instead of dynamic vehicular topology. GRank can identify vehicle which satisfied more user interests with higher aggregated per node throughput

and more cache hit rate compared to the other schemes due to their reachability and availability for efficient content access. Thus, the overall comparative analysis of GRank with different network ranking schemes proposed in the literature proved it as an efficient vehicle centrality scheme.

## V. Conclusions and Future Directions

In this work, we introduced an innovative vehicle ranking algorithm, "GRank", exploiting the vehicle autonomous decision making ability, allowing it to rank itself in a fully distributed VANET. GRank uses an information-centric approach to find important locations in the city based on user relevant interests. Results by comparing it with state of the art centrality schemes revealed that GRank is clearly best suited to efficiently identify important information facilitator vehicles in VANETs. Identification of popular vehicles can be helpful in different smart city applications such as urban sensing and vicinity monitoring. Such vehicles, equipped with sufficient storage and processing capabilities are recruited for efficient data collection, storage and distribution in VANETs. Designing such efficient schemes based on GRank 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, 2006.

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

[3] E. Estrada and N. Hatano, "Communicability in complex networks," *Physical Review E*, vol. 77, no. 3, p. 036111, 2008.

[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] 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.

[14] 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