# Autonomous Identification and Optimal Selection of Popular Smart Vehicles for Urban Sensing - An Information-centric Approach

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*—Vehicles today 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 cost. This invokes the need to identify the appropriate vehicles, important for different urban sensing tasks based on their natural mobility. This paper address this problem leveraging the self-decision making ability of a "Smart Vehicle" to measure its relative importance in the network. To do so, we present InfoRank as an Information-Centric algorithm for a vehicle to first autonomously rank different location-aware information. It then uses the information importance along its mobility pattern to find its importance in the network. We also present a selection algorithm to find the best ranked vehicles for urban sensing and vicinity monitoring to achieve a desired coverage within a limited budget. Our vehicle ranking system is the first step towards identifying the best information hubs to be used in the network for the efficient collection, storage and distribution of urban sensory information. We evaluate InfoRank under a scalable simulation environment using realistic vehicular mobility traces. Results show that the proposed ranking system efficiently identified socially important vehicles in comparison to other ranking schemes.**

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

## 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. A "Smart Vehicle" can now be considered as an instance of Internet of Things (IoT) aimed to harvest and share different sensory and multimedia data from urban streets to support various Intelligent Transportation System (ITS) applications such as efficient traffic management and urban environment sensing [1].

We observe daily thousands of such smart vehicles on the road in urban environment , with their high processing, storage and communication capabilities. A smart city administration aims to recruit a set of vehicles to perform urban sensing and vicinity monitoring in order to improve its citizen lifestyle while offering new services. Therefore, the paper targets two questions?

- What are the metric that classify a particular vehicle as a good candidate (eligibility) for different urban sensing operations?
- Which set of vehicles can be selected (out of the good candidate vehicles) as appropriate candidates to optimally achieve urban sensing under a given budget and a set of city-wide spatio-temporal coverage requirements?

To address this problem, we present an innovative vehicle ranking system for the identification and recruitment of the set of popular Information Facilitator Vehicles (IFVs) to efficiently gather, store and distribute urban sensing data. For example, buses, taxis, commuters available on frequent routes to address user interests in the network. Network analysis typically rely on different variants of centrality measures such as Degree, Closeness, Betweenness, Eigenvector centrality to identify important nodes. However, such schemes are difficult to use in the sporadic vehicular network topology.

To classify the vehicles eligibility, we define a novel algorithm as InfoRank, allowing the vehicle to autonomously classify itself based on the information importance as the measure of user interest for the cached content, while considering the content's popularity and the vehicle's mobility. Then, we devise an optimization algorithm for the selection/recruitment of a subset of these eligible vehicles for the above described smart-city usage under an associated budget and coverage requirements.

The major contributions to this paper can be summarized as follows:

- We propose a novel algorithm "InfoRank" [2] enabling a vehicle to autonomously (1) rank important location-aware information associated to it based on the satisfied user interests and the information validity and (2) rank itself based on the information importance and its mobility pattern in the city.
- We propose an algorithm to optimize the selection of the best ranked set of vehicles using InfoRank for urban sensing while maintaining city-wide coverage with the minimum cost.

Corresponding author: Junaid Ahmed Khan (khan@univ-mlv.fr)

- We validate the scalability of our ranking system by performing extensive simulations comprising around three thousand vehicles using realistic mobility traces to identify and select popular IFVs in time evolving vehicular networks.

The obtained results show that the proposed algorithms are well suited to help in the efficient identification and selection of the set of important vehicles 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. Section II provides a brief background on ICN and its usage in our case. In Section IV, we present InfoRank as the algorithm to find the information importance along the essential parameters to analytically compute the vehicle centrality. The recruitment of important vehicles with high centrality score under the constraints such as spatio-temporal coverage and dedicated budget limit is discussed in Section V. Section VI explains the performance evaluation of our ranking system towards the identification of important vehicles in the network. In Section VII we conclude the article along some insight for future research.

## II. Background

Recently we observe a shift towards Information-Centric Networking (ICN) [3] 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. ICN aims to decouple the service from the host, thus removing content association to any physical location. It also offers In-Network caching at intermediate nodes while forwarding and responding to subsequent user interests.

We adapt the Content Centric Networking (CCN) instance of the general ICN architecture to the requirement of vehicular environment while advocating the importance of the "information-centric or "content-centric" networking philosophy. We privilege the information user relevant importance rather than its localization where the ICN paradigm address the issue by decoupling the content provider-consumer and support in-network caching at intermediate nodes to improve content availability with minimum delays. Assuming an urban environment where vehicles constantly receive and satisfy interests for different content from the neighboring vehicles using multi-hop interest forwarding in a vehicular network. Each ICN enabled vehicle can maintain three routing parameters:

- Forwarding Information Base - FIB: It resembles a routing table which maps content name components to interfaces. Each vehicle FIB is populated by the routes discovered using name-based interest/content forwarding protocol.
- Pending Interest Table - PIT: It keeps track of all the incoming interests that the vehicle has forwarded but not satisfied yet. Each PIT entry records the content name carried in the interest, together with its incoming and outgoing interface(s).
- Content Store - CS: It is a temporary cache to store content each intermediate vehicle has received while forwarding content. Since a named-data packet is meaningful independent of where it comes from or where it is forwarded, it can be cached to satisfy future interests.

The named-data networking concept introduced by the information-centric networking paradigm is capable to co-exist with the mobility and intermittent connectivity challenge in mobile networks. The provider-consumer decoupling allows the consumer interest for a desired content forwarded to "any" content provider independent of the underlying network connectivity. We consider a publish-subscribe ICN model allowing a mobile node such as a vehicle in our case to subscribe for the following three roles:

*1) Information Provider:* An information provider vehicle acts as the content source to publish content. For example, it can subscribe itself to publish sensory information collected from urban streets using the vehicle embedded cameras and sensors.

*2) Information Facilitator:* Vehicle responsible to collect, cache and relay data generated by information provider vehicles as well as forwards the user interest for content to "facilitate" efficient content caching and distribution.

*3) Information Consumer:* The vehicle subscribed to request different content from the information facilitators/providers with in the vehicular network are considered as information consumers to "pull" content in an information-centric vehicular network.

The proposed ranking system 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.

## III. Related Work

We consider a smart city application of sensory data collection from urban streets using vehicles. This paper focus towards proposing an efficient information-centric ranking system for vehicles inspired from social networks. Therefore, the related work consists two parts a) How ICN paradigm is recently emerged along discussing different urban sensing schemes proposed in vehicular network and b) How social network analysis motivates the identification and selection of important vehicles in the underlying vehicular network.

### A. ICN meets Urban Sensing

Urban sensing and vicinity monitoring using vehicles has attracted lots of researchers in the past few years and several schemes are proposed [1] [4] [5], where sensor-equipped vehicles sense and share data in a vehicular network.

Similarly Recruitment of such vehicles for urban sensing is studied recently. For example, in [6], participants with high reputation are recruited to perform urban sensing. The idea is to cover an area of interest with a limited budget, however, the coverage metric is confined to particular road sections with the limitations of utilizing the infrastructure network. Moreover, the authors do not provide any metric to classify and identify the important participants. The ICN paradigm is recently suggested as a promising solution to cater the peculiarities of the vehicular environment, characterized by dynamic topologies, unreliable broadcast channels, short-lived and intermittent connectivity [7].

### B. Social Network meets Vehicular Network

Social Network Analysis (SNA) [8] is used to identify important nodes in a social network which 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].

In a recent survey on Vehicular Social Networks (VSNs) [10], different sociability, security and applicability aspects of VSN paradigm are discussed. The authors advocate the emergence of social networks in vehicular scenarios due to increase in novel crowd-sourcing applications and increasing services based on social networking content sharing. In [11], the authors provide an extensive survey on different social-aware routing protocols for Delay Tolerant Networks (DTNs) and suggest to adapt the social metrics according to application requirements which motivates the need to define a new metric for important vehicle identification.

### C. Discussion

Unlike the above mentioned applications, we exploit Information Centric Vehicular Networking for urban sensing and vicinity monitoring. In order to select "important" vehicles, there is a need for a vehicle centrality scheme to classify / rank vehicles as information producers, facilitators and consumers. However, It is unfeasible to use centrality-based popularity schemes to identify important information hubs in vehicular networks for multiple reasons.

First, the rapid topological changes due to the high mobility of vehicles requires a continuous time varying analysis of the network 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 vehicular network would be ignored. Second, centrality measures such as Betweenness, Closeness and Eigenvector centrality computation requires network wide parameters, while in a vehicular network, 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 network topologies does not ensure the existence of a stable path between nodes. Therefore, a new vehicle ranking system adapted to vehicular networks and enabling vehicles to decide their relative importance in the network by overcoming the above mentioned constraints need to be thought about.

## IV. InfoRank

InfoRank is a centrality measure enabling each vehicle to autonomously find the importance of different locations independent of a centralized database. We do not merely rely on the rapidly changing inter-vehicle contacts since such unstable behavior of the frequency and duration of vehicle contacts do not provide any useful information to decide a location importance in the time evolving vehicular network. InfoRank considers the user interests frequency for content associated to different location as a key metric to classify the importance of different locations with respect the the vehicle as it regularly responds to user interests. It also considers the information validity scope as a metric towards finding the importance for a particular information. Before describing the importance metrics, the following section defines the network model followed by the description of the information importance metrics.

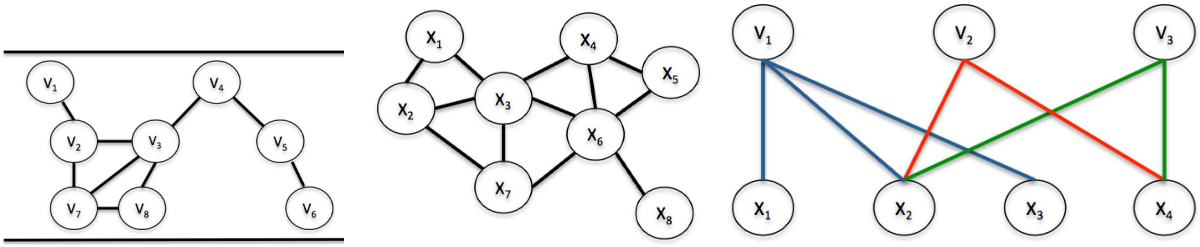
### A. Network Model

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 $j$ and $k$ 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 $\mathbb{X} = \{x\}$ represents location-aware content for 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. We define $x$ as a piece of location-aware content cached at a vehicle $v$ which reflect a content associated to a location in an urban environment. An exhaustive list of notations in presented in Table I stating all the used notations in the remaining of the article.

*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 information association decouple the dynamics of the vehicular graph by linking it with the stable nature of location graph. Despite the vehicles rapid mobility, $G(\mathbb{V}, \mathbb{X}, \mathbb{E})$ provides a relatively stable location-aware information association. 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.

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 $\overline{t_k} = [t_k, t_{k+1})$. Each vehicle finds its centrality




(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

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 content are associated to locations in the urban map.

The information distance $d(x, x_k)$ is the distance between the content location $x$ and the vehicle current position $x_k$ at time instant $t_{k+1}$, where $x, x_k \in X_v$. Information distance can be computed either as Curve-metric or Euclidean distance between the current vehicle's location and the information's location to assess its importance. We assume each vehicle knows the city map, i.e. the graph $G(\mathbb{X}, \mathbb{E}^x)$, but their cached-content related to this map is limited to the already visited locations or to the content received from neighboring vehicles. This is due to the limited storage capacity at vehicles as well as the limited vehicle coverage scope as the information it possesses is related to either its daily commute or associated neighborhoods.

### B. User Interests Satisfaction

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 the vehicle is capable of recording the time and position each time it responds as the content provider to a user interest. Therefore, it considers an information as popular if it observes an increase in the number of user interests for a particular location. 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 $I_x^v(\bar{t}_k) = \frac{r_x(\overline{t_k})}{R_x}$ 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 overall successful responds for the content $x \in X_v$ associated to the vehicle $v \in \mathbb{V}$.

### C. Information Validity Scope

The importance of each location-aware content is periodically updated based on the interest satisfaction frequency by the vehicle. Interest for each content specifies 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 $\tau$ where

$$\tau(t_{k+1}) = \begin{cases} 1 & t_{k+1} \leq t_x^f + \bar{t}_d \\ e^{-\delta \bar{t}_d} & t_{k+1} > t_x^f + \bar{t}_d \end{cases}$$

is the measure of the temporal information validity scope where $\delta \in [0, 1]$ is the tuning parameter depending on the application needs. It is used in the information timeliness metric where it tells us how quickly to decay the unnecessary increase in the information importance for the content which is not needed. An example would be the diminishing of the validity of the information regarding a road congestion which is now cleared, where delta defines the rapidness of the decay for such information. This decay can vary between different applications requiring different information diminishing time, thus we left it as a tunable metric.

For each information type, the information timeliness parameter $\tau$ 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, $\tau$ 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:

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

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 $I_x^v(\overline{t_k}) = 0$ avoids unnecessary increase in the information importance. The term $s_x^v \in [0, 1]$ represents the percentage of time the vehicle itself

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 |
| $\mathbb{E}$ | Set of edges between vehicles and locations |
| $\bar{t}_k$ | Time-slot $k$ for centrality computation |
| $t_k/t_{k+1}$ | Current time instant/next time instant |
| $X_v$ | Set of locations associated to vehicle $v$ |
| $d(x, x_k)$ | Location $x$ distance from current location $x_k$ |
| $I_x^v$ | Interests satisfaction frequency for location $x$ |
| $I_m^v$ | Ratio of unsatisfied to incoming user interests |
| $r_x$ | Number of successful responds for location $x$ 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 $x$ |
| $\bar{t}_d$ | Average interest deadline |
| $n$ | Total received interests in the previous slot |
| $t_x$ | Interest validity deadline for content $x$ |
| $\tau$ | Information timeliness |
| $\delta$ | Tuning parameter to limit information validity |
| $C_x^v$ | Importance of content $x$ associated to vehicle |
| $\lambda$ | Adjust importance based on distance from $x$ |
| $\epsilon$ | Adjust satisfied/received interest ratio |
| $s_x^v$ | vehicle reliability as content source for $x$ |
| $w_x$ | Information $x$ weight with respect to vehicle |
| $f_I^v$ | Information importance function |
| $C_v$ | Vehicle centrality |
| $\theta$ | Tune between previous and current centrality |
| $V \subset \mathbb{V}$ | Selected set of IFVs |
| $H_v$ | Vehicle coverage entropy |
| $A^{th}$ | Pre-defined area coverage threshold |
| $A_x$ | Individual location coverage vector for $x$ |
| $C^{th}$ | Pre-defined vehicle centrality threshold |
| $R^{th}$ | Overall vehicle centrality benchmark |
| $R_V$ | Overall vehicle centrality score |
| $B_V$ | Budget assigned for all vehicles |
| $B_v$ | Individual vehicle payoff |
| $B^{th}$ | Individual vehicle maximum dedicated budget |
| $\mathbb{B}$ | Overall maximum dedicated budget |
| $\omega$ | Adjust vehicle centrality contribution |
| $W$ | Set of vehicles not selected as IFVs |

acted as the original source for the content $x$. It is updated regularly to ensure the content relevant to the vehicle retains 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. It is used to let the application decide how much importance should be given to the distance from the information, it is useful for applications relying on location-based information and would like to increase/decrease information importance based on vehicle distance where the distance from information affects its importance.

### D. Vehicle Centrality

The vehicle considers its importance with respect to the associated information in order to measure its influence in the networks. Besides information importance in (1), we also consider the overall coverage scope as an important parameter to decide a vehicle importance in an urban environment.

***Definition 3:*** (Coverage Entropy) We define $H^v = -\sum_{x\in\mathbb{X}} p(x)\log p(x)$, as the coverage entropy of the vehicle periodically computed with respect to the entire city map (i.e vehicle associated sub-graph $X_v \in G(\mathbb{X}, \mathbb{E}^x)$). The probability $p(x)$ is the visiting frequency to each region $x \in \mathbb{X}$ before the importance computation time $t_{k+1}$.

The vehicle's coverage in the map can be represented as a set of mobility between regions since the urban map is divided into regions/zones where each vehicle travels between adjacent regions. Therefore, the overall vehicle mobility pattern is mapped as the set of visited regions depending on its daily commute. For example, the vehicles $A$ and $B$ coverage scope are bounded by the set of regions $M^A = \{x_3, x_2, x_2, x_3, x_5, x_2\}$ and $M^B = \{x_1, x_2, x_3, x_5, x_4, x_3\}$, where $x \in \mathbb{X}$. The vehicle A visits the regions $x_2$,$x_3$ and $x_5$ with probabilities $\frac{2}{6}$,$\frac{3}{6}$ and $\frac{1}{6}$, while B visits the regions $x_1$,$x_2$ $x_3$,$x_4$, and $x_5$ with probabilities $\frac{1}{6}$,$\frac{1}{6}$,$\frac{2}{6}$,$\frac{1}{6}$ and $\frac{1}{6}$ respectively. The corresponding coverage entropy for the mobility pattern as shown in Figure 2 is calculated as:

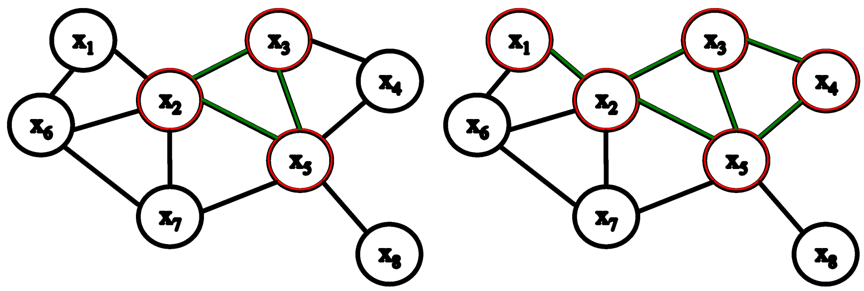
$$H^A = -\tfrac{2}{6}\log\tfrac{2}{6} - \tfrac{3}{6}\log\tfrac{3}{6} - \tfrac{1}{6}\log\tfrac{1}{6} = 0.439,$$
$$H^B = -\tfrac{2}{6}\log\tfrac{2}{6} - \left(\tfrac{1}{6}\log\tfrac{1}{6} * 4\right) = 0.639,$$

Vehicle A has a narrow coverage scope due to its limited geographical coverage, while B has a wider geographical coverage with respect to the urban map. Therefore, we consider coverage entropy as the coverage metric for the vehicle importance with respect to all locations in the city.

Algorithm 1 shows the steps allowing a vehicle to find the respective InfoRank. For a given location-dependent content in cache, the corresponding information importance is updated for the next time-slot at time instant $t_{k+1}$. The information-centric centrality function is given as:

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

For all contents $x \in X_v$ associated to $v$, $I_m^v(t_{k+1})$ are the ratio of missed interest to the total interests received by the vehicle while $\epsilon$ is the tuning parameter. Missed interest provides the vehicle reliability regarding successful respond to the incoming interests. $C_x^v(t_{k+1})$ is the respective content importance at time instant $t_{k+1}$, $w_x = \frac{R_x}{R_T}$ is the edge weight of information association graph $G(\mathbb{V}, \mathbb{X}, \mathbb{E})$ considering the interest satisfied for the content $x$ among all the contents in cache. $R_x$ is the number of responds for $x$ and $R_T$ is the number of responds




(a) Vehicle A city-wide mobility pattern (b) Vehicle B city-wide mobility pattern

Figure 2: An example of vehicle city-wide coverage scope as a metric of its mobility pattern

for all contents in the cache. $|X_v|$ is the cardinality of the sub-graph $X_v \subset \mathbb{X}$, all regions associated to the vehicle $v \in \mathbb{V}$. The term $\epsilon$ is used to decide to what extent the received interests which are not satisfied by the vehicle plays a role in its importance. For example, in case a service provider recruits vehicles as gateways where the vehicle naturally tends to receive relatively more interests than it satisfies, in such a case the service provider can tune epsilon to not depending much on the received interests not satisfied (missed) by the vehicle.

The vehicle centrality at the time instant $t_{k+1}$ is updated as the Exponential Weighted Moving Average (EWMA) function of the previous and current centrality score:

$$C_v(t_{k+1}) = (1-\theta)C_v(t_k) + \theta f_I^v(t_{k+1})$$

where $\theta$ is a tuning parameter to adjust the value for the past centrality score and the corresponding InfoRank in the current time-slot allowing the vehicle to decide how much past centrality score matters. It is needed in cases where a service provider is interested in real-time content recently recorded/collected from important vehicles on the road irrespective of their past centrality scores.

InfoRank 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 popularity in the network. However, the vehicle global mobility pattern in a city is bounded by the regions only known to the vehicle (visited before). Moreover, stale information is automatically deleted from the cache after some time due to the limited size storage buffer at vehicles.

**Algorithm 1** InfoRank

**INPUT:** $G(\mathbb{V}, \mathbb{X}, \mathbb{E})$ : information association graph
**OUTPUT:** Updated InfoRank 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 $I_x^v(\bar{t}) \leftarrow \frac{r_x(\bar{t})}{R_x}$
    **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**
  Find missed interests ratio $I_m^v(t_{k+1})$, Coverage entropy $H^v(t_{k+1})$
  Compute $f_I^v(t_{k+1})$ using (2)
**end for**
**return** $C_v(t_{k+1})$

## V. Urban Sensing: Best Ranked IFVs Selection

We recruit the best set of important IFVs identified by InfoRank for different urban sensing tasks while maintaining a certain level of city-wide coverage with the minimum cost. We assume a monetary reward is paid as an incentive to each IFV proportional to the contributions as the spatio-temporal coverage based on their centrality and the fuel cost. Therefore, we propose an optimal IFV recruitment scheme where the objective is to minimize the recruitment cost for the selected set of IFVs with an acceptable level of coverage for different parts of the city. We also define the centrality constraint for an individual vehicle centrality as well as the overall centrality of the selected set of IFVs to be higher than a certain threshold. Moreover, we consider a maximum threshold amount of budget as payoff possible to an individual vehicle to ensure fairness between IFVs. Similarly, the overall budget assigned for all IFVs should not exceed a certain threshold amount dedicated by the municipality.

### A. Problem Formulation

We formulate the IFV selection as a discrete optimization problem where the objective function F $:\mathcal{V} \rightarrow \mathbb{R}_+$ finds the set of important IFVs, $\mathcal{V} = \{V : V \subset \mathbb{V}\}$, that minimize the overall recruitment cost. The budget function is defined as $F_b(V) = \sum_{v \in V} H_v$ proportional to the fuel cost based on each vehicle coverage scope in the set of selected IFVs $V \subset \mathbb{V}$. We use the coverage entropy $H_v$ defined earlier which reflects the overall vehicle coverage spread in the city. For a vehicle, the more the coverage entropy, the more it costs to be recruited.

Similarly, we define the centrality function as $F_c(V) = \sum_{v \in V} (1 + C_v)^{-\omega}$, where $C_v$ is the respective vehicle centrality computed using Algorithm 1 and $\omega$ is the tuning parameter used to adjust the vehicle centrality contribution towards the centrality function. Therefore, the objective function comprises both the budget function and the centrality function as $F(V) = F_b(V) + F_c(V)$. The necessary constraints are defined as:

*1) Spatio-temporal Coverage:* We define the area coverage vector $A_x(\overline{t_k})$ as the number of vehicles covering the area $x \in \mathbb{X}$ at time-slot $\bar{t}_k$. The coverage requirement specifies the number of vehicles needed in location $x$ at a particular time-slot depending on the location importance. The Spatio-temporal coverage constraint is divided into two parts:

- Minimum Possible Coverage: A location $x$ should be covered by at least $A^{\text{th}}$ number of vehicles at time-slot $\overline{t_k}$, where $A^{\text{th}}$ ("th" stands for threshold) specifies the lower bound for the spatio-temporal coverage of all locations. For instance, we can define $A^{\text{th}}(\overline{t_k}) = 1$ as the lower bound requiring at least one vehicle covering each location at the each time-slot.
- Desired Coverage: Each location $x$ should be covered by $A_x^{\text{th}}$ number of vehicles at time-slot $\overline{t_k}$, where $A_x^{\text{th}}$ is specified as the desired number of vehicles required in location $x$ depending on its popularity.

*2) Vehicle Centrality:* To ensure the selection of the important information hubs, we define a two-tier vehicle centrality constraint as:

- Individual Centrality: Each of the selected IFVs should have a centrality $C_v$ higher than a given individual threshold centrality $C^{th}$.
- Overall Centrality: The overall centrality score $R_V$ of all the selected vehicles should be higher than a given centrality benchmark $R^{th}$, where $R_V = \sum_{v \in V} C_v$.

*3) Dedicated Budget:* The maximum payable amount for the selected IFVs is limited by an upper bound depending on the dedicated budget. Similar to the centrality constraint, we define a two-tier budget constraint as:

- Individual Budget: In order to achieve fairness between vehicles, The payoff $B_v$ for an individual vehicle should not exceed the maximum budget $B^{th}$ dedicated for a single vehicle.
- Overall Budget: The overall assigned budget $B_V$ should also not exceed the budget $\mathbb{B}$ dedicated for all IFVs, where $B_V = \sum_{v \in V} B_v$.

The optimization problem needs to take into account the two coverage possibilities, therefore we formulate $\text{OPT}_1$ for the worst case to achieve the minimum possible spatio-temporal coverage as follows:

$\underline{\text{OPT}_1}$ : Minimum Possible Coverage

$$\begin{aligned} \underset{V}{\text{minimize}} \quad & F(V) \\ \text{subject to} \quad & \min_{x \in \mathbb{X}} A_x(\bar{t}_k) \geq A^{\text{th}}(\bar{t}_k) && (\text{C}_1) \\ & \min_{v \in V} C_v \geq C^{\text{th}} && (\text{C}_2) \\ & R_V \geq R^{\text{th}} && (\text{C}_3) \\ & \max_{v \in V} B_v \leq B^{\text{th}} && (\text{C}_4) \\ & B_V \leq \mathbb{B} && (\text{C}_5) \end{aligned}$$

Similarly, we formulate $\text{OPT}_2$ as the optimization problem considering the possibility to achieve the desired level of coverage for each location as:

$\underline{\text{OPT}_2}$ : Desired Coverage

$$\begin{aligned} \underset{V}{\text{minimize}} \quad & F(V) \\ \text{subject to} \quad & A_x(\bar{t}_k) \geq A_x^{\text{th}}(\bar{t}_k),\, \forall\, x \in \mathbb{X}, \forall\, \bar{t}_k \quad (\text{C}_6) \\ & \text{and} \quad (C_2),\, (C_3),\, (C_4),\, (C_5) \end{aligned}$$

The constraint ($\text{C}_1$) in $\text{OPT}_1$ concerns the minimum possible number of vehicles able to cover the location $x$. The constraint ($\text{C}_2$) and ($\text{C}_3$) address the vehicle centrality constraint, where ($\text{C}_4$) and ($\text{C}_5$) deals with the budget constraints. Similarly, the constraint ($\text{C}_6$) in $\text{OPT}_2$ address the desired spatio-temporal coverage requirements.

**Algorithm 2** Optimized IFVs Removal

```
1:  INPUT: Vehicles set 𝕍
2:  OUTPUT: Selected IFVs set V ⊂ 𝕍
3:  Initialize W = ϕ,
4:  for each vehicle v ∈ 𝕍 do
5:      while (R_V ≥ R^th) and (B_V ≤ 𝔹) and (A_x(t̄) ≥ A^th),
        ∀t̄, ∀x ∈ X  do
6:          A_x(t̄) ← A_x(t̄) − A_{x_v}(t̄)
7:          R_V = R_V − C_v
8:          B_V = B_V + B_v
9:          W ← W ∪ v
10:     end while
11: end for
12: return V
```

### B. Algorithm: Optimized IFVs selection

The selection algorithm for the best ranked IFVs in the network are summarized in Algorithm 2 and 3. We assume vehicles $\mathbb{V}$ are subscribed as IFVs in order to provide different urban sensory operations and tasks. In Algorithm 2, we start removing each vehicle from the set of all IFVs $\mathbb{V}$ until sufficient IFVs are left as the subset $V \subset \mathbb{V}$. Line 5 ensures the overall centrality, budget and coverage constraints are respected. Line 6-9 removes the vehicle, subtract its centrality from overall centrality score as well as remove its cost from

**Algorithm 3** Optimized IFVs Revisiting

1: **INPUT:** Removed vehicles set $W \subset \mathbb{V}$, Selected IFVs set $V$
2: **OUTPUT:** Updated removed vehicles set $W \subset \mathbb{V}$, Selected IFVs set $V \subset \mathbb{V}$
3: Initialize step = 0, vehicles revisiting set $U = rand(u, step)$
4: step = step + 1
5: **while** ($R_V \geq R^{th}$) and ($B_V \leq \mathbb{B}$) and ($A_x(\bar{t}) \geq A^{th}$), $\forall \bar{t}, \forall x \in X$ **do**
6: **for** each vehicle $u \in U$ **do**
7: **if** ($|W| \geq 0$) and (step $\leq |V|$) and ($C_v \geq C^{th}$) and ($B_v \leq B^{th}$) **then**
8: $V \leftarrow V \cup u$
9: $A_x(\bar{t}) \leftarrow A_x(\bar{t}) + \sum_{i=1}^{step} A_{x_u}(\bar{t})$
10: $R_V = R_V + \sum_{i=1}^{step} C_u$
11: $B_V = B_V - \sum_{i=1}^{step} B_u$
12: $W \leftarrow W - U$
13: **end if**
14: **end for**
15: **end while**
16: **return** $V, W$

the overall cost. The removed vehicles are then added to the set $W$ until a local minima is reached.

In order to search for another local minima or the global minima, vehicles from the removed set $W$ are added back step-wise with an increase in the step size at each iteration as shown in Algorithm 3. The removed vehicles from the set $W$ are revisited to be added back to the selected IFVs set $V$ using the temporary revisiting vehicle set $U$. The set $U$ contains randomly selected vehicles from the removed set $W$ where the number of selected vehicles are specified by the step size. The condition in line 7 ensures the existence of vehicles in the removed set as well as defining the upper bound on the step size for the number of vehicles considered to add back at each iteration. The respective vehicle coverage, centrality and cost are updated for the set of vehicles added back in line 9-11. The algorithm continue to optimize until the desired set of vehicles $V$ which minimize the objective function is found.

## VI. Performance Evaluation

One of the basic requirement for evaluating the efficiency of our ranking system is scalability with the objective to find answers to the fundamental question: *How well can we identify the top IFVs?* Therefore, we use Network Simulator-3 (NS-3) as a scalable simulation platform for upto three thousand vehicles. The performance of the proposed algorithms is validated by implementing InfoRank 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 [12]. 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.

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 |

### A. Simulation Scenario

We simulate a VANET urban sensing scenario using the ndnSIM [13] 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. To incorporate the effect of buildings and environment, we implement a combination of Nakagami propagation loss model for fast fading and Log distance path loss propagation model with three distance fields (near, middle and far) with different exponents. The simulation scenario implements the following two applications:

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

*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 provider vehicles.

We assume the interests follow a Zipf distribution, where interests for popular contents are more frequent, therefore, we use Zipf distribution for interest generations in our simulations. We use Voronoi tessellations in our simulations to divide the urban map where randomly selected geographical coordinates are used to form different regions and the vehicles in the corresponding locations are associated to the formed Voronoi regions. It enables us to divide the entire map into logical regions which can be considered as different neighborhoods. Therefore, 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 provider acting as source for an area upon receiving the interest responds with the desired content. We perform each simulation upto ten times by analyzing different set of nodes (vehicles) randomly assigned as information providers and consumers.

In order to evaluate InfoRank as a generic vehicle ranking algorithm, the tuning parameters $\delta$, $\epsilon$, $\lambda$ and $\theta$ are set to $0.5$ in order to maintain generality since the significance of each parameter depends on the application requirements. We set their values to $0.5$ solely for experimentation purposes. However, all these tunable parameters can be defined or adjusted according the application/user requirements. For each simulation scenario, we rank the top information facilitator vehicles in the network by comparing their InfoRank scores with the respective Degree, Closeness, Betweenness and Eigenvector centrality score. For better performance analysis in different simulation scenarios, we consider the following performance metrics in comparison with the state of the art centrality schemes:

- Cumulative Satisfied Interests (CSI) for the top identified vehicles by each centrality scheme
- Comparison of top vehicles identified by each centrality scheme with their respective centrality scores.
- Average aggregated throughput of the identified important vehicles by each centrality scheme
- Cache hit rate for the top vehicles by each scheme to evaluate our algorithms along ICN in VANET mobility scenarios.

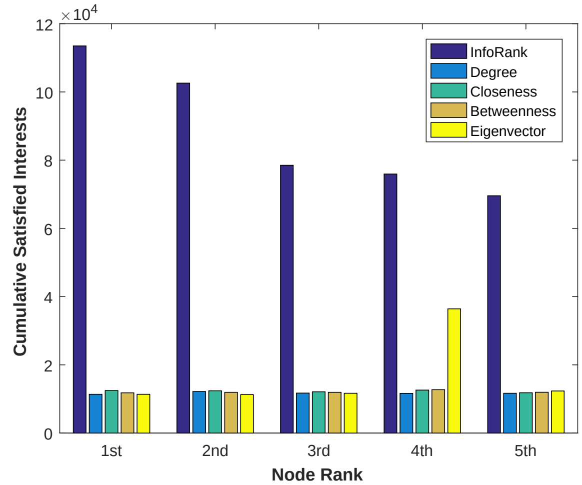



Figure 3: Comparison of the Cumulative Satisfied Interests by the top identified vehicles using each centrality scheme

### *B. Results: Individual Vehicle Ranking*

In each simulation, different vehicles are assigned randomly as providers for a location they already visited before. Similarly, different consumer vehicles are assigned randomly for vehicles planning to visit a location. The purpose is to show the persistence of InfoRank as well as its efficiency as a ranking scheme. It is important to show that it actually"ranks" with different set of providers and consumers in different scenarios. We normalize the score by assigning unity to the most important vehicle followed by the remaining vehicles to clearly show the ranking order. We will use the same convention to interpret results in the later sections. For each rank, the average score lies within a confidence interval of $0.01$ for a confidence level of $95\%$. In the first simulation, the vehicle $1115$ is identified to have the top InfoRank score among the selected vehicles in the network as it responds more frequently to the relevant user interests throughout the simulation.

It is to note that the number of "most important" vehicles depends on the application requirements, for example in the case of city-wide urban sensing, the municipality is limited by the budget and coverage constraints as discussed earlier in Section V.

*1) Cumulative Satisfied Interests:* The term Cumulative Satisfied Interests refers to the total number of user interests satisfied during the simulation duration. CSI is the measure of the vehicle importance with respect to the user interests. Increase in the number of satisfied interest implies a high vehicle association to the particular content for the incoming interests. Figure 3 compares the CSI of the five most prominent vehicles identified by InfoRank with those identified using different centrality schemes in an average of ten set of simulations.

Typical ranking schemes only takes into account physical topology towards computing a node importance in the network, ignoring the vehicle relevance with respect to the user interests. Nevertheless, vehicles identified by InfoRank as the top information facilitators satisfied five times more user interests than those identified using other schemes in all simulations, thus heavily outperforming all existing centrality metrics in the dynamic nature of vehicular environment. Such huge difference is due to the consideration of user interest satisfaction as a key metric towards information importance as well as the vehicle importance in the network.

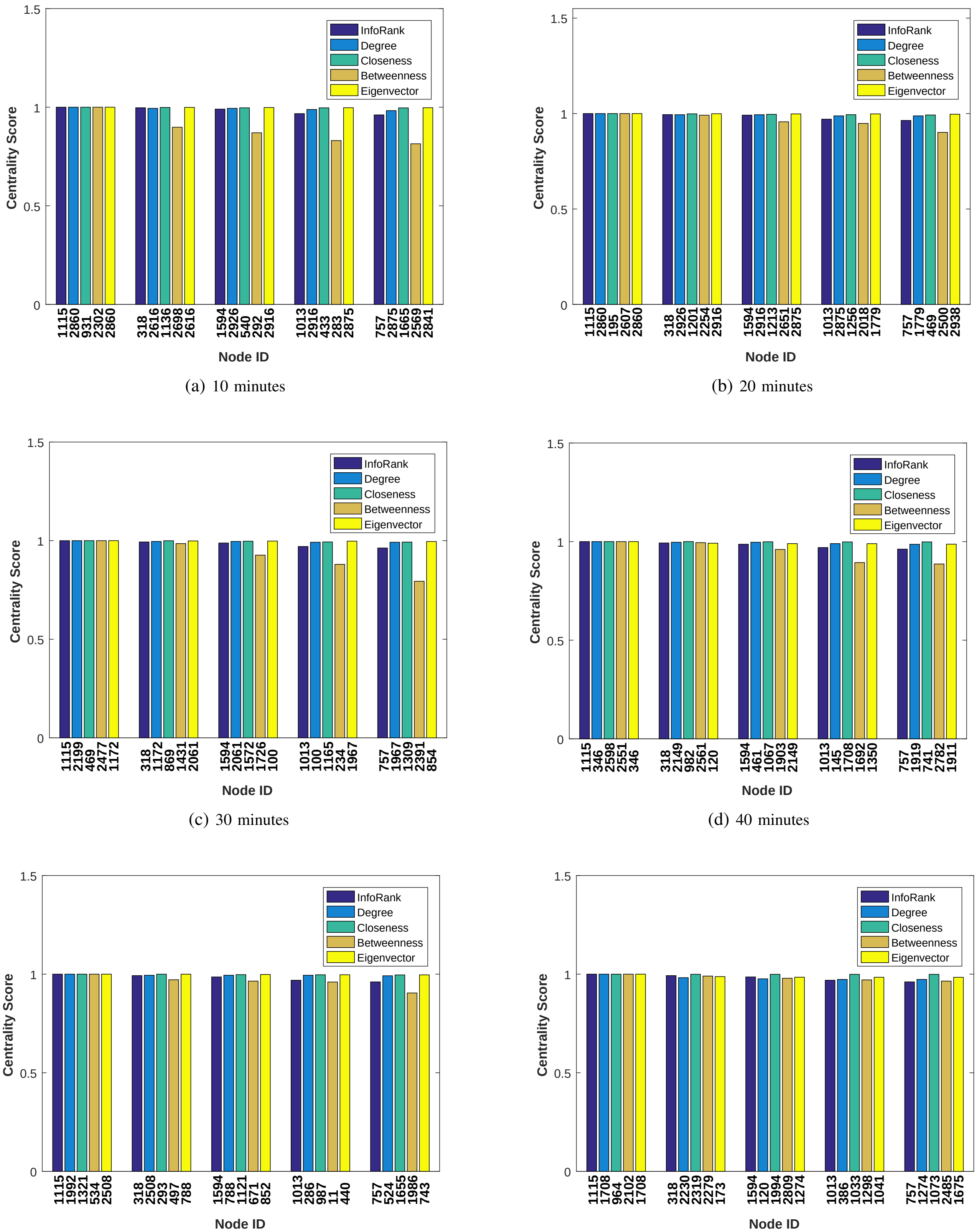
*2) Temporal behavior analysis of the top nodes:* It is important to efficiently analyze the time varying behavior of our algorithm due to the dynamic VANET environment to address two questions:

- Is InfoRank successful in identifying vehicles that can be relied for urban sensing over stable time duration?
- What happens if we use the state of the art schemes (metrics) to identify important vehicles? Why it is not feasible to use existing schemes?

The major purpose is to analyze the efficacy of InfoRank as a stable ranking algorithm in terms of its capability to rank vehicles. Secondly, we need to identify vehicle to be recruited for a relatively longer time period in order to be relied upon for different services (urban sensing in our case). Therefore we compare it with existing ranking systems with an analysis over time as a novel approach to study whether it is stable unlike the state of the art schemes.

Centrality score of the top five nodes identified by InfoRank and different centrality schemes are shown by the periodic network snapshots after each 10 minutes in Figure 4. We consider the top node identified by each scheme as a benchmark by assigning it a unity score. At the beginning, the vehicle $1115$ is ranked as the top vehicles by InfoRank, though the other schemes underrated it. Similarly we observe that overall a stable set of vehicles identified by InfoRank, i.e. the vehicle ID $1115$, $318$, $1594$, $1013$ and $757$, thus validating the feasibility of recruiting stable set of vehicles over longer periods.

After comparing the top nodes identified by each scheme, we observe that unlike InfoRank, other centrality schemes




(a) 10 minutes

(b) 20 minutes

(c) 30 minutes

(d) 40 minutes

(e) 50 minutes

(f) 60 minutes

Figure 4: Temporal snapshots after each 10 minutes comparing top identified nodes by each schemes - InfoRank

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. InfoRank ensures stable set of top ranked vehicles as it is clear from the time varying VANET analysis that they are not affected by the network dynamics. We are able to rank each vehicle considering relatively stable metrics, which is not the case for other schemes.

*3) Throughput:* We also evaluate InfoRank by analyzing the throughput achieved at the identified important nodes in the network. Figure 5 shows the aggregated per node throughput of the most prominent nodes identified by InfoRank compared to other centrality scheme. The average aggregated throughput (Kbps) is computed over the entire simulation duration for the ten set of simulations. The top nodes identified by InfoRank turns out to be an outlier for the amount of throughput compared to the nodes identified by other schemes. We observe in Figure 5 that the throughput of only the top most vehicles identified by Degree, Closeness and Eigenvector centrality is about half the amount of those identified by InfoRank. It is also clear that all other schemes results in a negligible amount of the throughput along the downward ranking order as the depend on the physical topology which is greatly affected by high vehicle mobility and intermittent connectivity. Moreover, we also notice that the nodes identified by betweenness centrality yields no substantial throughput along the top ranked order. Another reason for this behavior is the constant user interest satisfaction ratio from the best vehicles identified using InfoRank during the simulation duration.

*4) ICN Evaluation - In-Network Caching:* We evaluate the ICN built-in feature of In-Network caching at the intermediate nodes by computing the cache hit rate at the top nodes identified by each scheme as shown in Figure 6. 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 the ten set of simulations. The most important nodes identified by InfoRank in Figure 6 yield a higher hit rate of around 45% compared all the other schemes which resulted in negligible amount of around 2% on average cache hits during the simulation. This is because InfoRank considers information importance as a key factor, thus, the vehicle containing important information responds and subsequently cache more frequently compared to other vehicles. This proves that In-Network caching offered by ICN in InfoRank implementation overcomes the mobility and intermittent connectivity constraints in VANETs for efficient content access.

### *C. Results: Best set of IFVs selection*

In this section we discuss the results obtained after applying the optimization Algorithms 2 and 3 to identify the set of vehicles selected as the best set of IFVs for urban sensing under the constraints associated to the spatio-temporal coverage, vehicle centrality and the available budget.

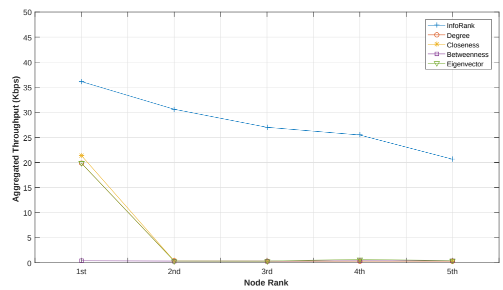



Figure 5: Comparison of the average aggregated per node throughput achieved by the top identified vehicles using each centrality scheme

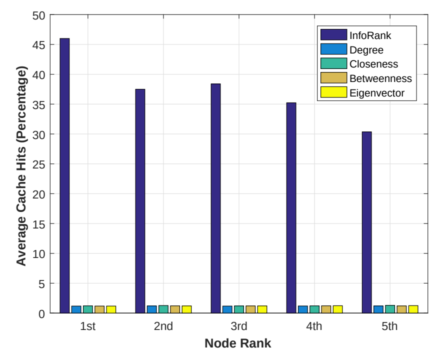



Figure 6: Comparison of the average cumulative cache hit rate at the top identified vehicles using each centrality scheme

*1) Cumulative Satisfied Interests:* In order to validate the optimized set of selected IFVs, we find the CSI by the vehicles identified as the best set of IFVs using our proposed IFV optimization stated in $OPT_1$ and $OPT_2$. We compare the CSI of the proposed optimizations algorithm with (1) the NP-Hard budgeted maximum coverage optimization problem [14], where the objective function targets to select the set of vehicles with the minimum cost (only low cost), while achieving the maximum coverage, (2) set of IFVs selected based on high vehicle centrality score as the objective function (only best ranked) and (3) the greedy approach of our optimization algorithm to select the best set of IFVs based on their high centrality score with the minimum cost (best ranked and low cost). We also compare the best set of IFVs identified by InfoRank by applying our optimization algorithms using the state of the art centrality schemes.

Figure 7 depict the ratio of the cumulative interests satisfied to the total received interests by the set of best ranked IFVs at different time instants. From the temporal analysis, we observe that the selected set of IFVs by InfoRank using our approach satisfied 60% user interests compared to an average of below 10% by other selection algorithms at different time-instants.

Additionally, results from applying our optimized IFVs selection algorithms to each centrality scheme are shown in

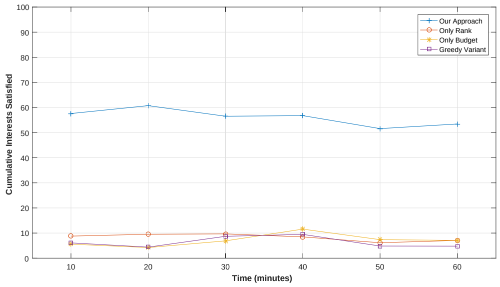



Figure 7: Temporal analysis of the average CSI by the optimized set of IFVs selected using each algorithm

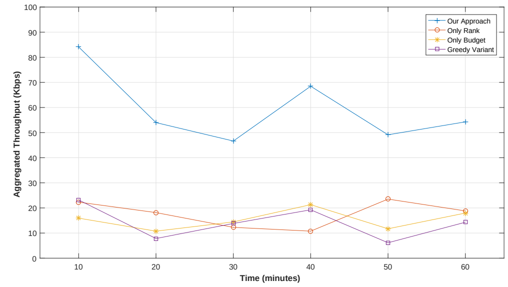



Figure 9: Temporal analysis of the throughput achieved by the optimized set of IFVs selected using each algorithm

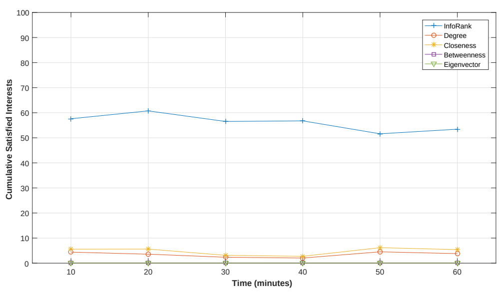



Figure 8: Temporal analysis of the average CSI by the optimized set of IFVs selected using each centrality scheme

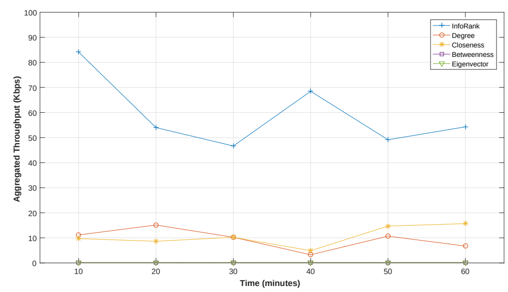



Figure 10: Temporal analysis of the throughput achieved by the optimized set of IFVs selected using each centrality scheme

Figure 8. Interestingly, compared to the 60% user interests satisfied by the set of nodes identified using InfoRank, none of the other centrality schemes succeeded in identifying a set of IFVs which satisfied more than 10% of the received user interest. We also observe that the set of IFVs identified by the famous eigenvector centrality failed to satisfy any interest generated in the network. InfoRank outperforms existing metrics as the optimal set of best ranked IFVs by InfoRank are co-located at important locations with respect to the user interests. Similarly, Algorithm 2 and 3 ensures that the best ranked IFVs by our scheme provide the desired coverage with the least possible cost. On the other hand, other centrality schemes ignore metrics associated to user interest and spatio-temporal coverage for the identifications of important nodes.

2) *Throughput:* In addition to individual per node throughput for the top ranked IFVs, we analyze the aggregated throughput of the set of best ranked IFVs selected for urban sensing. Figure 9 compares the throughput achieved by the set of best ranked IFVs identified using our optimization Algorithms 2 and 3 with other approaches. The efficiency of InfoRank using our proposed IFV selection algorithm is evident since we observe that the selected set of best IFVs achieved around three times more throughput compared to those identified by other approaches. It is due to high relevance to the user interests and the availability to satisfy interests (provide coverage) at different geographical locations at lower cost.

Similarly, the aggregated throughput of the selected set of vehicles by applying our optimized IFV selection algorithm using state of the art centrality schemes is shown in Figure 10. We observe that InfoRank results in more aggregated throughput (maximum 83 Kbps at around 10 minutes) compared to any other centrality scheme, where none of the existing metric yields a set of vehicles exceeding a maximum throughput of 20 Kbps. Thus, it clearly indicates that Algorithms 2 and 3 with InfoRank for selecting the appropriate set of IFVs outperforms the state of the art centrality schemes.

### *D. Discussion*

Finally, we are able to comment on the question we posed in the beginning of this section: *How well can we identify the top IFVs?* From the simulation results by implementing InfoRank and the optimized IFVs selection algorithm, following deduction can be made: A relatively stable set of top IFVs are identified by our vehicle ranking system compared to other schemes, thus we can easily identify and recruit important vehicles for various urban sensing tasks. We identified vehicles which satisfied more user interests with higher aggregated per node throughput and more cache hit rate compared to the other schemes.

Similarly, the best ranked set of IFVs selected using our approach resulted in higher user interests satisfaction ratio as well as higher aggregated throughput compared to other selection algorithms. At the same time, they provide efficient urban sensing facilities for the respective locations while achieving the desired spatio-temporal coverage with the minimum cost. Thus, the overall comparative analysis of our ranking system with different network ranking schemes in the literature suggests it as an efficient ranking system feasible in VANETs.

## VII. Conclusions and Future Directions

Identifying the appropriate vehicles for the collection, storage and distribution of sensory data from the fleet of vehicles on urban streets is a challenging task. This paper introduced an innovative vehicle ranking system comprising an information-centric algorithm, "InfoRank", allowing smart vehicles to rank themselves in a fully distributed VANET. It first ranks important location-aware information based on the user interests satisfaction and information validity scope and then uses the information importance to find the vehicle importance in the network. We also proposed an optimum selection algorithm for the set of best ranked vehicles for different city-wide urban sensing tasks. Results by comparing with state of the art schemes revealed that the proposed ranking system is best suited to efficiently identify important information facilitator vehicles in VANETs compared to other schemes.

Urban sensing and vicinity monitoring are the possible application of our ranking system as discussed before. Recruited set of vehicles with high centrality score can collaborate for the data collection, storage and distribution in VANETs. Designing such efficient schemes will be the subject of our future research.

## Acknowledgment

This work is funded by the French Ministry of Industry under the CarCoDe (ITEA2 11037) project.

## 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] J. A. Khan, Y. Ghamri-Doudane, and D. Botvich, "Inforank: Information-centric autonomous identification of popular smart vehicles," in *Vehicular Technologoy Conference VTC Fall, 2015 Proceedings IEEE*, pp. 1–6.

[3] 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.

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

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

[6] S. A. Hamid, H. Abouzeid, H. S. Hassanein, and G. Takahara, "Optimal recruitment of smart vehicles for reputation-aware public sensing," in *Wireless Communications and Networking Conference (WCNC), 2014 IEEE*. IEEE, 2014, pp. 3160–3165.

[7] P. TalebiFard, V. C. Leung, M. Amadeo, C. Campolo, and A. Molinaro, "Information-centric networking for vanets," in *Vehicular ad hoc Networks*. Springer, 2015, pp. 503–524.

[8] J. Scott, *Social network analysis*. Sage, 2012.

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

[10] A. M. Vegni and V. Loscri, "A survey on vehicular social networks," *Communications Surveys & Tutorials, IEEE*, vol. 17, no. 4, pp. 2397–2419, 2015.

[11] K. Wei, X. Liang, and K. Xu, "A survey of social-aware routing protocols in delay tolerant networks: applications, taxonomy and design-related issues," *Communications Surveys & Tutorials, IEEE*, vol. 16, no. 1, pp. 556–578, 2014.

[12] 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.

[13] 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

[14] S. Khuller, A. Moss, and J. S. Naor, "The budgeted maximum coverage problem," *Information Processing Letters*, vol. 70, no. 1, pp. 39–45, 1999.

**Junaid Ahmed Khan** (junaidakhan88@gmail.com) is currently a research fellow at the Computer Engineering Department, University of California, Santa Cruz, since June 2016. He has been a Ph.D candidate in computer science at the Gaspard-Monge Computer Science Laboratory (LIGM–UMR 8049) University of Paris-Est, Marne-la-Vallee, France since August 2013. He has also been a visiting researcher at the University of La Rochelle, France, from December 2013 till September 2016. He completed his B.S. degree in telecommunication engineering from the National University of Computer and Emerging Sciences, Islamabad, Pakistan, in June 2009. In May 2013, he received a Master's degree in Electrical Engineering from the School of Electrical Engineering and Computer Science, National University of Sciences and Technology, Islamabad, Pakistan. His research interests are information-centric networks, vehicular networks with focus on Autonomous/Smart Vehicles, Fog Networking, Internet of Things (IoT), and complex/social networks analysis.

**Yacine Ghamri-Doudane** is currently a full professor at the University of La Rochelle and director of its Laboratory of Informatics, Image and Interaction. Before that, he held an assistant/associate professor position at ENSIIE, a major French post-graduate school located in Evry, France, and was a member of LIGM–UMR 8049. From February 2011 to July 2012, he regularly visited the Performance Engineering Laboratory of University College Dublin, Ireland. He received an engineering degree in computer science from the National Institute of Computer Science, Algiers, Algeria, in 1998, an M.S. degree in signal, image, and speech processing from the National Institute of Applied Sciences, Lyon, France, in 1999, a Ph.D. degree in computer networks from University Pierre & Marie Curie, Paris 6, France, in 2003, and a Habilitation to Direct Research in computer science from Université Paris-Est in 2010. His current research interests lie in the area of wireless networking and mobile computing, with a current emphasis on topics related to IoT, wireless sensor networks, and vehicular networks. He holds three international patents, and has authored or co-authored seven book chapters, about 25 peer-reviewed international journal articles, and more than 80 peer-reviewed conference papers. Since 1999, he has participated in several national and European-wide research projects in his areas of interest, among them two regional research projects, two national research projects, nine European-wide research projects (one ongoing one in 2015–2016), as well as three EU COST Actions. As part of his professional activities linked to the computer networking research community, he also acted as Chair the IEEE ComSoc Technical Committee on Information Infrastructure & Networking from January 2010 to December 2013, and is currently chairing the IEEE ComSoc Humanitarian Communications Technologies Ad Hoc Committee (HCTC). He is an Editorial Board member of the IEEE IoT Journal, Springer AoT, Elsevier JNCA, and Wiley WCMC, is co-Editor-in-Chief of Springer/KICS CT Express, as well as the founding Editor-in-Chief of the IEEE ComSoc Ad Hoc and Sensor Network Technical Committee Newsletter. Among other conference involvements, he has acted as the TPC Chair of IEEE CCNC 2015, and Symposium Co-Chair of IEEE ICC 2009, 2010, and 2012 as well as IEEE GLOBECOM 2012 and 2015.