POST: Exploiting Dynamic Sociality for Mobile Advertising in Vehicular Networks

Jun Qin, Hongzi Zhu, *Member, IEEE*, Yanmin Zhu, *Member, IEEE*, Li Lu, *Member, IEEE*, Guangtao Xue, *Member, IEEE*, and Minglu Li, *Senior Member, IEEE*

Abstract—Mobile advertising in vehicular networks is of great interest with which timely information can be fast spread into the network. Given a limited budget for hiring seed vehicles, how to achieve the maximum advertising coverage within a given period of time is NP-hard. In this paper, we propose an innovative scheme, POST, for mobile advertising in vehicular networks. The POST design is based on two key observations we have found by analyzing three large-scale vehicular traces. First, vehicles demonstrate dynamic sociality in the network; second, such vehicular sociality has strong temporal correlations. With the knowledge, POST uses Markov chains to infer future vehicular sociality and adopts two greedy heuristics to select the most "centric" vehicles as seeds for mobile advertising. Extensive simulations based on three real data sets of taxi and bus traces have been carried out. The results show that POST can greatly improve the coverage and the intensity of advertising. For all the three involved data sets, it achieves an average gain of 64 percent comparing with the state-of-art schemes.

Index Terms—Vehicular networks, mobile advertising, dynamic sociality, social network analysis

1 Introduction

Vehicular networks are emerging as a new landscape of mobile ad hoc networks, aiming to provide a wide spectrum of safety and comfort applications to drivers and passengers. In vehicular networks, vehicles equipped with wireless communication devices can transfer data with each other via vehicle-to-vehicle (V2V) communications as well as with the roadside infrastructure via vehicle-to-roadside (V2R) communications. With vehicular networks, a wide range of new Intelligent Transportation System (ITS) applications are enabled, ranging from hazard warning, collision avoidance, and traffic management to navigation based on real-time traffic condition, trip planning and optimal route selection.

Among all the others, *mobile advertising* is an appealing application, where a small number of public vehicles such as taxis and buses, called *seed* vehicles, are chosen to propagate timely digital advertisements to other vehicles in the network. In such application scenarios, a seed can forward or "post" its advertisements to its neighboring vehicles or other mobile devices (e.g., smart phones) via short-range wireless communications when they approach to each other (called a *contact*). As a result, the advertisement information can be gradually spread out within the network. Considering the limited budget, the goal of the application is to select the best set of seeds paid to post a piece of timely advertisement so that the total number of vehicles seeing the advertisements within a given period of time is maximized. With an effective mobile advertising scheme, a great deal of timely and important information, such as instant municipal announcements, real-time traffic congestion information, and commercial promotion activities, can be fast propagated among mobile devices at very low cost.

To realize the mobile advertising application, however, is very challenging due to three reasons. First, as vehicles move, the topology of the network varies fast over time. It is very hard to know the exact future information of the network. Second, even if the network topology is known, we prove that the problem of choosing a given number of vehicles to post advertisements such that the network coverage is maximal within a given time period is NP-hard. Third, since a piece of advertisement may be time-critical, it should be spread as widely as possible while the content of the advertisement is still valid, which makes the problem even harder. One straightforward scheme might be to randomly choose a fixed number of vehicles as seeds to flood the network. It is simple but there is no guarantee that those randomly chosen seeds can always have the optimal performance.

In the literature, recent work has studied the influence maximization problem in the area of social network analysis [1], [2], [3]. Based on static social networks, individual influence has been measured using various centrality metrics and utilized to choose good candidate nodes to spread information. Although these studies shed the light on how to select preferable seeds, they cannot be directly applied in the mobile advertising problem as they are based on traditional static social network of which the topology is stable. In the context of data dissemination in vehicular and opportunistic networks, previous work came up with various routing mechanisms [4], [5], [6], originating from the field of conventional mobile ad hoc networks. Those studies mainly

- *J. Qin, H. Zhu, Y. Zhu, G. Xue, and M. Li are with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, P. R. China.*
  E-mail: {qin.jun, mlli}@sjtu.edu.cn, {hongzi, yzhu, xue-gt}@cs.sjtu.edu.cn.
- *L. Lu is with the School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, P. R. China. E-mail: luli2009@uestc.edu.cn.*

Manuscript received 12 Apr. 2015; revised 6 July 2015; accepted 9 Aug. 2015. Date of publication 11 Aug. 2015; date of current version 18 May 2016.
Recommended for acceptance by J. Chen.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TPDS.2015.2467392

focus on relatively short range data coverage and forwarding performance in terms of delivery ratio, delay and network overhead. In the mobile advertising problem, however, the major concern is how to select a best set of seed vehicles so that they can achieve as large coverage as possible within a given period of time. As a result, there is no successful solution, to the best of our knowledge, to addressing the mobile advertising problem in vehicular networks.

In this paper, we first take an empirical study on three real large-scale vehicular traces. By aggregating pairwise contacts, we find that vehicles demonstrate clear sociality in the extracted contact graphs, forming *vehicular social networks*. Considering the inherent dynamics of vehicular networks, we investigate how vehicular sociality, especially on three representative centrality metrics, changes over time by dividing time into slots of equal length under different scales of time. We have two key observations. First, the vehicular social network presents high dynamics which means the position or the "role" of individual vehicles in the network changes fast when it is observed under a small scale of time. Second, the dynamics of vehicular sociality show strong temporal patterns and correlations. With this knowledge, we propose an innovative scheme, POST, for mobile advertising in vehicular networks. The core idea of POST is to capture and utilize the temporal correlations of vehicular sociality to infer the contact behavior of the whole network in future, which can be further leveraged to improve the performance of mobile advertising. To this end, POST integrates three techniques: *capturing centrality correlations*, *inferring future vehicular centrality* and *selecting preferable seeds*. In particular, as the mobile advertising problem is NP-hard, POST adopts two greedy strategies to select the most "centric" vehicles to serve as seeds for mobile advertising. Moreover, the equivalence of the network position of those most centric vehicles is also considered to further improve the advertising coverage. Results of extensive trace-driven simulations demonstrate the efficacy of POST design. Comparing to the state-of-art algorithms, POST can achieve a coverage gain of 64 percent on average for all the three involved data sets.

We highlight our main contributions in this paper as follows:

- We first define the mobile advertising problem in vehicular networks and prove its NP-hardness, where the budget for hiring seed vehicles to post timely advertisements is limited.
- We conduct an extensive measurement study on social centrality characteristics of vehicles in the network. Through our measurements, we observe that the vehicular social centrality is highly dynamic and moreover shows strong temporal correlations.
- We propose an innovative scheme, POST, for addressing the mobile advertising problem in vehicular networks, which fully utilizes the temporal correlations of vehicular social centrality to effectively select a small set of seed vehicles to propagate an advertisement with the goal of maximizing the coverage of this advertisement within a short tenancy. We also demonstrate the effectiveness of POST through extensive trace-driven simulations.

The remainder of this paper is organized as following. Section 2 is the presentation of related work. We describe the system model and the mobile advertising problem in Section 3. In Section 4, we study the empirical trace data and analyze vehicular sociality. The findings on the high dynamics and temporal correlations of vehicular centrality are presented in Section 5. Section 6 is dedicated for the detailed design of POST. Section 7 presents the performance evaluation of POST through empirical trace-driven simulations. We conclude this paper in Section 8.

2 Related Work

The mobile advertising problem is most related to the influence maximization problem in the areas of social network analysis and online marketing. We introduce those studies and compare them with our work in this section.

In the area of social network analysis, the influence of individuals in spread processes is widely studied. Kitsak et al. [8] studied various effects of several classic centrality measures for choosing good spreaders in order to obtain the optimal performance in designing dissemination strategies throughout various complex networks. Their work suggests that in order to have the propagation area as large as possible, it is necessary to choose the most important nodes which are not connected directly in the early stage of the dissemination task. The work in [9] studied what is the key factor to influence a node in social contagion processes. Their detailed study shows that the decision made by a node in the network is highly influenced by its connected neighbors. In the work [10], it suggests that the importance of a node depends not only on its popularity but also the similarity between different nodes. In addition, similarity can be utilized to predict new linkage appearance in the future.

In the area of online marketing, as the Internet has expanded to the largest information platform, the influence maximization problem proposed by Domingos and Richardson [11], [12] has been studied as a fundamental algorithmic problem in related applications. Kempe et al. [1], [13] prove the NP-hardness of the maximization problem under the classic spread models in their work and have proposed their approximation algorithm based on influential node identification techniques in the social networks. In the following work, Karsai et al. [14] found that the information diffusion speed in social networks is usually slower than expected. They suggest that this phenomenon is because of the various correlations, e.g. the community structures embedded in the graph and topological correlations.

Although the above studies shed the light on how to select preferable seeds, they cannot be directly applied in the mobile advertising problem as they are based on traditional static social network where the topology of the network is relatively stable.

In vehicular and opportunistic networks, there are a large number of studies on data dissemination [4], [5], [6], [7]. Conventionally, various protocols originated from the field of traditional mobile ad hoc networks are on the contact-level. Some methods utilize local information about the groups of moving nodes, while some are based on dedicated infrastructures for content distribution, both increasing the opportunities of data forwarding. In recent years, the emerging routing and relaying solutions based on social network analysis techniques studied the problem from a new perspective on improving the dissemination efficiency. Daly and Haahr [15] proposed SimBet in their work. SimBet tries to utilize the similarity between nodes in a social network to increase the probability of successfully packet delivery via the most central nodes and community structures in the network. Similarly, Bubble Rap [16] also utilizes the importance of the nodes with their centrality in the relaying. The work in [17] explores geographic and social regularities of node mobility to design routes for greater chances of disseminating data in mobile social networks. ContentPlace [18] is a method that exploits social behaviors of the users in decentralized interest exchange. DelQue [19] utilizes both mobility and social relations of mobile users in the selection of relays for interest and response forwarding in information search schemes. Those studies focus on how to improve end-to-end delivery performance in terms of delivery ratio, delay and network overhead. In mobile advertising problem, however, the major concern is how to select a best set of seed vehicles so that they can achieve as large coverage as possible within a given period of time.

In conclusion, there exists no successful solution, to the best of our knowledge, to solving the mobile advertising problem in vehicular networks.

3 System Description and Problem Definition

3.1 System Description

We consider building the mobile advertising application upon urban vehicular networks, where the initial advertisements can be downloaded to seed vehicles via vehicle-to-roadside communications or other infrastructure-aided communications such as Wi-Fi and 2G/3G networks. Seeds may actively forward advertisements to encountered non-seed vehicles, while non-seeds do not carry out any forwarding tasks. The propagation of advertisements relies on opportunistic short-range wireless communications such as Bluetooth, Wi-Fi and DSRC [20].

The main advantage of using vehicles for advertising lies in three folds. First, as there is no need to deploy new billboards built as infrastructure, it can enormously save the deployment cost. Second, it also has much lower system maintenance cost than directly dispatching those advertisements to all vehicles via cellular networks (e.g., 2G/3G) since propagating advertisements via short-range wireless communications is free of charge. Third, it can also achieve extraordinary coverage utilizing the mobility of vehicles comparing to statically posted advertisements in tradition. We further consider using public commuting vehicles like taxis and buses served as seeds because they are public service vehicles and therefore have less privacy issues and have longer service time and larger areas comparing to normal vehicles.

3.2 Problem Definition and Its Difficulty

As many advertisements are time-critical, it is preferable to spread out those advertisements before their contents are out-of-date. Furthermore, we also consider the budget to deploy advertisements given a price for employing vehicles. Therefore, we define our mobile advertising problem as follows:

Definition 1 (Mobile Advertising, MA). *Given the budget B and price p for employing vehicles to propagate a piece of advertisement in a vehicular network within a given period of time T, how to select the best B/p vehicles in the network so that the total number of vehicles seeing the advertisement in the network is maximized?*

The mobile advertising problem is hard and we have the following theorem.

Theorem 1. *The mobile advertising problem is NP-hard.*

Proof. We prove the NP hardness by devising a polynomial reduction from a classic NP problem, *Max k-cover* [21], to our problem. The *Max k-cover* problem can be described as follows. Given a collection of subsets, $\mathcal{F} = \{S_1, S_2, \ldots, S_s\}$, of a set of n points, S, the objective is to select k subsets from $\mathcal{F}$ such that the total number of points contained in their union is maximized.

The reduction takes an instance of the *Max k-cover* problem as input. We construct an instance of the mobile advertising problem as follows. Assume that all future movements of all vehicles in the network are known. With this assumption, we know all future communication opportunities between any pair of vehicles within the given period of time. We can construct a graph, $\mathcal{G}(N, E)$, where N is the set of nodes and E is the set of edges. Each vehicle in the network is a node in the graph and there is an edge between a pair of nodes in the graph if the corresponding vehicles can communicate at least once within the given period of time. Denote the set of each node n_i and all its neighbors, i.e., the nodes covered by n_i, as $S_i \subseteq N$ for $i = 1, \ldots, |N|$. With this graph, the problem is to find $k = B/p$ different S_j for $j = 1, \ldots, k$ so that their union contains as many nodes as possible. Therefore, the mobile advertising problem is a NP-hard problem, which concludes the proof. □

As it is very hard, if not impossible to know all future information about the network, the mobile advertising problem can be even harder. We study this problem through an empirical methodology and elaborate the process in the following sections.

4 Sociality Analysis on Empirical Traces

4.1 Collecting Vehicular Traces

In order to understand realistic vehicular mobility and conduct informed design of mobile advertising schemes in vehicular networks, it is of great importance to study the empirical data in terms of frequency and temporal distribution of contacts among them. For this purpose, we use three data sets consisting of traces from two metropolises in China and two types of vehicles, i.e., buses and taxis. Key statistics of the traces are listed in Table 1.

Shanghai buses. The trace consists of GPS reports sent by 2,501 buses which serve on 100 routes and cover the whole downtown area in Shanghai between February 19 and March 5, 2007. A commuting bus periodically sends GPS reports back to a backend data center via GPRS channel. The specific information contained in such a report includes: ID, the longitude and latitude coordinates of the

TABLE 1
Main Statistics of Three Data Sets

Data Set	*Shanghai Taxi*	*Shanghai Bus*	*Shenzhen Taxi*
Vehicle number	2,109	2,501	8,291
From date	Feb. 1, 2007	Feb. 19, 2007	Oct. 1, 2009
Duration (day)	31	15	31
Granularity (sec.)	15*, 60**	60	60
Contact number	22,053,178	1,229,380	23,968,860

**vacant, **passengers onboard.*

current location, timestamp, moving speed, and heading direction. Due to the GPRS communication cost for data transmission, reports are sent at a granularity of around one minute.

Shanghai taxis. We have also collected the GPS trace of taxis in Shanghai between February 1 and March 3, 2007. We chose 2,109 taxis in the data sets which have consecutive GPS reports on each day during the 31 days. The information contained in a taxi GPS report is similar to that of bus except that taxis also report whether there are passengers onboard. The granularity of reports is one minute for taxis with passengers and about 15 seconds for vacant ones.

Shenzhen taxis. The trace collection of taxis in Shenzhen is similar to Shanghai taxi trace. We use the whole month trace in October, 2009. We chose 8,291 taxis which continuously send GPS reports during the whole period. Taxis in Shenzhen always send GPS reports on every one minute.

We choose taxis and buses for the study for three reasons. First, taxis and buses are two representative types of vehicles showing two distinct mobility patterns, namely, rather random and well scheduled, respectively. Second, as taxis and buses are public service vehicles, they commute in the city all the time and can cover a wide area, which makes them preferable candidates for mobile advertising. Third, the privacy problem is less concerned since they are public vehicles.

4.2 Constructing Contact Graph

In order to select the best set of vehicles as seeds to deploy advertisements with respect to the spread coverage, it is of great importance to understand the position of individual vehicles in the network.

For this reason, we first construct a *contact graph* $\mathcal{G}(N, E)$ for each trace by aggregating the pairwise contacts. Each vehicle i is a node of the graph, $n_i \in N$, and the edge $e_{ij} \in E$ represents node i and j have certain acquaintance between them. The key to establishing a meaningful contact graph is the metric used to aggregate contacts, which determines whether two nodes share a link and the strength of this connection if it exists. One straightforward way is to include an edge between two vehicle nodes in the contact graph if they have seen each other within the communication range. For the extreme case, edges may appear for those contacts happened only once. This way is simple, but the problem is that extra edges derived from random or "unexpected" contacts would be added in the established graph, which may blur the network structures. There are various metrics, such as the number of total contacts observed [16], the age of last contact [22], and the contact frequency and total duration [16], have been used to derive edge strengths. In our study, we use contact ratio [23] to extract edges so that there is an edge between two nodes in the contact graph if the ratio of time with contacts observed to the total period of a trace is higher than a threshold and the weight on this edge takes the contact frequency value.

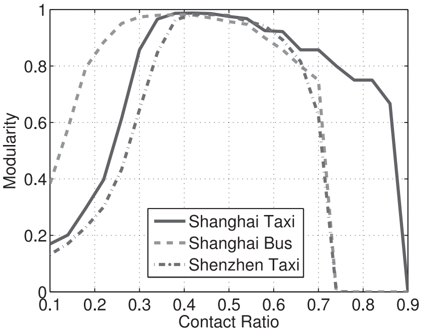


Fig. 1. Modularity derived under different contact ratio using Shanghai taxi trace on February 18, 2007.

In order to determine the appropriate contact ratio to extract contact graphs, we apply the Louvain algorithm [24] to find the community structure embedded in an established contact graph and evaluate the partition result using *modularity* [25], which is defined as

$$Q = \frac{1}{2m}\sum_{ij}\sum_{r}\left(A_{ij} - \frac{k_i k_j}{2m}\right)S_{ir}S_{jr}, \tag{1}$$

where m is the total number of edges, A_{ij} is the element of the adjacency matrix (if there is an edge between node i and j, A_{ij} is the contact ratio between node i and j; otherwise, $A_{ij} = 0$), k_i and k_j are the degrees of node i and j, respectively, and $S_{ir} = 1$ if node i belongs to group r and zero otherwise. We then examine the modularity of contact graphs derived under different contact ratios (shown in Fig. 1). It can be seen that the modularity first increases and then drops as the contact ratio increases. The reason is that with a small threshold, there are random contacts included in the contact graph; on the other hand, if a large threshold is used, then more "regular" relationships would be abandoned from the graph, which also causes the loss of valuable topology information. Modularity higher than 0.3 implies there are strong social structures in the graph [25]. In order to reduce the influence of random contacts and while being able to preserve the essential topology information, we use the minimum contact ratio which results the established contact graph with a modularity greater than 0.6. Fig. 2 illustrates a contact graph extracted with a contact ratio of 0.26 using Shanghai taxi trace on February 18, 2007, which contains 1,802 vehicle nodes. It can be seen that the community structure appears in this contact graph is very clear. With a corresponding modularity of 0.62, all the vehicles are divided into 32 communities. As relationships in the contact graph are extracted from pairwise contacts, it implies that a vehicle would have more chance to meet another vehicle within the same community than those outside the community.

4.3 Centrality Analysis on Contact Graph

With extracted contact graphs, we study the relative importance of individual vehicles within the network with respect to their *centrality* [26] as a more "centric" vehicle has higher

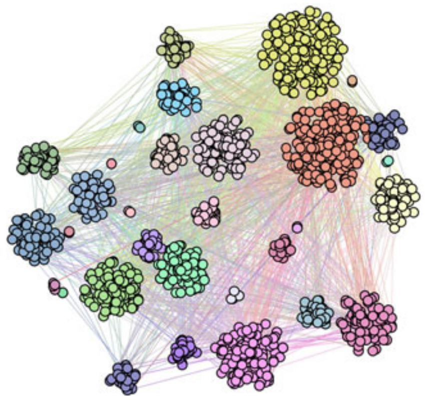


Fig. 2. Contact graph extracted from Shanghai taxi trace on February 18, 2007 containing 1,802 vehicles and 32 communities with a modularity of 0.62.

probability to meet more vehicles and therefore can cover a larger area of the network. As there are many classic centrality metrics introduced in the field of social network analysis [8], [26], [27], we study three most-related metrics:

1) *The degree centrality* [26]: is a natural way of measuring the importance of a node, which presents the number of other nodes in the graph that the node shares edges with. It presents the basic local structural property of nodes. In a contact graph, a vehicle with higher degree has greater probability to meet more vehicles, reflexing a higher popularity. The degree centrality of a given node in a contact graph is defined as the number of edges it is associated with.
2) *The closeness centrality* [26]: measures the reciprocal of the average distance of a node to all the other nodes in the network. A node with a higher closeness centrality implies it is closer to the other nodes in the network.
3) *The coreness centrality* [27]: measures the "depth" of a node within the network. The coreness of a node is k if it belongs to the k-core but not to the $(k+1)$-core, where the k-core of a graph is defined as a maximal sub-graph where each node has at least degree of k. A node with higher coreness is considered to be a better individual spreader in large-scale complex networks [8].

We examine the degree, closeness and coreness centrality of contact graphs derived from all available traces and plot their complementary cumulative density functions (CCDF) in Fig. 3. It can be seen that the CCDF of degree and coreness on all traces have exponential tails (i.e., linear under semi-logarithmic scale), which have been seen with different networks such as the power grid and railway networks [28]. The closeness, however, does not show any obvious distribution but is rather centralized. More specifically, we have two following key observations:

1) It shows that larger degree or coreness values do not mean larger closeness values, and vice versa, although the degree and coreness have similar distributions. Furthermore, the degree and coreness are good metrics to evaluate the centrality of vehicles with our traces while the closeness metric can hardly distinguish those "centric" vehicles from the rest as all vehicles have similar closeness values;
2) The exponential decay of both the degree and coreness metrics indicates that the portion of vehicles that have the largest degree or coreness values of all three types of vehicles appears small in both cities. This inspires us that it is possible to select only a small number of vehicles with the largest degree or coreness values to serve as candidate vehicles to perform mobile advertising.

5 Characterizing the Dynamic Sociality

Comparing to traditional social networks where edges are usually stable social relationships between people, links in contact graphs are extracted from highly dynamic contact information and therefore may vary significantly over time. In this section, we examine how contact graphs evolve with time.

5.1 Observing High Dynamics of Contact Graphs

In order to understand whether and how the network topology and the centrality properties of vehicles change over time, we divide time into slots of same length and construct a contact graph for each slot for all traces by aggregating all the contact events in that period using the graph extraction method as introduced in above section. We then compute all concerned centrality metrics for all vehicles in each contact graph.

Fig. 4a plots the time series of degrees of 100 randomly-chosen vehicles from the Shanghai taxi trace. Contact graphs are constructed using time slots of one hour from February 1 to February 4 (96 hours in total). Note that all

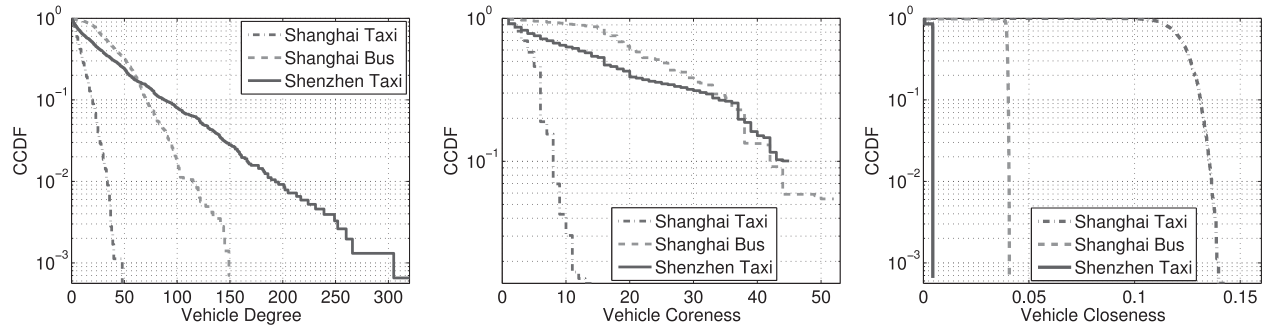


Fig. 3. CCDFs of degree, coreness and closeness in all contact graphs constructed from traces of Shanghai taxis, Shanghai buses and Shenzhen taxis; exponential tails have been found in plots of degree and coreness.

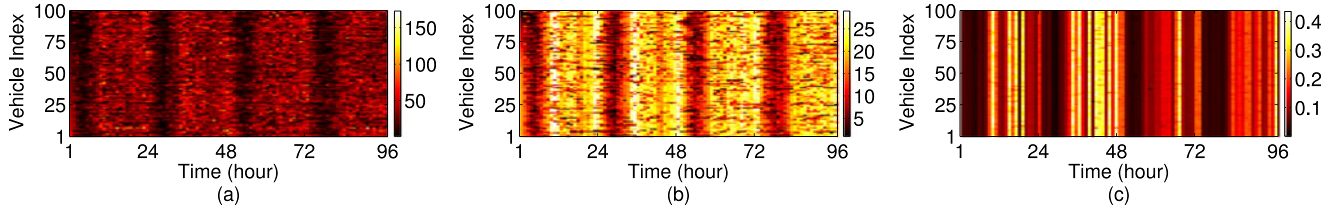


Fig. 4. Time series of (a) degrees, (b) corenesses and (c) closenesses of one hundred randomly-chosen vehicles using Shanghai taxi trace; first column in each plot is sorted and placed in the descending order.

experiment vehicles are ordered according to their degrees in the contact graph in the first hour (the first column shown in Fig. 4a). Three key observations should be pointed out. First, by checking each row, it can be seen that the degree of every vehicle varies enormously over time; second, by checking each column, it can also be seen that the relative importance (rank) among vehicles also changes between consecutive time slots; last, although the degrees of vehicles demonstrate high dynamics over time, there are obvious temporal patterns embedded in the time series. For example, a clear periodicity of one day can be seen in the figure. Similar observations can be found in Fig. 4b when examining the coreness centrality of vehicles. In contrast, although clear temporal patterns can also be found in the plot of closeness as in Fig. 4c, the absolute value and the relative rank of closeness among vehicles are rather stable which also explains the vertical drops in the CCDF shown in Fig. 3c.

We compute the coefficients of variation (CV) of different types of centrality of the 100 randomly chosen vehicles for each time slot in Fig. 4 and plot the results in Fig. 5. The CV is a standardized measure of dispersion of a probability distribution and defined as the ratio of the standard deviation to the mean. Two main observations can be seen from the figure: 1) the CV of degree and coreness show clear temporal patterns; 2) the CV of degree and coreness are much higher than that of closeness, suggesting wider diversities of such metrics among vehicles. As a result, the degree and coreness centralities are better metrics to study the dynamics of vehicular sociality.

With these observations, we make statements as follows. First, in order to gain the maximum coverage of advertising within a given period of time, seed vehicles for mobile advertising should be chosen according to the current or even future status of the network instead of using some particular static set of vehicles. Second, the dynamics of node sociality especially centrality are possible to be captured and utilized for selecting better seed vehicles as they have clear temporal patterns. Third, the degree and coreness centralities are good metrics to study the dynamics of vehicular sociality as they show both temporal patterns and wide diversity among vehicles. In the remainder of this paper, we select the degree centrality to study for demonstration.

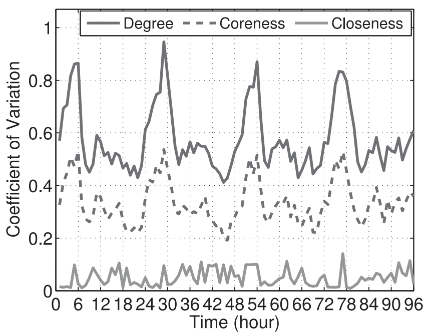


Fig. 5. Coefficients of variation of the degree, coreness, closeness centralities in each time slot.

5.2 Revealing the Temporal Correlations of Centrality

In order to capture the dynamics observed, we examine the entropy and conditional entropy of degree centrality measures.

Specifically, let X be the random variable representing the degree measures of a vehicle. If we have observed M measures, these measures can be presented by a vector $D = (d_0, d_1, \ldots, d_{M-1})$ where d_i, $0 \le i \le M-1$ denotes the ith degree measure during the ith time slot. The probability of the measure being j can be computed as x_j/M, where x_j represents the number of measure being j. Therefore, the entropy of D is:

$$H(X) = \sum_{j=0}^{\infty} (x_j/M) \log_2 \frac{1}{x_j/M}. \tag{2}$$

When $K = 1$, let X' be the random variable for the last measure of this vehicle given the measure X. X' and X have the same distribution when M is large enough. The vector D can be written as $Q = \{(d_i, d_{i+1}) : 0 \le i \le M-2\}$. Therefore, the joint entropy of X' and X can be computed as:

$$H(X', X) = \sum_{(x',x)\in Q} P(x', x) \log_2 \frac{1}{P(x', x)}, \tag{3}$$

where $P(x', x)$ is the number of times (x', x) appearing in Q divided by the total number of elements in Q. With $H(X)$ and $H(X', X)$, the conditional entropy of X given X' is:

$$\begin{aligned} H(X|X') &= H(X', X) - H(X') \\ &= H(X', X) - H(X). \end{aligned} \tag{4}$$

When $K = 2$, let X'' denote the random variable representing the distribution of the previous two measures given X. Similarly, we can compute the conditional entropy $H(X|X'')$ as:

$$\begin{aligned} H(X|X'') &= H(X'', X) - H(X'') \\ &= H(X'', X) - H(X', X). \end{aligned} \tag{5}$$

The cumulative distribution functions (CDF) of the entropy and the conditional entropy of degrees for $K = 1, 2$ and 3 over all vehicles in Shanghai taxi trace is shown in

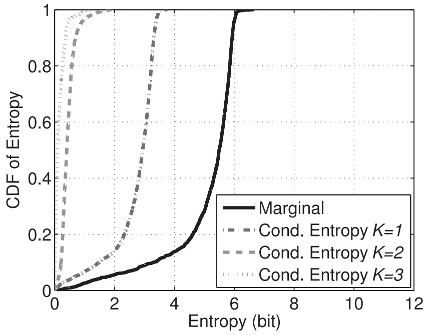


Fig. 6. CDF of entropies of degree centrality.

Fig. 6, using one hour to divide the trace and construct contact graphs. It can be seen that the conditional entropy when $K = 1$ is much smaller than the marginal entropy and the conditional entropy when $K = 2$ is much smaller than that when $K = 1$, which implies that the uncertainty about the current degree decreases when knowing the last degree of the same vehicle. We also show the mean entropy and the mean conditional entropy of all centrality metrics in Fig. 7 and have similar results. In summary, we conclude that the centrality of vehicles has strong temporal correlation. Generally, the more historical centrality information we know, the less uncertainty the current measure has.

5.3 Characterizing the Evolution of Vehicular Sociality

In order to characterize how vehicular sociality evolves along time, for a contact graph extracted in time slot t, denoted as $\mathcal{G}_t$, we study the distribution of contacts with other vehicles of each vehicle, and examine the correlation between the distribution in time slot t and that in time slot $t-n$, increasing n from one to a large number. We use *redundancy* to quantify the correlation.

In specific, the contacts between a vehicle v_i and all the other vehicles in time slot t forms a contact vector $C_t = (c_1, c_2, \ldots, c_{|N|})$, where $|N|$ is the total number of vehicles in $\mathcal{G}_t$ and c_j is the number of contacts that vehicle v_i has met with vehicle v_j in time slot t for $j = 1, \ldots, |N|$, where $c_j = 0$ if $i = j$. We also have the contact vector in time slot $t-n$, C_{t-n}. We compute the mutual information of C_t and C_{t-n}, $I(C_t, C_{t-n})$ via the joint entropy $H(C_t, C_{t-n})$ and the marginal entropy $H(C_t)$ and $H(C_{t-n})$ as follows:

$$I(C_t, C_{t-n}) = H(C_t) + H(C_{t-n}) - H(C_t, C_{t-n}). \tag{6}$$

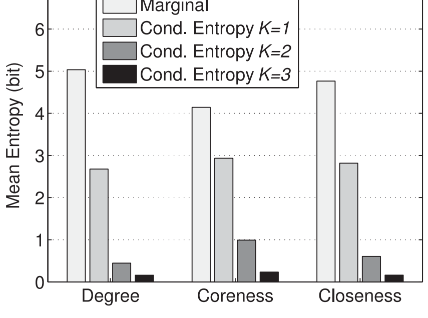


Fig. 7. Average entropies of centrality measures.

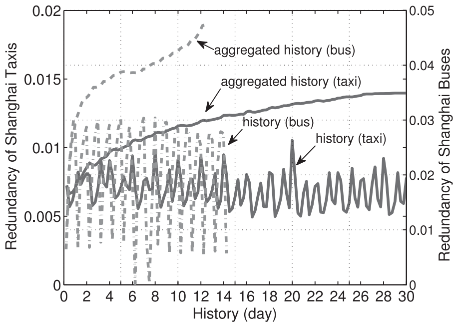


Fig. 8. Mean redundancy of contact layout.

We define the redundancy of C_t and C_{t-n} by

$$R(C_t, C_{t-n}) = \frac{I(C_t, C_{t-n})}{H(C_t) + H(C_{t-n})}. \tag{7}$$

We compute the mean redundancy averaged over all vehicles in Shanghai taxi and bus data sets. Time is divided into time slots of four hours. Fig. 8 shows the result for $n = 1$ to 120 (a period of one month). It can be seen that the layout of the contact relationship of vehicles has clear periodicities, i.e., a period of one day for buses and a period of two days for taxis. This might reflect the different shift rules of buses and taxis. In Shanghai, each bus is assigned to commute on a fixed route on every day. In contrast, taxi drivers usually shift every 24 hours. Thus, a taxi behaves very differently on every day but very similarly on every other day. In addition, buses have higher redundancy than taxis, which implies that contact relationship between buses is more predictable.

To better understand how much history data should be considered in capturing the vehicular sociality dynamics, we examine the redundancy between the layout of the contact vector in time slot t and the aggregated historical contact information from $t-1$ to $t-n$, i.e., $\sum_{i=1}^{n} C_{t-i}$. The average redundancy over Shanghai taxis and buses is also shown in Fig. 8. It is clear that the redundancy increases as the amount of history data increases and tends to stabilize. This implies that history information of four weeks should be sufficient for capturing temporal patterns of vehicular sociality.

6 Design of POST

6.1 Overview

From our analysis above, we learn the knowledge that vehicle networks show clear social structures when aggregating pairwise contacts and the vehicular sociality (particularly, the degree and coreness centrality) is highly dynamic but also demonstrates strong temporal patterns and correlations. This inspires our POST scheme to utilize the temporal correlations of vehicular sociality to infer the contact behavior of the whole network in future, which can be further leveraged to improve the performance of mobile advertising. More specifically, POST first collects the periodically-reported vehicular traces, from which the contact graphs corresponding to each time slot will be extracted. Then it uses Markov chains to capture historical temporal correlations of vehicular centrality and predict the expected centrality of each vehicle in the short future. With this information, POST adopts a greedy strategy to select the most "centric" or important vehicles to serve as seeds for mobile advertising. Moreover, in POST, the equivalence of the network position of those most centric vehicles is also considered when selecting seeds so that the seeds would have as little coverage overlap as possible.

We elaborate the three integrated components of POST in the rest of this section.

6.2 Capturing the Temporal Correlations of Centrality

As the class of finite-state Markov processes (Markov chain models) is rich enough to capture a large variety of temporal dependencies, we adopt Markov chains of kth order for capturing the temporal correlations of vehicular centrality. In Markov chain models, the current state of the process depends only on a certain number of previous values of the process, which is the order of the process.

Specifically, given the time requirement of a mobile advertising problem T, in order to capture the centrality dynamics under the time scale of the problem in the network, we divide time into slots of equal length T. Clearly, when T is relatively short, then more network dynamics can be seen between consecutive time slots but at the same time more random factors would involve in the observations which makes it hard to capture the temporal correlations of vehicular sociality and infer future network behavior. In contrast, when T is long, then the topologies of the extracted graphs tend to be stable and easy to estimate but lose most of the vehicular dynamics. We will extensively study the impact of the problem scale on the performance of mobile advertising in the section of performance evaluation.

Given a time series of contact graphs, for each vehicle v_i, we measure the centrality of each vehicle in each contact graph and get a sequence of centrality measures, denoted as $\{x_i\}_{i=1}^{n}$. By discretizing continuous measures, we can obtain a finite state space, denoted as $\mathcal{S}$. The k-order state transition probabilities of the Markov chain can be estimated for all $a \in \mathcal{S}$ and $\underline{b} \in \mathcal{S}^k$, $\underline{b} = (b_1, b_2, \ldots, b_k)$ as follows. Let $n_{\underline{b}a}$ be the number of times that state $\underline{b}$ is followed by value a in the sequence. Let $n_{\underline{b}}$ be the number of times that state $\underline{b}$ is seen and let $p_{\underline{b};a}$ denote the estimation of the state transition probability from state $\underline{b}$ to state $(b_2, \ldots, b_k, a)$. The maximum likelihood estimators of the state transition probabilities of the kth order Markov chain are:

$$p_{\underline{b};a} = \begin{cases} n_{\underline{b}a}/n_{\underline{b}}, & if\ n_{\underline{b}} > 0 \\ 0, & otherwise. \end{cases} \tag{8}$$

6.3 Inferring Future Vehicular Centrality

As the network is studied under the time scale of the mobile advertising problem, we only need to estimate the centrality of all vehicles in the next time slot with the established Markov chains.

In specific, let $\underline{b}_i$ denote the current state in the kth order Markov chain built for vehicle v_i. The estimated centrality of v_i in the next time slot $\varepsilon^i_{centrality}$ can be calculated as:

$$\varepsilon^i_{centrality} = \sum_{a \in \mathcal{S}} p_{\underline{b}_i;a} \cdot a. \tag{9}$$

In the process of inferring the future centrality for a node, two key parameters, i.e., the order of a Markov model and the length of historical data used to train the model, are critical to the accuracy of estimations. For a Markov chain model with a set of known states, simply increasing k can not necessarily fit for the temporal correlation included in a time series. Although the order of Markov chain can be evaluated via information content tests such as AIC and BIC [29], the practically appropriate value of order should be evaluated according to the specific application context, which is intensively studied in the evaluation section.

6.4 Selecting Preferable Seeds for Mobile Advertising

Since the mobile advertising problem with known future information is still NP-hard as proved in Section 3, we adopt greedy heuristics to select preferable seeds.

One greedy heuristic, called *straight*, might be to rank all vehicles according to their predicted centrality values and take the top B/p "centric" vehicles as seeds, where B is the budget and p is the price for employing one vehicle. The problem with this heuristic is that, if two centric vehicles have similar network positions, choosing both of them to serve may cause a large coverage overlap, which means that one of them is redundant, and not be able to gain as large coverage as possible.

Another strategy, called *exclusive*, is to add an extra requirement on the similarity distance between two potential seeds on the basis of *straight*. In *exclusive*, the current most centric vehicle in the vehicle ranking list is chosen as a seed. At the same time, it filters out all the other vehicles left on the list which have a distance closer than a given threshold. The procedure is repeated until the number of B/p is reached or there is not any node left. The purpose of *exclusive* is to try to disperse those selected seeds in order to gain larger coverage. The distance d_{ij} between vehicles v_i and v_j on a contact graph, which measures their social structural similarity, is defined as an Euclidean distance [30] between them:

$$d_{ij} = \sqrt{\sum_{k=1}^{|N|} \left[(x_{ik} - x_{jk})^2 + (x_{ki} - x_{kj})^2 \right]}, \tag{10}$$

where x_{ik} denotes the weight on the edge from v_i to v_k, and $i \neq k, j \neq k$. In order to infer the distance between a pair of nodes in the future time slot, we also build Markov chains for this purpose.

7 Performance Evaluation

7.1 Methodology

In this section, we evaluate POST scheme and compare with several alternative schemes. Since the primary concern of mobile advertising is the information propagation coverage and speed, we assume that vehicles have infinitely large memory and bandwidth, and messages can always be successfully transferred between vehicles. In all the schemes, only seeds may actively send advertisements to other vehicles via V2V communications.

We compare POST with several alternative seed selection schemes as follows:

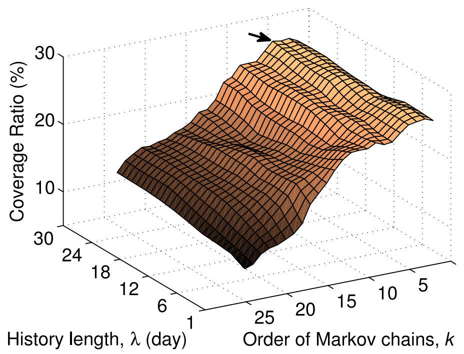


Fig. 9. Coverage ratio versus the order of Markov chains and the amount of history data under $T = 1$ h.

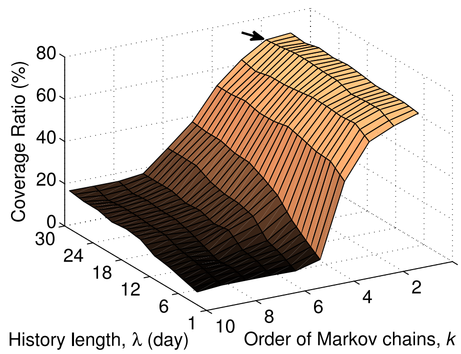


Fig. 10. Coverage ratio versus the order of Markov chains and the amount of history data under $T = 6$ h.

- **Random.** In this scheme, seed vehicles are randomly chosen for advertising. This scheme is simple and requires no extra information in selecting seeds.
- **Static with *straight* heuristic.** In this scheme [16], [18], seeds are chosen based on the rank of network position of all the vehicular nodes on a static contact graph, which is extracted using traces of a long period of time, e.g., all available traces.
- **Static with *exclusive* heuristic.** In this scheme, seeds are also chosen based on a static graph but an *exclusive* heuristic is also adopted. We implement the state-of-art algorithms in the field of social network analysis [1], [8], where the direct neighbors of a selected "centric" vehicle are cancelled from being seeds. Both the Static with *straight* and the Static with *exclusive* strategies can achieve an approximation ratio of $1 - 1/e \approx 0.63$ [31] which is considered as the best-possible polynomial time approximation for maximum coverage [21].

To evaluate the performance of our solution and the above schemes, we use *coverage ratio* as the metric, which refers to the ratio of the number of successfully posted vehicles (i.e., who have already seen the advertisement) to the total number of vehicles in the network at the end of the advertising period T. This metric is used to measure the gained coverage of an advertisement by a given set of seeds.

In the following simulations, we evaluate the above metrics of POST and all the alternative schemes, using real trace data of Shanghai taxis for demonstration. We use the contact records in the whole February of 2007 as the learning stage for all the alternative schemes and use the trace of the first three days in March for testing. In order to investigate different time scales of the mobile advertising problem, we examine the performance of all schemes under three different time scales, i.e., $T = 1$ hour, $T = 6$ hours and $T = 12$ hours. Under each time scale, we divide the trace and construct contact graphs accordingly.

7.2 Effect of Algorithm Parameters

We first examine the effects of protocol parameters on the network delivery performance. As we adopt Markov chains of kth order to capture temporal correlations of vehicular centrality and infer the expected centrality in future, we examine the effects of k and the length of history data λ on the advertising performance. Upon $T = 1$ hour, we vary k from one to 24; for $T = 6$ hours, we vary k from 1 to 10. Under all these time scales, we vary the amount of available history data λ from one day to 30 days. We set the quota of seeds to 10 and run POST with *straight* heuristic on all available testing trace and get the average.

Figs. 9 and 10 plot the average coverage ratios under time scales of one hour and six hours. It can be seen that, in general, with more history data to train the model, POST can get higher coverage ratio. It is observed that the highest coverage ratio is reached with all history data trained under both one hour and six hours scales. Under the time scale of one hour, the coverage ratio achieves the optimum when the order of Markov chains equals to five and that value under the scale of six hours is two. This implies that vehicular networks show higher dynamics under smaller time scales and should be captured using higher order Markov chains.

7.3 Performance Comparison

In these experiments, we compare POST with all the other alternative seed selection schemes under the time scale of one hour. We change the quota of seeds from 1 to 200 and run all schemes over all available testing trace and get the average.

Fig. 11 plots the average coverage ratio as a function of the number of seeds. It can be seen that two POST schemes outwit two corresponding Static schemes and the Random scheme. In particular, POST with *exclusive* heuristic has the best performance and achieves up to 35 and 60 percent performance gains on average comparing with POST with *straight* heuristic and Static with *exclusive* heuristic, respectively. It can also be seen that even though the coverage ratio increases as the number of seeds increases, the speed of such increment tends to decrease. For example, recruiting

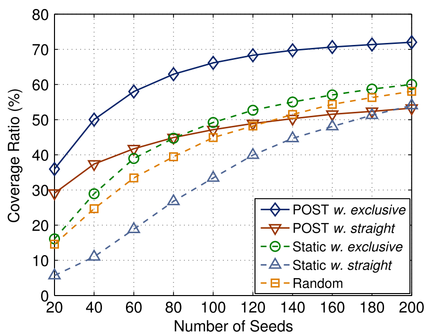


Fig. 11. Coverage ratio versus the number of seeds under $T = 1$h, with Shanghai taxi data set.

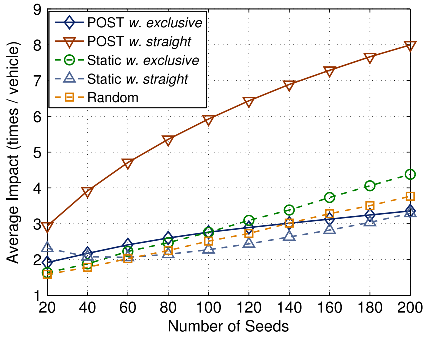


Fig. 12. Average impact versus the number of seeds under $T = 1$h, with Shanghai taxi data set.

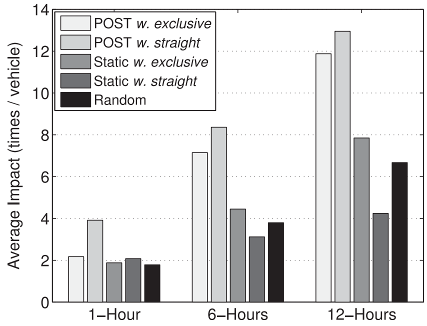


Fig. 14. Average impact versus time scale of the problem, with Shanghai taxi data set.

the first 20 seeds can achieve a coverage ratio of about 40 percent but adding another 20 seeds can only bring an extra 10 percent coverage ratio. In contrast, both the Random and Static schemes have better coverages than POST when the number of seed vehicles exceeds 120 and 190, respectively. The reason is that POST always selects those top vehicles with respect to centrality metric values as seed vehicles. As the number of seed vehicles increases, the probability of two selected vehicles having similar network positions also increases, which means they are redundant and therefore choosing both of them may not be able to gain as large coverage as possible. On the contrary, the Random scheme has larger probability to uniformly scatter seed vehicles among all the communities and therefore might achieve better coverage when the number of seed vehicles is large. The Static scheme has similar results as the static network structure is quite different from the current network topology. This result provides valuable information for advertisers to gain a better tradeoff between performance and investment.

In order to understand the rationale behind the above results, we examine the *average impact* of each scheme, which refers to the average number of times a vehicle has seen an advertisement. It can be calculated as the ratio of the total number of times that one of the seeds has posted the advertisement to another vehicle to the number of posted vehicles at the end of the advertising period T. The purpose of studying the average impact is to see whether a scheme can effectively reduce redundant dissemination with a given set of seeds. Fig. 12 plots the average impact as a function of the number of seeds. It can be seen that POST with *straight* heuristic extraordinarily outperforms all the others including POST with *exclusive* heuristic. This implies that although POST with *straight* heuristic has a moderate coverage ratio but it can achieve very good intensity of advertising, which may be more preferable under certain circumstances (e.g., expecting that the advertisement to be remembered longer). The reason is that POST with *straight* heuristic tends to select "centric" vehicles with more coverage overlap as seeds so that the intensity of an advertisement is enhanced as explained above.

7.4 Impact of Time Scale of Mobile Advertising

We further examine the impact of time scale of mobile advertising on the performance of POST. We set the seed quota to 40 and run all schemes using testing trace and get the average.

Figs. 13 and 14 plot the coverage ratio and average impact as a function of time scale of the mobile advertising problem, respectively. It can be seen that in both figures, as the time scale increases, POST always achieves best performance but the performance differences between POST and Static schemes also decrease. This verifies our discussion about the relation between network dynamics and the time scale of the problem. In general, as the time scale increases, the network dynamics tend to vanish and the network structure also tends to be stable even though the underlying vehicles are mobile.

7.5 Performance in Different Vehicular Networks

In order to validate the performance of POST, we show the experiment results with two more types of real vehicular traces as well, i.e., Shanghai bus data set and Shenzhen taxi data set. It should be noted that the whole area of Shenzhen is only one third of that of Shanghai, while the number of taxis is much greater than Shanghai. Therefore, the Shenzhen taxi data set has a higher network density than the

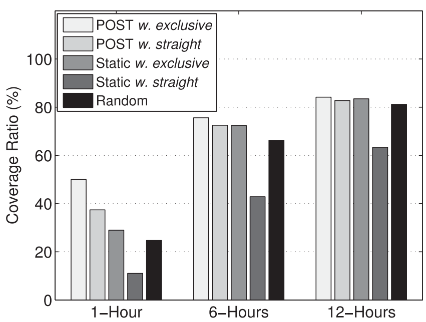


Fig. 13. Coverage ratio versus time scale of the problem, with Shanghai taxi data set.

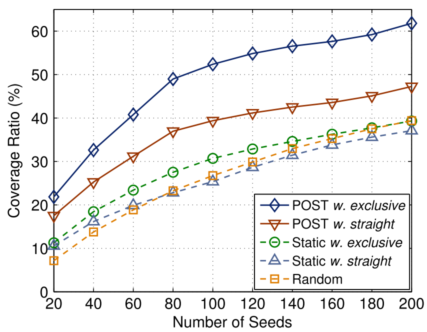


Fig. 15. Coverage ratio versus the number of seeds under $T = 1$h, with Shanghai bus data set.

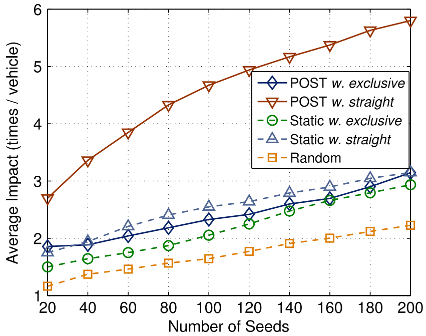


Fig. 16. Average impact versus the number of seeds under $T = 1$h, with Shanghai bus data set.

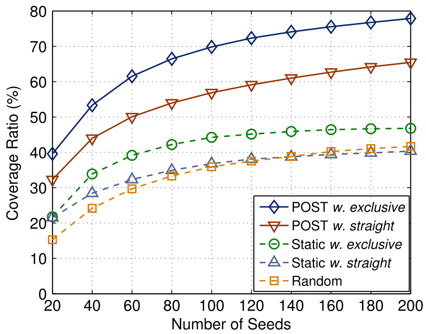


Fig. 17. Coverage ratio versus the number of seeds under $T = 1$h, with Shenzhen taxi data set.

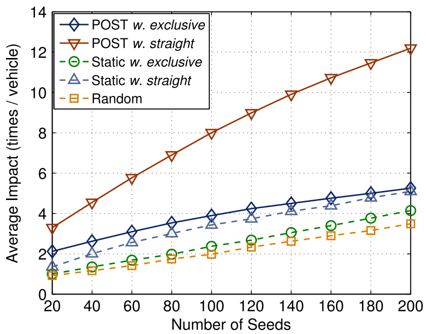


Fig. 18. Average impact versus the number of seeds under $T = 1$h, with Shenzhen taxi data set.

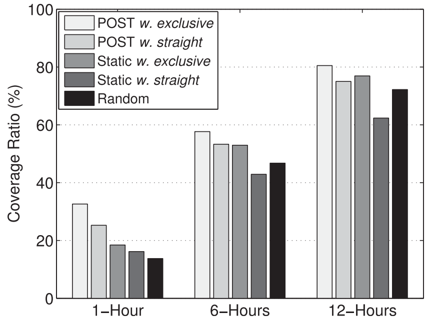


Fig. 19. Coverage ratio versus time scale of the problem, with Shanghai bus data set.

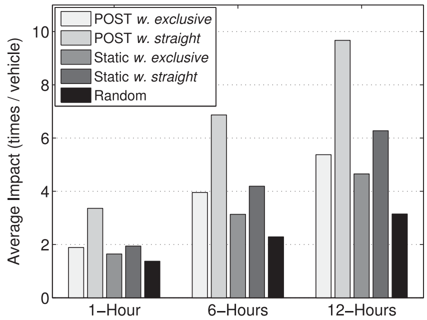


Fig. 20. Average impact versus time scale of the problem, with Shanghai bus data set.

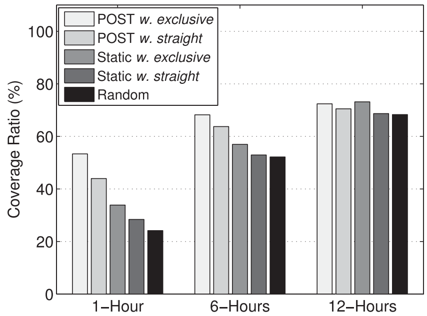


Fig. 21. Coverage ratio versus time scale of the problem, with Shenzhen taxi data set.

Shanghai taxi data set does. The Shanghai bus network has the lowest density because of the smallest vehicle number among these three data sets.

The performance of average coverage ratio as a function of the number of seeds for the five schemes is shown in Figs. 15 and 17. We can see the similar results that POST outperforms the rest of these schemes. Particularly, POST with *exclusive* heuristic also has the best performance. On average, in Shanghai bus network the performance gains of it are up to 31 and 70 percent comparing with POST with *straight* heuristic and Static with *exclusive* heuristic respectively, and up to 21 and 63 percent in Shenzhen taxi network. We also can find that POST with *straight* heuristic has better performance than the non-POST schemes constantly, while in Shanghai taxi network the average coverage of static with *exclusive* heuristic exceeds that of POST with *straight* heuristic when the number of seed vehicles is large. In addition, the performance of Random scheme is at nearly the bottom much of the time.

In Figs. 16 and 18, the performance of average impact as a function of the number of seeds is shown. We can see that the result is consistent with that in Shanghai taxi data set. In both Shanghai bus network and Shenzhen taxi network, the performance of POST with *straight* heuristic also exceeds all the other schemes including POST with *exclusive* heuristic. However, we also find that the average impact of POST in Shanghai bus network is smaller than that in Shanghai taxi network, while that in Shenzhen taxi network is the greatest among these three vehicular networks. This is because Shanghai bus network has the lowest density while Shenzhen taxi network has the highest one.

In Figs. 19 to 22, we show the impact of time scale of the mobile advertising problem on the performance of POST. Also, the seed quota has been set to 40. We run all the schemes using testing trace and get the average.

Figs. 19 and 21 show the coverage ratio as a function of time scale. We can see that in both Shanghai bus and Shenzhen networks, POST always achieves the best performance with the increasing time scale, but the performances of

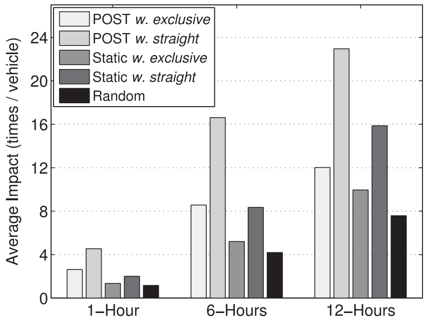


Fig. 22. Average impact versus time scale of the problem, with Shenzhen taxi data set.

POST and Static schemes are getting closer due to the fading of the network dynamics.

The average impact as a function of time scale is shown in Figs. 20 and 22. It can also be seen that POST with *straight* heuristic outperforms all the other schemes. In particular, the performances of it in Shenzhen taxi network are nearly twice as much as those in Shanghai taxi network with the time scales of 6 and 12 hours, respectively.

7.6 Discussion

From the experimental results, it can be seen that the performance using the Shanghai bus and Shenzhen taxi data sets are consistent with those using the Shanghai taxi data set. This confirms that the scheme of POST can still achieve better performance in the networks consisting of different types of vehicles. Moreover, the coverage ratio achieved by POST in experiments with different time scales is stably better than the other schemes. The reason why POST gain better coverage performance is twofold: first, POST captures the high dynamic of vehicular sociality in the network, and second, POST further excludes the redundant candidates in terms of social similarity, thus gains better information coverage by reducing contact overlap of the seeds.

In addition, POST can be adapted to different application requirements. POST fits better in the application scenarios where the dynamic of vehicular sociality is more significant. As the time scale for dissemination tasks increases, the coverage gain of POST than those of the other schemes is becoming smaller due to the vanishing of network dynamics. On the other hand, in some scenarios, the advertisers may require their advertisement to be seen more frequently, i.e., to gain a greater advertising impression. This requires greater average impacts by including those redundant candidates, i.e., the social similarity of the selected seed vehicles should be increased.

8 Conclusion

In this paper, we have studied the mobile advertising problem in vehicular networks and proved its NP-hardness. By analyzing three large-scale vehicular traces, we have found that the vehicles show clear sociality within the network and the vehicular sociality is highly dynamic and has strong temporal correlations. Based on the observations, we have proposed to use Markov chains to capture the patterned vehicular centrality and infer the expected future. We have also employed two heuristics to utilize the estimated centrality information to improve the performance of mobile advertising. We have carried out extensive trace-driven simulations based on the large data sets of real vehicular GPS traces of taxis and buses. Experiment results demonstrate that POST outperforms the state-of-art algorithms. For all the involved vehicular data sets, POST can achieve an average 64 percent performance gain overall against the state-of-art algorithms.

Acknowledgments

This research was supported in part by National Natural Science Foundation of China (Grants No. 61202375, 61373157, 61173171, 61170237, 61170238, 60903190, 61472255, 61420106010), the National High Technology Research and Development Program (2013AA01A601), the Fundamental Research Funds for the Central Universities (Grant No. ZYGX2012J072), Program for Changjiang Scholars and Innovative Research Team in Universities of China (IRT1158, PCSIRT).

References

[1] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proc. 9th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2003, pp. 137–146.
[2] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proc. ACM SIGKDD*, 2009, pp. 199–208.
[3] W. Chen, W. Lu, and N. Zhang, "Time-critical influence maximization in social networks with time-delayed diffusion process," in *Proc. 26th AAAI Conf. Artif. Intell.*, 2012, pp. 1–5.
[4] L. Wischhof, A. Ebner, and H. Rohling, "Information dissemination in self-organizing intervehicle networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 1, pp. 90–101, Mar. 2005.
[5] T. Kosch, C. Schwingenschlogl, and Li Ai, "Information dissemination in multihop inter-vehicle networks," in *Proc. IEEE 5th Int. Conf. Intell. Transp. Syst.*, 2002, pp. 685–690.
[6] V. Naumov and T. R. Gross, "Connectivity-aware routing (CAR) in vehicular ad-hoc networks," in *Proc. IEEE INFOCOM*, 2007, pp. 1919–1927.
[7] T. H. Luan, L. X. Cai, J. Chen, X. (Sherman) Shen, and F. Bai, "Engineering a distributed infrastructure for large-scale cost-effective content dissemination over urban vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 63, no. 3, pp. 1419–1435, Mar. 2014.
[8] 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 Phys.*, vol. 6, no. 11, pp. 888–893, 2010.
[9] J. Ugander, L. Backstrom, C. Marlow, and J. Kleinberg, "Structural diversity in social contagion," *Proc. Nat. Acad. Sci.*, vol. 109, no. 16, pp. 5962–5966, 2012.
[10] F. Papadopoulos, M. Kitsak, M. Ángeles Serrano, M. Boguñá, and D. Krioukov, "Popularity versus similarity in growing networks," *Nature*, vol. 489, no. 7417, pp. 537–540, 2012.
[11] M. Richardson and P. Domingos, "Mining knowledge-sharing sites for viral marketing," in *Proc. ACM SIGKDD*, 2002, pp. 61–70.
[12] P. Domingos and M. Richardson, "Mining the network value of customers," in *Proc. ACM SIGKDD*, 2001, pp. 57–66.
[13] D. Kempe, J. Kleinberg, and É. Tardos, "Influential nodes in a diffusion model for social networks," in *Proc. 32nd Int. Conf. Automata, Lang. Program.*, 2005, pp. 1127–1138.
[14] M. Karsai, M. Kivelä, R. K. Pan, K. Kaski, J. Kertész, A.-L. Barabási, and J. Saramäki, "Small but slow world: How network topology and burstiness slow down spreading," *Phys. Rev. E*, vol. 83, no. 2, p. 025102, 2011.
[15] E. M. Daly, and M. Haahr, "Social network analysis for routing in disconnected delay-tolerant manets," in *Proc. ACM MobiHoc*, 2007, pp. 32–40.
[16] P. Hui, J. Crowcroft, and E. Yoneki, "Bubble rap: Social-based forwarding in delay tolerant networks," in *Proc. ACM MobiHoc*, 2008, pp. 1576–1589.

[17] J. Fan, J. Chen, Y. Du, W. Gao, J. Wu, and Y. Sun, "Geocommunity-based broadcasting for data dissemination in mobile social networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 4, pp. 734–743, Apr. 2013.

[18] C. Boldrini, M. Conti, and A. Passarella, "ContentPlace: Social-aware data dissemination in opportunistic networks," in *Proc. 11th Int. Symp. Modeling, Anal. Simul. Wireless Mobile Syst.*, 2008, pp. 203–210.

[19] J. Fan, J. Chen, Y. Du, P. Wang, and Y. Sun, "DelQue: A socially aware delegation query scheme in delay-tolerant networks," *IEEE Trans. Veh. Technol.*, vol. 60, no. 5, pp. 2181–2193, Jun. 2011.

[20] J. Miček and J. Kapitulík, "Car-to-car communication system," in *Proc. IEEE Int. Multiconf. Comput. Sci. Inf. Technol.*, 2009, pp. 627–631.

[21] U. Feige, "A threshold of ln n for approximating set cover," *J. ACM*, vol. 45, no. 4, pp. 634–652, 1998.

[22] H. Dubois-Ferriere, M. Grossglauser, and M. Vetterli, "Age matters: Efficient route discovery in mobile ad hoc networks using encounter ages," in *Proc. ACM MobiHoc*, 2003, pp. 257–266.

[23] H. Zhu, M. Dong, S. Chang, Y. Zhu, M. Li, X. (Sherman) Shen, "ZOOM: Scaling the mobility for fast opportunistic forwarding in vehicular networks," in *Proc. IEEE INFOCOM*, 2013, pp. 2832–2840.

[24] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding the communities in large networks," *J. Statist. Mechanics: Theory Exp.*, p. P10008, 2008.

[25] M. E. J. Newman, "Modularity and community structure in networks," *Proc. Nat. Acad. Sci.*, vol. 103, no. 23, pp. 8577–8582, 2006.

[26] L. C. Freeman, "Centrality in social networks conceptual clarification," *Soc. Netw.*, vol. 1, no. 3, pp. 215–239, 1979.

[27] D. Miorandi and F. De Pellegrini, "K-shell decomposition for dynamic complex networks," in *Proc. IEEE 8th Int. Symp. Modeling Optim. Mobile, Ad Hoc Wireless Netw.*, 2010, pp. 488–496.

[28] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, "Classes of small-world networks," *Proc. Nat. Acad. Sci.*, vol. 97, no. 21, pp. 11149–11152, 2000.

[29] K. P. Burnham, and D. R. Anderson, "Multimodel inference: understanding AIC and BIC in model selection," *Sociol. Methods Res.*, vol. 33, no. 2, pp. 261–304, 2004.

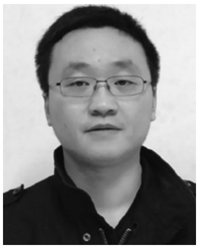
[30] S. Wasserman, *Social Network Analysis: Methods and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 1994.

[31] F. Chierichetti, R. Kumar, and A. Tomkins, "Max-cover in map-reduce," in *Proc. ACM 19th Int. Conf. World Wide Web*, 2010, pp. 231–240.

Jun Qin received the MSc degree in computer science from the Dresden University of Technology, Germany, in 2009. He is currently working toward the PhD degree in the Department of Computer Science and Engineering at Shanghai Jiao Tong University. His research interests include vehicular networks, ubiquitous computing, and service computing.

Hongzi Zhu received the PhD degree from Shanghai Jiao Tong University in 2009. He is an associate professor working in the Department of Computer Science and Engineering at Shanghai Jiao Tong University. His research interests include mobile computing, wireless networks, and vehicular ad hoc Networks. He is a member of the IEEE, IEEE Computer, and the IEEE Communication Society.

Yanmin Zhu received the BEng degree in computer science from the Xi'an Jiao Tong University in 2002 and the PhD degree in computer science from the Hong Kong University of Science and Technology in 2007. He is a professor in the Department of Computer Science and Engineering at the Shanghai Jiao Tong University. His research interests include ad-hoc sensor networks, mobile computing, grid computing, and resource management in distributed systems. He is a member of the IEEE and the IEEE Communication Society.



Li Lu received the PhD degree in the Key Lab of Information Security, Chinese Academy of Science, in 2007. He is an associate professor working at the School of Computer Science and Engineering, University of Electronic Science and Technology of China. His research interests include RFID technology and system, wireless network, and network security. He is a member of the IEEE, IEEE Communication Society, and ACM.

Guangtao Xue received the PhD degree from the Department of Computer Science and Engineering, Shanghai Jiao Tong University in 2004. He is an associate professor in the Department of Computer Science and Engineering at Shanghai Jiao Tong University, China. His research interests include vehicular ad hoc networks, wireless networks, mobile computing, and distributed computing. He is a member of the IEEE, IEEE Computer Society, and the IEEE Communication Society.

Minglu Li received the PhD degree in computer software from Shanghai Jiao Tong University in 1996. He is a full professor and the director of the Grid Computing Center in Shanghai Jiao Tong University. His research interests include grid computing, services computing, and sensor networks. He has published more than 100 papers in academic journals and international conferences. He is a member of the Executive Committee of the Technical Committee on Services Computing of the IEEE Computer Society and a senior member of the IEEE.