

# VIP Delegation: Enabling VIPs to Offload Data in Wireless Social Mobile Networks

Marco Valerio Barbera\*, Julinda Stefa\*, Aline Carneiro Viana<sup>‡</sup>, Marcelo Dias de Amorim<sup>§</sup>, and Mathias Boc<sup>†</sup>

\*Sapienza University of Rome, Italy, <sup>‡</sup>INRIA, France, <sup>§</sup>UPMC Sorbonne Universités, France, <sup>†</sup>CEA/LIST, France

{barbera, stef}a@di.uniroma1.it, aline.viana@inria.fr, marcelo.amorim@lip6.fr, michael.boc@cea.fr

**Abstract**—We propose the use of *opportunistic delegation* as a data traffic offload solution to the recent boost up of mobile data consumption in metropolitan areas, by investigating two main questions: (i) “How to gain insights into social mobile networking scenarios?” and (ii) “How to utilize such insights to design solutions to alleviate overloaded 3G networks?”. The purpose of our solution is to leverage usage of mobile applications requiring large data transfers by channeling the traffic to a few, socially selected important users in the network called *VIP delegates*. The proposed VIP selection strategies are based on social network properties and are compared to the optimal solution (that covers 100% of users with minimum number of VIPs). Our extensive experiments with real and synthetic traces show the effectiveness of VIP delegation both in terms of coverage and required number of VIPs – down to 7% in average of VIPs are needed in campus-like scenarios to offload about 90% of the traffic.

## I. INTRODUCTION

Since the modern smart-phones have been introduced worldwide, more and more users have become eager to engage with mobile applications and connected services. Simultaneously, smart-phone owners are using an increasing number of applications requiring the transfer of large amounts of data to/from mobile devices. Applications related to social networks [1], [2], global sensing [3], [4], and content distribution [5], [6] are just a few of the examples. As a consequence, the traffic generated by such devices has caused many problems to 3G network providers. AT&T’s subscribers in USA were getting extremely slow or no service at all because of network straining to meet iPhone users’ demand [7]. The company switched from unlimited traffic plans to tiered pricing for 3G data users in summer 2010. Similarly, Dutch T-Mobile’s infrastructure has not been able to cope with intense 3G traffic, forcing the company to issue refunds for affected users [8].

All these issues are bringing new technical challenges to the networking and telecommunication communities. In fact, finding *new ways to manage* such increased data usage is essential to improve the level of service required by the new wave of smart-phones applications. One of the most promising solutions to avoid overwhelming the 3G infrastructure is to rely on short-range opportunistic communications (Bluetooth or Wi-Fi) between wireless devices whenever possible.

In this paper, we propose *VIP delegation*, a solution to this problem based solely on the inherent social aspects of user mobility. Our idea is to exploit a *few, important* subscribers that with their movements and interactions are able to communicate with all the rest of the network users on a regular

basis. These VIP devices would act as a bridge between the network infrastructure and all the remaining of the network each time large amount of data has to be transferred. VIP delegation can help alleviate the network traffic in different scenarios. Collection of urban-sensing related data, distribution of content to users by service providers, and free updates of (light versions) mobile software’s ad pools are examples of applications that would directly benefit from the use of VIPs. As these are applications that induce large amounts network traffic, delegating the collection/distribution of such traffic to a few important nodes would result in immediate data offload.

After discussing related work in Section II, we propose, formalize, and evaluate two methods of VIPs selection: *global* and *neighborhood* VIP delegation (see Section III). While the former focuses on users that are *globally* important in the network (namely, *global VIPs*), the latter selects users that are important within their *social communities*. The importance of a user within the network is given in terms of well-known attributes such as centrality (betweenness, degree, and closeness) and page rank. In both cases, we observe that a short observation period (one week) is enough to detect users that keep their importance during long periods (several months). Selected nodes are then used to cover the entire network, on a daily basis, through solely direct wireless contacts with the remaining nodes (see Section IV).

Through extensive experiments on real-life and synthetic traces, we evaluate the performance of the global and neighborhood VIP delegation methods in terms of network coverage, by varying the number of VIPs chosen (see Section VI). We compare our solution with an optimal benchmark computed from the full knowledge of the system. The results reveal that our strategies get very close to the performance of the benchmark VIPs: Only 5.93% page-rank VIPs against almost 4% of the benchmark’s set to offload about 90% of the network in campus-like scenarios. Additionally we discuss on possible VIP incentives, the way VIPs offload the traffic accumulated to the network, and leveraged applications in Section VIII. Finally, we conclude with Section IX.

## II. RELATED WORK

Consumption of mobile data by the pervasive usage of smart-phones is forcing carriers to find ways to offload the network. As there is a strong belief that even more efficient technologies (e.g., 4G or WiMAX) will not be able to solve the problem [9], network providers have pushed more into

offloading to alternate networks such as femtocells and Wi-Fi [10]. Femtocells exploit broadband connection to the service provider’s network and leverage the licensed spectrum of cellular macro-cells to offer better indoor coverage to subscribers [11]. Besides from being localized (indoors only), such solution suffers from the non-proliferation of femtocells to subscribers’ homes, partially due to the user-charge for the necessary equipment (150 USD for AT&T’s microcell).

Carriers, rather than investing on a large-scale distribution of femtocells, are more willing to use more pervasive technologies, such as Wi-Fi access points and hot spots. More and more carriers in USA and worldwide are investing in this direction [12], by installing access points and hot-spots close to overloaded cellular towers, and by providing to clients Wi-Fi access within tiered monthly subscription. Wiffler is a solution in this context that focuses only on Internet access from moving vehicles [13]. It is worth mentioning that our solution is orthogonal to both Wi-Fi and femtocell offloading; nevertheless, it can be integrated to these methods to further help alleviate mobile data traffic.

Polat et al. suggest some sort of network members’ promotion to enhance network functionalities in a mobile ad hoc network [14]. Their solution makes use of the concept of connected message ferry dominating set (CMFDS), where ferry-members of the network are connected over space-time paths. Besides the differences in both the nature of the problem and the application scenario compared to our work, no social aspect/importance of the network members is considered in promotion.

Many research works targeting DTN’s (or opportunistic networks) make use of social ties between users to leverage network services ranging from routing [15]–[17] and multicasting [18], [19], to selfishness and network security [20], [21]. To the best of our knowledge, Han et al. are the first to exploit opportunistic communication to alleviate data traffic in cellular networks [22]. However, conversely from ours, their solutions only apply to information dissemination problems such as broadcasting in multi-hop opportunistic forwarding. In our scenario, where large amounts of data have to be transferred, multi-hop forwarding is not feasible – applications that require collection of sensing data would result highly expensive in terms of energy [3], [4]. Moreover, multi-hop forwarding requires collaboration of all users in the network. Even though such collaboration can be stimulated by incentive mechanisms [23], there is no guarantee on packet delivery. Rather, our solution relies on upgrading a crucial small set of users’ devices (down to 5.93% according to experiments with real campus-like data traces), that through direct contact with all network members help alleviate the data traffic in both upload and download, assuring that no packet is being lost.

### III. VIP DELEGATION IN A NUTSHELL

The movement of smart-phone users is not random; rather, it is a manifestation of their social behavior [16], [24]–[26]. This movement, along with wireless interactions among users, generates a social mobile network. The analysis of such

mobility patterns and the understanding of how mobile users “interact” (i.e., meet) play a critical role at the design of solutions/services for such kind of networks. In general lines, this paper investigates the following questions: (i) *how to gain insights into social mobile networking scenarios* and (ii) *how to utilize such insights to design solutions allowing alleviating the network traffic in the current overloaded 3G networks*.

Though the number of network users can be very large, just a few of them have an “important” role within the social graph induced by the encounters. The natural behavior of these VIP nodes, which are considerably fewer than the rest of the network, can be a valuable resource in both information dissemination and collection to/from the rest of the network. Motivated by the fact that opportunities for users to exchange data depend on their habits and mobility patterns, our idea is the following: Turn those *few* VIP nodes into bridges between regular users and the Internet, each time large amount of data is to be uploaded/downloaded by these latter ones. In a word, the VIP would act as delegates of the network infrastructure builder. As a side effect, this would immediately drop down the 3G network usage.

In our scenario, we assume that users download/upload large amount of data, thus making the use of multi-hop forwarding protocols unfeasible. Indeed, it is quite hard to convince the average user to act as a relay for others, even though to the closest access point, of such an overloading traffic. Rather, our solution relies on the upgrade of a small, crucial set of VIP nodes that regularly visit all network users and collect (disseminate) data to them on behalf of the network infrastructure. When this happens, we say that the *network is covered*. Similarly, when a VIP visits a user, we say that the *user is covered*.

Now the problem becomes the following: *how to choose the smallest VIP set that with their natural movement in the network cover all users daily?* We solve this problem by presenting two VIP selection methods that rely on either a global or a local view of the network (the methods are detailed in Section IV). We also present a benchmark solution for VIP delegation. The benchmark provides an optimal selection method that (i) requires total pre-knowledge of users’ behavior and (ii) is based on an adaptation of the well known NP-hard problem of the Minimum Dominating Set [27]. Such a method is clearly not feasible in real-life applications, since future cannot be available in advance, but useful to evaluate the goodness of our social-based VIP selection methods.

### IV. GLOBAL VS. HOOD VIPS SELECTION METHODS

Before presenting the different VIP selection strategies, we first need to identify when a user/node is “important” in the network and according to which structural attribute.

#### A. Who are the VIPs?

From mobility patterns and wireless interactions of users in a network, we establish a social undirected graph  $G(V, E)$  between users when they are socially related to each other. In this paper, we consider that there is a social tie between two

nodes if they are frequently in range (or contact) of each other (see a detailed description in Section VI-B). Using the social graph, we apply then several structural *attributes* to define the importance of a node in the network: betweenness centrality, closeness centrality, degree centrality, and page rank. All these are well-known attributes in social network theory [28], [29]:

- 1) *Betweenness centrality* measures the number of occurrences of a node in the shortest-path between pairs of others nodes. It thus determines “bridge nodes” that, with their movement, act as connectors between node groups. For a given node  $k$  it is calculated as:  $C_B(k) = \sum_{j \neq k}^N \sum_{i \neq k}^N \frac{g_{i,j}(k)}{g_{i,j}}$ , where  $N$  is the number of nodes in the network,  $g_{i,j}$  is the total number of shortest paths linking  $i$  and  $j$ , and  $g_{i,j}(k)$  is the number of those shortest paths that include  $k$ .
- 2) *Degree centrality* ranks nodes based on the number of their direct ties (i.e., neighbors) in the graph. It identifies the most popular nodes, also called *hubs* in social network theory, possible conduits for information exchange. Degree centrality is calculated as:  $C_D(k) = \sum_{i=1}^N a(k, i)$ , where  $a(k, i) = 1$  if  $k$  and  $i$  are linked, and  $a(k, i) = 0$  otherwise.
- 3) *Closeness centrality* ranks higher nodes with lower multi-hop distance to other nodes of the graph. It describes “independent nodes” that do not depend upon others as intermediaries or relayers of messages due to their closeness to other nodes. The closeness centrality for a node  $k$  is calculated as  $C_C(k) = \frac{N-1}{\sum_{i=1}^N d(k, i)}$ , where  $d(k, i)$  is the length of the shortest path between nodes  $k$  and  $i$ . To deal with disconnections it is computed within the subgraph induced by the elements of the connected component to which  $k$  belongs.
- 4) *Page rank*, the well known Google’s ranking algorithm, measures the likelihood of nodes in having important friends in a social graph [29]. In particular, page rank of a node  $i$  in the social graph is given by the equation  $PR(k) = \frac{1-d}{N} + d \sum_{i \in F(k)} \frac{PR(i)}{|F(i)|}$ , where  $d$  ( $0 \leq d \leq 1$ ) is the damping factor and  $F(k)$  is the set of neighbors of  $k$  in the social graph (the graph is undirected). The damping factor  $d$  controls the amount of randomness in page ranking: Values close to 1 will give high page rank to socially best-connected nodes.

### B. Global VIPs

The global VIP delegation strategy aims to select the smallest VIP set over the global social graph that is able to *daily cover the network* through direct contacts with network users. For this, the nodes are first ordered according to each of the earlier discussed rankings, and then one of the following VIP promoting ways is applied:

- *Blind global promotion*. It selects the top-ranked nodes not yet promoted, till the network is covered.
- *Greedy global promotion*. This is a set-cover flavored solution. In particular, it starts with promoting to VIP the top-ranked node. After this promotion, the nodes covered

by this VIP are dumped and ranking on the remaining nodes are re-computed. Again, the procedure is repeated till the network is covered.

### C. Hood VIPs

The second selection strategy, Neighborhood VIP delegation, is based on the intuition that repetitive meetings among people happen usually in the same places. The mobile social network generated by this behavior encompasses, besides contact locality, well tight social-community sub-structures. Thus, our strategy aims to cover each community at a time, selecting *hood VIPs* by their importance within their communities. Before doing so, we first detect social-communities using the *k-clique* community algorithm [30]. Afterwards, we rank nodes according to betweenness, closeness, or degree centrality and page rank. Then we start covering each community by promoting its members to VIPs similarly to the global VIPs methods:

- *Blind hood promotion*. It continuously selects the top-ranked nodes not yet promoted in the community, till the network is covered.
- *Greedy hood promotion*. The highest-ranked member in the community is promoted, nodes it covers within the community are dropped, and rankings are computed again in the remaining community graph.

In both promoting ways, when the whole community is covered, the procedure continues with another one, until all the communities are covered.

### D. VIP delegation in practice

The computation of previously discussed ranking metrics and of sub-communities in a social mobile network requires knowledge on user mobility. None of them is available when one has to pre-compute the VIPs set that will visit all the users in the future. We cannot expect to know every single user movement in advance! Nevertheless, the movement of users guided by their interests generates repeatability in their behaviors (e.g., go to work/school every day, hang out with the same group of friends). Thus, *observing users’ movement and meeting patterns for a short period reveals enough information to characterize the tightness of the social links in the network graph*.

In a real-life application, we could imagine the network infrastructure builder asking participating users to log their meetings for a certain time, called here as *training period*. These logs serve then to build the networks’ social graph on which the VIPs selection is made: the social graph  $G(V, E)$  is composed of vertexes representing nodes and edges describing their social ties (encounters). Two nodes are linked in the social graph if they meet frequently during the training period.

The results of our experiments with both real and synthetic traces show that this is a good strategy. Indeed, as we will see in Section VI, small sets of VIPs selected with our strategies on a training period of only *1 week* yields very good results in terms of user visiting, day by day, for all the remaining of the traces.

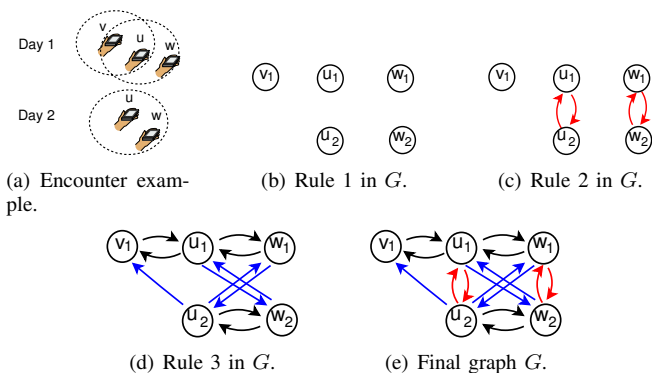


Figure 1. (a) Meeting between  $u$ ,  $v$ , and  $w$  during days 1 and 2. (b)-(d) Rules for the construction of graph  $G$ . (e) Final representation of graph  $G$ .

## V. BENCHMARK APPROACH

We propose a benchmark that gives an optimal solution to our problem: 100% of user coverage daily, with minimum number of VIPs. It is important to underline that the benchmark serves only for comparison purposes as it requires knowing the future to compute the exact set of VIPs. It is obtained by abstracting our application scenario to a formal representation. Consider that the network has to be covered daily by VIP delegates, for a period  $P$  during which the activity of all network users is known. Let also  $P$  be  $n$  days long. We construct a directed graph  $G = (V, E)$  through the following rules (a step-by-step generation of graph  $G$  is illustrated in Figure 1):

**Rule 1:** Graph  $G$  has a vertex  $u_i$  for each day  $i$  in which user  $u$  is active (i.e.,  $u$  has at least one contact during the day). This vertex impersonates  $u$  during day  $i$  in  $G$  and is referred to as the *image* of  $u$  during that day (see Figures 1(a) and 1(b)).

**Rule 2:** Every couple of images of the same user  $u$  in  $G$  is connected through a couple of directed edges, i.e., images of same user  $u$  are members of a clique in  $G$  (see Figure 1(c)).

**Rule 3:** If users  $u$  and  $v$  meet on day  $i$ , then every member  $u_t$  of the clique composed of  $u$ -images in  $G$  is connected to  $v_i$  through an edge  $(u_t, v_i)$ . Similarly, every member  $v_t$  of the clique composed of  $v$ -images in  $G$  is connected to  $u_i$  through an edge  $(v_t, u_i)$ . In particular,  $G$  also contains edges  $(v_i, u_i)$  and  $(u_i, v_i)$  representing that  $u$  and  $v$  met on day  $i$  (see Figure 1(d)).

The graph  $G$  constructed with the rules above represents users' behaviors in the network during the whole period  $P$ . Consider a certain user  $u$ . According to Rules 1 and 2, user  $u$  is "expanded" in  $G$  into a clique, containing images of  $u$  only for the days  $u$  is active (see Figures 1(b) and 1(c)). Moreover, if  $u$  meets  $v$  in day  $i$ , Rule 3 guarantees that all members of the clique representing  $u$  in  $G$  point to  $v_i$  (see Figure 1(d)). The intuition behind this rule is that outgoing edges from  $u$ 's clique indicate that " $u$  can be a delegate for  $v$  on day  $i$ ".

Rule 3 is applied to every day user  $u$  is active. Thus, all members of the  $u$ 's clique in  $G$  point to the same members of *other users' cliques*. Consequently, any image of  $u$  in  $G$  (any member of  $u$ 's clique) is enough to determine the set of users for which  $u$  can be a delegate, and on which days.

Intuitively, in order to cover all the network day by day, it is enough to select as delegates the members of a minimum out-dominating set of graph  $G$ . Moreover, such a set of delegates is the smallest set that can achieve full coverage. The following theorems, whose proofs are omitted due to space constraints, prove such intuition.

**Theorem 1:** Let  $MDS$  be a minimum out-dominating set of  $G$ . The set  $MDS$  can cover 100% of the active users for each day  $i \in P$ .

**Theorem 2:** Let  $MDS$  be a minimum out-dominating set of  $G$ . Let also  $S$  be the smallest set of VIP delegates able to cover, for every day  $i \in P$ , 100% of the active network users on day  $i$ . Then,  $|MDS| \leq |S|$ .

The above theorems indicate how to proceed to find the best possible solution to our problem: After constructing graph  $G$  according to Rules 1, 2, and 3, find a minimum out-dominating set of  $G$  and use the members of such set as benchmark VIPs.

The minimum dominating set is notably a NP-hard problem. Thus, to individuate our benchmark VIP delegates, we reduce our problem to Set Cover (equivalent to  $MDS$  under L-reductions [27]) for which a simple greedy algorithm is known to be only a logarithmic approximation factor away from the optimum [27]. To the best of our knowledge, there is provably no polynomial-time algorithm with a smaller approximation factor. The delegates obtained by this heuristic are then used as benchmark VIPs in our experiments.

## VI. EXPERIMENTAL SETTING

In this section, we give detailed information on both the real and synthetic data-sets we use for evaluation. Then, we present, step by step, how the social graph and the social communities are computed for each dataset.

### A. Data-sets

For the evaluation we use two real data-sets: Dartmouth [31] (movement of students and staff in campus) and Taxis [32] (movement of cabs in San Francisco)– and three synthetic datasets generated with the SWIM mobility model [26], shown to simulate well human mobility from statistical and social points of view. The vehicular mobility of the cabs is different from human mobility (Dartmouth). However, the purpose of using the taxis trace is to test our solution's extendibility to different contexts.

**Dartmouth.** Dartmouth includes SNMP logs from the access points across the Dartmouth College campus from April 2001 to June 2004. To generate user-to-user wireless contacts from the data-set, we follow the popular consideration in the literature that devices associated to the same AP at the same time are assumed to be in contact [24].

**Taxis.** The Taxi data-set contains GPS coordinates of 536 cabs collected over 24 days in San Francisco. Here, we assume that two cabs are in contact when their geographical distance is smaller than 250 m (according to Piorkowski et al. [32]).

**SWIM.** We first generate SWIM-500, a 2-month long, 500-node simulation of the Cambridge Campus data-set [26]. To

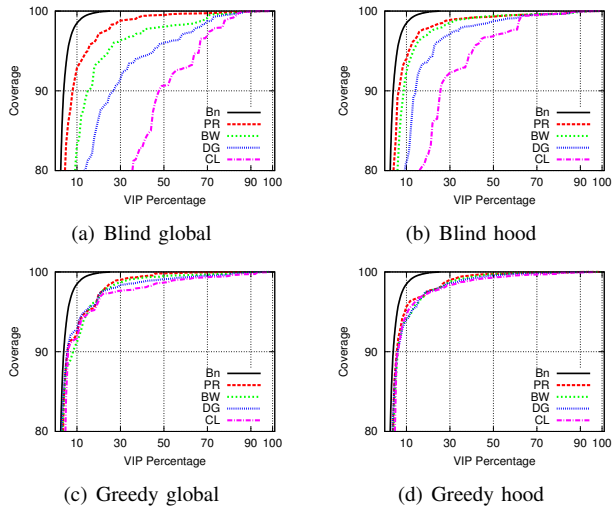


Figure 2. Performance of the selection strategies on the **Dartmouth** data-set. “Bn” refers to the benchmark, “PR” to the page-rank, “BW” to betweenness centrality, “DG” to degree centrality, and “CL” to closeness centrality.

emulate two possible network-enlargement scenarios—(1) an urban growing in both area and population and (2) a sudden over-population of a city— we generate two other traces of 1500 nodes. This is done by (1) keeping the density constant (D-SWIM-1500) and (2) keeping the area constant (A-SWIM-1500). Table I summarizes the details of the data-sets. Note that, although both Dartmouth and SWIM data-sets represent campus scenarios, they yield different activity per node per day as they used distinct technologies; respectively Wi-Fi in Dartmouth, and Bluetooth-like in SWIM.

### B. Training period and social graph

Our strategies *do not* require pre-knowledge of contact patterns among users. Rather, we use an observation/training period *as short as 1 week*, exploiting repeatability of users’ movement patterns and recurrence of contacts (i.e., short-range wireless connectivity) among them. The length of the training period is not casual. Our life and the activities we conduct are organized on a week-base, mostly having a common routine repeated day by day (e.g., go to work/school or have lunch in the same place). This also is reflected in the taxi movement, guided by the clients’ need to reach specific locations in the city. The repeatability of contacts is due to the popularity of geographical zones (e.g., city center, stations etc.), and to fixed tracks connecting them.

Our intuition on the length of the training period is also confirmed by the results shown in Table I. The properties of the training period are very close to the whole trace for all scenarios. This allows us characterizing social relationships among users and make prediction of future meetings easy. We are indeed able to generate a *social graph*, where two users are connected only if they have met with a certain frequency – that we call *social connectivity threshold* – during the training period. The social connectivity threshold depends on the scenario considered:

- In the Dartmouth data-set, social connectivity is due to university life (sharing the same classes, studying in the

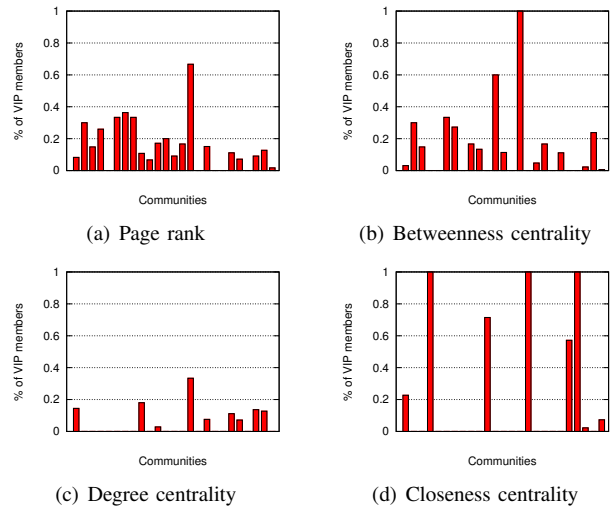


Figure 3. Distribution of VIPs per social attribute on the **Dartmouth** data-set with the *blind global* promotion strategy. The x-axis represents different communities detected.

same library, living in the same dorm), which generates lots of meetings between people during the week. We thus set the social connectivity threshold to at least 1 contact per day, for at least 5 days during a week.

- The social connectivity threshold in the Taxi data-set is higher due to higher speeds: At least 8 contacts per day during the training period’s week.
- SWIM-500 also represents a University campus, thus the social connectivity threshold value is the same as in the Dartmouth trace. In D-SWIM-1500 (constant density scaling) the threshold remains constant. Because of the increased density, it increases to at least 8 contacts per day for at least 5 days of the training week in A-SWIM-1500, the SWIM-500 scaled version with constant area.

Our *hood VIPs* selection strategies operate on a community basis and aim at covering single communities by selecting members that are important in the network. We determine communities on the social graph of both Dartmouth and SWIM traces through the  $k$ -clique algorithm [30], which is widely used in the area of social mobile networking [16]. The communities are well-knit and do not show much intersection between them: The average Jaccard similarity index [33] between intersecting communities is 0.038 in the Dartmouth case and about 0.025 in SWIM-500 and D-SWIM-1500 case, and 0.045 in A-SWIM-1500 where there is more overlapping. The Taxi data-set, due to the large number of contacts and the high mobility of nodes, does not present any community sub-structuring. When applying the  $k$ -clique algorithm, we only observe a huge community containing almost 80% of the nodes, whereas the remaining 20% do not belong to any community. Thus, we decided to apply only the global VIP selection strategies to this trace.

## VII. EXPERIMENTAL RESULTS

We analyze the performance of all our strategies in terms of coverage when applied to real and synthetic traces. We investigate the coverage trend with regard to an increasing number

Table I  
DETAILS ON THE DATASETS (DS) AND RESPECTIVE TRAINING PERIOD (TR).

Data set	Taxi	Dartmouth	SWIM-500	D-SWIM-1500	A-SWIM-1500
Total nodes	536	1142	500	1500	1500
DS AVG active/day	491	1060	499.98	1500	1500
TR AVG active/day	429	1061.5	500	1500	1500
DS AVG contacts/node/day	7804	292	128	130	380
TR AVG contacts/node/day	7656	284	131	129	378

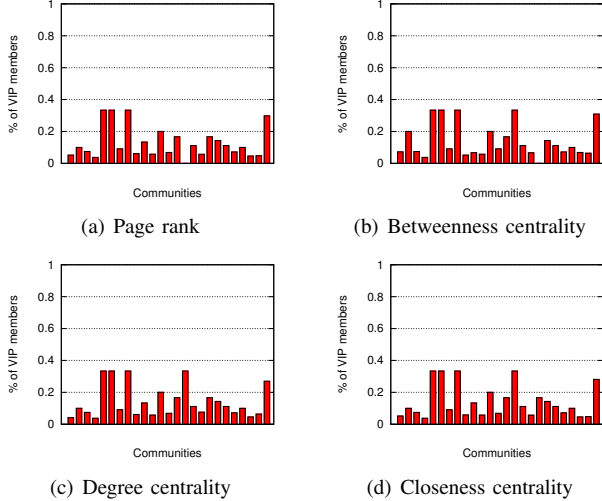


Figure 4. Distribution of VIPs per social attribute on the **Dartmouth** dataset with the *greedy global* promotion strategy. The x-axis represents different communities detected.

of the VIPs. The set used for coverage is updated from time to time following the order in which each strategy selects VIPs. For the sake of comparison, the results for the benchmark (“Bn”) are included in the plots. We use the same technique as above to build the benchmark’s trend: Updating the VIPs set and the corresponding network coverage, following the order in which the benchmark promotes VIPs.

It is worth noting that, for the page rank attribute, we have firstly analyzed the impact of varying the damping fact values (i.e.,  $d$ ) on the coverage of the selection strategies. We decided to use then  $d = 0.85$ , since it results in the best performance in terms of network coverage.

#### A. Performance results on real data-sets: Dartmouth case

**Blind promotion.** We show in Figure 2 the coverage obtained by each of the promotion strategies. The *blind* promotion in the global and hood VIP selection strategies yields the results presented in Figures 2(a) and 2(b). Notice that there is a coverage efficiency gap between page rank VIPs (referred as “PR” in the figure) and those of other centralities (referred as “BW”, “DG”, and “CL”). In addition, page rank is very close to the benchmark, even for small percentages of delegates considered. For instance, in the *global blind strategy*, to get to 90% of coverage, page rank only requires the promotion of 5.93% of nodes as delegates against 3.92% with the benchmark approach (see Table II).

Another consideration to be made is that hood selection is more effective than global selection. Hence, aiming to

Table II  
DELEGATES GIVEN BY EACH STRATEGY TO OBTAIN 90% COVERAGE ON DARTMOUTH. THE BENCHMARK NEEDS 3.92% OF NODES.

	G-Blind (%)	H-Blind (%)	G-Greedy (%)	H-Greedy (%)
PR	8.98	6.89	5.93	6.19
BW	15.96	9.16	8.98	6.19
DG	26.96	15.09	5.93	6.19
CL	47.993	26.0035	5.93	6.19

cover the network by forcing VIP selection within different communities seems to be a very good strategy. Nevertheless, there exist social attributes such as page rank that do not gain much from the hood selection. Indeed, global and hood page rank VIPs perform very similarly in both data-sets. This is because, on the one hand, page rank VIPs already target different communities, even in the global case. On the other hand, betweenness, degree, and closeness centrality tend to over-select VIPs from a few network communities, and consequently, leave uncovered many marginal ones. This is also confirmed by the resulting distribution of VIPs in communities after the global selection according to each attribute (see Figure 3). The tendency of these social attributes to target only a few communities is attenuated with the hood selection that boosts their efficiency in covering the network.

**Greedy promotion.** When applying the *greedy promotion*, the performance of all strategies improves considerably and gets much closer to the benchmark (see Figures 2(c) and 2(d)). In addition, VIPs obtained with each social attribute perform very similarly to each other, in both hood and global selections. This is due to the capacity of the greedy approach to *not* promote as VIPs nodes that are too close to each other in the social graph. Indeed, after every node’s promotion to VIP, all its neighbors in the social graph and their links are removed. Since communities are very well tight, only the promotion of one member can remove a large part of the community (if not all of it). Thus, attributes such as betweenness and closeness *do not* concentrate their selection on a few communities as in the global selection. This is also confirmed by Figure 4, where we show how the greedy strategy distributes VIPs among communities for different social attributes.

In addition, the results with 90% coverage presented in Table II confirm page ranks’s high performance ability when combined with every strategy.

#### B. Performance results on real data-sets: Taxi case

As already discussed, due to the high mobility of nodes in this dataset only a huge unique community containing 80% of nodes is detected, while the 20% remaining nodes do not belong to any community. Though this effect makes the global

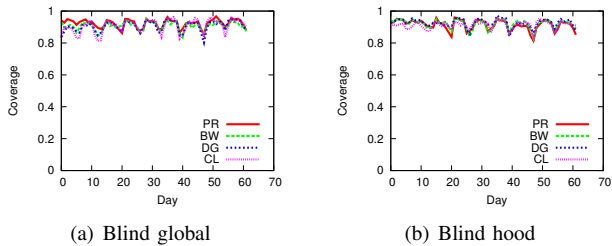


Figure 5. Coverage stability in time (**Dartmouth** data-set).

selection the only applicable, all strategies perform very well in this scenario. Moreover, the sets selected by each strategy to guarantee up to 90% of coverage have exactly the same (small) size: Only 0.93% of network nodes. The benchmark guarantees the same coverage with 0.2% of network nodes selected. Coverage trend related graphics are omitted due to lack of space.

### C. Performance results on synthetic data-sets: SWIM

Here we present results of our experiments done with the SWIM mobility model. We first generate SWIM-500, a 500 nodes simulation of the Cambridge University scenario [16], [26]. Then, to study VIP selection performance in two intuitive scenarios– a urban growing in both area and population (constant density), and, a sudden over-population of a city (constant area)– we generate two scaled versions of SWIM-500 with 1500 nodes, D-SWIM-1500 (constant density scaling) and A-SWIM-1500 (constant area scaling).

We start from the *Blind promotion*. Like in the Dartmouth scenario, page-rank VIPs result more efficient than those of other centralities (see Tables III, IV, and V containing the percentage of delegates needed by strategies to cover 90% of the network). The reason is the same discussed in the previous section. I.e. page-rank global VIPs are better distributed within communities with respect to VIPs of other centralities. Moreover, again aiming to cover the network by forcing VIPs to fall in different communities (hood selection) is a winning strategy.

The *Greedy promotion*, seamlessly to the Dartmouth case, boosts up the performance of all strategies by thus yielding much better coverage results with respect to the *Blind promotion* (see Tables III, IV, and V).

What’s interesting to notice here is the impact of the way of scaling on our strategies. When passing from SWIM-500 to D-SWIM-1500 (constant density) all strategies perform very similarly, in both blind and greedy promotions (see Tables III and IV). Conversely, in an emergency situation, where the network is suddenly much more overloaded as a result of the over-population of the network area (A-SWIM-1500), our strategies perform even better. Again, due to space constraints we omit coverage trend related graphics on SWIM. We want to stress however that their trend is very similar to the Dartmouth traces.

### D. Coverage stability

Here we study the stability of coverage of our strategies in time. We focus on VIP sets big enough (according to Table II)

Table III  
DELEGATES GIVEN BY EACH STRATEGY TO GET 90% COVERAGE ON SWIM-500. THE BENCHMARK APPROACH NEEDS 7.4%.

	G-Blind (%)	H-Blind (%)	G-Greedy (%)	H-Greedy (%)
BW	21	14	9	10.6
CL	48	25.4	12.8	11.6
DG	23	13.6	8	8.8
PR	10.8	10.2	9.8	9.6

Table IV  
DELEGATES GIVEN BY EACH STRATEGY TO GET 90% COVERAGE ON D-SWIM-1500. THE BENCHMARK APPROACH NEEDS 7.06%.

	G-Blind (%)	H-Blind (%)	G-Greedy (%)	H-Greedy (%)
BW	22	17.26	9	9.93
CL	59	42.06	12.93	11.6
DG	24	17.2	8	9.06
PR	10.9333	10.06	9	9.93

to cover 90% of the network for the whole period and we plot their actual daily performance in Figure 5. Due to the lack of space, we only present results related to the Blind strategies on the Dartmouth data set. We stress however that the same results hold for all the strategies over all the data-sets considered (also Taxi and SWIM). We observe that coverage is quite constant in time for every strategy. This reinforces our intuition on both the training period and the way the social graph is generated. With minimal information on the scenario and a very short observation of the network, our strategies are able to compute VIP sets that are small, efficient, and stable in time.

## VIII. INCENTIVIZING VIPs AND DATA TRANSFER

Being the human nature inherently selfish, it is more likely that no user would accept the promotion to VIP. However, the number of VIPs selected by our strategies to guarantee 90% coverage is quite low (8% in SWIM, 5.93% in Dartmouth, less than 1% in Taxi). In view of this, VIPs could have their devices upgraded to more fancy, recent ones, and get paid for carrying them around and "working" for the network provider/application builder. Considering the amount of funding that Governments worldwide are putting into global-sensing research [34]–[36] this incentive is more than real. Another possibility involves considering users’ traffic load at the delegates selection and use it to establish a maximum load threshold per delegate. Accordingly, combine it with the social attributes for delegate selection considering fairness and resource constraint among delegates.

Many applications could benefit by the VIPs: collection of urban-sensing related data; distribution of large content to users by service/software providers (e.g., software updates and recurrent security patches); free update of mobile software’s ad pools. All these are delay-tolerant, and would not suffer from the 1-day latency of the daily coverage of VIPs. Moreover, the data transfer between VIPs and the network could happen with different frequencies according to the time-sensitiveness of the data. Finally, it is clear that 3G networks cannot handle such traffic in the classic way (whenever VIPs like), because it would not be of any benefit to offloading. However, the 3G network can still be used in different moments of the day to transfer the data. So, the network load would result distributed in time rather than concentrated in highly congested

Table V

DELEGATES GIVEN BY EACH STRATEGY TO GET 90% COVERAGE ON A-SWIM-1500. THE BENCHMARK APPROACH NEEDS 2.53%.

	G-Blind (%)	H-Blind (%)	G-Greedy (%)	H-Greedy (%)
BW	10	6.73	4	4.4
CL	30	11.6	9	6.06
DG	7	5.13	4	3.53
PR	5	4.33	6	4.4

hours. Another possibility is to transfer the data through wired networks, whenever a VIPs device gets connected to a broadband network during the day. After all, if VIPs are being paid to perform such task, this assumption is more than reasonable.

Finally, in this work we focus on 90% network coverage because we empirically observed that this value leads to a good balance between coverage and number of VIPs. However, further investigation of the trade off between coverage and VIP number will be investigated in future work.

## IX. CONCLUSION

Dense metropolitan areas are suffering network overloading due to the data-traffic generated by the proliferation of smart-phone devices. In this paper, we describe VIP delegation, a mechanism to alleviate such traffic based on opportunistic contacts. Our solution relies on the upgrade of a small, crucial set of VIP nodes that regularly visit all network users and collect (disseminate) data to them on behalf of the network infrastructure.

VIPs are defined according to well known social network attributes (betweenness, closeness, degree centrality and page-rank), and are selected according to two methods: global (network-based) and hood (community-based) selection. All methods rely on a short network observation period of 1 week, and select VIP sets that result small, efficient, and stable in time. Extensive experiments with several real and synthetic data-sets show the effectiveness of our methods in offloading: VIP sets of about 7% and 1% of network nodes in respectively campus-like and vehicular mobility scenarios are enough to guarantee about 90% of network offload. Additionally, the performance of the VIPs selected by our methods is very close to an optimal benchmark VIPs set computed from the full knowledge of the system (i.e., based on past, present, and future contacts among nodes).

## REFERENCES

- [1] N. Eagle and A. Pentland, "Social serendipity: Mobilizing social software," *IEEE Pervasive Computing*, vol. 4, no. 2, pp. 28–34, 2005.
- [2] S. Gaonkar, J. Li, R. Choudhury, L. Cox, and A. Schmidt, "Microblog: sharing and querying content through mobile phones and social participation," in *ACM MobiSys*, 2008.
- [3] E. Miluzzo, N. D. Lane, K. Fodor, R. Peterson, H. Lu, M. Musolesi, S. B. Eisenman, X. Zheng, and A. T. Campbell, "Sensing meets mobile social networks: the design, implementation and evaluation of the cenceme application," in *ACM SenSys*, 2008.
- [4] S. B. Eisenman, N. D. Lane, E. Miluzzo, R. A. Peterson, G. Ahn, and A. T. Campbell, "Metrosense project: People-centric sensing at scale," in *ACM SenSys*, 2006.
- [5] S. Ioannidis, A. Chaintreau, and L. Massoulie, "Optimal and scalable distribution of content updates over a mobile social network," in *IEEE Infocom*, 2009.
- [6] L. McNamara, C. Mascolo, and L. Capra, "Media sharing based on colocation prediction in urban transport," in *ACM Mobicom*, 2008.

- [7] "Customers Angered as iPhones Overload AT&T," *NY Times*, <http://www.nytimes.com/2009/09/03/technology/companies/03att.html>, September 2009.
- [8] "iPhone overload: Dutch T-Mobile issues refund after 3G issues," *Ars Technica*, <http://arstechnica.com/tech-policy/news/2010/06/dutch-t-mobile-gives-some-cash-back-because-of-3g-issues.ars>, July 2010.
- [9] "Why 4G won't fix your mobile woes," *Infoworld*, <http://www.infoworld.com/d/mobilize/why-4g-wont-fix-your-mobile-woes-297>, Jan. 2010.
- [10] "Mobile Data Offload for 3G Networks," White Paper, IntelliNet Technologies, Inc., 2009.
- [11] V. Chandrasekhar, J. Andrews, and A. Gatherer, "Femtocell networks: a survey," *IEEE Communications Magazine*, vol. 46, no. 9, pp. 59–67, Sep. 2008.
- [12] "AT&T, Verizon Wireless join Wi-Fi interoperability group," *Cnet*, [http://news.cnet.com/8301-30686\\_3-20008476-266.html](http://news.cnet.com/8301-30686_3-20008476-266.html), Jun. 2010.
- [13] A. Balasubramanian, R. Mahajan, and A. Venkataramani, "Augmenting mobile 3g using wifi," in *ACM MobiSys*, 2010.
- [14] B. K. Polat, P. Sachdeva, M. H. Ammar, and E. W. Zegura, "Message ferries as generalized dominating sets in intermittently connected mobile networks," in *ACM MobiOpp*, 2010.
- [15] E. M. Daly and M. Haahr, "Social network analysis for routing in disconnected delay-tolerant manets," in *ACM MobiHoc*, 2007.
- [16] P. Hui, "People are the network: experimental design and evaluation of social-based forwarding algorithms," Ph.D. dissertation, UCAM-CL-TR-713. University of Cambridge, Comp.Lab., 2008.
- [17] A. Mtibaa, M. May, C. Diot, and M. Ammar, "Peoplerank: Social opportunistic forwarding," in *IEEE Infocom*, 2010.
- [18] F. Li and J. Wu, "LocalCom: a community-based epidemic forwarding scheme in disruption-tolerant networks," in *IEEE SECON*, 2009.
- [19] W. Gao, Q. Li, B. Zhao, and G. Cao, "Multicasting in delay tolerant networks: a social network perspective," in *ACM MobiHoc*, 2009.
- [20] Q. Li, S. Zhu, and G. Cao, "Routing in socially selfish delay tolerant networks," in *IEEE Infocom*, 2010.
- [21] A. Mei and J. Stefa, "Give2get: Forwarding in social mobile wireless networks of selfish individuals," in *ICDCS*, 2010.
- [22] B. Han, P. Hui, V. S. A. Kumar, V. M. Marathe, G. Peig, and A. Srinivasan, "Cellular traffic offloading through opportunistic communications: a case study," in *ACM CHANTS*, 2010.
- [23] A. Garyfalos and K. C. Almeroth, "Coupons: A multilevel incentive scheme for information dissemination in mobile networks," *IEEE Transactions on Mobile Computing*, vol. 7, no. 6, pp. 792–804, Jun. 2008.
- [24] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of human mobility on the design of opportunistic forwarding algorithms," *IEEE Transactions on Mobile Computing*, vol. 6, no. 6, pp. 600–620, Jun. 2007.
- [25] M. Gonzalez, C. Hidalgo, and A. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, pp. 779–782, 2008.
- [26] A. Mei and J. Stefa, "SWIM: A simple model to generate small mobile worlds," in *IEEE Infocom*, 2009.
- [27] V. Kann, "On the approximability of np-complete optimization problems," Ph.D. dissertation, Royal Institute of Technology, Stockholm 1992.
- [28] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social Networks*, vol. 1, no. 3, pp. 215–239, 1979.
- [29] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Comput. Netw. ISDN Syst.*, vol. 30, pp. 107–117, 1998.
- [30] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, no. 7043, pp. 814–818, 2005.
- [31] T. Henderson, D. Kotz, I. Abyzov, and J. Yeo, "CRAWDAD trace dartmouth/campus/movement/01\_04 (v. 2005-03-08)," Downloaded from [http://crawdadd.cs.dartmouth.edu/dartmouth/campus/movement/01\\_04](http://crawdadd.cs.dartmouth.edu/dartmouth/campus/movement/01_04), Mar. 2005.
- [32] M. Piorkowski, N. S.-Djukic, and M. Grossglauser, "A parsimonious model of mobile partitioned networks with clustering," in *COMSNETS*, 2009.
- [33] P. Jaccard, "Étude comparative de la distribution florale dans une portion des Alpes et des Jura," *Bulletin del la Société Vaudoise des Sciences Naturelles*, vol. 37, pp. 547–579, 1901.
- [34] CENS Urban sensing, <http://urban.cens.ucla.edu/projects>.
- [35] MIT Senseable City Lab, <http://senseable.mit.edu>.
- [36] EARSel Special Interest Groups, <http://www.earsel.org/?target=SIGs>.