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

Marco Valerio Barbera, Aline Carneiro Viana, Marcelo Dias de Amorim, Julinda Stefa

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

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

Marco Valerio Barbera — Aline Carneiro Viana — Marcelo Dias de Amorim — Julinda Stefa

N° 7563

March 2011

Thème COM

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--7563--FR+ENG

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

INRIA

centre de recherche
SACLAY – ÎLE-DE-FRANCE

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

Marco Valerio Barbera , Aline Carneiro Viana , Marcelo Dias de Amorim , Julinda Stefa

Thème COM — Systèmes communicants
Équipes-Projets Direction de l'INRIA Saclay - Ile de France

Rapport de recherche n° 7563 — March 2011 — 32 pages

Abstract: he recent boost up of mobile data consumption is straining 3G cellular networks in metropolitan areas and is the main reason for the ending of unlimited data plans by many providers. To address this problem, we propose the use of ***opportunistic delegation*** as a data traffic offload solution 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 important users in the network called ***VIP delegates***. To identify VIPs, we propose different strategies that only need short observation periods of the network (around 1 week) to cover several months of activity. 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). Through several experiments using real and synthetic traces, we 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.

Key-words: collaborative data collection or distribution; centrality; community detection; delay tolerant networks; social interactions

Centre de recherche INRIA Saclay – Île-de-France
Parc Orsay Université
4, rue Jacques Monod, 91893 ORSAY Cedex
Téléphone : +33 1 72 92 59 00

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

Résumé : Récemment, la large consommation de données par téléphones portables dans les zones métropolitaines constitue une rude épreuve aux réseaux cellulaires 3G et est la principale raison de la fin des forfaits de données illimité par de nombreux operateurs. Pour résoudre ce problème, nous proposons l'utilisation de ***délégation opportunistes*** comme une solution au déchargement des données de trafic. Pour cela, nous étudions deux questions principales: (i) "Comment obtenir un aperçu des scénarios sociaux et des réseaux mobiles" (ii) "Comment utiliser ces connaissances pour concevoir des solutions pour soulager les réseaux surchargé 3G?". Le but de notre solution est l'influence sur l'utilisation d'applications mobiles nécessitant d'importants transferts de données en canalisant le trafic à quelques-uns, les utilisateurs social important dans le réseau, appelé ici de ***délégués VIP***. Pour identifier les VIP, nous proposons des stratégies différentes qui ne nécessitent que des périodes d'observation de courte durée du réseau (environ 1 semaine) pour couvrir plusieurs mois d'activité. Les stratégies proposées pour la sélection de VIPs sont basées sur les propriétés des réseaux sociaux et sont comparées à la solution optimale (qui couvre 100 % des utilisateurs avec un nombre minimum de VIP). Grâce à plusieurs expériences en utilisant les traces réelles et de synthèse, nous montrons l'efficacité de notre délégation via VIPs à la fois en termes de couverture et le nombre requis de VIPs – jusqu'à 7% en moyenne de VIP sont nécessaires dans les scénarios de type campus pour décharger environ 90% du trafic.

Mots-clés : collecte ou distribution collaborative de données, centralité, détection de communautés, les réseaux tolérants aux délais, les interactions sociales

he recent boost up of mobile data consumption is straining 3G cellular networks in metropolitan areas and is the main reason for the ending of unlimited data plans by many providers. To address this problem, we propose the use of ***opportunistic delegation*** as a data traffic offload solution 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 important users in the network called ***VIP delegates***. To identify VIPs, we propose different strategies that only need short observation periods of the network (around 1 week) to cover several months of activity. 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). Through several experiments using real and synthetic traces, we 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.

1 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. This eagerness has boosted up sales in the market – more than 64% up annually worldwide in Q2 2010 [1]. 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 [2, 3], global sensing [4, 5], and content distribution [6, 7] 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 [8]. 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, by thus forcing the company to issue refunds for affected users [9].

All these issues are bringing new technical challenges to the networking and telecommunication community. 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 direct communication 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 movement and interactions are able to contact regularly all the rest of the network users. 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 [4, 5], distribution of content to users by service providers, and broadcasting are examples of applications that would directly benefit from the use of VIPs. As these examples are typical applications that induce large amount of traffic in the network, delegating the collection/distribution of such traffic to a few important nodes would result in immediate data offload.

After discussing related work in Section 2, we propose, formalize, and evaluate two methods of VIPs selection: *global* and *neighborhood* VIP delegation (see Section 3). While the former focus 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 (i.e., 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 contacts with the remaining nodes (see Section 4).

Through extensive experiments on both 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 6). We compare our solution with an optimal benchmark computed from the full knowledge of the system (i.e., based on past, present, and future contacts among nodes). 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 complementary solutions and future improvements in Section 8 and conclude with Section 9.

2 Background

In view of the scenario presented in the introduction, we propose an offload method based solely on the inherent social aspects of user mobility. Our idea is to detect subscribers that, with their inherent mobility, are able to *encounter* all the rest of the network users in a regular fashion. 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. Before this, we go through the related work in the area, discussing the most representative results on both data offloading and user-aided networking services.

2.1 Data offloading

Consumption of mobile data by the pervasive usage of smart-phones is forcing carriers to find ways to offload the network. Boosting the network was not enough for AT&T [10], so measures such as limited traffic plans with tiered pricing had to be taken [11]. The goal was twofold: attenuating the problem and educating application builders to write more efficient code. Not only this is perceived as a temporary solution to overloading, but there is a strong belief among network providers that the same problems will persist even with more efficient technologies such as 4G or WiMAX [12].

So far the most reasonable solution to the problem is offloading to alternate networks, such as femtocells and Wi-Fi [13]. 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 [14]. As a side effect, automatic switching of devices from cellular network to femtocells reduces the load of the network. Besides from being localized (indoors only), such solution suffers from the non-proliferation of femtocells to subscribers' homes. Moreover, charging users for the necessary equipment as the network providers are currently doing (150 USD for AT&T's microcell) will not help in this direction.

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. Offloading cellular traffic to Wi-Fi was seen as a challenge due to the incapability of former handsets to connect to such networks. However, the proliferation of modern Wi-Fi enabled smart-phones, together with the network providers' tendency towards already existing technologies has turn Wi-Fi offloading into a reality [13]. More and more carriers in USA and worldwide are investing in this direction [15], by installing access points and hot-spots close to overloaded cellular towers, and by providing to clients Wi-Fi access within tiered monthly subscription. In this direction, Balasubramanian et al. propose a system to augment access to 3G network through Wi-Fi offloading [16]. This system, called Wiffler focuses only on Internet access from *moving vehicles*. It leverages delay tolerance and fast switching of devices to overcome the poor availability and performance of Wi-Fi.

Even though offloading to Wi-Fi seems to be the best solution so far to cellular network overloading, the continuous increasing of mobile data-traffic demand suggests for integration of Wi-Fi with other offloading methods. Indeed, according to a report released from CISCO on February 2010 mobile data traffic is growing today 2.4 times faster than global fixed broadband one [17]. The mobile data consumed annually is expected to reach 40 exabytes by 2014, more than 90% of which will be driven by laptops' and smart-phones' *air cards*.

As we will show in this paper, our solution is essentially different from Wi-Fi and femtocell-based offloading; nevertheless, it can be integrated to these methods to further help alleviate mobile data overloads.

2.2 User-aided networking

Polat et al. suggest some sort of network members' promotion to enhance network functionalities [18]. They focus on providing multi-hop connectivity in a mobile ad-hoc network. 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 from the difference in both problematic and application scenario with our work, no social aspect/importance of the network members is considered in promotion.

Many research works targeting social mobile networks make use of social ties between users to leverage network services. The problems considered range from message multi-hop forwarding [19–21] and multicasting [22, 23], to selfishness and network security [24–26].

To the best of our knowledge, Han et al. were the first to exploit opportunistic communication to alleviate data traffic in cellular networks [27]. However, conversely from ours, their solutions only apply to information dissemination problems such as broadcasting. They focus on selecting k target users to which the information is first sent through cellular network. These target users will then, through *multi-hop* opportunistic forwarding, disseminate the information to all users in the network. In our scenario where large amount of data is to be transferred, multi-hop forwarding is not feasible: applications that require collection of sensing data would result highly expensive in terms of energy [4, 5]. Moreover, multi-hop forwarding requires collaboration of all users in the network. Even though such collaboration can be stimulated by incentive mechanisms [28], 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.

3 VIP Delegation in a Nutshell

The movement of smart-phone users is not random; rather, it is a manifestation of their social behavior [20, 29–31]. This movement, along with contact-based 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. Therefore, in general lines, this paper investigates the following questions:

- How to gain insights into social mobile networking scenarios?
- How to utilize such insights to design solutions allowing alleviating the network traffic in the current overloaded 3G networks?

In particular, though the number of network users can be very high, 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 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?* As previously mentioned, 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 4).

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 [32]. 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.

4 Global vs. Hood VIPs selection methods

We now present the different VIP selection strategies. To this end, we first need to identify when a user/node is "important" in the network and according to which structural attribute.

4.1 Who are the VIPs?

From mobility patterns and 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 contact with each other (see a detailed description in Section 6.3). 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 [33, 34].

Betweenness centrality measures the number of occurrences of a node in the shortest-path between pairs of others nodes. *Degree centrality* ranks nodes based on the number of their direct ties (i.e., neighbors) in the graph. *Closeness centrality* gives higher ranking to nodes with lower multi-hop distance to other nodes of the graph. Finally, *page rank*, the well known Google's ranking algorithm, measures the likelihood of nodes in having important friends in a social graph [34].

All these social attributes are very meaningful when speaking about social graphs. For instance, betweenness determines "bridge nodes" that, with their movement, act as connectors between group of nodes. Betweenness centrality for a given node k is calculated as:

$$C_B(k) = \sum_{\substack{j=1 \\ j \neq k}}^{N} \sum_{\substack{i=1 \\ j \neq k}}^{N} \frac{g_{i,j}(k)}{g_{i,j}}, \tag{1}$$

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. Degree identifies the most popular nodes, also called *hubs* in social network theory, which may act as a conduit for information exchange. Degree centrality is calculated as:

$$C_D(k) = \sum_{i=1}^{N} a(k, i), \tag{2}$$

where $a(k,i) = 1$ if there is a link between k and i, and $a(k,i) = 0$ otherwise.

Closeness describes "independent nodes" that do not dependent upon others as intermediaries or relayers of messages due to their closeness to other nodes. To deal with disconnections in the social graph, we calculate the closeness centrality as:

$$C_C(k) = \frac{N-1}{\sum_{i=1}^{N} d(k,i)}, \tag{3}$$

where $d(k,i)$ is the length of the shortest path between nodes k and i.

Finally, page rank tags users as important when they have social connections to other important users. In particular, page rank of a node i in the social graph is given by the equation:

$$PR(k) = \frac{1-d}{N} + d \sum_{i \in F(k)} \frac{PR(i)}{|F(i)|}, \tag{4}$$

where d ($0 \leq d \leq 1$) is the damping factor and $F(k)$ is the set of neighbors that links to k in the social graph. 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.

Nodes that are important according to all these social structural attributes match well human intuition about important nodes in a society. Thus, we chose to base our VIP selection methods on each of them.

4.2 Global VIPs

As we already mentioned, the global VIP delegation strategy aims to select the smallest VIP set over the global social graph, able to *daily cover the network* through direct contacts with network users. For this, the network 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 continuously 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 firstly starts with promoting to VIP the top-ranked node. After this promotion, the nodes this VIP covers are dumped and ranking on the remaining nodes are re-computed. Again, the procedure is repeated till the network is covered.

4.3 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: we encounter work friends at the office, gym friends at the gym, our family at home, and our neighbors in our neighborhood. The mobile social network generated by this behavior encompasses, besides contact locality, also well tight social-community sub-structures. Thus, our strategy here aims to cover each community at a time, selecting *hood VIPs* by their importance within their communities. To detect social-communities among network members we use the *k-clique* community algorithm [29].

As in the global VIP delegation, the *hood* VIP selection also requires nodes to be ranked according to betweenness, closeness, or degree centrality and page rank. After such ranking is provided, we start covering each community by promoting its members to VIPs. This promotion is performed as for in the global VIPs method:

- *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, till all the network's communities are covered.

4.4 VIP delegation in practice

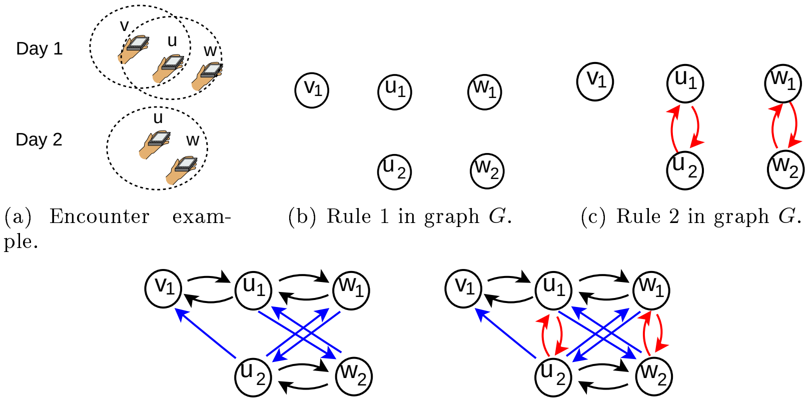
The computation of previously discussed ranking metrics and of sub-communities on a social mobile network requires knowledge on user mobility. Clearly, none of them is available from scratch when one has to pre-compute the VIPs set that repeatedly will visits all the network users in the future. On the other hand, we cannot pretend to know every single's user movement in advance. Nevertheless, dealing with human movement and the properties that it encompasses helps us circumvent such a problem. Indeed, 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 6, 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.

5 Benchmark approach

The solutions presented in the previous section are based on social network properties and require only a short-term observation of the network (i.e., a short training period of one week). In order to evaluate the efficiency of our strategies, we propose a benchmark approach that gives the optimal solution: 100% of user coverage daily, with minimum number of VIPs. It is important



(a) Encounter example. (b) Rule 1 in graph G. (c) Rule 2 in graph G.

(d) Rule 3 in graph G. (e) Final graph G.

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.

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. In this section, we describe such abstraction as well as the benchmark delegate selection.

5.1 Application scenario's abstraction

Suppose 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 above rules represents users' behavior in the network during the whole period P. Take for example 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 on which user u is active. As a consequence, all members of the $u's$ clique in G point to the same members of *other users' cliques*. Thus, any image of user u in G (any member of $u's$ clique) is enough to determine the set of users v for which u can be a delegate, and on which days.

5.2 Benchmark delegates selection

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 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 $i \in P$.*

Proof. First recall that according to Rule 2, images of a same user form a clique in G. Since *MDS* is minimum, it contains *at most* one image for every user u. When members of such a set are promoted to delegates, we get at most one delegate-instance per user.

Suppose, without loss of generality, that user v is active during day i, i.e., vertex v_i is present in graph G. As *MDS* is a dominating set, either of the following cases might happen: (i) some image v_t of v is in *MDS* or (ii) there is at least one other node u's image $u_l \in MDS$ such that the edge (u_l, v_i) is in G. In case (i), since $v_t \in MDS$, v is promoted to a delegate and is covered by itself. Case (ii) can only happen if edge (u_l, v_i) was added by Rule 3, i.e., u and v met during day i. Given that $u_l \in MDS$, u is promoted to a delegate. Thus, v is necessarily covered on day i. □

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, $|MDS| \leq |S|$.*

Proof. Suppose, on the contrary, that $|S| < |MDS|$. By construction, and with a similar reasoning used in the proof of Theorem 1 it's easy to see that S is an out-dominating set of G. Then, by the minimum cardinality of *MDS*, we are done. □

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 VIP delegates.

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 [32]) for which a simple greedy algorithm is known to be only a logarithmic approximation factor away from the optimum [32]. The delegates obtained by this heuristic are then used as benchmark VIPs in our experiments.

Table 1: Details on the datasets and respective training period.

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

6 Experimental setting: From data-sets to social graphs

In this section, we give detailed information on the data-sets we use for evaluating our strategies. For our experiments, we use both real and synthetic data-sets. Then, we present, step by step, how the social graph is created from each of the sets (Section 6.3) and how the social communities necessary for the *hood VIPs* selection strategies are computed (Section 6.4).

6.1 Real data-sets

For the evaluation we use two real data-sets: Dartmouth [35] (movement of students and staff in a college campus) and Taxis [36] (movement of taxi cabs in San Francisco). 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 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 [37]. We consider activities from the 5th of January to the 6th of March 2004, corresponding to a 2-month period during which the academic campus life is reasonably consistent. We chose to work with the set of nodes that have a fairly stable activity in time: at least 500 contacts per week with *any* other device. This results in a set of 1,146 nodes with an average of 1,060 daily active devices and 292 daily contacts in average per device.

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 (following suggestions of Piorkowski et al. [36]). This yields an average of 491 active nodes per day and 7,804 daily contacts per node.

6.2 Synthetic data-sets

Synthetic traces are generated using the SWIM model [38], shown to simulate well human mobility from statistical and social points of view. It also predicts well the performance of protocols in various scenarios (e.g., conference and university campus) and is shown to scale.

We use SWIM to simulate a 500-node version of the Cambridge Campus data-set (of only 36 Bluetooth enabled iMotes, 11 days long) [38]. We call this trace SWIM-500. It simulates user activity during 2 months, yielding 128 daily contacts per node in average. Then, we scale up to 1,500 nodes in two ways: (i) by keeping density constant (D-SWIM-1500) and (2) by keeping the area constant (A-SWIM-1500). The purpose of the two different scalings is to study the behavior of our strategies in two scenarios: D-SWIM-1500 simulates an urban growing in both area and population and A-SWIM-1500 refers to a sudden over-population of a given city.

Table 1 summarizes the details of the data-sets. Note that, although both data-sets represent campus scenarios, they yield different activity per node per day as they used distinct technologies (Wi-Fi capable APs in Dartmouth and Bluetooth-like characteristics in SWIM) in the two data-sets.

6.3 Training period and social graph

As we already discussed in Section 4.4, 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 among them. The length of the training period is not casual. Usually, 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). Such repetition also infers the common meetings generated by those activities.

In the case of the Taxi data-set, the repeatability of contacts is due to several factors including the popularity of geographical zones in the city (e.g., center, stations, and airports), the fixed tracks leading to such zones, and the common city areas covered by groups of taxis. As shown by Piorkowski et al., popularity of areas generates clusters of connectivity among cabs [36]. Taxis' movements are guided by clients' (humans) necessity to reach a specific geographic location. Thus, one week observation is again enough to predict future meetings.

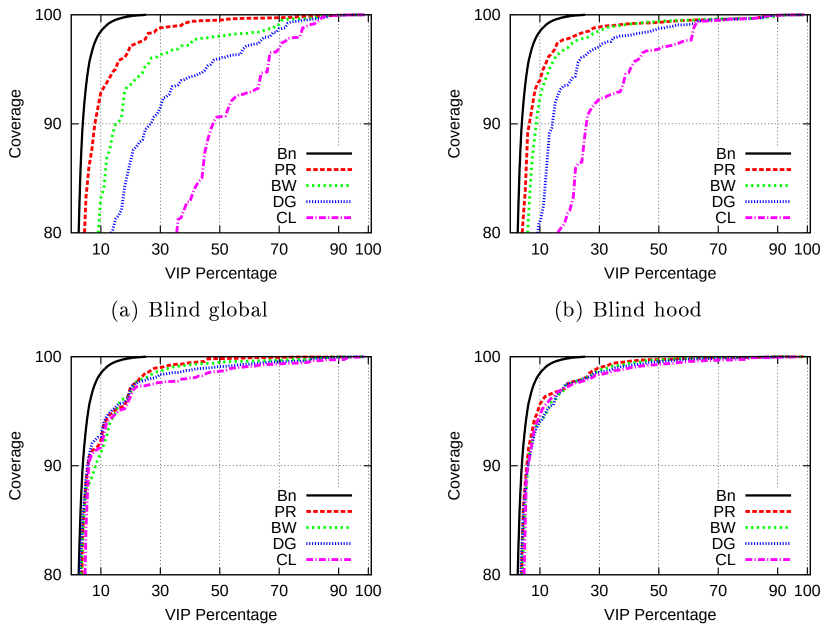
Our intuition on the length of the training period is also confirmed by the results shown in Table 1. Indeed, the properties of the training period are very close to the whole trace, for each considered scenario. Hence, this makes prediction of future meetings easy: the training period we have chosen allows us characterizing social relationships among users. 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 in mostly due to the frequentation of the same classes, or studying in the same library, or living in the same dormitory. All these activities generate lots of meetings between people during the week. We thus set the social connectivity threshold in this case to be 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 were considered.

- As the SWIM-500 trace also represents a University campus, we use the same social connectivity threshold of the Dartmouth trace: at least 1 contact per day for at least 5 days of the training week. This leads to a set of 498 nodes. When scaling up with constant density (D-SWIM-1500) the social connectivity threshold remains constant. It increases to at least 8 contacts per day for at least 5 days of the training week when scaling up with constant area (A-SWIM-1500).

The social graphs generated by the social connectivity thresholds are then used to individuate the VIP delegates, according to each of the strategies described in Section 4.



(a) Blind global (b) Blind hood

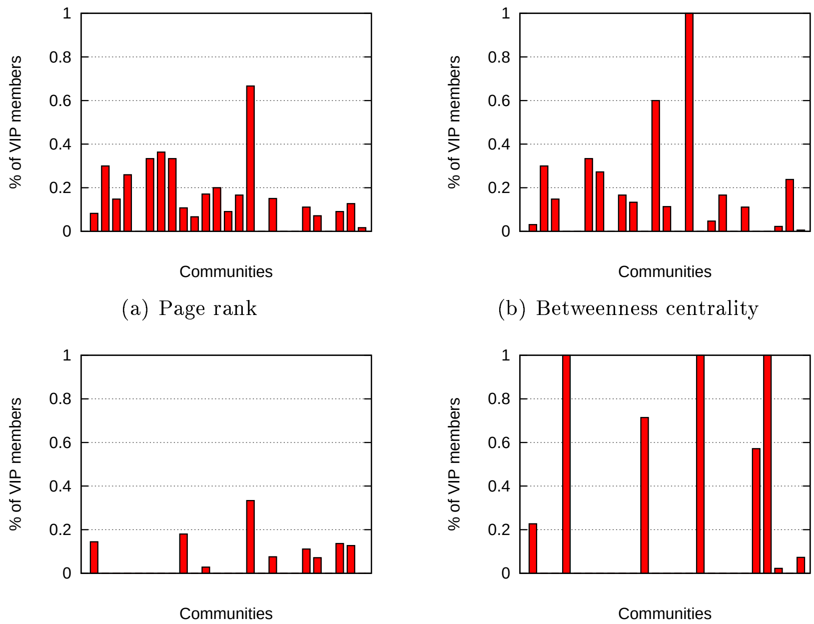
(c) Greedy global (d) Greedy hood

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.

6.4 Community detection

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 through the k-clique algorithm [29], which is widely used in the area of social mobile networking [20].

With respect to the campus-like scenarios, we have the following parameters: the Dartmouth data set has 24 communities of 41 members in average, the SWIM-500 trace has 16 communities with 32 members in average, the D-SWIM-1500 trace has 39 communities with 39.6 members in average, and the A-SWIM-



(a) Page rank (b) Betweenness centrality

(c) Degree centrality (d) Closeness centrality

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.

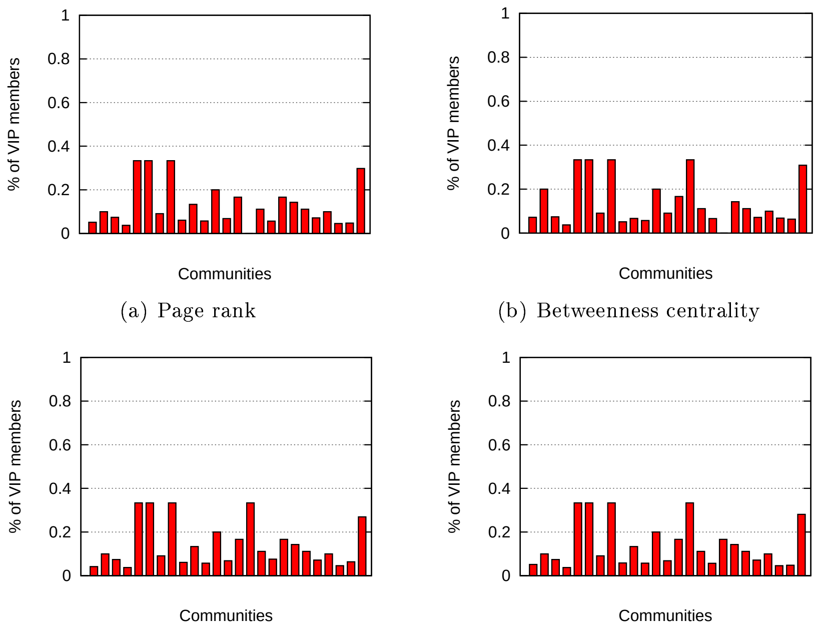
1500 has 35 communities with 44 members in average. Note that constant-area scaling yields less, bigger communities.

To study the overlapping between communities we use the Jaccard similarity index [39], which, for two sets A and B is computed as: $J_{A,B} = \frac{|A \cap B|}{|A \cup B|}$. The communities are well-knit and do not show much intersection between them. Indeed, the average Jaccard similarity index between intersecting communities is 0.038 in the Dartmouth case and about 0.025 in SWIM-500 and D-SWIM-1500 case. This result supports recent findings on universities' communities detected with the k-clique algorithm [20]. Conversely, in the constant-area scaling of A-SWIM-1500, the communities have a higher overlapping: the Jaccard similarity index in this case is 0.045.

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.

7 Experimental Results

We analyze the performance of all our strategies in terms of coverage when applied to real and synthetic traces (see, respectively, Sections 7.1, 7.2, and 7.3).



(a) Page rank

(b) Betweenness centrality

(c) Degree centrality

(d) Closeness centrality

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

For better understanding the quality of the VIPs selected by each strategy, we investigate the coverage trend with regard to an increasing number 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 nodes to VIPs.

For the page rank attribute we noted that, varying the damping factor in the interval $[0.51; 0.99]$ does not change the performance of page rank VIPs with respect to the VIPs selected according to other centralities. However, we decided to use $d = 0.85$, since, for page rank, it results in the best performance in terms of network coverage.

7.1 Performance results on real data-sets: Dartmouth case

Blind promotion. We show in Fig. 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 Fig. 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

Table 2: VIP sets cardinality to get 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

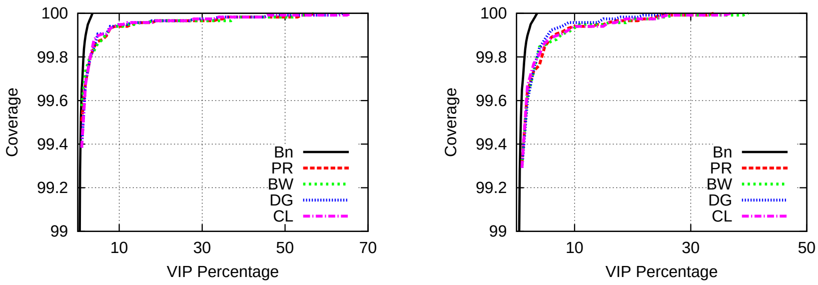
delegates considered. For instance, in the *global blind strategy*, to get to 90% of coverage, page rank only requires the promotion of 5.93% of nodes as delegates against 3.92% with the benchmark approach (see Table 2).

Another consideration to be made is that hood selection is more effective than global selection. Hence, aiming to 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. 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. In Fig. 3, we show how the global strategy distributes VIPs among communities of different centralities.

Greedy promotion. When applying the *greedy promotion*, the performance of all strategies improves considerably and gets much closer to the benchmark (see Fig. 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 Fig. 4, where we show how the greedy strategy distributes VIPs among communities for different social attributes.

7.2 Performance results on real data-sets: Taxi case

As discussed in Section 6.4, the community sub-structuring of the Taxi data-set is flat. This means that, due to the high mobility of nodes, a huge unique community containing 80% of nodes is detected and the 20% remaining nodes do not belong to any community. Thus, only the global selection strategy is applicable to this data-set. Fig. 5 shows the performance of blind and greedy global selection strategies in terms of coverage for the Taxi data-set. As we can see, due to the high mobility of nodes, all strategies perform very well in this scenario. Moreover, the sets selected by each strategy to guarantee up to 90% of



(a) Blind global (b) Greedy global

Figure 5: Performance of blind global and greedy global selection strategies on the **Taxi** 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.

coverage are exactly of the same (small) size: Only 0.93% of network nodes. The benchmark guarantees the same coverage with 0.2% of network nodes selected.

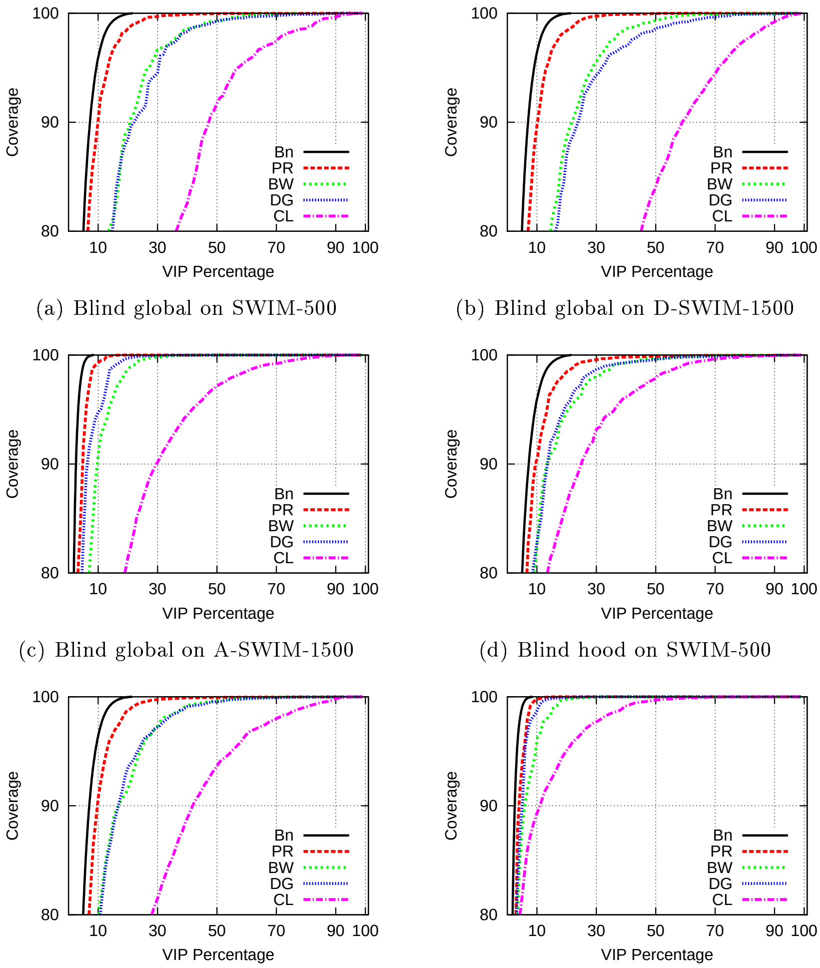
7.3 Performance results on synthetic data-sets: SWIM

We now present the results for the SWIM mobility model. As discussed in Section 6.2, starting from SWIM-500 (a 500-node simulation of the Cambridge University scenario [40]), we generate two scaled versions with 1,500 nodes: D-SWIM-1500 (constant density scaling) and A-SWIM-1500 (constant area scaling). Our purpose is to study the reaction of the different strategies in two cases: an urban growing in both area and population (constant density) and a sudden over-population of a city (constant area).

We start from *blind promotion* (see Fig. 6). First notice that, again, like in the Dartmouth scenario, page rank VIPs are more efficient than VIPs of other centralities. The reason is the same as discussed in the previous section, i.e., page rank global VIPs are better distributed within communities with respect to VIPs obtained with other centralities. This is also confirmed by Fig. 8 where such distribution is shown for the trace D-SWIM-1500 (the relative figures for traces SWIM-500 and A-SWIM-1500 are omitted due to space constraints). Once again, aiming to cover the network by forcing VIPs to fall in different communities (hood selection) is a winning strategy.

Results related to *greedy promotion* are presented in Fig. 7. As in the Dartmouth case, the performance of all strategies is boosted up by the greedy selection of VIPs, yielding a better distribution of delegates within communities (see Fig. 9) and thus, much better coverage results with respect to the *blind promotion*.

What is interesting to notice here is the impact of the way of scaling in our strategies. When passing from SWIM-500 to D-SWIM-1500 (constant density), all strategies perform very similarly in both blind and greedy promotions. 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 (see Fig. 6(c), 6(f), 7(c), and 7(f)). This is also confirmed by the results shown in Tables 3, 4, and 5 that contain,



(a) Blind global on SWIM-500

(b) Blind global on D-SWIM-1500

(c) Blind global on A-SWIM-1500

(d) Blind hood on SWIM-500

(e) Blind hood on D-SWIM-1500

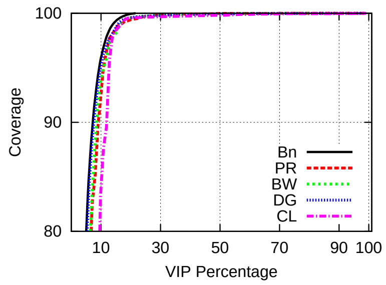
(f) Blind hood on A-SWIM-1500

Figure 6: Performance of (a)-(c) blind global and (d)-(f) blind hood selection on SWIM. "Bn" refers to the benchmark, "PR" to the page-rank, "BW" to betweenness centrality, "DG" to degree centrality, and "CL" to closeness centrality.

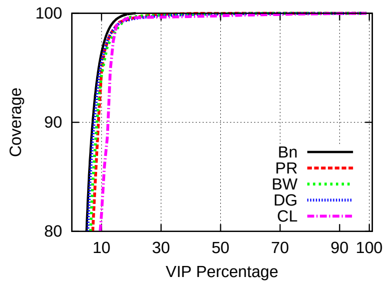
Table 3: Delegates given by each strategy to get 90% coverage on SWIM-500. The benchmark approach needs 7.4%.

	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

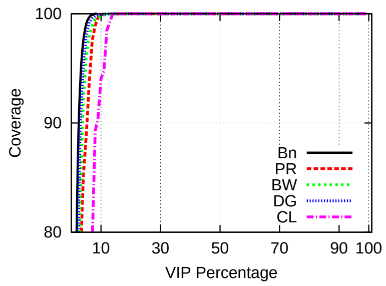
for each data-set, the percentage of delegates needed by the different strategies to cover 90% of the network.



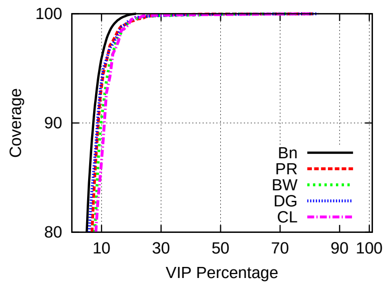
(a) Greedy global on SWIM-500



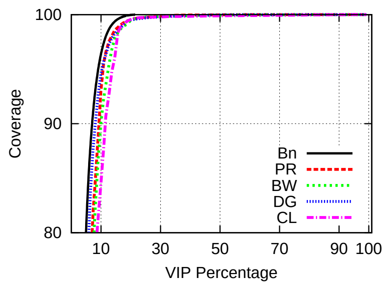
(b) Greedy global on D-SWIM-1500



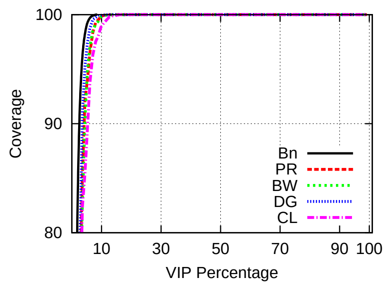
(c) Greedy global on A-SWIM-1500



(d) Greedy hood on SWIM-500



(e) Greedy hood on D-SWIM-1500

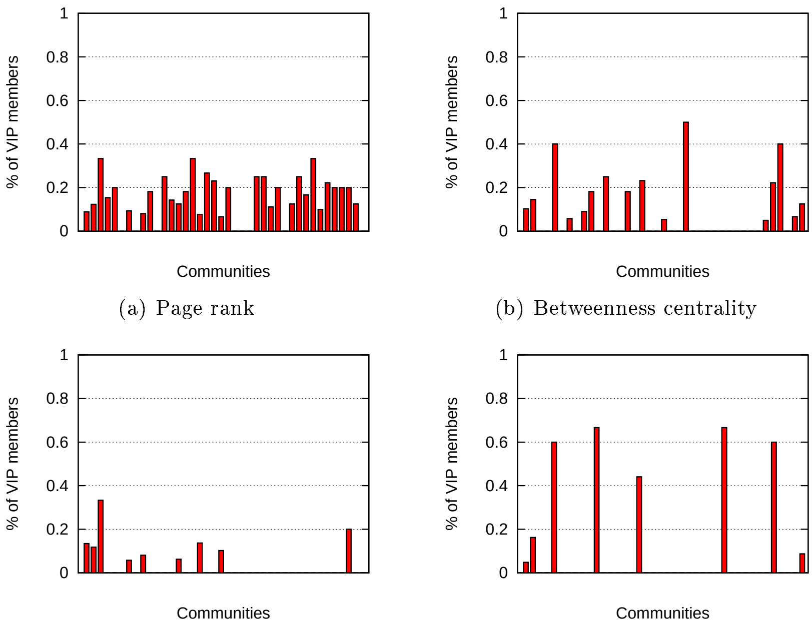


(f) Greedy hood on A-SWIM-1500

Figure 7: Performance of (a)-(c) greedy global and (d)-(f) greedy hood selection on SWIM. "Bn" refers to the benchmark, "PR" to the page-rank, "BW" to betweenness centrality, "DG" to degree centrality, and "CL" to closeness centrality.

Table 4: VIP sets cardinality 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



(a) Page rank (b) Betweenness centrality

(c) Degree centrality (d) Closeness centrality

Figure 8: Distribution of VIPs per social attribute on the **D-SWIM-1500** dataset with the *blind global* promotion strategy. The x-axis represents different communities detected.

Table 5: VIP sets cardinality 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

7.4 Coverage potential

To complete our study, we investigate the *coverage potential* of the first 10% of nodes promoted to delegates according to each strategy. To this end, we measure, for each delegate, the ratio of non-delegates neighbors on the social graph (i.e., the number of non-delegates neighbors of delegate i over the total number of neighbors in the social graph). We average then the result over the set of all delegates chosen by the corresponding strategy. Intuitively, the bigger this value, the larger the coverage potential of the strategy, and vice-versa. In Tables 6–8 we present the results for every strategy/social attribute for respectively the Dartmouth, Taxi and the D-SWIM-1500 trace. Because of space constraints we omit the tables related to SWIM-500 and A-SWIM-1500.

Note that in the *global blind* selection, page-rank is the one with the highest value, followed by betweenness, degree, and finally closeness. This again sup-

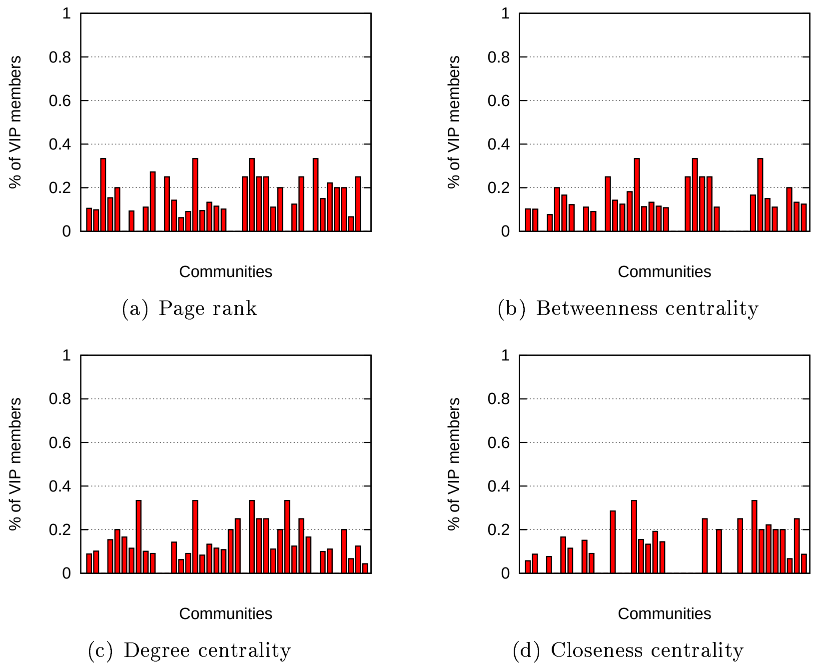


Figure 9: Distribution of VIPs per social attribute on the **D-SWIM-1500** dataset with the *greedy global* promotion strategy. The x-axis represents different communities detected.

Table 6: Coverage potential for each strategy on Dartmouth. The Benchmark's potential is 0.91.

	G-Blind	H-Blind	G-Greedy	H-Greedy
PR	0.830	0.069	1.0	0.831
BW	0.657	0.067	1.0	0.969
DG	0.530	0.061	1.0	0.890
CL	0.144	0.059	1.0	0.886

Table 7: Coverage potential for each strategy on TAXI. The Benchmark's potential is 0.96.

	G-Blind	G-Greedy
PR	0.760	1.0
BW	0.758	1.0
DG	0.742	1.0
CL	0.745	1.0

ports the results of Fig. 2(a). The potential falls drastically when considering the *hood blind* selection (second column of Table 6): delegates are forced to be in the same community, in a blind way. Thus, with high probability, they are socially connected with each other. However, page rank remains the attribute with the highest value, supporting the results of Fig. 2(b).

Table 8: Coverage potential for each strategy on D-SWIM-1500. The Benchmark's potential is 0.99.

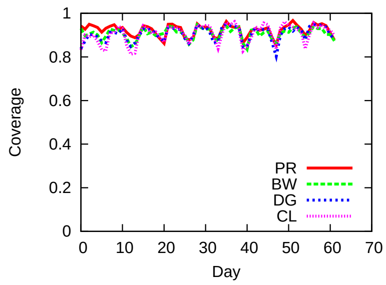
	G-Blind	H-Blind	G-Greedy	H-Greedy
PR	0.888	0.558	1.0	0.992
BW	0.688	0.451	1.0	0.993
DG	0.607	0.288	1.0	0.994
CL	0.180	0.094	1.0	0.993

The *global greedy* selection naturally yields the highest coverage potential for every attribute: after each node promotion its social connected neighbors are eliminated from the graph, thus, the ratio of non-delegates neighbors of a node is 1. The *hood greedy* selection (fourth column of Tables 6 and 8) yields lower results with respect to the global greedy one. This is because selection is done on a community basis, and *only* community neighbors are eliminated after promotion. Since communities are not totally distinct, it might happen that two VIP neighbors in the social graph belong to different communities and, consequently, are eliminated after the promotion, decreasing thus the coverage potential of the strategy. This effect is smaller for high betweenness nodes: they tend to belong to the same group of communities (the ones that they connect). Closeness/degree attributes suffer less from this effect, as they select nodes that are central to communities. Finally, page rank is the one that suffers most: high page rank nodes are well distributed within the community to which they belong (being communities well-knit). Thus, they are more likely to have high page rank neighbors belonging to other communities (that the hood greedy selection does not eliminate).

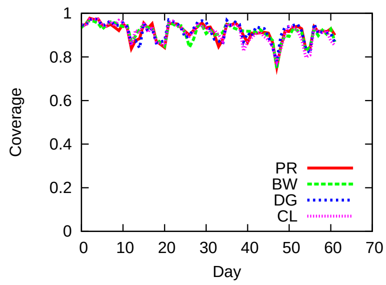
It is worth to note that the coverage potential just gives a hint on the real coverage power of a method: It does not affect the real ability of the selection method/attribute in covering the network. Indeed, for all traces (see Tables 6, 7, and 8) the coverage power of the benchmark in all traces is less than all the values related to the global greedy selection. Regardless of the coverage potential, the benchmark performs better with respect to every strategy. In addition, the results with 90% coverage presented in Tables 2–5, confirm page ranks's high performance ability when combined with every strategy.

7.5 Coverage stability

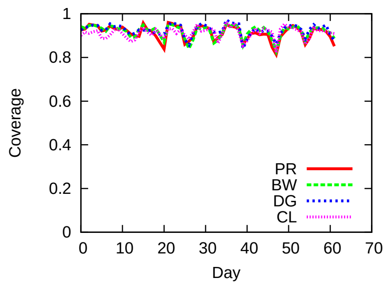
Finally, we investigate the stability of coverage of our strategies in time. We focus on the delegates set needed to reach, in average, 90% coverage for each strategy on all tracers. In Fig. 10 and 11, we plot the coverage per day. Due to the lack of space, we only present results related to the Dartmouth and D-SWIM-1500 data sets. We stress however that the results are similar also for the omitted traces. 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.



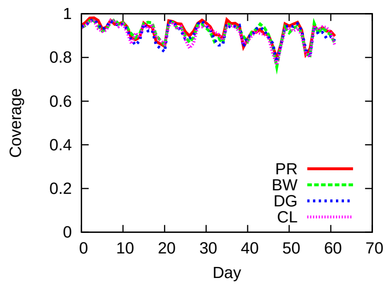
(a) Blind global



(b) Greedy global

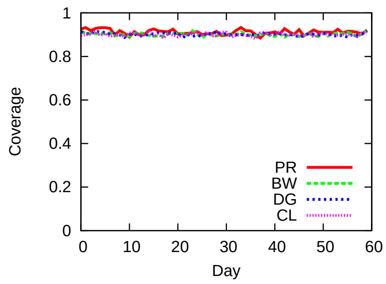


(c) Blind hood

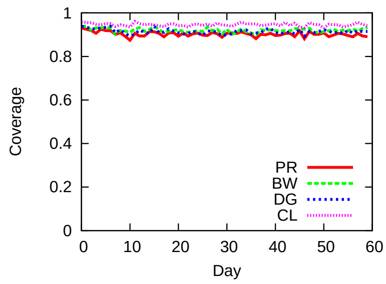


(d) Greedy hood

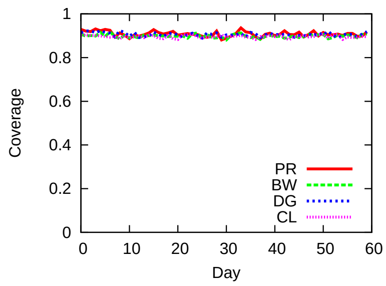
Figure 10: Coverage stability in time for all strategies (**Dartmouth** data-set).



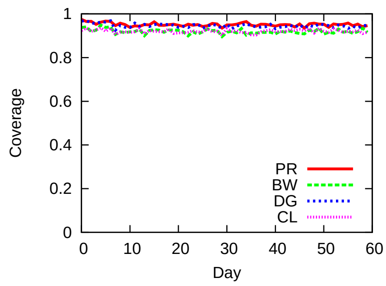
(a) Blind global



(b) Greedy global



(c) Blind hood



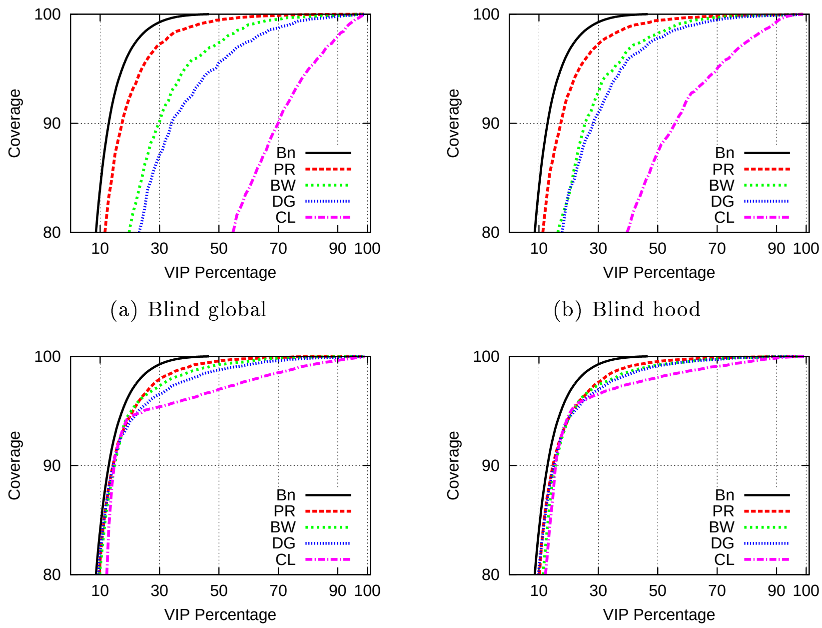
(d) Greedy hood

Figure 11: Coverage stability in time for all strategies (**D-SWIM-1500** data-set).

Table 9: VIP sets cardinality to get 90% coverage on D-SWIM-1500. The benchmark approach needs 13%. Half-day coverage interval.

	G-Blind (%)	H-Blind (%)	G-Greedy (%)	H-Greedy (%)
BW	30	25.53	15	15.63
CL	70	56.33	15	15.69
DG	34	28.66	15	15.65
PR	18	17.46	15	15.6

7.6 Coverage intervals VS VIPs

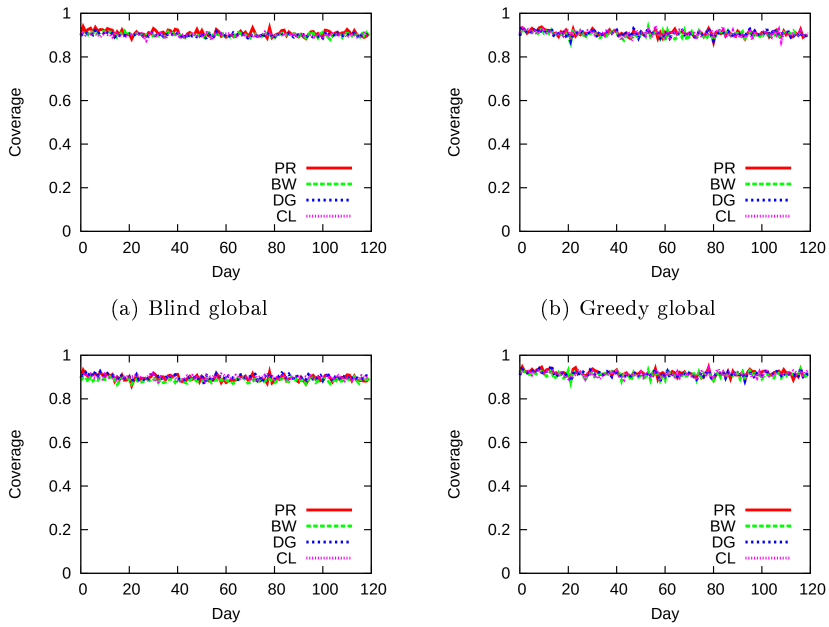


(a) Blind global (b) Blind hood

(c) Greedy global (d) Greedy hood

Figure 12: Performance of the selection strategies on the **D-SWIM-1500** dataset. “Bn” refers to the benchmark, “PR” to the page-rank, “BW” to betweenness centrality, “DG” to degree centrality, and “CL” to closeness centrality. Half-day coverage interval.

The VIPs selected by our strategies are expected to cover all nodes every day of the data-trace by carrying data traffic from/to the users. Many applications could benefit by these 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. But what happens for applications that require coverage intervals different from 1-day? How does the length of the coverage interval impact the selection of VIPs? Clearly, if the coverage interval is longer, the 1-day coverage VIPs are a superset of the required number of dele-



(a) Blind global (b) Greedy global

(c) Blind hood (d) Greedy hood

Figure 13: Coverage stability in time for all strategies (**D-SWIM-1500** dataset). Half-day coverage interval.

gates. Indeed, for a coverage interval long e.g. 2 days, the 1-day coverage VIPs would perform as good as in the 2-day coverage case. However, if the coverage interval required is smaller, the VIP set required to cover the network is likely to change. To quantify such change we have also studied the half-day network coverage for all the traces. For both the Taxi and the A-SWIM-1500 traces we noted absolutely no difference from the 1-day coverage case. We believe this is due to the high mixing and speed of movement of cabs in the Taxi case, and due to the high node density in the A-SWIM-1500 trace. Recall that A-SWIM-1500 is obtained scaling SWIM-500 with constant area.

In Dartmouth, SWIM-500 and D-SWIM-1500 we observed a growth in the VIPs number required to assure 90% of network coverage (due to lack of space here we only show results related to the largest trace: D-SWIM-1500. However, we stress that both the Dartmouth and SWIM-500 traces yield similar results). Intuitively, this is because the meeting patterns of the first half of the day are different from those of the second half. The growth on the number of delegates required to cover 90% of the network in the half-day coverage case is also reflected in the benchmark's VIPs, which are almost doubled with respect to the 1-day coverage interval case (see Table 9). Hence, one would expect that the same should happen also with the VIP sets selected by our strategies. However, from the comparison of the 1-day coverage interval results of Table 4 with the half-day coverage interval results of Table 9 we note that the VIP sets have increased of about 60%. This again means that selecting VIPs according to their "importance" in the network is a good strategy: Indeed, most of the important people during the morning remain so also during evening. However, the coverage interval length indeed does impact the cardinality of VIPs. This suggests for application developers or network infrastructure builders to trade off between data transfer frequency and number of VIPs.

To conclude, Fig. 12 and 13 show respectively, the coverage trend, and the performance of the different selection strategies for the case of half-day interval coverage. Again we note that page-rank wins over the other centralities, and the hood selection strategy wins over the global selection one. Finally, Fig. 13 confirms the stability in time of our VIPs, regardless of the length of the coverage interval.

8 Discussion

The VIPs selected by our strategies are expected to cover all nodes every day of the data-trace by carrying data traffic from/to the users. Even though many situations require immediate connectivity (e.g., status update on social-networks), it is reasonable to assume that many others applications are more tolerant to delay. Example of such applications are: collection of urban-sensing related data [4, 5]; 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. Traffic corresponding to these applications requires VIPs to have sufficient resources in both battery and memory. To account for this issue, two propositions could be considered: (i) load distribution mechanisms could be applied at the delegates selection and/or (ii) nodes selected as delegates could have their devices upgraded.

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 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.

At the first proposition and in the case of data collection applications, the idea is to estimate the future potential traffic load generated by nodes during the training period of one week. This traffic load could be then used to establish a maximum load threshold per delegate and, accordingly, combined with the social attributes for delegate selection. In this way, the number of contacts per delegate could be limited by nodes' traffic volume. The load consideration during the delegate selection could then allow the selection of additional delegates to collect data from the yet uncovered nodes. In this way, collection fairness and resource constraint among delegates could be considered.

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 [41–43] 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.

Another consideration to be made on our strategies is the following: "How to handle the traffic of 10% of network users that remain uncovered by our delegates?" As discussed above, the coverage of 90% of nodes requires the promotion as delegates of very few and constant in time network members. This confirms the advantage of our opportunistic delegation approach for covering a high percentage of nodes in a daily basis. Additionally, we claim that the impact of the few 10% non-covered nodes on the 3G network will be small and typically generated by nodes that are marginal to the network (e.g., people frequenting peripheral areas of a city). Usually, nodes having a high activity or mostly visiting central areas in the network will be represented in the constructed social graph, stressing their frequent encounters. In this way, we believe that such more active nodes will be mostly responsible for the traffic overloading previously mentioned and will be covered by the selected delegates, with high probability. Therefore, to answer the previous consideration, the few remaining uncovered nodes could directly transfer their data using 3G cellular networks, at the end of the day, once no delegate visit was detected.

9 Conclusion

The mobile data-traffic generated by the proliferation of smart-phone devices is overloading the cellular network in highly dense, metropolitan areas. The problem, very likely to grow worse in the future, is forcing network providers to come up with new efficient offloading techniques. In this paper, we describe VIP delegation, a mechanism to alleviate the mobile 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 several well known attributes such as betweenness, closeness, degree centrality and page-rank. Each attribute is combined to two main selection 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. Indeed, our 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.

Lastly, we also compare our methods with an optimal benchmark VIP set, computed from the full knowledge of the system (i.e., based on past, present, and future contacts among nodes). The performance is very close to the benchmarks', for all the scenarios considered.

References

[1] "Canalys report," http://www.canalys.com/pr/2010/r2010081.html, 2010.

[2] N. Eagle and A. Pentland, "Social serendipity: Mobilizing social software," *IEEE Pervasive Computing*, vol. 4, no. 2, pp. 28–34, 2005.

[3] S. Gaonkar, J. Li, R. Choudhury, L. Cox, and A. Schmidt, "Micro-blog: sharing and querying content through mobile phones and social participation," in *ACM MobiSys*, 2008.

[4] 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.

[5] S. B. Eisenman, N. D. Lane, E. Miluzzo, R. A. Peterson, G. Ahn, and A. T. Campbell, "Metrosense project: People-centric sensisng at scale," in *ACM SenSys*, 2006.

[6] S. Ioannidis, A. Chaintreau, and L. Massoulie, "Optimal and scalable distribution of content updates over a mobile social network," in *IEEE Infocom*, 2009.

[7] L. McNamara, C. Mascolo, and L. Capra, "Media sharing based on colocation prediction in urban transport," in *ACM Mobicom*, 2008.

[8] "Customers Angered as iPhones Overload AT&T," NY Times, http://www.nytimes.com/2009/09/03/technology/companies/03att.html, September 2009.

[9] "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.

[10] "AT&T Beefing Up Network for iPad and iPhone," Reuters, http://www.reuters.com/article/idUS182677699120100130, Jan. 2010.

[11] "Limited 3G bandwidth is good for mobile apps," Infoworld, http://www.infoworld.com/d/mobilize/limited-3g-bandwidth-good-mobile-apps-676, October 2010.

[12] "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.

[13] "Mobile Data Offload for 3G Networks," White Paper, IntelliNet Technologies, Inc., 2009.

[14] V. Chandrasekhar, J. Andrews, and A. Gatherer, "Femtocell networks: a survey," *IEEE Communications Magazine*, vol. 46, no. 9, pp. 59 –67, Sep. 2008.

[15] "AT&T, Verizon Wireless join Wi-Fi interoperability group," Cnet, http://news.cnet.com/8301-30686_3-20008476-266.html, Jun. 2010.

[16] A. Balasubramanian, R. Mahajan, and A. Venkataramani, "Augmenting mobile 3g using wifi," in *ACM MobiSys*, 2010.

[17] "Cisco visual networking index forecast predicts continued mobile data traffic surge," http://newsroom.cisco.com/dlls/2010/prod_020910b.html, February 2010.

[18] 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.

[19] E. M. Daly and M. Haahr, "Social network analysis for routing in disconnected delay-tolerant manets," in *ACM MobiHoc*, 2007.

[20] 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.

[21] A. Mtibaa, M. May, C. Diot, and M. Ammar, "Peoplerank: Social opportunistic forwarding," in *IEEE Infocom*, 2010.

[22] F. Li and J. Wu, "LocalCom: a community-based epidemic forwarding scheme in disruption-tolerant networks," in *IEEE SECON*, 2009.

[23] W. Gao, Q. Li, B. Zhao, and G. Cao, "Multicasting in delay tolerant networks: a social network perspective," in *ACM MobiHoc*, 2009.

[24] Q. Li, S. Zhu, and G. Cao, "Routing in socially selfish delay tolerant networks," in *IEEE Infocom*, 2010.

[25] A. Mei and J. Stefa, "Give2get: Forwarding in social mobile wireless networks of selfish individuals," in *ICDCS*, 2010.

[26] M. Conti, R. Di Pietro, A. Gabrielli, L. V. Mancini, and A. Mei, "The smallville effect: Social ties make mobile networks more secure against the node capture attack." in *ACM MobiWac*, 2010.

[27] 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.

[28] 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.

[29] 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.

[30] M. Gonzalez, C. Hidalgo, and A. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, pp. 779–782, 2008.

[31] C. Boldrini, M. Conti, and A. Passarella, "The sociable traveller: human traveling patterns in social-based mobility," in *ACM MobiWac*, 2009.

[32] V. Kann, "On the approximability of np-complete optimization problems," Ph.D. dissertation, Royal Institute of Technology, Stockholm 1992.

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

[34] 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.

[35] T. Henderson, D. Kotz, I. Abyzov, and J. Yeo, "CRAWDAD trace dartmouth/campus/movement/01_04 (v. 2005–03–08)," Downloaded from http://crawdad.cs.dartmouth.edu/dartmouth/campus/movement/01_04, Mar. 2005.

[36] M. Piorkowski, N. S.-Djukic, and M. Grossglauser, "A parsimonious model of mobile partitioned networks with clustering," in *COMSNETS*, 2009.

[37] 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.

[38] A. Mei and J. Stefa, "SWIM: A simple model to generate small mobile worlds," in *IEEE Infocom*, 2009.

[39] 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.

[40] J. Leguay, A. Lindgren, J. Scott, T. Riedman, J. Crowcroft, and P. Hui, "CRAWDAD trace upmc/content/imote/cambridge (v. 2006–11–17)," Downloaded from http://crawdad.cs.dartmouth.edu/upmc/content/imote/cambridge, Nov. 2006.

[41] CENS Urban sensing, http://urban.cens.ucla.edu/projects.

[42] MIT Senseable City Lab, http://senseable.mit.edu.

[43] EARSeL Special Interest Groups, http://www.earsel.org/?target=SIGs.

Contents

Centre de recherche INRIA Saclay – Île-de-France
Parc Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 Orsay Cedex (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399