



Data Offloading via Optimal Target Set Selection in Opportunistic Networks

Prince Sharma¹ · Shailendra Shukla² · Amol Vasudeva¹

Accepted: 6 April 2021 / Published online: 29 April 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

The rapid rate of dependence over internet usage using digital devices also results in enormous data traffic. The conventional way to handle these services is to increase the infrastructure. However, it results in high cost of implementation. Therefore, to overcome the data burden, researchers have come up with data offloading schemes using solutions for NP-hard Target Set Selection (TSS) problem. Our work focuses on TSS optimization and respective data offloading scheme. We propose a heuristics-based optimal TSS algorithm, a distinctive community identification algorithm, and an opportunistic data offloading algorithm. The proposed scheme has an overall polynomial time complexity of the order $O(k^3)$, where k is the number of nodes in the primary target set for convergence. However we have obtained its realization to linear order for practical reasons. To validate our results, we have used state-of-the-art datasets and compared it with literature-based approaches. Our analysis shows that the proposed Final Target Set Selection (FTSS) algorithm outperforms the greedy approach by 35% in terms of traffic over cellular towers. It reduces the traffic by 20% as compared to the heuristic approach. It has 23% less average latency in comparison to the community-based algorithm.

Keywords Mobile data offloading · Target set selection · Ad-hoc networks · Overlapping communities · Opportunistic communications · Data forwarding

1 Introduction

The rise in use of mobile phones, tablets, Personal Digital Assistants (PDAs), and internet based applications for services like news, online video streaming, gaming, and massive file transmissions has shown significant growth in internet traffic. The dependence over mobile internet data has been exponentially rising due to its widespread significance and frequent usage. The pandemic

like situations of COVID-19 have further added to this dependence. The conventional way to handle these internet-based services is to rely on external infrastructures. However, the current demand for data, if equipped with these devices, would incur a very high cost of implementation. It requires a high financial investment, long process of development, low return value, and a high maintenance cost. There is an exponentially rising need for attention for mobile data traffic and its offloading alternatives.

As per the white paper report from [1], nearly 15% of the global population of internet users is expected to rise by the year 2023 than from the year 2018. The number of devices connected to IP networks is likely to be three times its global population. More than 70% of the global population is expected to have mobile connectivity with about four times rise in machineto-machine connections. The analysis forecasts that the global mobile data traffic would rise to five times to reach nearly 164EB per month by the year 2025 [2]. In order to reduce the traffic load corresponding to one junction, an optimal solution is to use ad-hoc networks. For such mobile data offloading, WiFi and Bluetooth technologies have been described as a promising

✉ Prince Sharma
prince.sharma@mail.juit.ac.in

Shailendra Shukla
ss@mnnit.ac.in

Amol Vasudeva
amol.vasudeva@juit.ac.in

¹ Jaypee University of Information Technology, Wakanaghat, Solan, Himachal Pradesh, PIN 173234, India

² Motilal Nehru National Institute of Technology, Allahabad, Uttar Pradesh, PIN 211004, India

solution [3]. Opportunistic communication [4] through WiFi based common hotspots serve similar objectives.

To identify offloading structures for such optimizations, the static opportunistic contact analysis is done [5–11]. However, the significance of continuously evolving networks is yet to be identified. The literature suggests comparative analysis of greedy, heuristic, and random approaches in [8, 12] for static target sets. It suffers the drawback of being unrealistic due to the dynamic behavior of the network graph. The majority of research work is focused on the identification of users, who behave similarly on some similar grounds. It is achieved through heuristics based TSS [13]. TSS is the procedure of selecting a limited set of nodes, which can share the duplicate data, which is otherwise meant for each node via access points. Such a solution is NP-hard [13] in order to approximate minimum or maximum participants for set identification. The associations are static and independent of time constraints, and the majority of the literature suggests modeling for fixed networks, which does not vary with time. However, we need to consider the opportunistic associations concerning time constraints.

We intend to optimize the answer to this NP-hard TSS problem in this paper and suggest the identification of an optimized target set to propose access point level offloading. Therefore, TSS detection can be activated to minimize the final set to give its customers cellular network offloading at different times as against [14]. As illustrated in Fig. 1, we focus to target a limited set of users within the range of the same access point. Unfortunately, the access point has to deliver similar content frequently to many users, which renders a bottleneck at a superior level and lays the foundation for data offloading needs. The problem is simplified by offering a differentiation technique of data offloading. The proposed solution includes optimization of a dynamic target set identification at three sub-levels. It consists of the initial selection of the target set, the portion

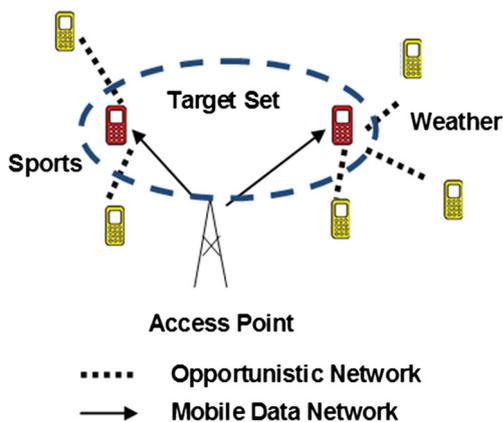


Fig. 1 Offloading target sets for different interests

of the secondary target set and the optimized selection of the target set for complex networks. Our significant contributions can be summarized as:

- We aim optimal target set using three-phase optimizations for TSS limiting the users’ overlapping communities.
- We guarantee the restricted membership of each node to minimize the possibility of conflicting groups in the event of different interests.

2 Literature survey

The majority of the literature related to mobile data offloading proposes several algorithms for feasible optimizations. However the research work is inter-related, but we try to categorize the literature survey broadly into three sub-categories. The research over data offloading is divided into the area of application, identification of data offloading parameters and study of TSS. So we review the existing literature for data offloading on its basis in the following subsections.

2.1 Types of ad-hoc connections

The most favorable strategy to migrate data traffic from cellular networks to device-to-device networks is by using Delay Tolerant Networks (DTNs). The limited capacity of the DTN based devices and the varied interest of user constraints for their limited storage has been studied in [15]. DTNs with real traces of humans and vehicles are transformed into maximization function problem using optimized 0-1 Knapsack with linear constraints. The authors have proposed a Greedy Algorithm (GA) for general scenarios, an Approximation Algorithm (AA) for shorter lifetime scenarios, and an Optimal Algorithm (OA) for homogeneous contact rates and buffer size. However, their work assumes the occurrence of contacts between any two nodes to be following Poisson rates. Thus the implementation is limited to such models related to DTN traces only. The topic of offloading maximization is also discussed in [8] using a greedy method based on heuristic documents. It requires the approximation of shorter-lived record associations. An optimal algorithm for heterogeneous contacts is proposed in it. Their work is compared with greedy and approximated algorithms in [16]. There is a proposal for an adaptive transporting algorithm based on Lyapunov optimization, which can release part of the application’s computing to a dedicated server and adapt to evolving environments as in [17]. Special consideration has been given to heterogeneous networks, making them more practical by categorizing

users as helpers and subscribers. Kempe et al. approximated the upper bound for TSS [18] and Chen determined its lower bound [10]. The authors assume the users to mutually agree to share and get the same information on sharing basis. This could reduce their load for payment and also may share their resources to avoid congestion. In the case of Vehicular Ad-hoc Networks (VANETs), these Wi-Fi-based access points are assumed to be associated with the public mode of transport [19]. The users are assumed to be belonging to a limited community at similar time instances across the day intervals. This helps them to be considered on priority for all such networks. The offloading using the Wi-Fi-based access points proposed in [7] lowers the transmission cost. It is intended to be lower than the transmission cost via cellular networks. The suggested algorithm in [9] explores the network-based interaction between nodes and tests whether connectivity exists between them. Based on this association within a common transmission range, the authors categorize a few nodes as sub-network deterministic nodes. Even after finalizing these nodes, few nodes may fall in the range of more than one deterministic node. The authors have identified such scenario of multiple communities in [20]. In such cases, the final priority of a node needs to be determined on the basis of some significant characteristics. The authors have proposed a data forwarding algorithm namely Social Attraction and Infrastructure Support (SAIS) which uses the property of social networking for implementation. The major contribution of their work is the realistic addressing of small ratio of access points in comparison to the total number of mobile users. The authors have also addressed the property of graph cliques for network realization.

2.2 Parameters of data offloading

Few authors use encounter frequency within the same classroom to consider offloading the cellular traffic [11]. They have used social community and encounter-based frequency to analyze the performance of data forwarding. The authors have proposed data forwarding on the basis of encounter, Rarest First (RF) meeting, and random strategies and compared them. The findings show the comparative analysis is outperformed by the RF algorithm. The authors use the material latency of distribution in the RF algorithm to assess the effect of a set deadline for the inclusion of sources in the group. The findings illustrate optimization based on the balance of frequency and latency of offloading. The main consideration is the expectation of regularity of human mobility to demonstrate social engagement as a key factor of assessing the opportunistic discharge of mobile data based on contact. The problem has been proved to be sub-modular followed by the application of greedy algorithms and heuristics of human mobility patterns.

However, the assumption of some content dissemination controller to decide which nodes content must be sent is also partial as in [21]

The literature includes [12] with different probabilities of system discovery to achieve the opportunistic communication of moving cell phones for short contact time. The comparative results in [9] show that the greedy algorithm is better than heuristic and random algorithms.. It derives a framework to exchange small data during short contact periods. In [5], there is a device model for the implementation of traffic offloading using motion predictions that uses the collection of different matrices to evaluate the neighboring node coverage zones and the likelihood of meeting. The likelihood of the meeting is used as the heuristic parameter for evaluating the coverage relationship in the graph-based coverage calculation. For the ad-hoc network, the network was simulated over the ns-2 network simulator. The majority of the research work considers the use of online social networks to identify the social participation and activity status using social networks like Facebook or Twitter in [21]. [22] suggests acceptable worst-case solution using tree-based transformation of static graphical network. The authors propose the Adaptive Finding Overlapping Community Structures (AFOCS) algorithm comparing it with C finder [23] and COPRA methods [24] for dynamic group detection. In order to locate the local community and then merge the overlapping populations, their initial work enforces the relationship under the basic community structure. The group optimization AFOCS algorithm causes new nodes to be inserted and removed. It also controls the addition and removal of them into the group. The authors in [25] have demonstrated the correlation between space-crossing community detection and its influence in data forwarding in mobile social networks. They have proposed a data forwarding algorithm namely Social Attraction and Access Point Spreading (SAAS) to improve data forwarding efficiency addressing the properties of delivery ratio and delay.

2.3 Target set selection problem

The issue of determining which nodes are more helpful to speed up the dissemination process has been aimed at TSS [6, 21]. Reinforcement learning based approach of actor-critic modeling is suggested for solving TSS in [21]. The results for the actor-critic approach and Derivative Re-injection to Offload Data (DROiD) [6] are compared in opportunistic networks for single community scenario and multiple community scenarios. The researchers also targeted the issue in order to define the cap on the number of nodes to be allocated to disseminate the material. In order to define the heuristic behavior and evaluate the nodes that are more useful for spreading the material, the acknowledgment message with additional information is used. It is achieved

using the learning principle of temporal difference. It uses the content distribution stage, the ratio of nodes used to the nodes available, and the percentage of time remaining until the panic zone for content delivery. In [26], VIP delegation is suggested on the basis of the social dimensions of user mobility. The authors use frequency of meetings as the significant attribute to determine the strength of social ties. The VIP promotion techniques have been classified using random and greedy approaches into Blind global promotion and Greedy global promotion techniques respectively. The authors have used the attributes of between centrality, proximity centrality, degree centrality, and page rank to determine the social strengths for VIP neighborhood selection. The authors define the value of a node in the graph. Small nodes are identified in few research studies to send the same data on the basis of certain deterministic characteristics to the maximum number of neighboring node [27, 28]. The major issue is that the access point would have to provide some equal or unequal offering for all the users irrespective of their unequal significance [8, 9, 11]. Thus incentive determination becomes significant. The major drawback of the present literature is the ignorance of continuously evolving network composition and limiting the analysis of static network-based communities and less attention to overlapping communities. This can be achieved by the collaborative effort of users along with their network service providers, which we utilized using tree-based graphs in our scenario. We have tried to merge the greedy and overlapping community approaches to the dynamic level.

3 System model and assumptions

In this section, we summarize the model and the assumptions laid with the inclusive definitions and notations explained. Our model is a hybrid of basic mobile data network along with the associated opportunistic network supporting the infrastructure based requirements for efficient mobile data offloading.

3.1 System model

We assume a network of mobile users with different interests within the range of an access point to belong to a single community initially. For simplicity, we presume all users have the same capabilities, at least the minimum of all. All users are mobile and interconnected using wireless links. Also, before its Time-to-live (TTL) expires, we have to consider that the data transmission is efficient. TTL is the data item's transmission date regardless of download or upload. Our aim is to derive minimal but optimal subset of users. Our system model comprises of a group of communities represented by the set of nodes or the

Mobile Users (MUs) for a specific service represented by edges in between. The dynamic behavior of MUs has been observed in the data set pre-processing in reference to different time instances. As illustrated in Fig. 1, since one or more nodes are connected to some other nodes in similar communities, we may try to offload the traffic to it from the access point. In this section, we summarize the model and the assumptions laid with the inclusive definitions and notations explained. Our model is a hybrid of basic mobile data network along with the associated opportunistic network supporting the infrastructure based requirements for efficient mobile data offloading.

We consider a service, such as Sports News or Weather Update, subscribed to by n users in the range of the access point and transmitting the relevant data to the whole population. Thus, the overall network load handled by the access point is calculated as the product of the number of nodes and the corresponding individual load for each node. We may need to find out the heuristic pattern of k users within the range of access point S . This helps to find out the dynamic patterns of these nodes. Another way out could be to find relations between these nodes to obtain subsets using certain characteristics of similar subscriptions. We reflect $C[i]$ as the cost of these data transmissions through the access point of the cellular network and $c[i]$ as the cost of these transmissions through Wi-Fi hotspots found in the immediate vicinity. In terms of data bytes, the expense is observed. The data record has to be kept by the cellular network's access point. By taking the ratio $c[i] : (C[i])$, the improvisation can be determined. Our aim is to identify every neighboring node j for each node i and assign the matrix attribute $I_f(i)j = 1$. The value of $I_f(i)j$ is unity if node i is in direct contact with node j , and is otherwise 0. In general, the network nodes are divided into different classes and our focus of observation is focused on a single S community based on one form of subscription. In a larger subset, this is determined as a local target set range. We refer to this as the Optimal Goal Set (OTS) derivative based on the values of SI_{n_i} , BI_{n_i} [29] and $Depth_{n_i}$. The availability of users within the connectivity spectrum of data access via Wi-Fi ensures data offloading in real life. In addition, it is also time-bound for the period for which the consumer is in the deliverable range. In addition, there is a fixed size of the content to be shared across these data access points. Using a tuple-based offloading incentive function $S'n_i = [\alpha, \beta]$. The lower bound of α is 0. The value of α defines the length of the node i within the Wi-Fi hotspot proximity range. The value of $\beta = [0/1]$ indicates the possibility of the Wi-Fi based connection for the downloading service depending on the cost of data entry and outflow. The capacity of the transmission is determined by using the speed and the time that the consumer stays connected. The final set is achieved by the heuristic greedy method. The optimal derivation S'

contains k' nodes in such a way that $k' \ll k$. The list of notations and symbols used in this paper are enlisted in the Table 1 below.

3.2 Definitions

In order to understand the system model we need to define the following terms at first:

Definition 1 Community Selection: Traditionally, the term Community is defined as a group of users who have a common belief or behavior to ensure that they are tightly knit nodes, with more internal links than external links [20]. Based on this definition, we use the term Community initially to identify users within the range of one access point. Thus $S_t = [n_1, n_2, \dots, n_k]$ is the initial community of users. We determine the sub-communities also on the basis of common user interests as in [30] to ensure strong internal links. We define a small subset S' of S nodes that could be targeted to deliver data based on short-range communications made only at the user level to the entire collection S . The optimal set and final target set are derived from it. After the sub-community determination, when one node is selected from the major superset S_t for content

delivery of interest item i , it is also termed as a Community. Thus we have used the term Community interchangeably, in reference to the interest-based subgroups for the users and significant offloading users for the access points.

Definition 2 Overlapping Community: Based on the earlier definition of Community itself, when the access point observes a user to belong to more than one user group then, then we use the term the Overlapping Community. In other words, we say if a user is selected for more than one interest item, then the communities may overlap among themselves. Considering the user u_x being interested in item i_a and for item i_b also, we have the user u_x to belong to more communities a and b , simultaneously.

3.3 Assumptions

We have assumed that all the users within the range of one access point belong to the same community. Also, the classification of all nodes in any set is on the basis of a similar category of subscribers for the same service. All nodes are ready to replicate the similar interest data within a defined time limit. Each community has dynamic interactivity based interconnections. Every node is also presumed to agree to share its list of neighbors with their interests in the form of Summary Vector components similar to the cache enabled scenario as in [31].

Table 1 Description of notations used

Sr.No.	Notation	Description
1	S	Collection of devices within the range of a single access point
2	k	Total number of users who subscribe similar interests in S
3	SI_{n_i}	Similarity impact for node n_i
4	BI_{n_i}	Betweenness Impact for node n_i
5	fBI	Betweenness impact function
6	PTS	Initial primary target set identified
7	ONS_{n_j}	Optimal neighbor set identified for node n_j
8	$SV - I$	Similarity vector containing common list of subscriptions
9	$SV - II$	Similarity vector containing compressed data
10	Nbr_{n_i}	Set of neighbors of n_i
11	$If(i)_j$	Knapsack identification factor of node i w.r.t. node j .
12	D	Set of interests $d_1, d_2, d_3, \dots, d_z$ of independently different sizes
13	SK_{max}	Set of nodes with maximum influence value
14	W	Variable window size for limited number of nodes
15	$Utility_{n_i}$	Utilization function for node n_i
16	$row_{wt}(n_y)$	Matrix row weight of node n_y

4 Problem statement

Our problem is to identifying a subset of users belonging to the same or different community and discharging the data which is otherwise intended for the access point of the cellular network. The problem of data offloading relies on selecting a limited set of nodes and then forwarding the data to its identified target subset. The major goal remains to select a subset of vertices in the graph which in turn could satisfy some other vertices on the basis of some common attributes. This objective is achieved by the identification of common subscription-based communities, followed by targeting only limited nodes from the identified subset which belong to the same community over a fixed span of time δt and hence is dynamic. Our model focuses to prioritize nodes for an optimized subset in case of overlapping interest-based scenarios. Earlier to our work, the literature suggests social network-based static community derivation. We do also address the overlapping communities by limiting them to belong to one user interest at a time. In order to belong to one subset of S' , we accept groups dependent on the same facilities. The solution to this problem is subdivided into different stages of recognition and subset selection optimization. In order to derive the

complete route for the data packet to reach through the maximum number of nodes, the neighbor prediction for each set is followed. Several authors mentioned in the literature also encounter the same form of problem [4, 8, 10, 11], but the approach proposed typically suffers the disadvantage because the model is static and partial or predetermined. It becomes more unrealistic because when we consider the offload for mobile data subscribers, the users are mobile in our model. The relationship is extracted from some of the solutions based on experiences in history that establish a predetermined connection. Taking into account these limitations of previous work, we seek to achieve dynamic allocation [17, 32] for each goal set. Centered on changing the allocation of nodes to many target sets, we suggest a more complex algorithm. We exploit two feasible optimization scenarios: firstly, the target set selection within a single community is restricted to the selection of subscribers to one service and secondly, the offloading of neighbors across different communities. This model of network-based graph is similar to the assignment problem of a knapsack. In order to represent a set to be belonging to a community or not, we declare every matrix entry to be unity if it lies in the community and zero otherwise. Consequently, the aim of this study is to define the number of initial users to be identified as the primary and secondary target sets. This helps to derive optimized target sets. The level of interactivity in the overlapping population needs to be detected after the selection of the optimal target sets. This is done to define the path via the optimized target source set.

5 Proposed algorithm

We divide the entire procedure into three sub-algorithmic steps to yield the optimized final target set. The first sub-algorithm provides the nodes for the primary target set which are further optimized to a limited target set in the second sub-algorithm. The final sub-algorithm determines the route optimized data forwarding scheme for data offloading.

Algorithm 1 We start with the Primary Target Set (PTS) identification algorithm. This algorithm aims to use a limited set of nodes from a single community on the basis of similarity index values and the optimum threshold values for the nodes within the sets. The nodes are checked for their similar choices of subscription derived based on set of interests. The nodes which fall in the range of the access point at time instance t are identified and compared with the nodes at instance, $t + \delta t$. For each node, n_i all the neighbors m are identified. Corresponding to these neighbors, we evaluate the similarity in data subscription as in [33]. We evaluate Betweenness Impact BI_{n_i} for each node n_i , using

betweenness centrality BC_{n_i} [29], and its total number of neighbors ρ_{n_i} using

$$BI_{n_i} = \frac{BC_{n_i}}{\rho_{n_i}} \tag{1}$$

Here the value of betweenness centrality BC_{n_i} for the node n_i is given by

$$BC_{n_i} = \sum_{j=1}^{N-1} \sum_{k=1}^{j-1} \frac{g_{jk}(n_i)}{g_{jk}} \tag{2}$$

where $g_{jk}(n_i)$ identifies the number of paths between nodes n_j and node n_k passing through the node n_i . Here g_{jk} is the total number of paths connecting node n_j and node n_k . For each node and its immediate neighbor, we evaluate the similarity in data for them. We also use the influence function similar to the k-truss used by the authors in [34]. It ensures maximum usage of cut vertex-based nodes. It is based on the concept of influence sub-graphs in graph theory. According to it, a K -influence subgraph $n_i(K)$ for a graph G is defined as the largest sub-graph with all interconnected edges belonging to at least $K - 2$ triangles. Thus each edge for the node n_i has an influence value, $n_i(ij) = K$, if it belongs to $n_i(K)$ but does not belong to $n_i(K + 1)$. This is equivalent to the clique of order K . The nodes are ranked on the basis of Influence Size, INS_{n_i} derived using

$$INS_{n_i} = \max |n_i(ij)| \tag{3}$$

It helps to derive the maximum impact for different influence values among all nodes in the graph.

We derive the fraction in the influence $F = R_W^{K_{max}}$ for a variable size window W , using the following equation

$$R_W^{K_{max}} = \frac{\text{Number of nodes in } n_i(K) / |S|_{k_{max}}}{|W| / |S|} \tag{4}$$

The final preference for edges is determined on the basis of combined value of *EgoBetweenness* and *Influence* for each node, which is calculated using

$$Utility_{n_i} = (\alpha \times F \times INS_{n_i}) + (\beta \times BI_{n_i}) \tag{5}$$

Here α and β are tunable constants with $\alpha + \beta = 1$ to give priority. The n_i node is chosen for inclusion in the PTS set based on the maximum value of $Utility_{n_i}$. The same protocol is updated for $BI \geq 0.5$ nodes. This ensures the priority given to the impact of similarity using tunable constants. We divide the set of users into two halves based on the value of BI , with the values to be either $BI < 0.5$ or $BI \geq 0.5$. Although the complexity of this algorithm is more, yet it is responsible for the major improvisation achieved in our results.

Algorithm 1 *PTS*, an algorithm for initial set selection by access point.

INPUT: $S=[n_1, n_2, \dots, n_k]$, $D = [d_1, d_2, \dots, d_{int}]$

OUTPUT: *PTS*

```

1:  $a \leftarrow 1, W \leftarrow 1, \alpha \leftarrow 0.1$ 
2: for  $a \leq i$  do
3:   evaluate  $BI_{n_a}$ 
4:   for  $b \leq \text{int}$  do
5:     evaluate  $R_W^{K_{max}}$ 
6:     evaluate  $INS_{n_i}$ 
7:   end for
8:   for each  $\alpha$  and  $\beta$  do
9:     evaluate  $Utility_{n_i}$ 
10:  end for
11: end for
12: if  $Utility_{n_i} \geq 0.5$  then
13:   update  $PTS \leftarrow n_i$ 
14: end if
15: return PTS

```

Algorithm 2 The target set is optimized further when we optimize the sets on the basis of interests prioritized by the nodes. We aim to identify the nodes which should be preferred more over the rest of the nodes, similar to the social network connections on the basis of frequent interactivity in terms of activity status governed by the access point. These nodes are referred to as the optimal nodes. The data is replicated to its neighboring nodes by them. An access point needs to prioritize a small number of nodes in any group in which a user encounters V neighbors for data retrieval using several edges E . The *ONS* algorithm does this identification of neighbors. It uses the breadth-first search and depth-first search approaches to determine the set of progressive nodes. Such nodes share the utility values from the previous algorithm to their neighbors. The compressed message carrying two summary vectors is expected to carry all nodes: summary vectors $SV - I$ and $SV - II$. The $SV - I$ includes each node's list of subscription interests, and $SV - II$ stores the data in compressed form. For the adjacent node, the data are given on the basis that the subscriptions across the summary vectors are equivalent. If the initially available data is relatively low than the data in the main node, the data is transmitted to the nearest node. Based on a similar form of subscription, we continue to classify individual populations. This problem is defined as optimum neighbor set selection. In this algorithm, we propose to share the data to the neighboring nodes in the form of summary vectors. It considers the nodes which have multiple belongings to more than a single communities identified through the channel. We determine the overlapping on the basis of its matrix representation for each matrix containing a node that belongs to more than one community. The weights of common interests across the interest matrix ensure the selection of non-overlapping communities.

Algorithm 2 *ONS*, an algorithm for optimal neighbor selection algorithm for each user.

INPUT: $PTS=[n_1, n_2, \dots, n_j]$

OUTPUT: ONS_{n_j}

```

1: for  $e \leq j$  do
2:   if  $h \in Nbe_{n_e}$  then
3:     share the  $SV - I_u$  &  $SV - II_u$  with  $h$ 
4:   end if
5: end for
6: for all  $e \in Pj$  do
7:   if  $SV - I_e \leq SV - I_i$  and  $SV - II_e \leq SV - II_j$  then
8:     update  $IF(i)_e = 1$ 
9:      $ONS_{n_j} \leftarrow n_e$ 
10:  else
11:    update  $IF(i)_e = 0$ 
12:  end if
13: end for
14: return  $ONS_{n_j}$ 

```

Algorithm 3 We transform the network visualization into a tree data structure. This helps in avoiding any cycles and reduces the number of directed connecting links. Such a transformation helps us to impose the phenomenon of shortest path application across the minimum spanning tree using the depth attribute of each node. In order to evaluate depth, the graph users need to be connected. Thus the row matrix sum is used to obtain the depth based relation. The maximum depth evaluated using $row_{wt}(n_y)$ determines the maximum utility of minimum number of nodes associations with lesser delay tolerance and assuring maximum portion of the network covered. The nodes with the best available ad-hoc approach are selected.

Algorithm 3 *FTSS*, Final target set selection algorithm.

INPUT: $PTS = [n_1, n_2, \dots, n_j]$, $ONS = [ONS_{n_1}, ONS_{n_2}, \dots, ONS_{n_m}]$

OUTPUT: ONS_{n_j}

```

1: for all  $i$  do
2:   Initialize  $row_{wt}(n_i) = 0$ 
3: end for
4: for  $y \in ONS$  and  $x \in PTS$  do
5:   for  $x \in ONS, x \neq y$  do
6:     if  $IF(y)_x = 1$  then
7:       update  $row_{wt}(n_y) = row_{wt}(n_x) + 1$ 
8:     end if
9:   end for
10: end for
11: update  $FTSS = PTS \cap ONS$  according to descending order  $row_{wt}(n_y)$ 
12: return FTSS

```

Complexity analysis for algorithm The target set selection is an NP-hard problem [13, 35] to approximate the maximization and minimization variants. The *PTS* algorithm has a complexity of order $i \times int$. The number of interests *int*, are far less in comparison to *i* number of users. Hence we can consider $O(i \times int) \approx O(i)$. Also, the second algorithm is dependent on the number of neighbors of nodes. The number of primary neighbors for any node are also very less in comparison to the total number of users. Hence for *ONS* algorithm, the complexity can also be approximated to be of the order $O(j)$. The final algorithm *FTSS* is dependent on the output of the previous *PTS* and *ONS* algorithms. *ONS* is repeated for each user identified from *PTS*. Hence the overall order for *FTSS* is $(i \times j)$. However, the average number of neighbors is also very less in comparison to total number of users. Hence we may consider $O(i \times j) \approx O(i)$. Thus the overall complexity of our algorithm is linear for practical reasons.

6 Simulation and performance evaluation

The proposed optimization is compared with literature strategies for its implementation using MATLAB. Our results are authenticated for data forwarding in case of limited sizes of the target sets involving more significant nodes. The simulation is evaluated over reality mining dataset from MIT and bluetooth dataset from NUS. These datasets have been used to identify the social communities and groups on the basis of the identification of bluetooth enabled devices in the proximity for static and dynamic associations which evolve with time. In this section, we present the simulation results of the greedy heuristic community-based algorithm and compare them with the naive *FTSS* algorithm. We consider a scenario of transmission of a fixed size message of 10 Kbs for our purpose of simulation. Much like newspaper distribution by a hawker, each delivery of data packets consists of a single packet. The goal is to determine the most effective target set to unload cellular data on the basis of opportunistic

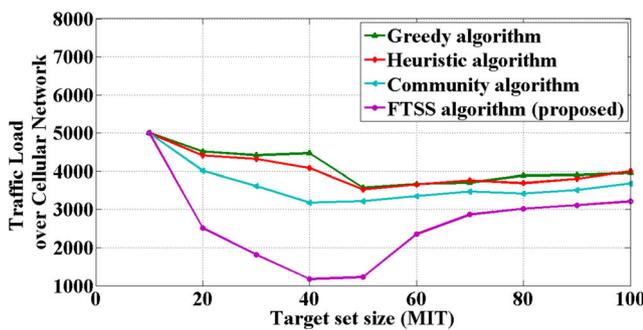


Fig. 2 Traffic load over cellular network access point for MIT dataset

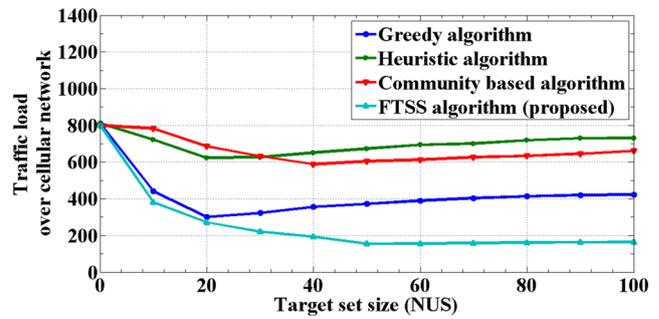


Fig. 3 Traffic load over cellular network access point for NUS dataset

communications that are available at various times and then route the data packet through it.

6.1 Traffic load comparison

We have considered only 1000 nodes from the MIT dataset in the simulation setup initially. We compare the literature based algorithms with our algorithm for a fixed size target set with an upper bound of 50 nodes. We considered a 20-second time limit for each subscriber to retain and exchange the data with their neighbors. Otherwise, the network access point would send data to all nodes in its range automatically after 20 seconds. For more users, we have repeated the same method from 1000 to 5000 in the reach of cellular networks as shown in Fig. 2.

Our algorithm is used to estimate the proportion of users who can access data by means of the limited number of users in the targeted sets. *FTSS* based goal set selection is found to be more optimal than previous algorithms. If more subscribers are interested in having the same data, our algorithm provides optimal results for the optimal percentage of users. As the number of mutual interest subscribers increases, the percentage of happy users using *FTSS* increases. The rationale for this improvisation is the possibility that subscribers would have more chances in a wider opportunistic network to contact others. In this simulation, 800 nodes from the NUS dataset are studied and the findings are shown in Fig. 3. We pick a 10 Kb data packet to be sent to all these nodes with a choice of

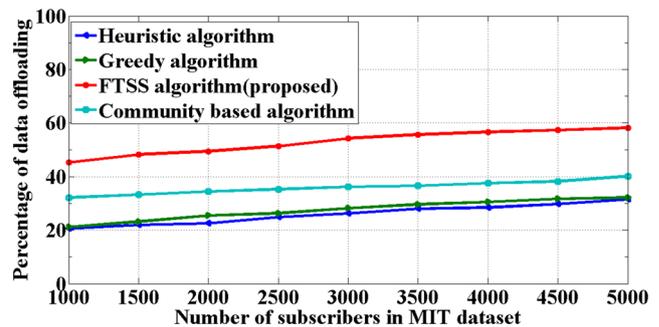


Fig. 4 Comparison of data offloading phenomenon over MIT dataset

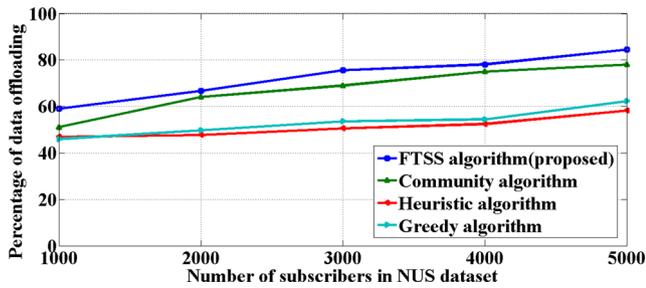


Fig. 5 Comparison of data offloading phenomenon over NUS dataset

10-100 node goal sets of different sizes. Initially, the entire traffic is handled by the access point itself when there is no subscriber in the target range. Therefore, $800 \times 10 = 8000$ Kbs of data must be transmitted. The amount of traffic managed by the access point is reduced as we encourage more users to assist in offloading as target set users.

6.2 Data offloading comparison

Figures 4 and 5 illustrate the extent of offloading percentage, which rise with the increase in the number of subscribers from respective MIT and NUS traces. We increase the participation of subscribers from 1000 to 5000 and observe nearly 20% more data offloading in comparison to literature based algorithms for MIT dataset. However, in simulation over NUS dataset we observe 10-25% more data offloading for similar extent of contribution of subscribers.

6.3 Average latency comparison

We depict the latency observations in our simulation using Figs. 6 and 7. We observe that average latency is also reduced using FTSS algorithm based implementation for both datasets. The average latency is also reduced nearly 10-12 milliseconds for varying sizes of target sets. As we go on to increase the sizes of target sets from 100 to 1000, although the average latency is reducing. But the results using FTSS shows less latency in comparison to literature based algorithms.

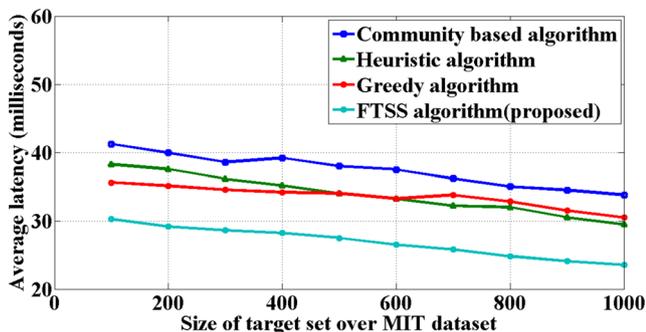


Fig. 6 Impact of using FTSS over latency for MIT dataset

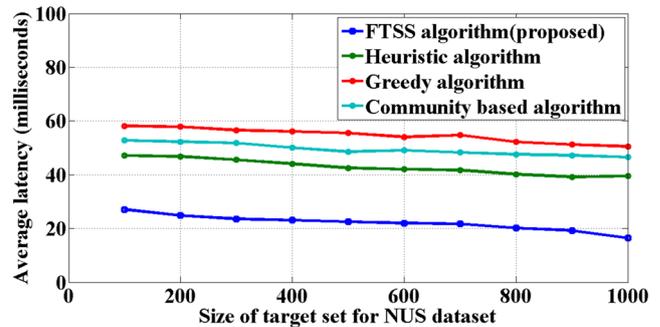


Fig. 7 Impact of using FTSS over latency for NUS dataset

6.4 Performance gain comparison

At last we check the performance of our algorithm for different message sizes. For fixed latency of 20 milliseconds, the overall performance gain reduces. The results for MIT and NUS datasets have been shown in Figs. 8 and 9 respectively. The results in Fig. 8 show that for a message size of about 50 Kbs, there is nearly 20% performance gain. However, the gain is less for smaller as well as larger message size for MIT datasets. Similarly the results in Fig. 9 for NUS dataset, we obtain the best optimal size of message size of about 40 Kbs. However, the results show less performance gains for smaller as well as larger message sizes.

7 Conclusion and future scope

Instead of completely using the access point resources of network providers, minimization of data traffic using inbuilt service capacities of the users, yield optimized results for data routing. PTS sub-algorithm has the complexity of the order of $O(k^2)$. ONS algorithm has $O(k^3)$ complexity whereas the FTSS has $O(k^2)$ complexity. Thus we have provided a heuristic based hybrid solution of $O(k^3)$ order, with a limited set of constraints. The approach assumes that all users are ready to share their identity and interests with the access points for cooperation. Also, every node

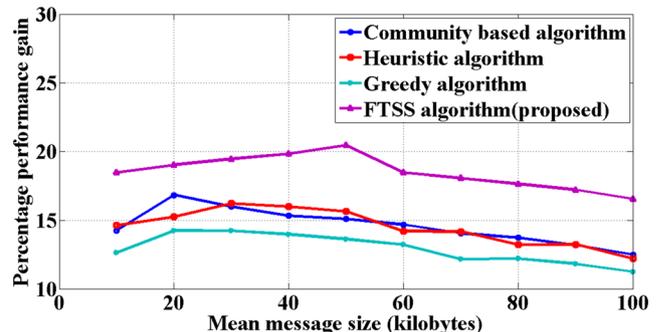


Fig. 8 Performance gain comparison for MIT dataset

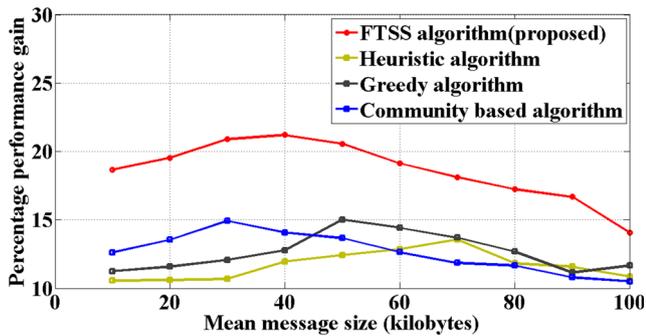


Fig. 9 Performance gain comparison for NUS dataset

has similar information about its immediate neighbors. This lays the foundation for the determination of optimal target selection in opportunistic networks such as VANETs or DTNs. Analysis of our results shows that the hybrid *FTSS* algorithm outperforms the greedy approach by 35% in terms of traffic offloading over cellular towers, 20% less as compared to the heuristic approach, and 23% less average latency when compared to the community-based algorithms. The algorithm yields at least 5–6 % less offloaders in the target sets in comparison to the heuristics-based networks. Since all nodes in the network may or may not be trustworthy amongst a network. Thus the impact of trust determination for such an evolutionary network has been excluded in the current work which will be explored in the future. The vehicular hotspot based access points and determination of incentives for each of the helper, nodes are the future orientations for research in this area. The delay tolerance intervals can also be varied which could be considered along with the determination of incentives for helpers to offer services using them. Our results of optimization render efficient usage of users and reduction in data traffic. The overall load in limited geographic scenarios is minimized using our modeling and implementation.

References

- Cisco Visual Networking Index (2017) Global mobile data traffic forecast update, 2016–2021 white paper. <http://goo.gl/yITuVx>,
- Cerwall P et al (2020) Ericsson mobility report june 2020. Ericsson com
- Dawy Z, Saad W, Ghosh A, Andrews JG, Yaacoub E (2017) Toward massive machine type cellular communications. *IEEE Wirel Commun* 24(1):120–128
- Zhou H, Chen X, He S, Zhu C, Leung VCM (2020) Freshness-aware seed selection for offloading cellular traffic through opportunistic mobile networks. *IEEE Trans Wire Commun* 19(4):2658–2669
- Baier P, Dürr F, Rothermel K (2012) Tomp: Opportunistic traffic offloading using movement predictions. In: Local computer networks (LCN), 2012 IEEE 37th conference on, pp 50–58. IEEE
- Valerio L, Bruno R, Passarella A (2015) Cellular traffic offloading via opportunistic networking with reinforcement learning. *Comput Commun* 71:129–141
- Costa P, Mascolo C, Musolesi M, Picco GP (2008) Socially-aware routing for publish-subscribe in delay-tolerant mobile ad hoc networks. *IEEE J Sel Areas Commun* 26(5):748–760
- Han B, Hui P, Kumar VSA, Marathe MV, Shao J, Srinivasan A (2012) Mobile data offloading through opportunistic communications and social participation. *IEEE Trans Mobile Comput* 11(5):821–834
- Han B, Hui P, Kumar VS, Marathe MV, Pei G, Srinivasan A (2010) Cellular traffic offloading through opportunistic communications: a case study. In: Proceedings of the 5th ACM workshop on challenged networks, pp 31–38. ACM
- Wang X, Chen M, Han Z, Wu DO, Kwon TT (2014) Toss: Traffic offloading by social network service-based opportunistic sharing in mobile social networks. In: INFOCOM, 2014 Proceedings IEEE, pp 2346–2354. IEEE
- Chuang Y-J, Lin KC-J (2012) Cellular traffic offloading through community-based opportunistic dissemination. In: Wireless communications and networking conference (WCNC), 2012 IEEE, pp 3188–3193. IEEE
- Andreev S, Pyattaev A, Johnsson K, Galinina O, Koucheryavy Y (2014) Cellular traffic offloading onto network-assisted device-to-device connections. *IEEE Commun Mag* 52(4):20–31
- Nichterlein A, Niedermeier R, Uhlmann J, Weller M (2013) On tractable cases of target set selection. *Soc Netw Anal Min* 3(2):233–256
- Alsharif N, Xuemin S (2016) *i car-ii*: Infrastructure-based connectivity aware routing in vehicular networks. *IEEE Trans Veh Technol* 66(5):4231–4244
- Li Y, Su G, Hui P, Jin D, Su L, Zeng L (2011) Multiple mobile data offloading through delay tolerant networks. In: Proceedings of the 6th ACM workshop on challenged networks, pp 43–48. ACM
- Jiang J, Zhang S, Bo L, Li B (2016) Maximized cellular traffic offloading via device-to-device content sharing. *IEEE J Sel Areas Commun* 34(1):82–91
- Huang D, Wang P, Niyato D (2012) A dynamic offloading algorithm for mobile computing. *IEEE Trans Wirel Commun* 11(6):1991–1995
- Kempe D, Kleinberg JM, Tardos É (2015) Maximizing the spread of influence through a social network. *Theory Comput* 11(4):105–147
- Li Y, Jin D, Hui P, Chen S (2016) Contact-aware data replication in roadside unit aided vehicular delay tolerant networks. *IEEE Trans Mob Comput* 15(2):306–321
- Li Z, Wang C, Yang S, Jiang C, Ivan S (2015) Space-crossing: Community-based data forwarding in mobile social networks under the hybrid communication architecture. *IEEE Trans Wirel Commun* 14(9):4720–4727
- Valerio L, Bruno R, Passarella A (2014) Adaptive data offloading in opportunistic networks through an actor-critic learning method. In: Proceedings of the 9th ACM MobiCom workshop on challenged networks, pp 31–36. ACM
- Nguyen NP, Dinh TN, Tokala S, Thai MT (2011) Overlapping communities in dynamic networks: their detection and mobile applications. In: Proceedings of the 17th annual international conference on Mobile computing and networking, pp 85–96. ACM
- Palla G, Derényi I, Farkas I, Vicsek T (2005) Uncovering the overlapping community structure of complex networks in nature and society. [arXiv:physics/0506133](https://arxiv.org/abs/physics/0506133)
- Gregory S (2010) Finding overlapping communities in networks by label propagation. *New J Phys* 12(10):103018
- Li Z, Wang C, Yang S, Jiang C, Stojmenovic I (2014) Improving data forwarding in mobile social networks with infrastructure support: a space-crossing community approach. In: INFOCOM, 2014 Proceedings IEEE, pp 1941–1949. IEEE

26. Barbera MV, Viana ACx, de Amorim MD, Stefa J (2014) Data offloading in social mobile networks through vip delegation. *Ad Hoc Netw* 19:92–110
27. Kouyoumdjieva ST, Gunnar K (2016) Energy-aware opportunistic mobile data offloading under full and limited cooperation. *Comput Commun* 84:84–95
28. Filippo R, De Amorim MD, Conan V, Passarella A, Bruno R, Conti M (2014) Data offloading techniques in cellular networks: A survey. *IEEE Commun Surv Tutor* 17(2):580–603
29. Daly EM, Haahr M (2007) Social network analysis for routing in disconnected delay-tolerant manets. In: *Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing*, pp 32–40. ACM
30. Cui P, Chen S, Camp J (2019) Greenloading: Using the citizens band radio for energy-efficient offloading of shared interests. *Comput Commun* 144:66–75
31. Mei H, Huimin L, Peng L (2019) Data offloading in cache-enabled cross-haul networks. *Comput Commun* 142:1–8
32. Sharma P, Shailendra S (2018) Optimal target set selection via opinion dynamics. In: *Fifth International Conference on Parallel, Distributed and Grid Computing (PDG)*, pp 806–811. IEEE, p 2018
33. Malliaros FD, Rossi M-EG, Vazirgiannis M (2016) Locating influential nodes in complex networks. *Sci Rep* 6:19307
34. Cohen J (2008) Trusses: Cohesive subgraphs for social network analysis. *National Security Agency Technical Report*, 16
35. Eyal A, Oren B-Z, Guy W (2010) Combinatorial model and bounds for target set selection. *Theor Comput Sci* 411(44–46):4017–4022

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Prince Sharma is pursuing his Ph.D. as a research scholar from Jaypee University of Information Technology, Himachal Pradesh. Earlier to this, he has done his Master of Technology and Bachelor of Technology from Himachal Pradesh University. His research interests include Computer Algorithms, Computer Networks, Theoretical Computer Science and Bioinformatics. He has more than five years of teaching and research experience.



Shailendra Shukla is an Assistant Professor at the Computer Science and Engineering Department of the Motilal Nehru National Institute of Technology Allahabad. He received the MS-(Information Security) from the “Indian Institute of Information Technology Allahabad and then completed a Ph.D. from the Indian Institute of Technology Patna in computer science. His doctoral work is based on “On Boundary Detection and Localization in Wireless Sensor Networks”. In this work, he proposed a collection of networking algorithms that address the security problems like routing in the Internet of Things, localization, boundary node detection (surveillance), virtual coordinate assignment (Geography routing/localizations), cyber-physical systems, monitoring, and surveillance. He has published articles in various publication houses like Elsevier, Springer, IEEE. Previously, he was associated with Jaypee University for 5 years.



Amol Vasudeva is working as an Assistant Professor (Senior Grade) in the Department of Computer Science and Engineering, Jaypee University of Information Technology, Wagnaghat, India. He completed his M. Tech. degree in Computer Science from Jaypee University of Information Technology, Wagnaghat, Solan, Himachal Pradesh, India in 2009. In 2018, he completed his Ph.D. degree in Computer Science from Himachal Pradesh University, Summer Hill, Shimla, Himachal Pradesh, India. He is an IEEE and ACM Member. His research area includes security in wireless networks and Internet of Things.