

**Energy efficient target set selection and buffer management for D2D mobile data offloading**Prince Sharma<sup>a\*</sup><sup>a</sup>Jaypee University of Information Technology, India**CHRONICLE***Article history:*

Received: September 10, 2020

Received in revised format:

October 9, 2020

Accepted: December 12, 2020

Available online: December 12, 2020

*Keywords:*

Target set selection,

Data offloading

D2D

Summary vector

Betweenness

**ABSTRACT**

Data offloading offers a significant solution to the problem of explosive rise in mobile data traffic. A naive approach would be to utilize the infrastructure (cellular tower, WiFi, femtocell) or other mobile devices to offload data. However, increasing the number of a cellular towers, WiFi, or femtocells is costlier deal for data delivery. Recently, device-to-device (D2D) paradigm of data communication has emerged out as one of the most promising solutions to deal with cost effective cellular traffic offloading. D2D communication provides a direct communication link between closely located mobile users. Another significant feature of D2D is its content centric nature, which makes it useful in data offloading. In this paper, we have addressed the issue of data offloading in mobile devices and proposed a hybrid model of D2D communication with ad-hoc nature. The paper also considers the issues like memory constraints of the devices, pruning of replicated messages and energy efficiency to increase the lifetime of the battery. Considering all the constraints and trade off, we have modeled our problem into optimal target selection problem and distributed community detection problem, both of which are NP-hard. We propose a clustering algorithm to optimize the cooperative mobile nodes. The proposed algorithm uses the betweenness centrality and k-means for optimizing target set section. Our proposed algorithm requires less time in terms of computational complexity with limited space. We compare it with the community-based approach in terms of load transferred for varying target set sizes for validation. The simulation results of the suggested algorithm may reduce the energy requirements up to 16.7% and is able to accommodate 80% more traffic as compared to the community-based algorithm.

© 2021 by the authors; licensee Growing Science, Canada.

**1. Introduction**

The demand for data and internet has increased in the last few years due to the cheap and easily available smartphones and personal digital assistants. Around 71% growth in global mobile data usage was witnessed in the year 2017 which has grown approximately 17 times over the previous five years from 2012 (Index, C.V.N., 2017). Also the overall mobile traffic is expected to be reaching about 77 exabyte by 2022. In 2017, 54% of mobile data traffic has been released with technologies such as WiFi or femtocells. The study indicates that by 2022 smartphones could reach 90% of mobile data traffic. The WiFi offload is expected to more dependent over fourth-generation (4G) and fifth-generation (5G) networks than the slower networks. The reason behind the exponential growth in data is because of mobile social applications and multimedia streaming. The data plans are expected to be more generous due to the introduction of 5G networks and the respective offloading percentage is estimated to be around 71% by 2022. The network burden that the service providers will bear is likely to be depleted very soon. Due to the present COVID-19 pandemic like situations, more and more work is relying over data networks through distant networking solutions. There is a tremendous need for technically and economically feasible alternatives to satisfy these high demands of mobile data and prevent blocking networks. There are several solutions which are provided to overcome the limited bandwidth and heavy load situations for networks. An intuitive approach is to build new base stations (BS), increase cellular network capacity by adding smaller cell size to BS or upgrade the cellular network to advanced networks of next generation such as Wi-Max and Long-Term Evolution (Kumar et al., 2013) to handle large volumes of data transmission.

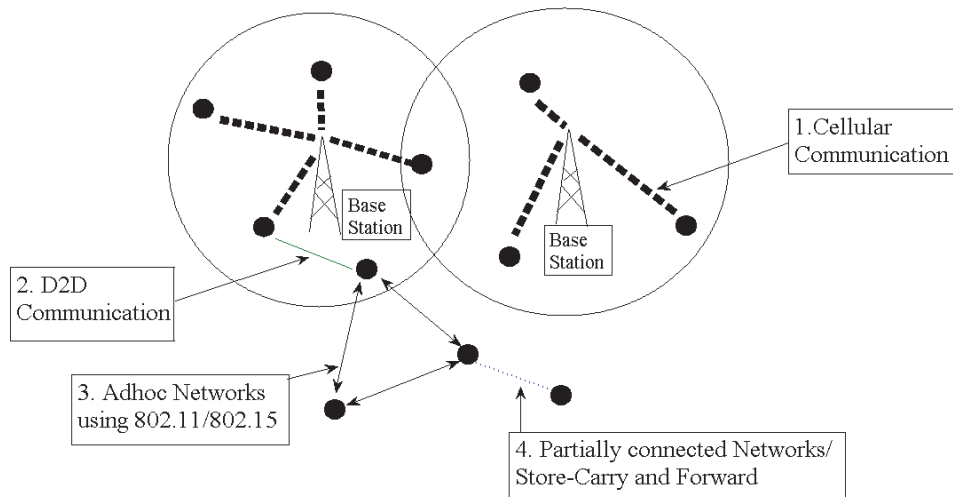
\* Corresponding author.

E-mail address: [prince.sharma@mail.juit.ac.in](mailto:prince.sharma@mail.juit.ac.in) (P. Sharma)

Another solution could be, to use broadband cellular network technologies like 4G/5G to accommodate massive data traffic. However, because of their higher energy usage (Akyildiz et al., 2010) and high settlement cost, these solutions are not adaptable enough to sustain today's mobile internet services. Moreover, the practical need of the hour is to keep pace with the rapid rise in data demands with the available network supply. Other substitutes like WiFi offloading are impressive methods to move from the cellular network naturally, and it has shown a significant improvement in offloading (Fortetsanakis & Papadopouli, 2013). Immediate offloading and deferred offloading are used frequently for offloading. The unlimited availability of WiFi and the sharing of information on-site uses immediate offload. Most advanced mobile phones currently choose WiFi over the cellular interface to transmit on-site data. Data has a due time duration (Li et al., 2011) in deferred offloading and it resumes repeatedly for data offloading until the exchange finishes. In deferred offloading delay may happen because of the WiFi association or movement of the mobile device in and out from the WiFi region. The only limitation of WiFi offloading and cellular offloading is that it requires high financial investment, long process of infrastructure development (Aijaz et al., 2013) low return, and high maintenance cost. Such drawbacks can be overcome by using partially connect (Vahdat et al., 2000) or delay tolerant networks (Jain et al., 2004). Thus, the D2D communication (Kaufman et al., 2013) using the ad-hoc property of nodes, as illustrated in Figure 1 may substitute a cheaper and adaptable solution. It shows that if association like store carry and forward approach (Doppler et al., 2009) among the mobile users can be established then opportunistic communication (Han et al., 2011) and Delay Tolerant Networks (DTN) (Li et al., 2011) can be deployed for data offloading. The major advantage of using D2D network is the less monetary cost (Vahdat et al., 2000) and optimal resource utilization (Jain et al., 2004). However, the detection of optimal target-set selection (Han et al., 2011) plays an important role in data offloading. The other challenges are restricted storage on any device, caching of data and unwanted replication of data.

The important benefit of our proposed mobile data discharge method is that it uses the existing framework and the social activity of the mobile node nearly without any financial costs. In order to minimize the mobile data traffic by the vector classification, we aim to analyze how to choose the original target for  $k$  users only. The targets set nodes help all subscribed users through their social involvement to transmit data further. For example, the nodes in the transmission range may communicate with each other opportunistically. The information can also be spread by non-target users after it is received by target users or others. To validate our proposed work, we compare our work with real-world data set of MIT (Eagle et al., 2006), INFOCOM (Hui et al., 2005), and NUS (Srinivasan et al., 2006). We compare our proposed protocol with community based approach, greedy approach and heuristic approach in our implementation in Sharma et al. (2018), but restrict our comparison in this article only with community-based approach since being the better. This paper makes a valuable contribution as follows:

1. It proposes a hybrid approach for data offloading using D2D, infrastructure-based nodes and mobile nodes.
2. The distributed algorithm is prone to redundant messages at subscriber level.
3. We also propose community formation, so that data dissemination can take less time to propagate the information to the rest of the subscribed users.



**Fig. 1.** Hybrid model of D2D Communication with ad-hoc nature in partially connected networks

The remainder of this paper is organized as follows. In Section 2, we have described briefly the relevant work related to data offloading and opportunistic networks. The device model, problem formulation for our work is outlined in Section 3 and Section 4 respectively. Section 5 deals with our offered strategies for data offloading. Section 6 and Section 7 cover the results of the simulation and experimental results respectively. Finally, we finish our conclusions with future scope in Section 8.

## 2. Theoretical background

Numerous researchers and practitioners have been interested in wireless network analysis as well as several studies on mobile data transmission have been published. In the last few years, there has been little knowledge of methods and calculations that improvise the overloading conditions of data offloading. We have divided our related work based on different data communication techniques. These techniques are cellular networks, ad-hoc networks using Bluetooth, WiFi offloading, D2D networks and opportunistic networks. The cellular network-based data offloading approach uses GSM, CDMA, AMPS, LTE, or small cells like Microcell, Picocells, Femtocell (Ghosh et al., 2012). The authors in (Ioannidis et al., 2009) propose a subscription-based dynamic-content distribution service with the help of a service provider for mobile users. The paper proposes a new approach for optimum bandwidth allocation by the service provider to upload user content. The research in Zhuo et al., (2011) concentrates mainly on stimulating customers to discharge cellular traffic via Delay Tolerant Network (DTN). The authors in Chuang et al. (2012) use encounter-based frequency and community detection for data forwarding within the same classroom area. The contribution of this work is the selection criterion of primary sources disjointed communities. The authors compare their work with different forwarding strategies like encounter-based forwarding, rarest first forwarding, and random forwarding. To prove their claim, they used NUS dataset. The paper uses social network property for community detection. The limitation of the reviewed work is that the authors consider that the community develops on the basis of location and network from a fully connected graph, which is impractical in the real world. The second technique is the utilization of Bluetooth or WiFi (Ioannidis et al., 2009) based offloading. The researchers in Baier et al. (2016) develop a definition of traffic offload by reducing cell traffic by transferring to local networks such as WiFi. They offer a TOMP model that analyses prospective inter-device connectivity using mobility predictions from smartphone users. The encounter probability is computed as a heuristic parameter to determine the coverage relation. In the research work Lee et al. (2014), the economic benefits were discussed by modeling a two-stage sequence game of the interaction between one client and one source. The authors concentrate on the key issue of the WiFi offload for several sensitive time frames with limited capacity of APs. They used the game theory strategy of Nash equilibrium to solve the bargaining problem. The authors in Jung et al. (2014) propose a solution to increase the per-user throughput while considering the deadline constraints. For improving mobile network capacity, the authors of Osseiran et al. (2009) and Asadi et al., (2014) propose to use a network coding scheme with D2D communication. They assume a cooperative network where two D2D communications are used to interchange the uplink data among D2D users before the message is transmitted to the base station. The proposed solution was only limited to uplink they have not considered the downlink data management. Bao et al. (2014) have proposed a caching based approach for data transmission in which each device caches the available content. When the device reaches the area containing dense clusters of mobile devices, we identify it as a data spot. In such a scenario the operator instructs the mobile devices to share similar data. Their approach does not consider the redundant data pruning and memory constants. In Jiang et al. (2015), the authors propose the solution on how content sharing by D2D communication can be maximized. The authors have used specified content caching mechanism. It is used to identify the best matching pairs in between the senders and receivers. Such optimization problem is defined in terms of knapsack using selective cache. The sender-receiver problem is solved using bipartite graph properties using maximum weighted match problem. The limitation of this work is they have not considered the partially connected networks and pruning of unwanted messages. Data can be downloaded from the end-user or the ISP (Aijaz et al., 2013). This allows defined operators to reduce the cellular system clog, while offering cost investment funds for the end user, which are focused processed data management and greater accessibility to data transmission. The operators can increase the asset utilization of accessible systems and reduce the congestion of operators, thereby identifying their system assets to different clients by uploading the movement of portable information to accessible correlative system. There are two types of infrastructures based on offloading namely immediate offloading and deferred offloading featured in Valerio et al. (2014). Immediate offloading means using WiFi connectivity without restrictions and on-the-spot sharing of information. Deferred offloading is related to a deadline and continues to share information until the share is completed. This time delay may occur because of problems with infrastructures or because a customer transfers into and out of the area of infrastructure. This delay could occur because of problems with infrastructure associations or because a customer transfers in and out of the infrastructure area. In the worst case, the cell will have to adjust the working phenomenon if the exchange of information does not take place in time. The fourth technique is to utilize the opportunistic communication for data offloading. The work done in Li et al. (2013) proposes a data offloading solution for disruption tolerant networks. This article defines the issue as the 0-1 knapsack with linear restrictions maximizing problem. The problem of maximization is solved by a greedy method based on heuristic record or approximation of records of less time duration. The proposed algorithm is compared with greedy, approximated, and optimal algorithms. The authors in Han and Srinivasan (2012) and Han et al. (2013) exploit the importance of target set selection problem using opportunistic communications. They have proposed greedy algorithm, heuristics-based algorithm, and random algorithm using mobile social networks for data offloading. They used the regularity of migration flows as a key factor of opportunistic communication to suggest social interaction. The authors justify the selection problem as sub-modular and then introduce the greedy algorithm before applying the heuristic approach based on the patterns of human mobility. The comparison among greedy, heuristic and random implementation reveal the greedy approaches to be best in terms of practical implementation and the global information needs. The authors in Valerio et al. (2014) have identified two major sub-problems for data offloading. The first problem is deciding the number of nodes to be assigned for the diffusion of content and, second, which nodes are more efficient and useful for accelerating the dissemination process. The authors have used a reinforcement learning method based on actor-critic approach to solve the first problem. For the second problem, they have used the acknowledgment messages information to establish the heuristic behavior for message dissemination. In Barbera et al. (2014)

the authors propose an opportunistic communication based data offloading. In this article the authors choose a subset of important VIP clients on the basis of investigation of contact design between end-clients using centrality and page-rank. These VIP delegations help in data forwarding using social importance and mobility. It uses VIP methods that are listed as reliant on global blind promotion and global greed. The issue of the neighborhood VIP delegation is also taken into consideration, based on the k-clique group algorithm, that most repetitive meetings between people usually take place at the same place. The significance of a node is evaluated on the basis of between centrality, closeness, centrality and page ranking. To validate their result, they have used real data sets of Dartmouth and Taxis. Other works like Han and Srinivasan (2012) and Wang et al. (2014) propose traffic offloading approaches being helped by social network services (SNS) via opportunistic sharing in mobile social networks. The emphasis of Zhuo et al. (2011) is over stimulating clients to unload mobile traffic using DTNs. The authors have proposed an incentive framework for down-linking mobile traffic offloading based on an auction mechanism.

The major drawback identified in the literature survey is to focus over target set selection for heterogeneous networks focusing on the problem of multiple communities of users simultaneously in a real time environment. We try to utilize the social behavior based realistic environments using D2D communications to enable opportunistic communications amongst users.

### 3. System model and problem definition

Our concern is to model the practical behavior of the opportunistic network which may or may not be completely connected taking into consideration the energy dissipation across them. We enable both opportunistic communications using the devices.

#### 3.1 Issues, Controversies and Problems

The literature suffers the shortcoming of suitable well suited model which could relate all types of opportunistic communications. We have considered a hybrid of two scenarios, infrastructure-based sparsely connected networks. We assume that cellular towers  $AP = AP_1 AP_2, AP_3, \dots AP_n$  form a connected graph. To make scenario simplified, we have considered a scenario of single access point  $AP_1$  accounting for the connection for  $k$  mobile users visualized as a set  $M = (M_1, M_2, \dots, M_k)$ . The mobile users who are in the range of access point  $AP_1$  can download/upload data bytes at  $t_1$ . We denote the data bytes as  $D = d(1), d(2), d(3), \dots, d(i)$  where  $d(i)$  is defined back as the composite function such that  $d(i) = [b(i), \mu(i)]$  where  $1 \leq i \leq n$  and  $b(i)$  represents the respective message size held by node  $M_i$  in terms of bytes and  $\mu(i)$  denotes the time taken it to transfer data  $d(i)$ .

We need to consider for the practical observations, that the data items have been organized in the ascending order without the loss of generality. We assume that each object of the data is indivisible, at the same time. In addition, we must consider the transfer of data in terms of successful upload and download before the expiry of the TTL time, which can be visualized as the transmission deadline of such data items. The cellular coverage is assumed to have clusters, especially in indoor where the cellular tower is deployed in such a way that it covers all the mobile users. We have considered it because sometimes the coverage may be weak in the center of the massive structures. We have excellent coverage on the street like cases, but when moving further indoor, the cellular signal deteriorates quickly. Although inadequate indoor coverage can be solved by adding small cells, it requires a considerable amount of time to implement and becomes costly. Hence, we propose that each device is capable of D2D communication for indoor communication or spotty scenarios. For rural environment, we assume that the cellular towers are sparsely deployed and sometimes create a partially connected network. In such a scenario, we assume that nodes are capable of forming an ad-hoc network. We enlist our assumptions as follows:

1. All mobile devices form a disconnected undirected graph. Vertices of the graph represent the mobile devices.
2. In common with all nodes, the data are replicated if available prior to a specified delivery date.
3. Every network has interconnections focused on adaptive interactivity.
4. The list of its neighbors can be shared by every node.

The initial step of any data offloading procedure is target selection problem as they are crucial to start the dissemination data properly. The community-based source selection problem is one of the approaches for source selection. Community-based source selection problem is NP-complete based on the reduction from the dominating set problem. To overcome this issue, heuristic approach is being used. The next objective is pruning of message as data flooded in, before TTL. Due to the limitations of local storage, sometimes it becomes infeasible to cache all messages. Apart from it, there are scenarios where multiple devices that can send/receive requests, which will lead to multiple messages of the same time. Hence the main challenge is how to improve the message, offloaded by using D2D communication.

### 4. Proposed algorithm

#### 4.1 Optimal Target set Selection Algorithm

Our problem is selecting a subset of users belonging to the same or different communities as defined in system model and offloads the data traffic otherwise meant for the cellular network. We consider a single community  $S$  of users based on the subscription of the same service. Our aim is to subdivide them at various levels of identification and optimize the subset

selection procedure. Several authors described in the literature also encounter the similar kind of problem, but the solution proposed generally has the limitation of static nodes. The existing solution becomes impractical to realize since the users are mobile, and in most cases the authors inferred a relationship based on the history of experiences, which identifies a heuristic relationship. Taking these flaws into account, we concentrate on realization of an optimal and realistic implementation for allocation of target sets. In accordance with the change in nodes assignment to many target classes, we suggest a dynamic algorithm. TSS optimization is categorized into two phases: the first phase involves the nodes within a single access point and the second phase involves TSS optimization avoiding the overlapping communities and use message pruning. It is helped on the basis of selection of subscribers for similar service. We describe the algorithmic optimization for identifying initial target set collection on the basis of identical interests opted for subscription. The second step of the algorithmic description consists of the detecting the subscribers with more than one target sets category. After the optimized target set for a single group is determined, the findings for the overlapping communities are extended for multiple groups. Initially, a set of mobile users  $N = [n_1, n_2, n_3, \dots, n_m]$  are assumed to be connected with access point. The inter-connection between mobile devices is represented as the graph where each node denotes the individual mobile device and edge denotes the communication link between them. All mobile devices create a list of neighbors *Nbrlist* which contains the information about devices in proximity. Mobile devices periodically share the *Nbrlist* information with access points. The clustering of nodes is done using k-means algorithm, input parameters of betweenness and cut-vertex. Then the cluster based CV node uses the overlapping node to find out the other clients for message dissemination. The access point on receiving this information, computes the  $BI(n_i)$  Betweenness Centrality (Daly et al., 2008) of all mobile devices using the following Eq. (1).

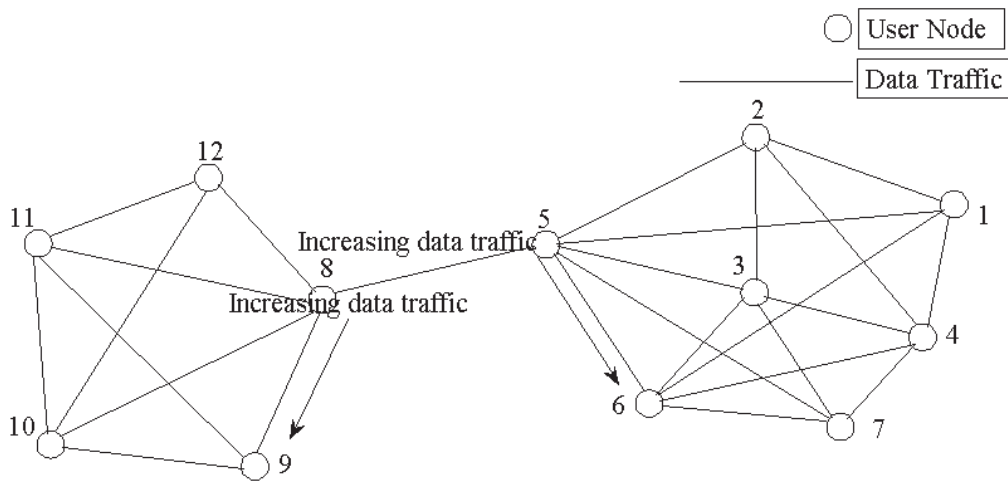


Fig. 2. Increasing traffic in undirected graph of nodes

$$BI(n_i) = \sum_{j=1, k=1}^{j=N, k=j-1} g_{jk}(n_i) / g_{jk} \tag{1}$$

where  $g_{jk}$  represents the number of edges connecting node  $n_j$  & node  $n_k$  and  $g_{jk}(n_i)$  identifies the number of paths which include  $n_i$ . In addition to betweenness centrality, a function cut-vertex  $CV$  is also computed using similar  $k$  – *truss* algorithm used in (Cohen et al., 2008). It is derived on the basis of the graph theory, where the  $k$  – *influence* subgraph  $T_K$  for a graph  $G$  is defined as the largest subgraph with all interconnected edges belonging to at least  $K - 2$  triangles. The betweenness value of each node is saved in data structure *Bet* in sorted manner and  $CV$ . In the next step, algorithm creates a clusters of  $N$  using  $k$ -means clustering. On the basis of betweenness centrality and location  $L$  of each nodes, access point creates a  $d$  dimensional observation vector  $X = x_1, x_2, x_3, \dots, x_n$ . The  $k$ -means clustering partition the  $n$  observations into  $k$  sets  $C = C_1, C_2, C_3, \dots, C_k$  so as to minimize the within cluster sum of squares. The insight of  $k$ -means clustering is evaluated using Equation 2 as follows:

$$\arg \min_{i=1, c \in C_i} \sum_{i=1}^k |x - \mu_i|^2 = \arg \min_{i=1, i \in k} |C_i| Var C_i \tag{2}$$

where  $\mu$  is the mean of points in  $C$ . Provided that the overall variance is constant, this implies that the square based deviations between points in different clusters are maximized. Thus, the equivalence is derived using the identity in Equation 3 as follows:

$$\sum_{c \in C_i} |x - \mu_i|^2 = \sum_{x \neq y_i \in C_i} (x - \mu_i)(\mu_i - y) \tag{3}$$

The *CutVertex* values are used to find the optimal target set  $T$ , The subset of clusters  $C$  with most number of *CutVertex* nodes are highly capable of disseminating information between different communities. This is accomplished by the following algorithm.

**Algorithm 1. Algorithm executed at access point for optimized target set identification**

**Input:** Set of users  $N$ , Neighbor list of nodes  $Nbrlist$ , Location of nodes  $L$ .

**Output:** Clusters  $C$  and  $T$ .

**Step 1:** Initialize:  $C \leftarrow Null$  and  $T \leftarrow Null$ .

**Step 2:** For each user in the range, the access point accesses  $Nbrlist(n_i)$

**Step 3:** For each user in the range, the AP evaluates  $BI(n_i)$ .

**Step 4:** The AP computes  $CV$  using  $Nbrlist$  and  $k - truss$  method.

**Step 5:** Create  $d$  - dimensional observations vector  $X$  using location  $L$  and  $BI(n_i)$ .

**Step 6:** Create cluster  $C$  applying  $k$ -means method over matrix  $X$ .

**Step 7:** Update  $T \leftarrow C$ , with maximum  $CV$ .

**Step 7:** Return  $T$  and  $C$

*4.2 Overlapping node detection and Summary vector computation*

The second algorithm works for the detection of overlapping node (*OL*) detection and summary vector (*SV*) computation. Since the network is dynamic, so the key point of this approach is to use only local information of the network as global information requires high communication cost, processing, and energy drainage. Egocentric betweenness is used in the algorithm as it detects belonging of one node to multiple communities yielding overlapping nodes. The major benefit of egocentric betweenness is that each node does not need global knowledge of the network, but needs connectivity information only up to 2 hop neighbors. Table 1 shows the values of nodes and their overlapping nodes depicted in Fig. 2.

**Table 1**  
Overlapping nodes

Node $N_i$	Egocentric Betweenness $\sigma_i$	Threshold $\zeta(i)$	Overlapping nodes $OLs$
1	25	54	4, 3, 6
2	25	134	5
3	83	102	5
4	83	102	5
5	4	141	11
6	83	102	5
7	0	68	4, 3, 6
8	433	93	5
9	0	163	8
10	33	116	4, 3, 6
11	33	116	8
12	0	166	8

The concept of a *SV* is based on a bloom filter that is used to reduce the excessive message and redundant messages. Each node allocated to a buffer keeps track of the latest message exchange with the rest, so that no redundant communications are available. The buffer stores messages that come from other mobile nodes and collects them. The indexes of all messages need to be saved in a hash table and held in a single identifier known as a summary vector in order to manage every node effectively. When a node enters the proximity of another node, the summary vectors are shared. The message requested from a neighbor determines each summary vector exchange. The nodes also save the time when a certain node was exchanged to prevent redundant message exchange. The algorithm for computation of ego node and pruning of redundant message is defined in the following steps:

**Algorithm 2. Overlapping node detection and summary vector based message pruning algorithm**

**Step 1:** Each node  $i$  evaluates its egocentric betweenness  $\sigma_i$ .

**Step 2:** Node  $i$  exchanges  $\sigma_i$  with all of its neighbors in the  $Nbrlist$  and computes the threshold value:  $\zeta(i) = \sum_{j=1}^n \sigma_j$ , where  $(j = 1, 2, 3, \dots, n)$  represents the  $j$ th neighbor of node  $i$  and  $\sigma_j$  represents the egocentric betweenness of neighbor  $j$ .

**Step 3:** Node  $i$  finds the overlapping node from its neighbor set on the basis of comparisons of  $\zeta(j)$  and  $\zeta(i)$ . If  $\zeta(j) > \zeta(i)$ , then the node  $j$  is declared as overlapping node (*OL*) or the vice versa.

**Step 4:** The identified overlapping nodes (*OL*) from the previous step share  $SV_1$  with each node  $i$  and compare them. If there exist common entries in between, then ignore the final target sets  $T$ , else share the data of uncommon entries.

**Step 5:** Return final target set  $T$ .

We explain the algorithm to be implemented by the access point now. In the first step every node  $i$ , finds the ego betweenness  $\sigma_i$ , to determine its node usability for its neighbors in the list  $Nbrlist$ . Egocentric betweenness is characterized as having the probability of a node lying between two nodes. More the  $\sigma_i$  value, therefore it is more likely for a node to be linked to other nodes.  $A_{i,j}^2$  contains some walks of length 2 between node  $i$  and  $j$  considering  $A$  as the adjacency matrix of a node. We need not to evaluate the smaller associations like walks of unit length since it will not contribute to betweenness. In the end  $A^2[1 - A]_{i,j}$  is computed, with the number of geodesic connections of length 2 in between the node  $i$  and node  $j$ . Each node  $i$  exchanges the ego betweenness value given by  $\sigma_i$  with its neighbors. Each node then decides the nodes between its neighbours, which may have a greater effect on the data dissemination in the network. It is termed as sensitive *OL*. Finally, the cut-vertex from Algorithm 1 and OLs periodically update *SV* message sharing process in anti-entropy session. Presuming two nodes  $N_1$  and node  $N_2$  encounter each other. Let  $SV_1$  be a summary vector of node non-overlapping node and  $S - V_2$  be a summary vector of an overlapping node. Upon encounter, they exchange their summary vectors and compares it with buffered messages and accordingly requests the missing data.

### 4.3 Complexity analysis

The first phase of the algorithm computes the betweenness centrality using *Nbrlist*. This step requires  $O(n^2)$  space complexity for matrix creation and  $O(n^3)$  as runtime, which can be reduced to  $O(n^2)$  using the randomized algorithm. The second phase of algorithm computes *Cutvertex* using *k-truss* algorithm which  $O(n + E)$  as run time,  $E$  is the edge in the graph created by  $n$  nodes at a particular instance. Finally, *k-means* algorithm for clustering, which needed  $O(n^{dk+1})$  time complexity, where  $n$  is the number of mobile nodes  $k$  is the numbers of clusters created and  $d$  is the dimension constant. So the dominating complexity in all the three phases is  $O(n^{dk+1})$  as  $dk \geq 2$ , space complexity is  $O(n^2)$  and communication cost is  $O(1)$ . In Algorithm 2 the communication cost is the same as of the first phase of Algorithm 1, and the space complexity depends on the size of the bloom filter used for the creation of the summary vector.

## 5. Simulation Results

We have divided our simulation work into two parts; in the first part, we have evaluated our proposed algorithm standalone for the analysis of D2D communication, and in the second part, a comparative analysis is done with a community-based approach (Valerio et al., 2014).

### 5.1 Standalone analysis

In this subsection, we have evaluated the following parameters for analyzing the performance:

1. Energy utilization in terms of battery consumption.
2. Delays involved with different target set sizes.
3. Signal-to-noise ratio-based unpredictability and mean square error.

**Table 2**

Simulation scenario

Total number of nodes	Range	Area	Physical layer	Buffer
100-1000	250 meters	1000 × 1000	Bluetooth, WiFi, D2D	50 Mb

A total of 1000 nodes has been used in the networks in an area of 1000 × 1000 meters varied from 100 to 1000. Clustering of nodes is done using *k-means* algorithm and input parameters of betweenness and *cutvertex*. Then the cluster node uses the overlapping node to find out the other clients for message dissemination. The simulation parameter is illustrated in Table 2 and the simulation is performed using MATLAB.

**Table 3**

Battery consumption in different scenarios

Using Bluetooth	Using WiFi	Using WiFi sleep	Using D2D
70.34272	3.092963	30.86594	1.53723
70.43231	2.184556	30.64314	1.56564
70.49282	2.168105	30.39878	1.57687
70.53174	3.949731	30.15915	1.63434
70.47921	3.941977	30.57173	1.76577
70.50166	3.65894	30.80652	1.87668
70.96901	2.39671	30.71554	1.75466

### 5.2 Energy Consumption

In order to compare the energy drainage of the clients, their battery usages are studied. The Table 3 depicts that Bluetooth consumes maximum energy, whereas D2D consumes the least. The table also indicates that additional data is taken from other devices while WiFi is in sleep mode. The results indicate such implementation to use more energy. The difference in battery consumption is approximately 95% less in case of D2D over that of the Bluetooth. The energy consumption of Bluetooth is significantly more than D2D because it uses the internal power of the device.

### 5.3 Delay with respect to load transfer

The Fig. 4 depicts the transfer load delay in milliseconds against different target sets. In this figure,  $x$  - axis indicates the different sizes of the target set, whereas the  $y$  - axis indicates the heuristics-based transfer load. It shows the variation of load delay for fixed target of 1000 nodes during a 6 hours session. We have depicted the results for five delay scenarios of half an hour, an hour and its double followed by three hours and its double. However, we need to re-iterate for the remaining session periods in order to check the results for similar time duration. Thus, we have double iterations with 3 hours delay. We repeat the iterations thrice during 2-hour repetitions. Likewise, we repeat six iterations with 1-hour delay allowed and twelve such iterations for 30-minute reminders. The Fig. 4 clearly depicts that the delay is minimum for the load with lesser delay scenario. We observe that the load delay has no significant contribution with respect to different delays allowed. Hence, we focus over other aspects for comparison.

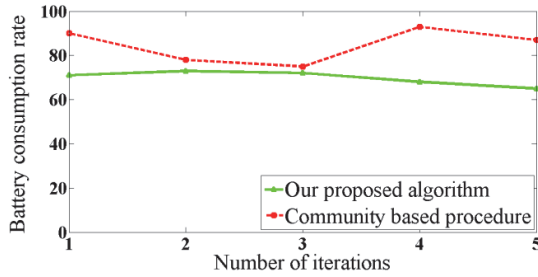


Fig. 3. Rate of battery consumption

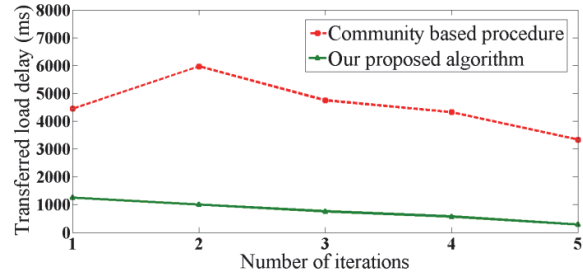


Fig. 4. Delay during load transfer

### 5.4 Mean Square Error(MSE) and Signal to Noise Ratio(SNR)Variance

Battery consumption is defined as the deterioration of energy because of usage. We have compared it for five time-slabs alike as in the previous subsection.  $MSE$  stands for the mean square error, and mathematically it is evaluated using  $MSE = (ngs - nngs)/NSC$ , where  $ngs$  is noisy group strength,  $nngs$  is non noisy group strength, and  $NSC$  represents the total number of nodes within a single cluster.  $SNR$  means the signal to noise ratio which assesses the unpredictability of the total group against its total strength. Mathematically, it can be defined as  $SNR = MaxBitValue^2 / MSE$ . The Fig. 5 illustrates different load variations and their deviations for the mean square error. The  $MSE$  rises with the increase in load. The results show that the increase in load begins steadily to increase until the third load. But, the  $MSE$  switches suddenly faster as it enters the fourth load variance. This is because several noisy nodes are attached to it at the third load variation. This makes it much quicker to intensify. The instance where  $MSE$  rises abruptly, the  $SNR$  also gets lowered. The maximum value of  $MSE$  is approximately  $14 \times 10^7$ , whereas the minimum value of  $MSE$  is about unity. The  $x$ -axis is a load variation and the  $y$ -axis is a value for the  $MSE$  or  $SNR$  respectively. The analysis is done for the number of four load variations. The value of  $SNR$  ranges in between 0.4 and 2. The Fig. 5 depicts that  $SNR$  steadily increases until the third load variation and then decreases because it gets extra noisy. The mean square error at the 3rd iteration is high and then  $SNR$  begins to decrease as shown in the figure. For the third load we obtained the highest  $SNR$  1,79, but for the first and second load variations the lowest possible  $SNR$  is measured.

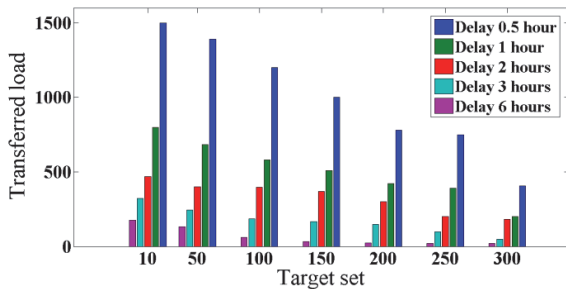


Fig. 5. Target set with transferred load

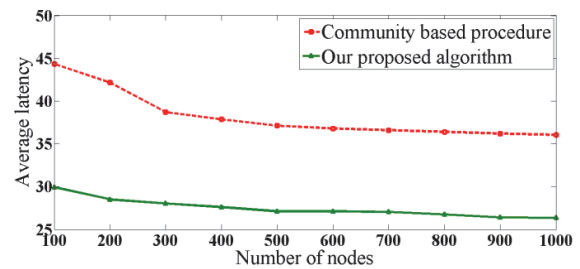


Fig. 6. Latency changes with different target set sizes

### 5.5 Comparative Analysis

We have contrasted our proposed clustering algorithm with the community-based algorithm for comparative study. The comparison is made on the factors of load transfer, latency and buffer usage. The obtained result is illustrated in the Figs. 6, 7, 8 respectively. The simulation results reveal our proposed algorithm offers about 75% to 80% less delay in data traffic offloaded from the access point on an average with approximately 16.7% more energy optimization in five iterations. Result in Fig. 7 shows that our proposed algorithm also outperforms the community-based approach in terms of latency as the number of



nodes steadily increases from 100 to 1000. The maximum latency is observed as 30% for smaller target set size which reduces further as we go on to increase the count of nodes in the target set. The reason for this downfall of latency is our proposed approach which uses the concept of a summary vector and overlapping nodes helping in faster delivery of data. In this simulation, we have considered the buffer size of each mobile node fixed of around 50 MB. However, we have also analyzed the impact over latency with varying buffer size as shown in the Fig. 8. We gradually increase the size of the message, which is to be disseminated from 10MB to 26MB. To make the simulation delay-tolerant, we considered that each of the data could be delayed up to 2 hours. As we go on to increase the buffer capacity, the latency also tends to increase. The result shows that with the enhanced buffer size the latency increases. The results also show that our proposed approach can perform better than the community-based approach for known size of target set community.

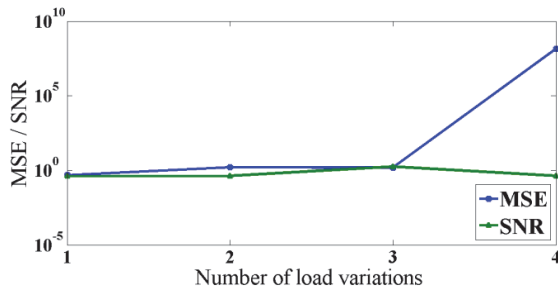


Fig. 7. Variance in MSE and SNR

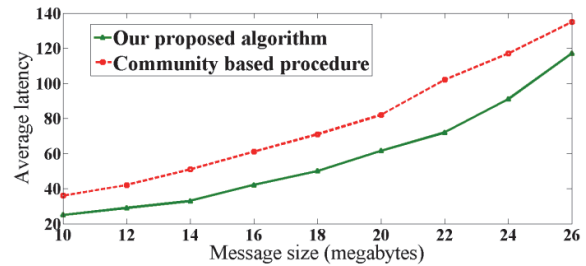


Fig. 8. Latency and message load

## 6. Conclusion and future scope

In this research paper, we have proposed an efficient D2D communication using  $k$ -means clustering-based algorithmic approach for data offloading. The proposed work not only efficiently offloads the mobile data but also optimizes the network performance in terms of energy consumption, buffer, and latency. The limitation of this work is that it does not cover the mobility aspect of the nodes in depth. However, it has fair applications to offloading in limited geographical situations such as conferences, educational institutes, residential locations especially in the pandemic like situations of COVID. The future aspect of our work will also carry the usage of meta-heuristic algorithms to optimize target set selection automatically and to reduce the delivery time of data items in disconnected data. The aspect of all users to be available as helpers needs to be exploited for the trust-based evaluation and incentive evaluation.

## References

- Aijaz, A., Aghvami, H., & Amani, M. (2013). A survey on mobile data offloading: technical and business perspectives. *IEEE Wireless Communications*, 20(2), 104-112.
- Akyildiz, I. F., Gutierrez-Estevez, D. M., & Reyes, E. C. (2010). The evolution to 4G cellular systems: LTE-Advanced. *Physical Communication*, 3(4), 217-244.
- Asadi, A., Wang, Q., & Mancuso, V. (2014). A survey on device-to-device communication in cellular networks. *IEEE Communications Surveys & Tutorials*, 16(4), 1801-1819.
- Baier, P., Dürr, F., & Rothermel, K. (2012). TOMP: Opportunistic traffic offloading using movement predictions. In *37th Annual IEEE Conference on Local Computer Networks* (pp. 50-58). IEEE.
- Bao, X., Lee, U., Rimac, I., & Choudhury, R. R. (2010). DataSpotting: offloading cellular traffic via managed device-to-device data transfer at data spots. *ACM SIGMOBILE Mobile Computing and Communications Review*, 14(3), 37-39.
- Barbera, M. V., Viana, A. C., de Amorim, M. D., & Stefa, J. (2014). Data offloading in social mobile networks through VIP delegation. *Ad Hoc Networks*, 19, 92-110.
- Bulut, E., & Szymanski, B. K. (2013). WiFi access point deployment for efficient mobile data offloading. *ACM SIGMOBILE Mobile Computing and Communications Review*, 17(1), 71-78.
- Chen, J., Jiang, A. A., Kanj, I. A., Xia, G., & Zhang, F. (2011). Separability and topology control of quasi unit disk graphs. *Wireless Networks*, 17(1), 53-67.
- Chuang, Y. J., & Lin, K. C. J. (2012). Cellular traffic offloading through community-based opportunistic dissemination. In *2012 IEEE Wireless Communications and Networking Conference (WCNC)* (pp. 3188-3193). IEEE.
- Cohen, J. (2008). Trusses: Cohesive subgraphs for social network analysis. *National Security Agency Technical Report*, 16, 3-1.
- Daly, E. M., & Haahr, M. (2008). Social network analysis for information flow in disconnected delay-tolerant MANETs. *IEEE Transactions on Mobile Computing*, 8(5), 606-621.
- Dimatteo, S., Hui, P., Han, B., & Li, V. O. (2011). Cellular traffic offloading through WiFi networks. In *2011 IEEE eighth international conference on mobile ad-hoc and sensor systems* (pp. 192-201). IEEE.
- Doppler, K., Rinne, M., Wijting, C., Ribeiro, C. B., & Hugl, K. (2009). Device-to-device communication as an underlay to LTE-advanced networks. *IEEE Communications Magazine*, 47(12), 42-49.

- Eagle, N., & Pentland, A. S. (2006). Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, 10(4), 255-268.
- Fortetsanakakis, G., & Papadopoulou, M. (2016). How beneficial is the WiFi offloading? A detailed game-theoretical analysis in wireless oligopolies. In *2016 IEEE 17th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)* (pp. 1-10). IEEE.
- Ghosh, A., Mangalvedhe, N., Ratasuk, R., Mondal, B., Cudak, M., Visotsky, E., & Dhillon, H. S. (2012). Heterogeneous cellular networks: From theory to practice. *IEEE Communications Magazine*, 50(6), 54-64.
- Han, B., Hui, P., Kumar, V. A., Marathe, M. V., Shao, J., & Srinivasan, A. (2011). Mobile data offloading through opportunistic communications and social participation. *IEEE Transactions on Mobile Computing*, 11(5), 821-834.
- Han, B., & Srinivasan, A. (2012). Your friends have more friends than you do: identifying influential mobile users through random walks. In *Proceedings of the thirteenth ACM international symposium on Mobile Ad Hoc Networking and Computing*, (pp. 5-14).
- Hui, P., Chaintreau, A., Scott, J., Gass, R., Crowcroft, J., & Diot, C. (2005). Pocket switched networks and human mobility in conference environments. In *Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking*, pp. 244-251.
- Index, C. V. N. (2017) Global mobile data traffic forecast update, 2016–2021.
- Ioannidis, S., Chaintreau, A., & Massoulié, L. (2009). Optimal and scalable distribution of content updates over a mobile social network. In *IEEE INFOCOM 2009* (pp. 1422-1430). IEEE.
- Jain, S., Fall, K., & Patra, R. (2004). Routing in a delay tolerant network. In *Proceedings of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications* (pp. 145-158).
- Jiang, J., Zhang, S., Li, B., & Li, B. (2015). Maximized cellular traffic offloading via device-to-device content sharing. *IEEE Journal on Selected Areas in Communications*, 34(1), 82-91.
- Jung, B. H., Song, N. O., & Sung, D. K. (2013). A network-assisted user-centric WiFi-offloading model for maximizing per-user throughput in a heterogeneous network. *IEEE Transactions on Vehicular Technology*, 63(4), 1940-1945.
- Kaufman, B., Lilleberg, J., & Aazhang, B. (2013). Spectrum sharing scheme between cellular users and ad-hoc device-to-device users. *IEEE Transactions on Wireless Communications*, 12(3), 1038-1049.
- Kouyoumdjieva, S. T., & Karlsson, G. (2016). Device-to-device mobile data offloading for music streaming. In *2016 IFIP networking conference (IFIP Networking) and workshops* (pp. 377-385). IEEE.
- Kumar, K., Liu, J., Lu, Y. H., & Bhargava, B. (2013). A survey of computation offloading for mobile systems. *Mobile Networks and Applications*, 18(1), 129-140.
- Lee, J., Yi, Y., Chong, S., & Jin, Y. (2014). Economics of WiFi offloading: Trading delay for cellular capacity. *IEEE Transactions on Wireless Communications*, 13(3), 1540-1554.
- 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).
- Li, Y., Qian, M., Jin, D., Hui, P., Wang, Z., & Chen, S. (2013). Multiple mobile data offloading through disruption tolerant networks. *IEEE Transactions on Mobile Computing*, 13(7), 1579-1596.
- Osseiran, A., Doppler, K., Ribeiro, C., Xiao, M., Skoglund, M., & Manssour, J. (2009). Advances in device-to-device communications and network coding for IMT-advanced. *ICT Mobile Summit*, 1-8.
- Rebecchi, F. (2015). Device-to-device data offloading: from model to implementation (Doctoral dissertation).
- Sandonaris, A., Erkip, E., & Aazhang, B. (2003). User cooperation diversity-Part II: Implementation aspects and performance analysis. *IEEE Transactions on Communication*, 51(11), 1939-1948.
- Sharma, P., & Shukla, S. (2018). Optimal Target Set Selection via Opinion Dynamics. In *2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC)* (pp. 806-811). IEEE.
- Shukla, S., Kumar, N., & Misra, R. (2014). Impact of bloom filter on infection rate in epidemic forwarding for ICNs. *Wireless personal communications*, 75(4), 2165-2180.
- Srinivasan, V., Motani, M., & Ooi, W. T. (2006). Analysis and implications of student contact patterns derived from campus schedules. In *Proceedings of the 12th annual international conference on Mobile computing and networking* (pp. 86-97).
- 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).
- Wang, X., Chen, M., Han, Z., Wu, D. O., & Kwon, T. T. (2014). TOSS: Traffic offloading by social network service-based opportunistic sharing in mobile social networks. In *IEEE INFOCOM 2014-IEEE Conference on Computer Communications* (pp. 2346-2354). IEEE.
- Zhuo, X., Gao, W., Cao, G., & Dai, Y. (2011). Win-Coupon: An incentive framework for 3G traffic offloading. In *2011 19th IEEE international conference on network protocols* (pp. 206-215). IEEE.

