

Article

# A Novel Weighted Clustering Algorithm Supported by a Distributed Architecture for D2D Enabled Content-Centric Networks

Saad Aslam \* , Fakhrul Alam , Syed Faraz Hasan and Mohammad Rashid

Department of Mechanical & Electrical Engineering, SF&AT, Massey University, Auckland 0632, New Zealand; f.alam@massey.ac.nz (F.A.); f.hasan@massey.ac.nz (S.F.H.); m.a.rashid@massey.ac.nz (M.R.)

\* Correspondence: s.aslam@massey.ac.nz

Received: 10 July 2020; Accepted: 23 September 2020; Published: 25 September 2020



**Abstract:** Next generation cellular systems need efficient content-distribution schemes. Content-sharing via Device-to-Device (D2D) clustered networks has emerged as a popular approach for alleviating the burden on the cellular network. In this article, we utilize Content-Centric Networking and Network Virtualization to propose a distributed architecture, that supports efficient content delivery. We propose to use clustering at the user level for content-distribution. A weighted multifactor clustering algorithm is proposed for grouping the D2D User Equipment (DUEs) sharing a common interest. The proposed algorithm is evaluated in terms of energy efficiency, area spectral efficiency, and throughput. The effect of the number of clusters on these performance parameters is also discussed. The proposed algorithm has been further modified to allow for a tradeoff between fairness and other performance parameters. A comprehensive simulation study demonstrates that the proposed clustering algorithm is more flexible and outperforms several classical and state-of-the-art algorithms.

**Keywords:** D2D communication; distributed architectures; 5G; content-sharing; clustering algorithm; content-centric networking; network virtualization

---

## 1. Introduction

Unprecedented demand for multicast applications is driving a move towards content-centric cellular networks. However, to accommodate such services, future networks must go beyond the capabilities of their current generation counterparts [1,2]. Cellular users are actively engaged in generating and sharing the content of various types [3]. Therefore, future networks will require to handle significantly higher multimedia services [4,5]. Existing centralized architectures and mechanisms may not be able to meet the content-sharing demands [2,3]. Decentralized mechanism and load mitigation in cellular networks are required to meet the rising demand of content sharing. One of the techniques that can effectively address load mitigation is device-to-device (D2D) communication [5]. It is different from the conventional cellular communication where all communication goes via the core network irrespective of proximity of the devices [6]. Various architectural frameworks have been proposed in the literature to support D2D-based content sharing. The concept of clustering users in proximity sharing a common interest has been very popular for multicasting scenarios [3]. Typically, an intermediate node, termed as cluster head (CH) fetches the content from the Base Station (BS) and delivers it to several content requestors [7–9]. Medium Access Control (MAC) strategies utilizing clustering for reducing energy consumption specific to multicasting scenarios are being proposed [8–10]. It is reported in the literature that clustering improves the D2D caching efficiency and communication [11,12]. We presented the effectiveness of utilizing clustering in a D2D multicasting scenario in one of our previous works [13]. Formation of appropriate clusters is significant as it

influences the performance of the underlay cellular networks [3]. Considering all these, we propose a multifactor weighted clustering algorithm in this article. The algorithm utilizes proximity, channel gain, and channel variance (details in later sections) to form the clusters.

To realize clustering that supports content-sharing via D2D, a suitable architecture is necessary that not only conforms to the standards of future cellular networks but is also distributive in nature. Moreover, it should be capable of handling a high user density. Therefore, we propose a decentralized architecture for content-sharing that is suited for 5G and future cellular network. This architecture utilizes Content-Centric Networking (CCN) and Network Virtualization (NV), key technologies of 5G architecture [14–19]. Our proposal also employs the frame structure utilized in the published article on 5G [20–23]. The details regarding the proposed architecture can be found in Section 3.

Recent literature suggests that the geographical distribution of mobile users play a vital role in successful content-caching [24]. Moreover, to ensure nearby availability of content, caching at a D2D device should consider social ties and requests pattern. In a real-scenario, different social events such as sporting events and concerts can significantly influence the clustering schemes and the multicasting scenario. Consider the example of a concert where users are interested in the same videos of the artist. In such cases where users share a strong social relationship, a user (e.g., CH) can easily contribute to the distribution of the video. Therefore, it is important to identify users with common social characteristics. In this study, the proposed decentralized architecture is supported by simple hash-based functions that have been previously used in multimedia broadcast networks for identifying users with a common interest.

The major contributions of this research work are as follows:

- A distributed architecture is proposed that is effectively supported by hash functions to identify the socially connected users. This is in contrast to the majority of the published works on D2D multicasting that do not consider distributed architecture along with content-identification.
- A novel multifactor weighted clustering has been proposed. The performance of the proposed algorithm is shown to be superior compared to five benchmarked algorithms. In addition, the weights of the algorithm can be adjusted to suit the system's requirements. This flexibility in trading off the performance with respect to various parameters is not available for existing algorithms.
- The benchmarked algorithms are tested for throughput fairness which has not been reported in the literature on clustering. Moreover, different from the existing works, the impact of the number of clusters on the energy consumption and area spectral efficiency is also demonstrated.
- To the best of the author's knowledge, reported work in the literature considers either the spatial distribution of users or users' social ties for their respective clustering algorithms. We propose to include both to make the clustering process comprehensive and evaluate its impact on the system's performance.

The rest of the manuscript is organized into five sections. Relevant literature and related work have been summarized in the next section. Section 3 presents the details of the proposed architecture and content identification technique. The proposed clustering algorithm is described in Section 3 as well. Section 4 presents the system model including the simulation setup. It explains all the major assumptions of the simulation environment, and mathematical models of the performance parameters. Section 5 discusses the performance of the proposed scheme benchmarked against existing methods. The findings of this research work are summarized in the Conclusion section while discussing the future directions of the proposed study.

## 2. Related Work

With exponentially increasing network devices, it is becoming difficult to fulfill the QoS requirements of multimedia services [25,26]. The growth of cellular devices has been addressed with the concept of dense networks having a large number of small cells. However, the limited

capacity of the backhaul becomes a bottleneck in such a scenario [27]. It has been proposed that the significant growth in 5G networks and beyond can be accommodated by investing more in Content-Centric Networks (CCN). The notion behind a CCN is to present a scalable and efficient mechanism for content delivery [28]. The techniques developed around CCN are expected to reduce the transmission delay by caching the data within the network. Data caching is performed closer to the group of mobile devices/content requesters by exploiting social ties (e.g., shared interest in content) or connections [11,29]. Device-level caching can be facilitated by D2D communication. D2D has been effectively used to disseminate data in various network scenarios. However, the network architectures presented in many of these works [11,29–31] considering the D2D multicasting scenario are centralized in nature and require massive message passing to make the whole scheme work. Such schemes do not meet the requirements of dense future cellular networks. Contrary to this, we have proposed a decentralized architecture that can effectively support the D2D multicasting scenario.

Researchers have investigated the technologies that address the scalability of the network and the ability to cater to the growth in wireless traffic and services. Network virtualization concept has been widely used [27–29]. Virtual Private Networks (VPN) is one of the examples. Network virtualization aims to slice the resources of cellular architecture into virtual resources to be shared among multiple users. It should be noted that by cellular resources we mean a licensed spectrum and infrastructure e.g., Core Network and Radio Access Network, etc. [29]. All the signaling and message passing that needs to take place to set up the virtual network is well researched and presented in the literature [27–29]. Our study proposes a network architecture that combines the concepts of CCN and NV. This is different to the schemes found in literature, as we merge both the technologies. There are significant advantages to the proposed merger. One of the most important features is that not only the cellular resources but the actual contents can be shared. Duplicate transmissions exhaust the cellular resources. The content-sharing among the networks with the aid of virtualization can significantly reduce these redundant transmissions. Details of the architecture are presented in the next section.

Clustering complements the proposed architecture. However, when a multicasting scenario is considered in the literature within the context of clustering, most works assume that users sharing a common interest have already been identified. Furthermore, if a system model is presented, how a typical cellular architecture can support such a model is not shown or glossed over. Most articles either discuss social tie/social interest modelling or clustering algorithm in detail but not both [32,33]. In [32], a clustering algorithm has been proposed and social metric is also considered as an important factor. However, it does not provide any details around how the users having the same interest are identified. The architecture that supports their system model is not presented either. Research work presented in [33] discusses the social ties/social attributes in detail and effectively describes the mechanism behind modelling social metrics. It also considers clustering for a multicasting scenario but the clustering itself is assumed to have taken place by placing the users in a certain grid. Therefore, clustering and its effects on the performance parameters have not been discussed. Moreover, it is important to note that [32,33] uses centralized mechanisms. In contrast to these works, we chose to address both content-identification (through the hashing concept) supporting a decentralized architecture and a clustering algorithm to complete the big picture.

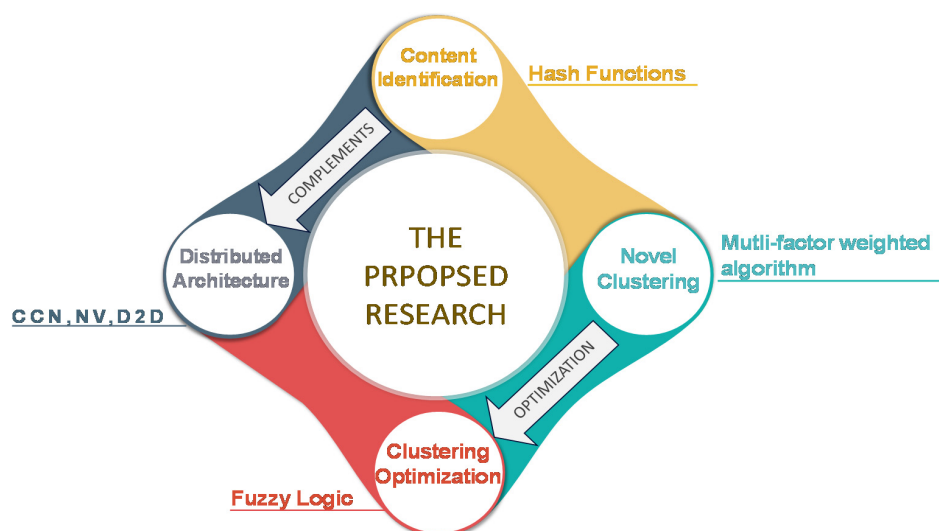
The consideration of distributed architecture, different performance parameters, and the flexibility to trade-off the performance for fairness is not seen in the recent and relevant literature. Different research works mentioned in Table 1 show the stated fact. For instance, a few works focus on throughput, whereas others provide details on energy consumption. On the other hand, only a few articles have considered Area Spectral Efficiency (ASE) and Fairness. Most importantly, all these works rely on centralized architectures. As opposed to this, our work provides the details on all these parameters while proposing a distributed architecture.

**Table 1.** Summary of related research articles.

Research	Year	Distributed Architecture	Performance Parameters			
			Throughput	Energy Consumption	ASE (Area Spectral Efficiency)	Fairness
Asadi et al. [34]	2016	×	✓	✓	×	✓
Zhang et al. [35]	2017	×	×	✓	×	×
Yang et al. [12]	2018	×	×	✓	×	×
Huang et al. [36]	2018	×	✓	×	×	×
Rahman et al. [37]	2018	×	×	✓	×	×
Pizzi et al. [38]	2019	×	✓	×	×	×
Aslam et al. [13]	2019	×	✓	✓	✓	×
Shi et al. [39]	2019	×	×	✓	×	×
Wu et al. [40]	2019	×	✓	×	×	×
Wang et al. [41]	2019	×	×	×	✓	×
Zhou et al. [42]	2020	×	×	✓	✓	×
Our Proposal	2020	✓	✓	✓	✓	✓

### 3. Proposed Distributed Architecture and Clustering Algorithm

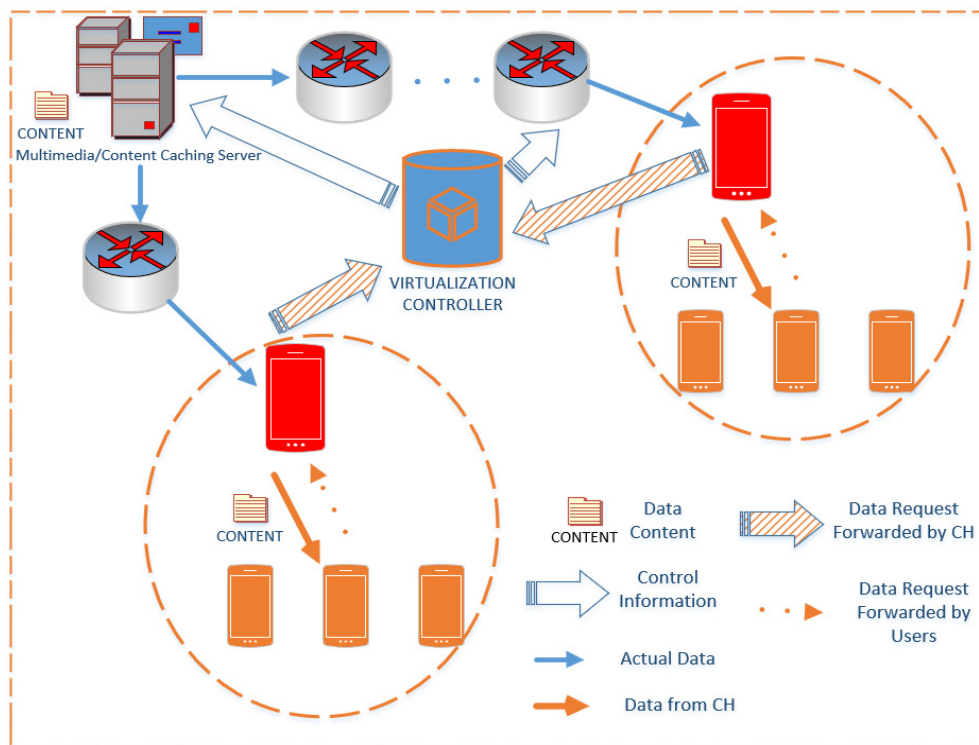
The proposed concept is illustrated in Figure 1. The utilization of clustering for content-sharing is the key concept of the proposed architecture. The architecture is based on CCN and NV. Hash functions identify users with common interest who are then organized in clusters based on the proposed multifactor algorithm. The clustering algorithm is optimized using fuzzy optimization which is useful in optimizing clustering as well as other parameters of a cellular network [43–46]. Each cluster has a cluster head (CH) that are responsible for multicasting the required information to the cluster members. After clustering, all the content-requests from any given cluster traverse through the CH and users are served via the CH utilizing D2D communication.



**Figure 1.** Summary of the Proposed Mechanism. The four critical aspects are: distributed architecture implementation using Content-Centric Networking (CCN) and Network Virtualization (NV), identification of users with common interest using hash functions, cluster formation with the proposed multifactor algorithm and optimizing the clustering algorithm using fuzzy logic.

The concept of CCN is predicated on the requested content reaching the requester without having the need to reach the content publisher/provider [29]. Therefore, caching the requested content at an intermediate node will enable the content-delivery with reduced energy consumptions and latency. Once the intermediate node has cached the content, it can be provided to several requesters.

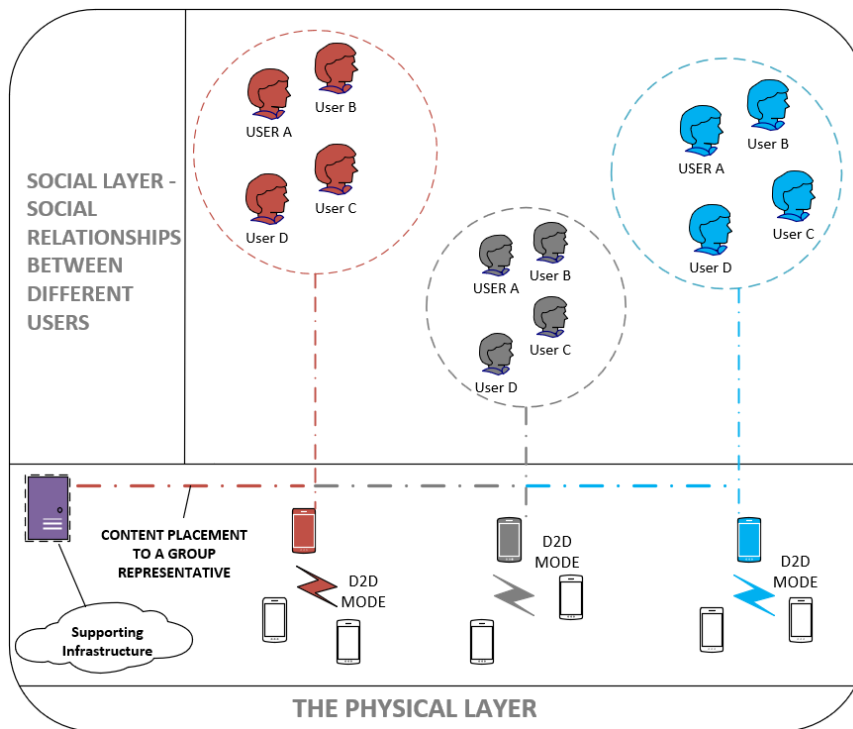
Architectures supporting Content-Centric Networking have been proposed in the literature [28,29,47,48]. However, these works consider centralized mechanisms. The architectures presented in [28,29] involve BS, and significant signaling is required to take place between BS and the D2D nodes, before the content-delivery. On the other hand, [47] does not provide any details on the architecture. The work presented in [48] does consider decentralized mechanisms, but it does not explicitly show any architecture that supports their mechanism. Our research considers a similar approach as presented in [28,29] with necessary modifications to accommodate the clustering of users for content-sharing scenarios and making the scheme distributive. Figure 2 shows the network model of the proposed content-centric architecture. It is different from a conventional wireless network such as Internet Protocol (IP). The basic difference lies in the establishment of the connection. IP based networks first establish the connection between the requestor and the provider before the content is delivered. In contrast, the content is requested without the establishment of the connection with the host/content-provider in CNN. The proposed architecture utilizes Content-Centric Networking as well as network virtualization. The controller for virtualization, shown in Figure 2, is responsible for providing the location of the content-holder as well as setting the virtual infrastructure components for content-delivery. A mobile user requests certain content without the information of the host which holds that content. Another important entity of the proposed architecture is the caching server. It is an integral part of the network which caches popular contents and reduces duplicate transmissions.



**Figure 2.** The proposed distributed network architecture. Each red device represents a cluster head that forwards the requested content to the virtualization controller that connects with the multimedia servers to fetch the contents.

Figure 3 shows the layered network. The first layer, termed as a social layer, represents the social ties that exist among different groups of users. The physical layer represents the mobile devices that model

the communication taking place in various clusters represented by a CH. It also shows the supporting infrastructure required to set up the communication and making the content-delivery possible.



**Figure 3.** The visualization of a layered network showing the interaction between the social users and the corresponding physical layer.

### 3.1. Content Identification Using Hash Functions

One of the important aspects of the proposed architecture is the mechanism that addresses the content-identification. We propose to utilize hash functions [49] for content-identification. Hash functions perform the mapping between the given data and hash of a specific length. The size/length of the output of a hash function does not depend on the length of the input. Hash can be regarded as a “signature” for a given text [49–51]. One of the major applications of hash functions lies in the field of multimedia broadcast networks, as a content identifier [49–51]. The hash function aids the network by providing content identification to easily determine which content has been broadcasted, timing information, and to what station. Several hashing algorithms exist in the literature; we suggest using SHA-256 due to its reduced complexity and speed [52,53].

The binary sequence generated by the hash functions for a particular “text” or “name” will always be the same as shown in Figure 4. Therefore, if we produce the hash of the different contents at the content-servers, the hash value can be matched with the one generated by the content requestors. A match means that the same content is being demanded. Groups requesting the same content can, therefore, be identified based on the hash values. It is clear that hashing will not only help in identifying the content but also the group of users sharing the same interest. Therefore, we believe that hashing is a natural choice.

### 3.2. The Proposed Clustering Algorithm

Clustering commences once users demanding the same content have been identified. The user clustering process consists of three main steps: selection of appropriate clustering metrics, identification of the devices suitable for being a CH, and finally, associating the cluster members with their respective CHs. The overall clustering process is shown in the flow chart of Figure 5.

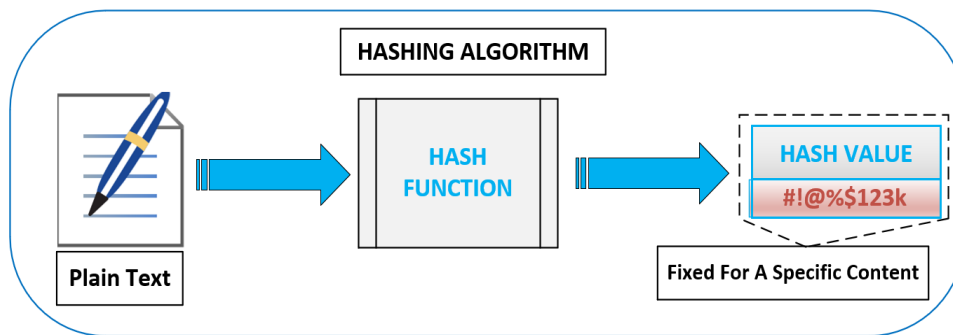


Figure 4. The hash function: plain text to hash value.

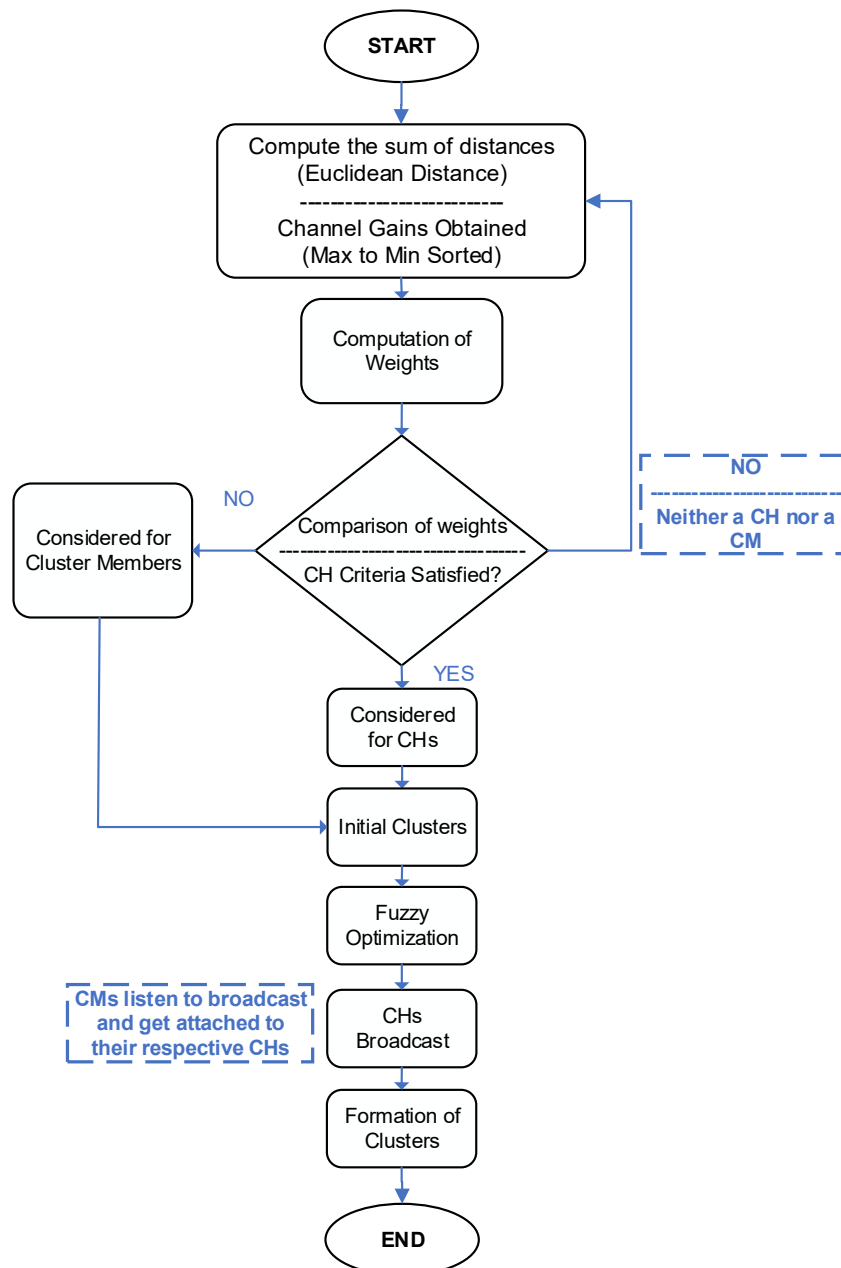


Figure 5. Flow chart of the proposed clustering algorithm.

### 3.2.1. Weighted Clustering Approach

CHs are selected on a per-frame basis. The duration of one frame is 10 ms following the relevant literature. All the users are considered CHs for the first frame; therefore, the clustering algorithm is implemented for the next frames. It is assumed that every node is capable of being a CH and has enough energy [12]. The position of the nodes remains the same during one frame. However, for the next frame user distribution/placement of users change and, therefore, every simulation represents a different user distribution. This is in accordance with the standard literature relevant to multimedia multicasting scenarios [11,12].

After the initialization, the algorithm gathers the information about the clustering metrics and clusters are formed, details of which can be found in the subsequent sections. Before the clustering takes place, the distance among the devices and the channel conditions are obtained and conveyed to all the neighbors during the discovery phase as explained in the next subsection. Based on the information received, CHs announce its cluster members. It is assumed that multiple users can be detected simultaneously by the CH. All the members listen to the broadcast of the CHs and get attached to the one that serves them the best considering distance and channel conditions.

### 3.2.2. Device Discovery

Though the device discovery is out of the scope of this research, we utilized the information obtained through the device discovery. Therefore, the authors are describing the discovery process within the context of the proposed algorithm.

Before the clustering takes place, it is necessary to discover the devices and create a neighbor list. We assume, as is common practice in the literature [54–58], that the neighbor list is available with the nodes. For these tasks, we propose to use the Peer Discovery Resource (PDR). PDR represents a resource unit, used to transmit the discovery signal or beacon signal. Two of the standard PDR structures that are used in published literature are LTE-A and FlashlinQ. Literature suggests that a considerable amount of information can be conveyed using either of these structures [34,56,57]. Moreover, different research works have utilized the PDR to send the clustering-related information [55,56]. We propose to utilize the same concept and use the PDR to send the information regarding the predefined clustering metric detailed in the subsequent section. Therefore, the signaling load for the proposed clustering scheme will be accommodated by standard signaling taking place in a D2D network.

The users in proximity to one another receive the discovery signals. Devices decode this signal containing information such as device or user ID and its link characteristics (such as SINR, channel conditions) with the user. Based on these characteristics, a device decides which of the users whose signal it has received can be classified as neighbors. There are various advanced channel estimation algorithms and processes for 5G networks [58] that can be utilized for this purpose. Once the neighbor detection has taken place, every user possesses the list of its neighbors.

### 3.2.3. Clustering Metrics

The selection of clustering metrics significantly impacts the system's performance. The two factors of the proposed clustering algorithm include distance and channel conditions among the users. Each factor is assigned to its respective weight. These metrics are selected due to their effect on target performance parameters such as throughput, area spectral efficiency, and energy consumptions. The significance of this selection will be further highlighted in the results section.

#### 1. The Distance among the Nodes

Recent literature has identified the significance of the spatial distribution of users as it directly influences the caching efficiency [24]. Furthermore, there is a high probability of successful D2D transmission if the devices are in proximity [11]. Hence, we chose distance among the nodes as an important metric for forming appropriate clusters. It is also important since users that do not



exist in proximity are not ideal candidates for being a part of the same cluster even with a strong social relationship.

## 2. Channel Conditions

Since we are considering a multicasting scenario where a CH will be communicating with the rest of the cluster members, it is important for the cluster members to have good links with the CH. If we ignore these conditions, both inter-cluster and intra-cluster communication might be impaired. Therefore, we believe that channel conditions between the prospective CH and its cluster members is an important metric.

### 3.2.4. Cluster Head Selection

During the CH selection process, devices use PDR to broadcast beacons continuously. These beacons include a predefined metrics (e.g., distance, channel conditions). Every device decoding the beacon stores the corresponding metric and its identifiers (see Section 3.2.2). This information is vital for the devices to select a CH and delegating the control to it for further communication. If a certain device is not able to receive a beacon signal, it might be out of reach of another device, and it can self-select itself as CH. Once the metric information is received from the beacons, all the devices compare their metrics. The devices with the lowest metric values are identified as CHs. The remaining becomes the cluster members. It should be noted here that all the users need to fulfill the predefined criterion to be considered for clustering. It is based on the social relationship among the devices. The following steps summarize the proposed algorithm.

Step 1: Determine the neighbors of each node using D2D discovery. Parameters of interest are stored.

Step 2: Determine the nodes sharing the same interest/content using the hash function.

Step 3: Compute the sum of distances (Euclidean Distance) for all the nodes against all their neighbors.

$$D_{(a,b)} = \sqrt{(a_x^2 - b_x^2) + (a_y^2 - b_y^2)} \quad (1)$$

where  $a$  and  $b$  represent any two neighboring devices.

Step 4: Weights of the nodes are accumulated as follows:

$$W_T = I \cdot \left[ w_1 * D_{(a,b)} + w_2 * \left( \frac{1}{h_{ab}} \right) \right] \quad (2)$$

$w_1$ ,  $w_2$  represents the weights given to distance, channel gains, respectively.  $h_{ab}$  represents the channel gain between the nodes  $a$  and  $b$ . The weights represented in Equation (2) are such that  $\sum_{f=1}^2 w_f = 1$ . The node with the minimum  $W_T$  is chosen as the CH.

Since only nodes sharing a common interest should be considered for clustering, the total weight is being multiplied with a binary interest-factor denoted by " $I$ ", so that if  $z = w_1 * D_{(a,b)} + w_2 * \left( \frac{1}{h_{ab}} \right)$ , then,

$$W_T = \begin{cases} z, & I \neq 0 \\ 0, & I = 0 \end{cases} \quad (3)$$

Step 5: Compare the weights for each node and select the cluster head corresponding to the smallest  $W_T$ .

Step 6: For the remaining devices, repeat steps 3 and 4, until each node is either selected as a CH or a CM.

Step 7: Clustering optimization using Fuzzy (details in Section 3.2.6).

### 3.2.5. Feature Scaling for the Clustering Metric

The two different factors for clustering (Equation (2)) do not have the same range of values. Therefore, data normalization was performed [59]. There are different normalization techniques available in the literature such as Min-Max normalization, decimal scaling, and Z-score scaling [59,60]. Min-Max and decimal scaling do not handle outliers very well. Therefore, we chose to utilize the z-score normalization given by Equation (4).

$$Z_{score} = \frac{x - \bar{x}}{\sigma} \quad (4)$$

where  $x$  is the original value (e.g., channel condition) for which we calculated the z-score, and  $\bar{x}$  and  $\sigma$  represent the mean and the standard deviation of  $x$ , respectively.

### 3.2.6. Fuzzy Optimization of Clustering

The initial clusters formed based on the proposed algorithm need to be optimized. Therefore, a fuzzy optimization technique was applied. Fuzzy optimization partitions the users into  $C$  clusters based on the proposed criterion of clustering. Each input to this function is attached to an attribute such as the weights in our study. Fuzzy optimization is based on a partition matrix [61]  $P \in Q_{FO}$  where;

$$Q_{FO} = \left\{ P \in [0,1]^{C \times N} \mid \sum_{i=1}^c p_{ik} = 1, k=1, \dots, N, \sum_{i=1}^c p_{ik} > 0, i=1, \dots, c \right\} \quad (5)$$

The objective function of Fuzzy Optimization is given by Equation (6) [61]:

$$OF(P, V) = \sum_{i=1}^c \sum_{k=1}^N p_{ik} \|X_k - V_i\|^2 \quad (6)$$

Equation (6) represents the objective function where  $C$  is the number of clusters and  $V$  is the set of cluster centers.  $N$  represents the number of samples (users in our case) and  $X_k$  is the  $k$ th calculated sample where  $\|\cdot\|^2$  represents the Euclidean norm, and  $p_{ik}$  denotes the membership of  $X_k$  to cluster  $i$ . Each element of the partition matrix is a measure of the extent to which a particular user belongs to a certain cluster. The complete optimization process is explained in the flow chart given in Figure 6.

### 3.2.7. Communication

After the selection of CHs, they broadcast a message containing their IDs. They can use the same PDR used for neighbor detection to broadcast the results once the broadcast is received, and all the non-CH devices select those CHs to which they are closest and receive better channel conditions. The cluster members then associate themselves with a certain cluster, and the formation of the clusters is complete. The above-mentioned procedure is completely decentralized which is very important for dense networks. Once the clusters are formed, communication of all the cluster members goes via the CH. The operating phases of the proposed algorithm are shown in Figure 7. The next frame follows the same activities.

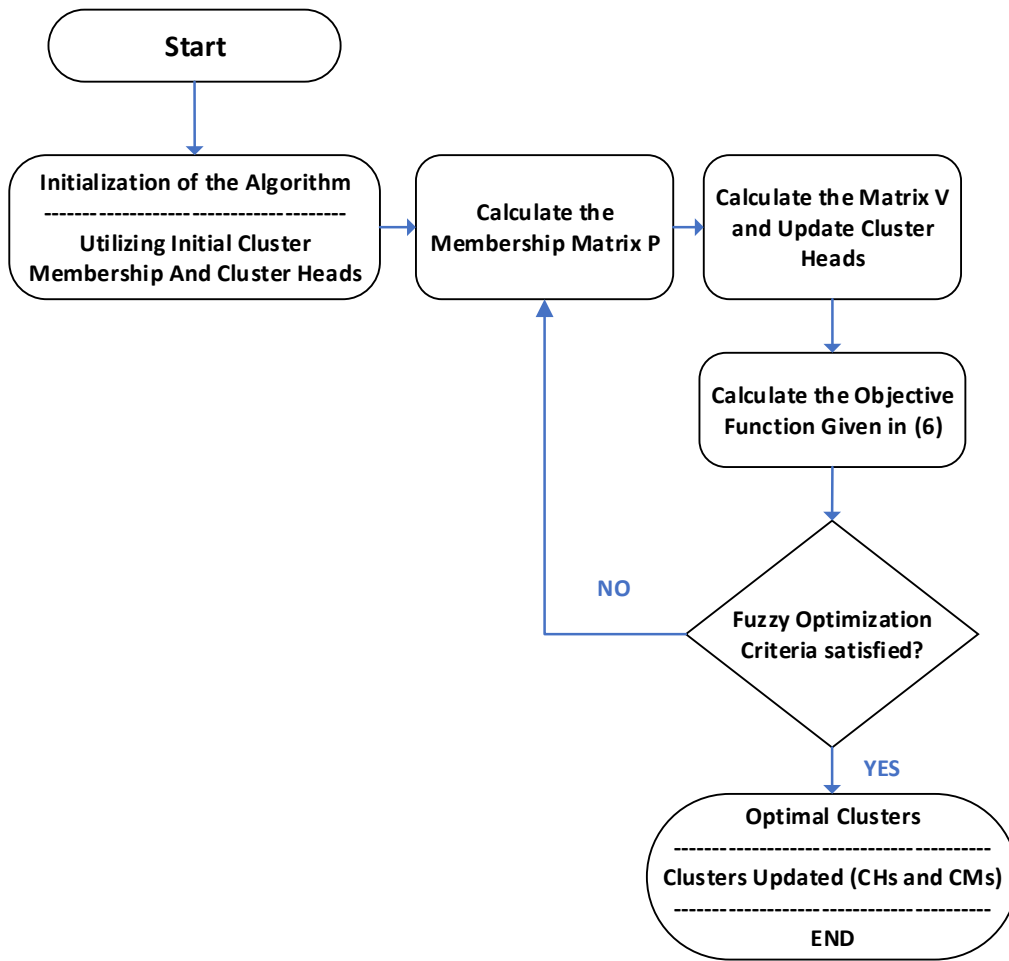


Figure 6. Optimization flow chart.

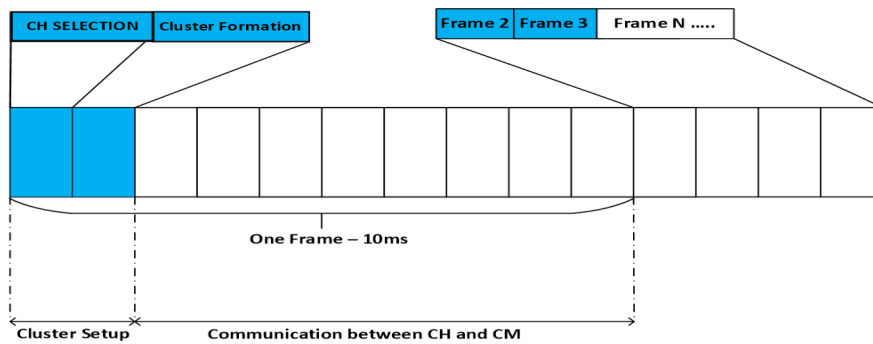


Figure 7. Frame structure for clustering.

#### 4. System Model and Simulation Setup

We consider a single cell where users are randomly distributed. In-band D2D communication using the underlying concept is considered. In this case, D2D reuses cellular resources. These techniques are well researched [12,62,63]. The reason for considering the underlying concept is that reutilizing the resources improve spectral efficiency. However, it creates interference and, therefore, was considered in our simulation scenario. Conventionally, BS provides the requested content; however, it comes at the expense of increased energy consumption and usage [62,63]. In contrast, the CH is responsible for delivering the contents to the requestors as depicted in Figure 2. Once the data have been fetched by the CH, the requested content is distributed utilizing the D2D multicast communication.

#### 4.1. Mathematical Models for Performance Parameters

##### 4.1.1. Achievable Rates for Cluster Head and Cluster Members

There are total  $N$  users in the network which constitute the set  $N = \{m_1, m_2, m_3, \dots, m_N\}$ . The CHs and cluster members are indexed as  $j$  and  $k$ , respectively. For clarity, all the other symbols are summarized in Table 2.

**Table 2.** List of Symbols.

Symbol	Representation
$N$	Set comprises of all the users
$k$	Index of cluster member
$CH$	Cluster Head
$R_{CH_j}$	Achievable Rate of $CH_j$ when receiving the contents from the Base Station (BS)
$SNR_{CH_j}$	Signal-to-Noise Ratio of a $CH_j$
$N_o$	Noise Spectral Density
$B$	Bandwidth of the Transmission Channel
$h_{BS,CH_j}$	Channel Gain between the BS and the $CH_j$
$P_B$	Transmit Power of the BS
$R_{m_k}$	Achievable Rate of cluster member $m_k$
$h_{m_k,CH_j}$	Channel Gain between the cluster member $m_k$ and $CH_j$
$P_{CH_j}$	Transmit Power of the $CH_j$
$F_S$	File Size (size of the demanded content)
$P_{chrx}$	Power consumed by the CH to receive the contents from BS
$P_{mrx}$	Power consumed by the cluster member to receive the content from cluster head (CH)

The achievable rate at the CH can be written as:

$$R_{CH_j} = B \log_2(1 + SNR_{CH_j}) \quad (7)$$

where the SNR of the  $CH_j$  is given by:

$$SNR_{CH_j} = \frac{P_B h_{BS, CH_j}}{N_o B} \quad (8)$$

Therefore, we may write Equation (7) as:

$$R_{CH_j} = B \log_2 \left( 1 + \frac{P_B h_{BS, CH_j}}{N_o B} \right) \quad (9)$$

Since we are considering a multicasting scenario, the achievable rate depends on the worst physical link. Otherwise, the successful reception of the content for all cluster members cannot be made certain. Therefore, the achievable rate at the cluster member  $m_k$  can be written as follow:

$$R_{m_k} = B \log_2 \left( 1 + \frac{P_{CH_j} h_{m_k, CH_j}}{N_o B} \right) \quad (10)$$

It should be noted that  $R_{m_k}$  is the minimum achievable rate to make sure that all the cluster members receive the content.

#### 4.1.2. Energy Model

Downlink energy consumption is considered in this study. We utilized the energy consumption model presented in [64]. We assumed that the content demanded by the users is a file of size “ $F_S$ ” bits. Suppose this file needs to be transmitted from  $CH_j$  to cluster member  $m_k$  with an achievable data rate of  $R_{m_k}$ . The time required to transmit this file is  $\left(\frac{F_S}{R_{m_k}}\right)$  seconds. Therefore, energy consumption  $E_C$  in one of the clusters “ $C$ ” can be written as:

$$E_c = \frac{F_s P_{chrx}}{R_{CH_j}} + \frac{F_s P_{CH_j}}{R_{m_k}} + \sum_{\substack{j \neq m \\ \forall m}} \frac{F_s P_{mrx}}{R_{m_k}} \quad (11)$$

Equation (11) represents the sum of three independent terms. Energy consumption of CH to receive data is represented by the first term, whereas the second term represents the energy consumed by the CH to transmit the data to their cluster members. The sum of the energy consumed by the cluster members to receive the demanded content is shown by the last term in Equation (11).

#### 4.2. Simulation Setup

The simulation environment was built on MATLAB. We considered a network model similar to Figure 2. A single cell of 1 sq.km area was considered. For the conventional cellular communication scenario, the BS was placed at the center of the cell. Moreover, it is important to mention that we explored the performance of a multimedia application (content). The packet size is 100 kB as suggested by relevant literature [65]. This simulation can easily be extended for any other multimedia application (e.g., video broadcast, eHealth, etc.) by varying the file size and packet interarrival rates [65]. We selected the weights empirically, which can be adjusted according to the system requirements. The number of clusters formed to produce all the results were chosen using the Calinski–Harabasz criteria [66]. Various user densities have been considered to produce the results. The optimum number of clusters for various user densities is different, and hence, a specific number is not explicitly mentioned. All the simulation parameters of interest are detailed in Table 3. Parameters related to channel and energy consumption are adapted from the relevant literature [64,67,68].

**Table 3.** Simulation parameters.

Parameters	Value
Simulation Platform	MATLAB
Channel Model	Rayleigh Distributed
User Placement	Uniformly Distributed
Node Density	100 to 1000
Cluster Size	Variable
Number of Clusters	Variable
Transmit Power of CH	1.425 Joules/s
Power required to receive data from BS	1.8 Joules/s
Power required to receive data from CH	0.925 Joules/s
Content Considered	A file of size 100 kBits
Classical benchmarked Schemes	K-Medoids (KM), Genetic Algorithm (GA) and Fuzzy C-Means (FCM)
State-of-the-art benchmarked Schemes	Proposed in [69]. (referred in this document as benchmarked I) Proposed in [70]. (referred in this document as benchmarked II)
Number of Simulation Runs	10,000

## 5. Results and Discussion

### 5.1. Impact of Clustering and Social-Interest

The proposed algorithm takes clustering and social interest into account, as both have a significant impact on the system. To demonstrate this impact, we consider three different scenarios. In the first scenario, conventional cellular communication takes place that does not involve D2D mode and clustering. The other two cases consider the proposed clustering algorithm, explained by the following text.

- Clustered D2D users with no interest factor

In this case, we assume that users do not share a common interest i.e., all of them are not interested in a single file (content). Users demand files of different sizes varying from 10 to 100 kB in a random manner. Though this scenario does not consider the social-factor, we still clustered them, as the literature suggests that even without the social-factor, clustering yields significant throughput gains [3,8,11,23]. The clustering criteria for these nodes are the same as mentioned in Equation (2) except that the interest-factor “ $I$ ” is not considered. The weights selected are as follows:  $w_1 = 0.4$ ,  $w_2 = 0.6$ . These were empirically selected to maximize throughput performance.

- Clustered D2D users with interest factor

The third scenario considers the social interest i.e., all the users in a given cluster are interested in a single file of size 100 kbits. This emulates social gatherings such as a concert or a stadium, where there is a large gathering, interested in a similar video/content. This scenario was implemented using the proposed algorithm. The value of the two weights remains the same as discussed in the previous scenario.

Figure 8 shows the result of aggregate throughput versus the number of users. It clearly shows the impact of social awareness as the aggregate throughput was maximum when it was considered. On the other hand, aggregate throughput was considerably low when social awareness was ignored. At the user density of one hundred, the difference between the two curves was approximately 19%. The throughput for a conventional cellular network with no clustering remained considerably low compared to the other two scenarios. This result, therefore, shows that clustering does play a vital role in enhancing the system’s performance. Furthermore, it can be seen that both social-interest and physical parameters (e.g., spatial distribution and channel gains) should be considered while modelling a system as it may bring significant benefits for the users as well as the whole network.

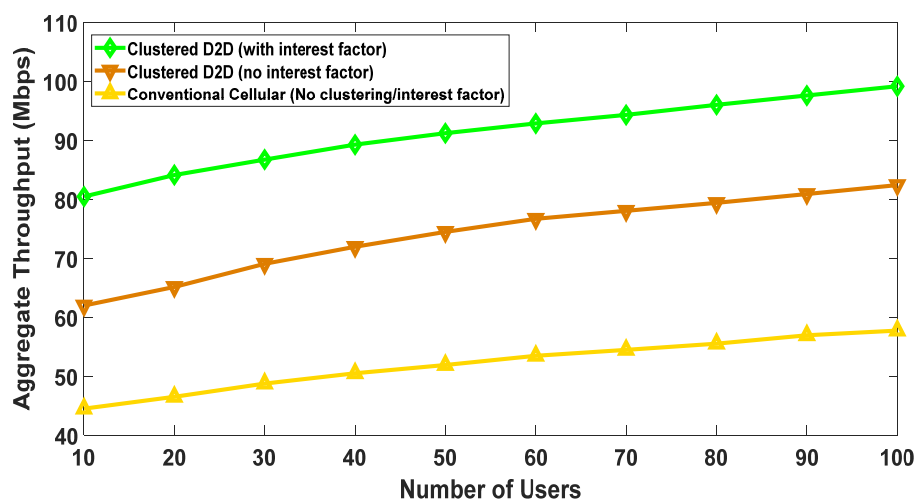


Figure 8. The impact of clustering and social-interest on throughput.

## 5.2. Benchmarking against Existing Algorithms

We selected five algorithms to benchmark against. Three of these are classical algorithms that are widely found in the literature, namely K-Medoids, Fuzzy C-Means (FCM), and Genetic Algorithm (GA) based clustering. These three algorithms have not been investigated and benchmarked within the context of D2D clustering and content-sharing applications, though an initial investigation was performed in our previous work [13]. The remaining two are the state-of-the-art and recently proposed algorithms. “Benchmarked I” has been proposed by Tulu.M.M et al. [69]. This algorithm applies the concept of entropy of betweenness centrality (EBC) to select CHs for content-sharing. The entropy of betweenness is based on the social relationship between the nodes and the shortest paths that exist between the nodes. “Benchmarked II” is proposed by Kazez C.A et al. [70]. This algorithm takes the neighbors and distance among the users as inputs for the selection of CHs.

### 5.2.1. Throughput Comparison

The following result shows the comparison of the throughput performance. The proposed algorithm utilizes the social interest and physical parameters of the users to enhance the system’s performance. This was discussed in the previous result, and it is further elaborated in Figure 9, as it demonstrates that the proposed algorithm performs approximately 7% better than the next best algorithm (Benchmarked I) at one thousand nodes. Benchmarked algorithms I and II utilize the social interest, but they do not consider both distance and channel conditions among the users for clustering the users. Our result shows that consideration of both metrics does have a positive effect on the system’s throughput. This is because many users that are in proximity to each other may not have better channel conditions due to various factors (e.g., shadowing).

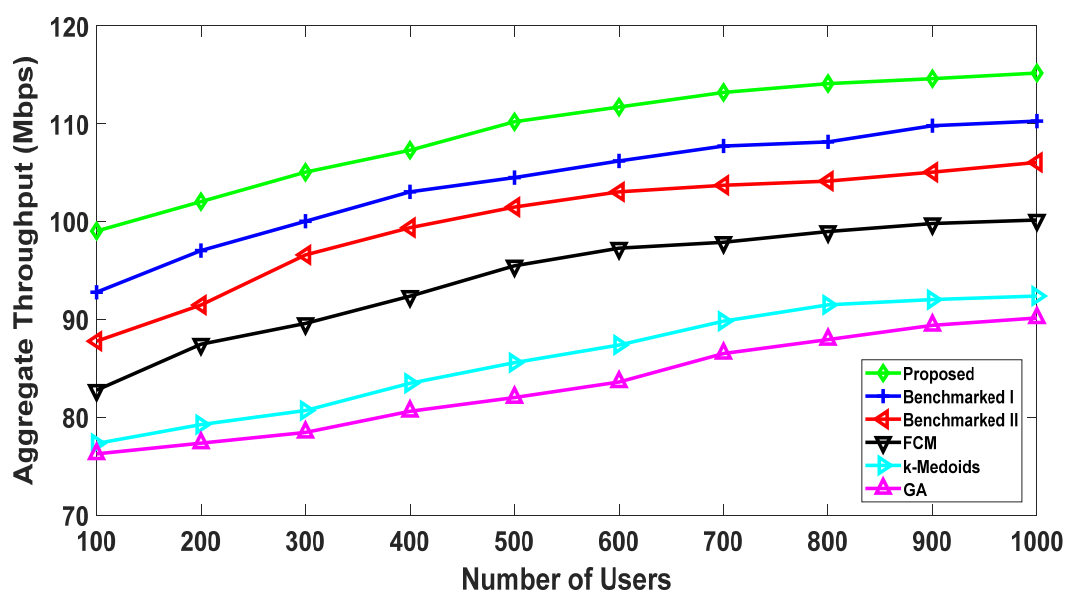


Figure 9. Aggregate throughput: comparison with the benchmarked ( $w_1 = 0.4$ ,  $w_2 = 0.6$ ).

### 5.2.2. Energy Consumption of Users

The result shown in Figure 10 represents the energy consumption of the nodes in Joules with a varying number of users. It is evident from Figure 9 that we achieved better throughput as compared to the rest of the algorithms. If the file size of 100 kbits is constant, then the energy consumptions will be significantly dependent on the transfer rate. Consequently, the proposed algorithm performed the best (demonstrated by least energy consumptions) at different user densities as compared to the other algorithms. At one thousand nodes, the proposed algorithm approximately consumed 6% less energy as compared to the second-best algorithm.

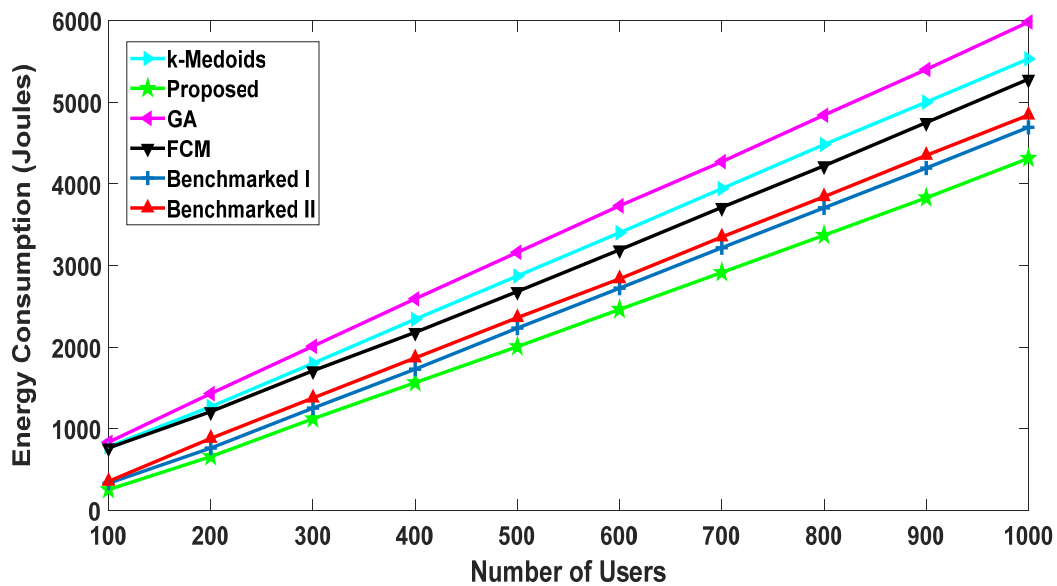


Figure 10. Energy consumption: comparison with the benchmarked ( $w_1 = 0.4$ ,  $w_2 = 0.6$ ).

The energy consumption of the proposed algorithm is further elaborated in Figure 11. The Cumulative Distribution Function (CDF) of the energy consumption is presented for the proposed algorithm and the benchmarked clustering algorithms at a user density of one thousand. We can observe that even at the node level, the proposed algorithm outperformed the benchmarked algorithms in all quartiles with regard to energy consumption. Therefore, the overall lower energy consumption was not achieved by favoring a few nodes to a large extent while disregarding the others.

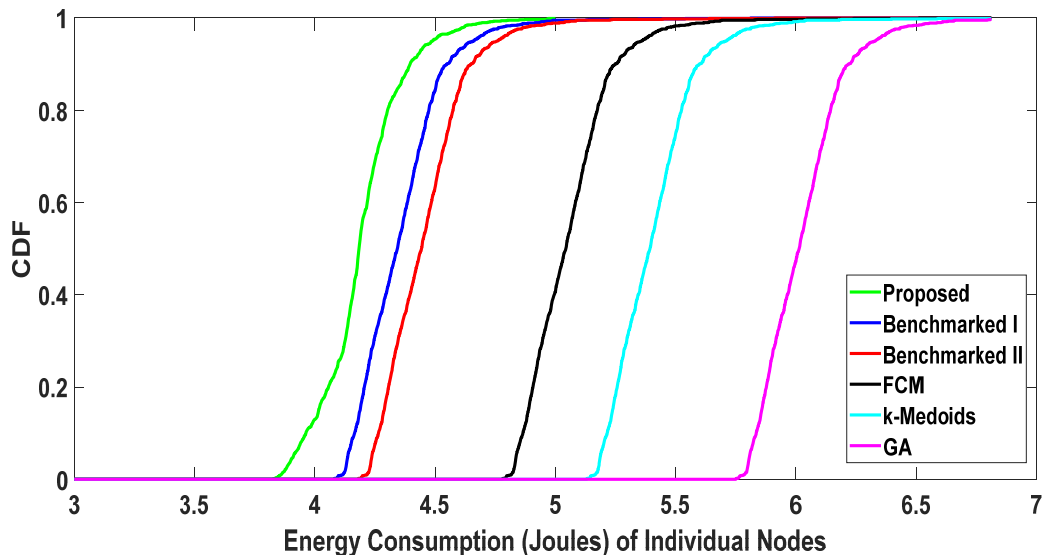


Figure 11. Cumulative Distribution Function (CDF) of energy consumption ( $w_1 = 0.4$ ,  $w_2 = 0.6$ ).

### 5.2.3. Area Spectral Efficiency

Area Spectral Efficiency represents the sum of average achievable rates per unit bandwidth per unit area [71]. To the best of the authors' knowledge, ASE has not been evaluated for all the five benchmarked algorithms. It can be observed in Figure 12 that the ASE of the proposed algorithm was better than all the benchmarked algorithms. ASE depends significantly on average rates of the users if the area and per unit bandwidth remain constant. Therefore, the proposed algorithm has higher ASE. Furthermore, the performance was better than all the other benchmarked schemes. It is also



encouraging to observe that the performance improved for the proposed algorithm as the user density increased. This shows the scalability of the proposed algorithm. The proposed algorithm showed approximately 3% improvement in ASE at the node density of one thousand, as compared to the benchmarked scheme I that showed the second-best performance. The classical algorithms for generic clustering are not purpose-built for a D2D scenario and are far inferior to the proposed, “Benchmarked I” and “Benchmarked II” algorithms.

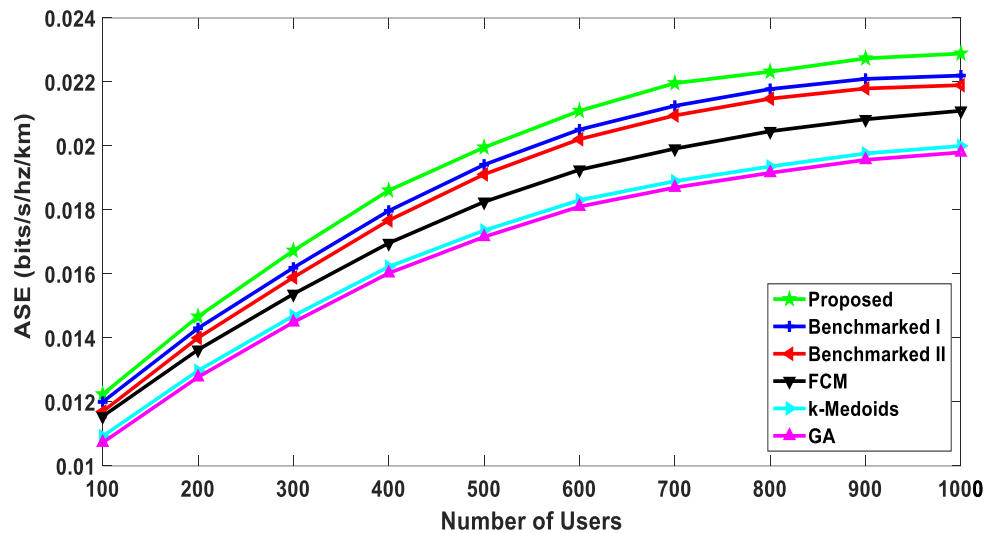


Figure 12. ASE: comparison with the benchmarked ( $w_1 = 0.4$ ,  $w_2 = 0.6$ ).

### 5.3. The Optimal Number of Clusters

We investigated the effect of the number of clusters on the energy consumption and ASE. It has not been reported in the literature considering D2D Content-Centric Networks. The selection of the number of clusters significantly affects the clustering performance. A trade-off always exists when it comes to selecting the number of clusters. Increasing the number of clusters up to a certain extent will bring benefits but at the expense of increased signaling and complexity. The clustering metrics that we selected for the proposed algorithm can vary significantly; thus, it is not easy to predetermine the cluster size. Therefore, the size of the cluster is variable.

However, there should be a criterion that can help determine the number of clusters that can be formed based on a given scenario such as user distribution, values of the clustering metrics, etc. In this study, the Calinski–Harabasz (Cal–Har) criterion [66] was selected. It is also termed as the variance ratio criterion. Mathematically, it can be defined as:

$$Cluster\ Size_{(Cal-Har)} = \frac{V_B}{V_W} * \frac{(N - C)}{(C - 1)} \quad (12)$$

In Equation (12), between-cluster and within-cluster variance are represented by  $V_B$  and  $V_W$ , respectively. The total number of users is denoted by  $N$ , whereas  $C$  is the number of clusters against which this criterion will be judged. Clustering metrics determine the variance between and within clusters. To find the optimal solution, Equation (12) needs to be maximized with respect to the number of clusters. As the ratio of the variances given in Equation (12) increases, user segregation becomes more precise which leads to the optimal number of clusters.

For a user density of one thousand, the Cal–Har criterion result is depicted in Figure 13. The criterion value was highest when the number of clusters was seven. Therefore, for the given user distribution and node density, the optimal number of clusters should be seven. We then investigated whether this is the optimal choice when considering energy consumption, and ASE. Results for both parameters for a various number of clusters are presented in Figures 14 and 15, respectively. A trade-off

can be seen for the choice of the number of clusters. As shown in Figure 14, the result for energy consumption aligns with the Calinski–Harabasz criterion, as the lowest energy consumption occurred when the number of clusters was seven to ten. On the other hand, the optimal number of clusters for ASE appears to lie between 10 and 15.

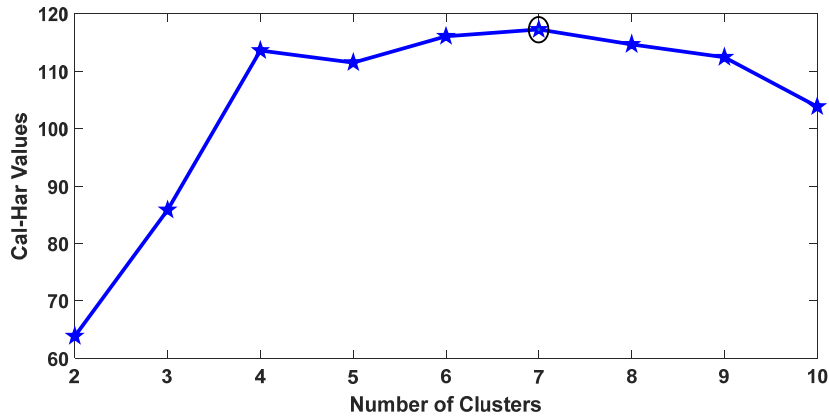


Figure 13. Optimal number of clusters: Cal–Har criterion.

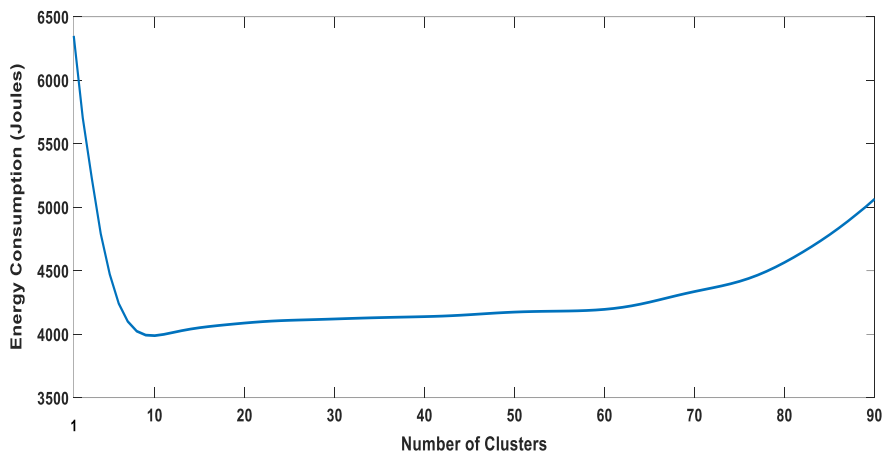


Figure 14. Energy consumption and the number of clusters: a comparison.

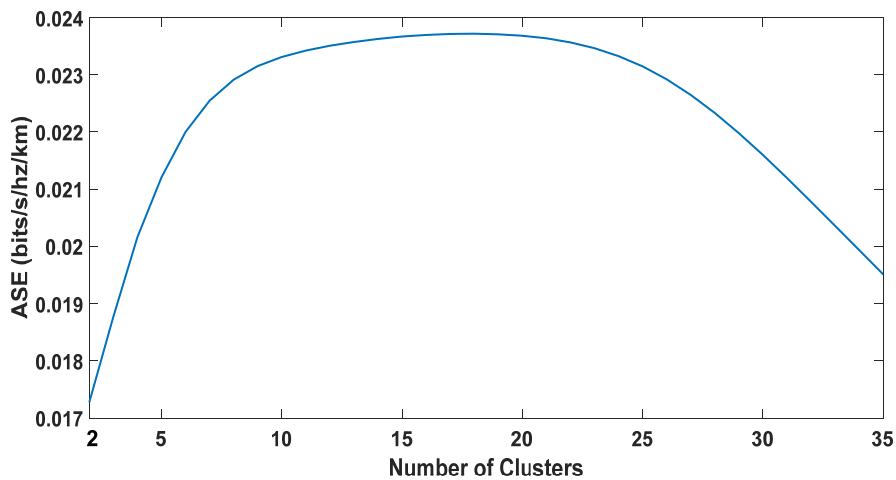


Figure 15. ASE and the number of clusters: a comparison.

It is not straight forward to suggest a certain number of clusters based on the above-mentioned results. It is important to consider a few additional factors that might influence the selection.

The signaling overhead and complexity of cluster maintenance increase with the increase in the number of clusters. Moreover, we can observe from the results presented in Figures 14 and 15 that ASE did not differ significantly when forming seven to ten clusters as opposed to forming 10–15. Therefore, seven is likely to be a better choice, since it represents the lowest energy consumption while sacrificing minimal ASE gain for lower signaling overhead and complexity.

Various external factors influence the choice of the number of clusters as well. As discussed earlier, our study considers forming clusters only for those users that are interested in content sharing. Therefore, in some situations, only a few users might be interested in content-sharing, and hence, making a certain number of clusters is not necessary. On the other hand, a scenario can build-up where a large group of users is interested in content-sharing, but even in this case, the physical location of the users might influence the choice of the number of clusters. A large number of closely packed users at a concert or a sports event only needs a few clusters. As opposed to this scenario, users sharing a common interest might be dispersed in a geographical area, requiring a higher number of clusters. More clusters can also be formed in a scenario where users might be in close vicinity, but they have different content of interest. In that case, users sharing the same interest are expected to be in one cluster, whereas the rest of the users should form a separate cluster. Moreover, some users may be better served by the BS, and it should not be mandatory for all the users to be considered for the clustering.

#### 5.4. Throughput Fairness

The benchmarked algorithms considered in this study do not consider throughput fairness as a potential performance parameter. The literature suggests that fairness is crucial for evaluating a cellular network [72,73]. Therefore, we extended the proposed algorithm to suit the throughput fairness as well. To target the fairness, we introduced another clustering metric—the variance among the channel conditions of the users.

Considering only absolute channel conditions might disregard many users having not so good channel conditions. Therefore, in that case, there would be a significant difference in the throughputs of the individual users. Owing to this reason, we introduced variance of channel conditions, and our assumption is validated by the results demonstrating throughput fairness, shown later in this section.

Moreover, since we are considering a multicasting scenario, we are looking for approximately similar channel conditions with each node. The reason being, in a multicasting scenario, if each device receives the transmission at significantly different rates, then the complexity of the system would increase and might become infeasible [74].

Let us take an example of a scenario where a video stream needs to be broadcasted to a group of users. Assume that there are a few users with higher rates as compared to the others. Since we are considering broadcasting, the maximum achievable rates are determined by the worst physical link in the group. Therefore, users having higher rates will face long delays waiting for the other users to catch up. This reason makes it even more significant to have variance in channel conditions as an important factor in addition to just the absolute channel conditions. Though, owing to the degradation in the throughput performance, variance in channel conditions cannot be selected as the sole criterion. Therefore, the modified clustering algorithm considers all three different metrics which are attached to their respective weights.

The introduction of variance among channel conditions modifies the Equation (2), to the following;

$$W_T = I. \left[ w_1 * D_{(a,b)} + w_2 * \left( \frac{1}{h_{ab}} \right) + w_3 * var(h) \right] \quad (13)$$

$w_3$  represents the weights given to variance of the channel gains. The term “ $Var(h)$ ” represents the variance among the channel conditions of the users. The weights represented in Equation (13) are such that  $\sum_{f=1}^3 w_f = 1$ . The node with the minimum  $W_T$  is chosen as the CH.

Jain's fairness model [75] was used to evaluate the fairness performance of the proposed algorithm. The Jain's fairness index denoted by  $J(x)$  is represented by Equation (14).

$$J(x) = \frac{(\sum_{i=1}^M x_i)^2}{M \sum_{i=1}^M x_i^2} \quad (14)$$

$x_i$  represents the throughput of the  $i$ th user, given that there are total  $M$  users.

We simulated the proposed algorithm with the weights as follows:  $w_1 = w_2 = 0.1$ ,  $w_3 = 0.8$ . The weights selected for this result were empirically adjusted to enhance fairness. So, the maximum weight is given to the variance of channels. Though the selected weights did not yield the best energy consumption and ASE, it did outperform all the benchmarked algorithms when it comes to fairness. This is depicted in Figure 16. At the user density of one thousand, the proposed algorithm performed approximately 7% better than the benchmarked scheme I, which performed the best among the existing algorithms. This result demonstrates the flexibility of the proposed algorithm. By simply adjusting the weights, it is possible to achieve better fairness. It should be noted that the works reporting the benchmarked algorithms I and II do not discuss fairness. These algorithms also do not have any parameter to perform this trade-off.

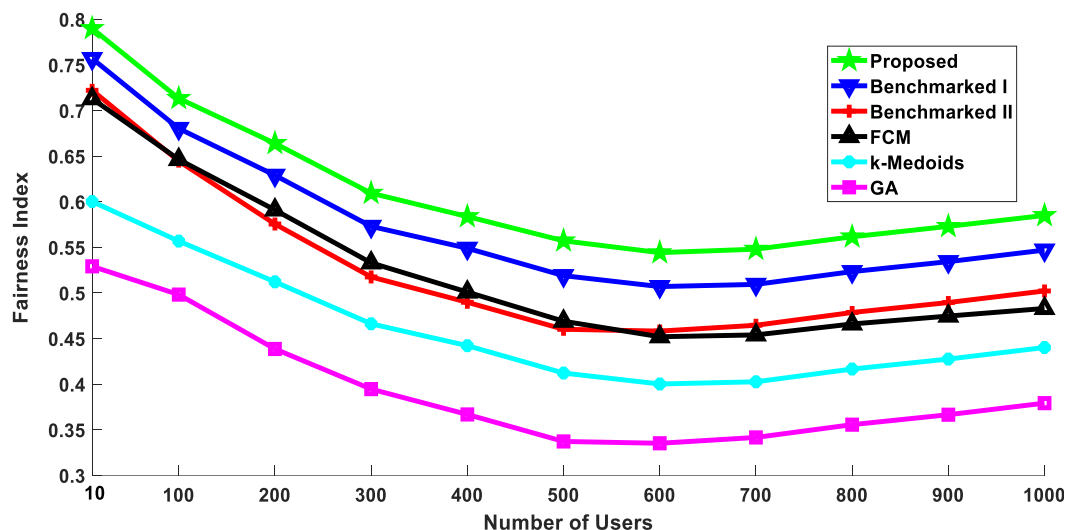


Figure 16. Jain's Fairness Index: Comparison with the benchmarked ( $w_1 = w_2 = 0.1$ ,  $w_3 = 0.8$ ).

### 5.5. The Trade-Off among Different Performance Parameters

The energy consumption and ASE for the weights  $w_1 = w_2 = 0.1$  and  $w_3 = 0.8$  are presented in Figures 17 and 18, respectively. It can be observed that the cost of improved fairness is a slight degradation in performance with regard to energy consumption and ASE. However, the performance of the proposed algorithm is satisfactory in the sense that it is better than the other four benchmarked algorithms and there is only a very small performance gap with the best scheme. The proposed algorithm is able to trade-off energy consumption and ASE with fairness, which is not possible in any of the benchmarked algorithms. Our algorithm provides this flexibility by adjusting the weights to enhance the desired performance parameter accordingly. It can achieve the best performance for a given parameter while providing a satisfactory performance with respect to the rest.

To the best of the author's knowledge, the algorithms considered for benchmarking in this study have not been investigated for all the three performance parameters i.e., energy consumption, ASE, and fairness.

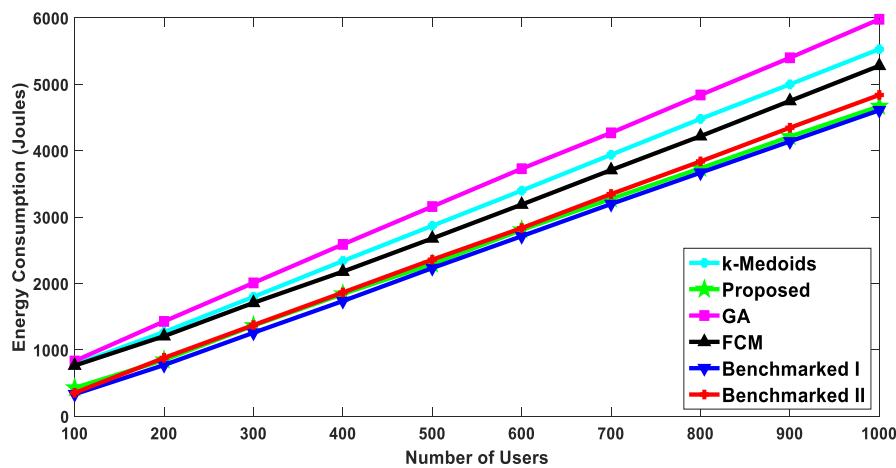


Figure 17. Energy consumption: comparison with the benchmarked schemes ( $w_1 = w_2 = 0.1$ ,  $w_3 = 0.8$ ).

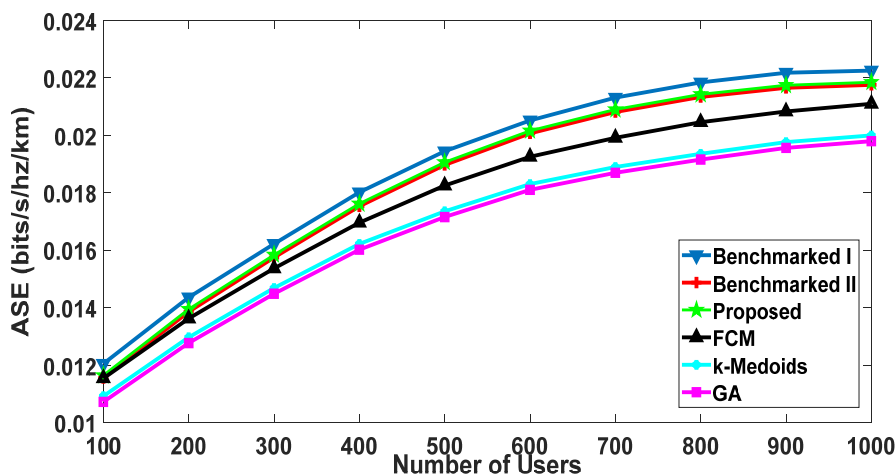


Figure 18. ASE: comparison with the benchmarked schemes ( $w_1 = w_2 = 0.1$ ,  $w_3 = 0.8$ ).

## 6. Conclusions

This paper presented a content-sharing framework for D2D communication in a multicasting scenario. We utilize Content-Centric Networking and Network Virtualization to propose a distributive architecture. This study showed the significance of spatial distribution and social-ties on the throughput's performance and established that both are vital for enhancing the performance of the Content-Centric Network.

A novel weighted clustering algorithm was incorporated in the proposed architecture. It is evident from the results that clustering enhanced the system's performance. The performance of the proposed algorithm was thoroughly investigated against different popular clustering algorithms. The proposed algorithm shows a 6% improvement in energy consumption while achieving 3% better ASE as compared to the best-benchmarked algorithm. The effect of the number of clusters on the energy consumption of users and ASE was also investigated. A trade-off exists between the two metrics in the selection of the number of clusters. The optimal energy consumption was achieved at a smaller number of clusters as compared to ASE. It is suggested that signaling overhead required to set up more clusters should be considered while selecting the number of clusters. Hence, the formation of a smaller number of clusters showing optimal energy consumptions at the cost of marginal degradation in ASE is acceptable. A slight modification in the algorithm and weight adjustment improved throughput fairness up to 7%. On the contrary, all the benchmarked algorithms do not consider fairness.

For future research, it would be important to apply machine learning algorithms to optimally select the weights according to the system's requirements. The authors also intend to investigate the

performance of the proposed scheme for various multimedia applications. Additionally, the comparative study of the signaling overhead required for the proposed algorithm needs to be explored as well. Moreover, while this study provides an approach to finding the number of clusters, one of the future directions could be a comprehensive study of optimal selection of the number of clusters.

**Author Contributions:** S.A. and F.A. conceived and conceptualized the research. S.A. implemented the research, gathered the results, and produced the first draft. F.A., S.F.H., and M.R. provided feedback and helped revise the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** The authors acknowledge Massey University for the resources provided to conduct this research. S.A. also acknowledges the support received through the Massey University Doctoral Scholarship.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Penttinen, J.T. *3GPP Mobile Communications: LTE/SAE and LTE-A*; Wiley: Hoboken, NJ, USA, 2013.
2. Benchaabene, Y.; Boujnah, N.; Zarai, F. 5G cellular: Survey on some challenging techniques. In Proceedings of the 2017 18th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT), Taipei, Taiwan, 18–20 December 2017; IEEE: Piscataway, NJ, USA, 2017.
3. Khan, M.F.; Yau, K.-L.A.; Noor, R.M.; Imran, M.A. Survey and taxonomy of clustering algorithms in 5G. *J. Netw. Comput. Appl.* **2020**, *154*, 102539. [[CrossRef](#)]
4. Index, C.V.N. Global mobile data traffic forecast update, 2015–2020. In *Cisco White Paper*; Cisco: San Jose, CA, USA, 2016; p. 9.
5. Ericsson, A. *Ericsson Mobility Report: On the Pulse of the Networked Society*; Tech. Rep. EAB-14; Ericsson: Stockholm, Sweden, 2015; p. 61078.
6. Asadi, A.; Wang, Q.; Mancuso, V. A survey on device-to-device communication in cellular networks. *IEEE Commun. Surv. Tutor.* **2014**, *16*, 1801–1819. [[CrossRef](#)]
7. Cao, Y.; Jiang, T.; Chen, X.; Zhang, J. Social-aware video multicast based on device-to-device communications. *IEEE Trans. Mob. Comput.* **2015**, *15*, 1528–1539. [[CrossRef](#)]
8. Chang, Z.; Hu, Y.; Chen, Y.; Zeng, B. Cluster-oriented device-to-device multimedia communications: Joint power, bandwidth, and link selection optimization. *IEEE Trans. Veh. Technol.* **2017**, *67*, 1570–1581. [[CrossRef](#)]
9. Feng, L.; Zhao, P.; Zhou, F.; Yin, M.; Yu, P.; Li, W.; Qiu, X. Resource allocation for 5G D2D multicast content sharing in social-aware cellular networks. *IEEE Commun. Mag.* **2018**, *56*, 112–118. [[CrossRef](#)]
10. Jiang, J.; Zhang, S.; Li, B.; Li, B. Maximized cellular traffic offloading via device-to-device content sharing. *IEEE J. Sel. Areas Commun.* **2015**, *34*, 82–91. [[CrossRef](#)]
11. Lin, P.; Khan, K.S.; Song, Q.; Jamalipour, A. Caching in heterogeneous ultradense 5G networks: A comprehensive cooperation approach. *IEEE Veh. Technol. Mag.* **2019**, *14*, 22–32. [[CrossRef](#)]
12. Yang, L.; Wu, D.; Xu, S.; Zhang, G.; Cai, Y. Social-energy-aware user clustering for content sharing based on D2D multicast communications. *IEEE Access* **2018**, *6*, 36092–36104. [[CrossRef](#)]
13. Aslam, S.; Alam, F.; Hasan, S.F.; Rashid, M. Performance Analysis of Clustering Algorithms for Content-Sharing Based D2D Enabled 5G Networks. In Proceedings of the 2019 29th International Telecommunication Networks and Applications Conference (ITNAC), Auckland, New Zealand, 27–29 November 2019; IEEE: Piscataway, NJ, USA, 2019.
14. Akyildiz, I.F.; Nie, S.; Lin, S.-C.; Chandrasekaran, M. 5G roadmap: 10 key enabling technologies. *Comput. Netw.* **2016**, *106*, 17–48. [[CrossRef](#)]
15. Morgado, A.; Huq, K.M.S.; Mumtaz, S.; Rodriguez, J. A survey of 5G technologies: Regulatory, standardization and industrial perspectives. *Digit. Commun. Netw.* **2018**, *4*, 87–97. [[CrossRef](#)]
16. Bega, D.; Gramaglia, M.; Bernardos Cano, C.J.; Banchs, A.; Costa-Perez, X. Toward the network of the future: From enabling technologies to 5G concepts. *Trans. Emerg. Telecommun. Technol.* **2017**, *28*, e3205. [[CrossRef](#)]
17. Feng, Z.; Qiu, C.; Feng, Z.; Wei, Z.; Li, W.; Zhang, P. An effective approach to 5G: Wireless network virtualization. *IEEE Commun. Mag.* **2015**, *53*, 53–59. [[CrossRef](#)]

18. Hossain, E.; Hasan, M. 5G cellular: Key enabling technologies and research challenges. *IEEE Instrum. Meas. Mag.* **2015**, *18*, 11–21. [[CrossRef](#)]
19. Ateya, A.A.; Muthanna, A.; Makolkina, M.; Koucheryavy, A. Study of 5G services standardization: Specifications and requirements. In Proceedings of the 2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), Moscow, Russia, 5–9 November 2018; IEEE: Piscataway, NJ, USA, 2018.
20. Pérez, G.O.; Hernández, J.A.; Larrabeiti, D. Fronthaul network modeling and dimensioning meeting ultra-low latency requirements for 5G. *IEEE/OSA J. Opt. Commun. Netw.* **2018**, *10*, 573–581.
21. Lien, S.-Y.; Shieh, S.-L.; Huang, Y.; Su, B.; Hsu, Y.-L.; Wei, H.-Y. 5G new radio: Waveform, frame structure, multiple access, and initial access. *IEEE Commun. Mag.* **2017**, *55*, 64–71. [[CrossRef](#)]
22. Rappaport, T.S.; Sun, S.; Shafi, M. Investigation and comparison of 3GPP and NYUSIM channel models for 5G wireless communications. In Proceedings of the 2017 IEEE 86th Vehicular Technology Conference (VTC-Fall), Toronto, ON, Canada, 24–27 September 2017; IEEE: Piscataway, NJ, USA, 2017.
23. Sanusi, I.O.; Nasr, K.M.; Moessner, K. Channel assignment and power control for D2D-enabled cellular networks. In Proceedings of the 2019 International Conference on Computing, Electronics & Communications Engineering (iCCECE), London, UK, 22–23 August 2019; IEEE: Piscataway, NJ, USA, 2019.
24. Yang, P.; Zhang, N.; Zhang, S.; Yu, L.; Zhang, J.; Shen, X.S. Content popularity prediction towards location-aware mobile edge caching. *IEEE Trans. Multimed.* **2018**, *21*, 915–929. [[CrossRef](#)]
25. Semenciuc, E.; Pastrav, A.; Palade, T.; Puschita, E. Performance evaluation of a Cloud-based QoS support mechanism. In Proceedings of the 2016 International Conference on Communications (COMM), Bucharest, Romania, 9–10 June 2016; IEEE: Piscataway, NJ, USA, 2016.
26. Puschita, E.; Palade, T.; Moldovan, A.; Colda, R.; Vermesan, I. An innovative qos paradigm based on cognitive in-network management of resources for a future unified network architecture: I-name qos model. In Proceedings of the 2010 Second International Conference on Advances in Future Internet, Venice, Italy, 18–25 July 2010; IEEE: Piscataway, NJ, USA, 2010.
27. Wu, Y.; Yao, S.; Yang, Y.; Zhou, T.; Qian, H.; Hu, H.; Hamalainen, M. Challenges of mobile social device caching. *IEEE Access* **2016**, *4*, 8938–8947. [[CrossRef](#)]
28. Liang, C.; Yu, F.R.; Zhang, X. Information-centric network function virtualization over 5G mobile wireless networks. *IEEE Netw.* **2015**, *29*, 68–74. [[CrossRef](#)]
29. Zhang, X.; Zhu, Q. Information-centric virtualization for software-defined statistical QoS provisioning over 5G multimedia big data wireless networks. *IEEE J. Sel. Areas Commun.* **2019**, *37*, 1721–1738. [[CrossRef](#)]
30. Zhou, Z.; Gao, C.; Xu, C.; Zhang, Y.; Mumtaz, S.; Rodriguez, J. Social big-data-based content dissemination in internet of vehicles. *IEEE Trans. Ind. Inform.* **2017**, *14*, 768–777. [[CrossRef](#)]
31. Zhou, Z.; Yu, H.; Xu, C.; Zhang, Y.; Mumtaz, S.; Rodriguez, J. Dependable content distribution in D2D-based cooperative vehicular networks: A big data-integrated coalition game approach. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 953–964. [[CrossRef](#)]
32. Zhu, Y.; Qin, X.; Zhang, P. An efficient multicast clustering scheme for D2D assisted offloading in cellular networks. In Proceedings of the 2018 IEEE/CIC International Conference on Communications in China (ICCC), Beijing, China, 16–18 August 2018; IEEE: Piscataway, NJ, USA, 2018.
33. Wu, D.; Yan, J.; Wang, H.; Wu, D.; Wang, R. Social attribute aware incentive mechanism for device-to-device video distribution. *IEEE Trans. Multimed.* **2017**, *19*, 1908–1920. [[CrossRef](#)]
34. Asadi, A.; Mancuso, V. Network-assisted outband D2D-clustering in 5G cellular networks: Theory and practice. *IEEE Trans. Mob. Comput.* **2016**, *16*, 2246–2259. [[CrossRef](#)]
35. Zhang, Z.; Wang, L.; Zhang, J. Energy efficiency of D2D multi-user cooperation. *Sensors* **2017**, *17*, 697. [[CrossRef](#)]
36. Huang, H.; Liu, B.; Hu, M.; Tao, Y.; Xiang, W. User-information-aware D2D multicast file distribution mechanism. *Sensors* **2018**, *18*, 3389. [[CrossRef](#)]
37. Rahman, M.A.; Lee, Y.; Koo, I. Energy-efficient power allocation and relay selection schemes for relay-assisted d2d communications in 5g wireless networks. *Sensors* **2018**, *18*, 2865. [[CrossRef](#)]
38. Pizzi, S.; Rinaldi, F.; Molinaro, A.; Iera, A.; Araniti, G. Energy-Efficient Multicast Service Delivery Exploiting Single Frequency Device-To-Device Communications in 5G New Radio Systems. *Sensors* **2018**, *18*, 2205. [[CrossRef](#)]

39. Shi, R.; Wang, B.; Shi, F.; Piao, D.; Hu, J. A Design of D2D-Clustering Algorithm for Group D2D Communication. In Proceedings of the International Conference on Communications and Networking in China, Shanghai, China, 29 November–1 December 2019; Springer: Cham, Switzerland, 2019.
40. Wu, Y.; Wu, D.; Yang, L.; Shi, X.; Ao, L.; Fu, Q. Matching-coalition based cluster formation for D2D multicast content sharing. *IEEE Access* **2019**, *7*, 73913–73928. [[CrossRef](#)]
41. Wang, Y.; Zhu, Q. Performance analysis of clustered device-to-device networks using matern cluster process. *Wirel. Netw.* **2019**, *25*, 4849–4858. [[CrossRef](#)]
42. Zhou, L.; Wu, Y.; Yu, H. A Two-Layer, Energy-Efficient Approach for Joint Power Control and Uplink–Downlink Channel Allocation in D2D Communication. *Sensors* **2020**, *20*, 3285. [[CrossRef](#)]
43. Padmavathi, P.; Eswaramurthy, V.; Revathi, J. Fuzzy Social Spider Optimization Algorithm for Fuzzy Clustering Analysis. In Proceedings of the 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT), Coimbatore, India, 1–3 March 2018; IEEE: Piscataway, NJ, USA, 2018.
44. Khan, A.A.; Abolhasan, M.; Ni, W.; Lipman, J.; Jamalipour, A. A hybrid-fuzzy logic guided genetic algorithm (H-FLGA) approach for resource optimization in 5G VANETs. *IEEE Trans. Veh. Technol.* **2019**, *68*, 6964–6974. [[CrossRef](#)]
45. Cardoso, E.; Silva, K.; Francês, R. Intelligent handover procedure for heterogeneous LTE networks using fuzzy logic. In Proceedings of the 2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC), Valencia, Spain, 26–30 June 2017; IEEE: Piscataway, NJ, USA, 2017.
46. Alhammadi, A.; Roslee, M.; Alias, M.Y.; Shayea, I.; Alriah, S.; Abas, A.B. Advanced handover self-optimization approach for 4G/5G HetNets using weighted fuzzy logic control. In Proceedings of the 2019 15th International Conference on Telecommunications (ConTEL), Graz, Austria, 3–5 July 2019; IEEE: Piscataway, NJ, USA, 2019.
47. Kim, D.; Kim, J.-h.; Moon, C.; Choi, J.; Yeom, I. Efficient content delivery in mobile ad-hoc networks using CCN. *Ad Hoc Netw.* **2016**, *36*, 81–99. [[CrossRef](#)]
48. Wang, Q.; Chen, D.; Zhang, N.; Qin, Z.; Qin, Z. LACS: A lightweight label-based access control scheme in IoT-based 5G caching context. *IEEE Access* **2017**, *5*, 4018–4027. [[CrossRef](#)]
49. Fritter, M.; Ould-Khessal, N.; Fazackerley, S.; Lawrence, R. Experimental evaluation of hash function performance on embedded devices. In Proceedings of the 2018 IEEE Canadian Conference on Electrical & Computer Engineering (CCECE), Quebec City, QC, Canada, 13–16 May 2018; IEEE: Piscataway, NJ, USA, 2018.
50. Varna, A.L.; Wu, M. Modeling and analysis of content identification. In Proceedings of the 2009 IEEE International Conference on Multimedia and Expo, New York, NY, USA, 28 June–3 July 2009; IEEE: Piscataway, NJ, USA, 2009.
51. Malhotra, J.; Bakal, J. A survey and comparative study of data deduplication techniques. In Proceedings of the 2015 International Conference on Pervasive Computing (ICPC), Pune, India, 8–10 January 2015; IEEE: Piscataway, NJ, USA, 2015.
52. Padhi, M.; Chaudhari, R. An optimized pipelined architecture of SHA-256 hash function. In Proceedings of the 2017 7th International Symposium on Embedded Computing and System Design (ISED), Durgapur, India, 18–20 December 2017; IEEE: Piscataway, NJ, USA, 2017.
53. Qiuyun, X.; Ligang, H.; Qiming, L.; Shuqin, G.; Jinhui, W. The Verification of SHA-256 IP using a semi-automatic UVM platform. In Proceedings of the 2017 13th IEEE International Conference on Electronic Measurement & Instruments (ICEMI), Yangzhou, China, 20–22 October 2017; IEEE: Piscataway, NJ, USA, 2017.
54. Jaffry, S.; Zaidi, S.K.; Shah, S.T.; Hasan, S.F.; Gui, X. D2D neighborhood discovery by a mobile device. In Proceedings of the ICC 2019-2019 IEEE International Conference on Communications (ICC), Shanghai, China, 20–24 May 2019; IEEE: Piscataway, NJ, USA, 2019.
55. Wang, Y.; Li, Y.; Wang, W.; Song, M. A locality-based mobile caching policy for D2D-based content sharing network. In Proceedings of the 2016 IEEE Global Communications Conference (GLOBECOM), Washington, DC, USA, 4–8 December 2016; IEEE: Piscataway, NJ, USA, 2016.
56. Lin, X.; Andrews, J.G.; Ghosh, A. Spectrum sharing for device-to-device communication in cellular networks. *IEEE Trans. Wirel. Commun.* **2014**, *13*, 6727–6740. [[CrossRef](#)]
57. Mancuso, V.; Asadi, A.; Jacko, P. Tie-breaking can maximize fairness without sacrificing throughput in D2D-assisted networks. In Proceedings of the 2016 IEEE 17th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), Coimbra, Portugal, 21–24 June 2016; IEEE: Piscataway, NJ, USA, 2016.



58. Kuerbis, M.; Balasubramanya, N.M.; Lampe, L.; Lampe, A. On the use of channel models and channel estimation techniques for massive MIMO systems. In Proceedings of the 2016 24th European Signal Processing Conference (EUSIPCO), Budapest, Hungary, 29 August–2 September 2016; IEEE: Piscataway, NJ, USA, 2016.
59. Wimmer, H.; Powell, L. Principle component analysis for feature reduction and data preprocessing in data science. In Proceedings of the Conference on Information Systems Applied Research, Las Vegas, NV, USA, 6–9 November 2016.
60. Obaid, H.S.; Dheyab, S.A.; Sabry, S.S. The Impact of Data Pre-Processing Techniques and Dimensionality Reduction on the Accuracy of Machine Learning. In Proceedings of the 2019 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON), Jaipur, India, 13–15 March 2019; IEEE: Piscataway, NJ, USA, 2019.
61. Runkler, T.A.; Katz, C. Fuzzy clustering by particle swarm optimization. In Proceedings of the 2006 IEEE International Conference on Fuzzy Systems, Vancouver, BC, Canada, 16–21 July 2006; IEEE: Piscataway, NJ, USA, 2006.
62. Antonopoulos, A.; Verikoukis, C. Multi-player game theoretic MAC strategies for energy efficient data dissemination. *IEEE Trans. Wirel. Commun.* **2013**, *13*, 592–603. [[CrossRef](#)]
63. Antonopoulos, A.; Kartsakli, E.; Verikoukis, C. Game theoretic D2D content dissemination in 4G cellular networks. *IEEE Commun. Mag.* **2014**, *52*, 125–132. [[CrossRef](#)]
64. Yaacoub, E.; Kubbar, O. Energy-efficient device-to-device communications in LTE public safety networks. In Proceedings of the 2012 IEEE Globecom Workshops, Anaheim, CA, USA, 3–7 December 2012; IEEE: Piscataway, NJ, USA, 2012.
65. Ometov, A.; Orsino, A.; Militano, L.; Moltchanov, D.; Araniti, G.; Olshannikova, E.; Fodor, G.; Andreev, S.; Olsson, T.; Iera, A. Toward trusted, social-aware D2D connectivity: Bridging across the technology and sociality realms. *IEEE Wirel. Commun.* **2016**, *23*, 103–111. [[CrossRef](#)]
66. Łukasik, S.; Kowalski, P.A.; Charytanowicz, M.; Kulczycki, P. Clustering using flower pollination algorithm and Calinski-Harabasz index. In Proceedings of the 2016 IEEE Congress on Evolutionary Computation (CEC), Vancouver, BC, Canada, 24–29 July 2016; IEEE: Piscataway, NJ, USA, 2016.
67. 3rd Generation Partnership Project (3GPP). *3GPP TR 25.814 3GPP TSG RAN Physical Layer Aspects for Evolved UTRA, v7. 1.0*; 3GPP: Sophia Antipolis, France, 2006.
68. Mahmud, K.; Inoue, M.; Murakami, H.; Hasegawa, M.; Morikawa, H. Energy consumption measurement of wireless interfaces in multi-service user terminals for heterogeneous wireless networks. *IEICE Trans. Commun.* **2005**, *88*, 1097–1110. [[CrossRef](#)]
69. Tulu, M.M.; Hou, R.; Li, C.; Amentie, M.D. Cluster head selection method for content-centric mobile social network in 5G. *IET Commun.* **2017**, *12*, 402–408. [[CrossRef](#)]
70. Kazez, A.C.; Girici, T. Clustering-based device-to-device cache placement. *Ad Hoc Netw.* **2019**, *84*, 170–177. [[CrossRef](#)]
71. Jung, H.; Lee, I.-H. Coverage probability and area spectral efficiency of clustered linear unmanned vehicle sensor networks. *Sensors* **2017**, *17*, 2550. [[CrossRef](#)] [[PubMed](#)]
72. Cheng, H.T.; Zhuang, W. An optimization framework for balancing throughput and fairness in wireless networks with QoS support. *IEEE Trans. Wirel. Commun.* **2008**, *7*, 584–593. [[CrossRef](#)]
73. Huaizhou, S.; Prasad, R.V.; Onur, E.; Niemegeers, I. Fairness in wireless networks: Issues, measures and challenges. *IEEE Commun. Surv. Tutor.* **2013**, *16*, 5–24. [[CrossRef](#)]
74. Seppälä, J.; Koskela, T.; Chen, T.; Hakola, S. Network controlled device-to-device (D2D) and cluster multicast concept for LTE and LTE-A networks. In Proceedings of the 2011 IEEE Wireless Communications and Networking Conference, Cancun, Quintana Roo, Mexico, 28–31 March 2011; IEEE: Piscataway, NJ, USA, 2011.
75. Jain, R.K.; Chiu, D.-M.W.; Hawe, W.R. *A Quantitative Measure of Fairness and Discrimination*; Eastern Research Laboratory, Digital Equipment Corporation: Hudson, MA, USA, 1984.

