

A Query-Based Greedy Approach for Authentic Influencer Discovery in SIoT

Farah Batool¹, Abdul Rehman², Dongsun Kim^{2,*}, Assad Abbas¹, Raheel Nawaz³ and Tahir Mustafa Madni¹

¹COMSATS University Islamabad, Islamabad, Pakistan

²School of Computer Science and Engineering, Kyungpook National University, Daegu, 41566, Korea

³Department of Operations, Technology, Events and Hospitality Management, Manchester Metropolitan University, United Kingdom

*Corresponding Author: Dongsun Kim. Email: darkrsw@knu.ac.kr

Received: 29 June 2022; Accepted: 13 October 2022

Abstract: The authors propose an informed search greedy approach that efficiently identifies the influencer nodes in the social Internet of Things with the ability to provide legitimate information. Primarily, the proposed approach minimizes the network size and eliminates undesirable connections. For that, the proposed approach ranks each of the nodes and prioritizes them to identify an authentic influencer. Therefore, the proposed approach discards the nodes having a rank (α) lesser than 0.5 to reduce the network complexity. α is the variable value represents the rank of each node that varies between 0 to 1. Node with the higher value of α gets the higher priority and vice versa. The threshold value $\alpha = 0.5$ defined by the authors with respect to their network pruning requirements that can be vary with respect to other research problems. Finally, the algorithm in the proposed approach traverses the trimmed network to identify the authentic node to obtain the desired information. The performance of the proposed method is evaluated in terms of time complexity and accuracy by executing the algorithm on both the original and pruned networks. Experimental results show that the approach identifies authentic influencers on a resultant network in significantly less time than in the original network. Moreover, the accuracy of the proposed approach in identifying the influencer node is significantly higher than that of the original network. Furthermore, the comparison of the proposed approach with the existing approaches demonstrates its efficiency in terms of time consumption and network traversal through the minimum number of hops.

Keywords: Online social network; influencer search; query-based approach; greedy search; social internet of things (siot)

1 Introduction

Several Online Social Networks (OSNs), such as Facebook, Twitter, Linked In, etc. have closely connected Web users all over the world. People may participate directly in these networks to exchange



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ideas, knowledge, experiences and perspectives [1,2]. Moreover, social media users also discover and distribute knowledge in a variety of formats, including text, image, audio and video. Businesses that previously used to rely on conventional marketing strategies, advertisements in magazines, newspapers and television have now turned to social media for rapid and effective promotions because the majority of the people have joined social networks [3]. The use of OSNs not only connects users with their friends and family members but also serves as a quick way to keep users up to date on local and global news and events. The research in the OSN domain helps us understand how information, experiences, opinions and inventions spread across the world. Moreover, the analysis of OSNs offers insights into how certain communities operate. Furthermore, it can be seen as a way of identifying and analyzing the interests or activities of each community around the globe.

Although the increasing number of OSNs has significantly improved the connectivity of social entities across the borders, however, all social networks lack the criterion of measuring their trustworthiness. A single user can make several fake IDs and any user can send or receive friendship requests. Consequently, the intertwining of duplicate local and/or global users from various cultures or communities not only increases the complexity but also makes the network uncertain. Moreover, the OSNs do not check the authenticity of the users or the contents. Such networks may produce a massive amount of unstructured data and spurious information [4]. Consequently, it is difficult to find authentic information about a specific field or community. It has been observed that the pervasive dissemination of deceitful information on social media has become a global menace. The reason is that social media communities have no criteria to assess the legitimacy of the links. Consequently, these communities may possess assorted, insignificant and spurious ties. Therefore, evading such perils and the identification of credible influencers or information is an imperative task. The influencers play an important role in any network by exchanging information rapidly. Consequently, people in this network value the information they receive from the influencers.

So far it has been a challenging and crucial problem to efficiently identify the authentic and influential spreaders of information in any group or community. As there are a large number of nodes and connections in a social network, therefore, identification of any sort of information or individuals is computationally expensive. The reason is that the existing techniques exhaustively traverse the networks to identify the desired individual or influencer. Moreover, accurate identification of influencers is equally important. Therefore, to enhance the effectiveness of social media platforms, techniques with the capability to efficiently identify the influencers in social networks while preserving the identification accuracy are needed. The objective of the research is to efficiently identify the influencer nodes that spread authentic information about a specific community or field in a network. The task of identification of the influencers in the coarsely established inadequate network is an arduous task and can be achieved by revamping the network in consonance with an individual's requirements. To that end, this research work proposes an informed search greedy approach for accomplishing the aforesaid task. The proposed approach examines the influence capacity of each user or node and uses it to distinguish the influential and non-influential users.

The Social Internet of Things (SIoT) devices mimic human behavior and search for friends from the real social network of the owner with respect to some common objectives and reliability measures. SIoT is based on the use of Social Network Science (SNS) concepts for improving the network navigability and Information/Service discovery in IoT [5,6]. Our proposed approach augments the existing techniques by fending off the weaknesses and deficiencies in [5,6]. An important question is how the algorithm obtains the information or service if the information provider lies at the seventh or higher layer of the network and the algorithm travels maximum up to the sixth layer. Besides, the algorithm terminates if any node or path has already been traversed, which causes irregular

stoppage and may create a hindrance in information retrieval. In contrast, our proposed algorithm traverses the entire graph based on priority and finds the most accurate and informative node in significantly less time. For the ease of reading, throughout the terms, Information and services will be used interchangeably.

It is important to mention that the proposed algorithm is greedy in nature because it chooses the highest-ranked node in each layer and traverses that node rather than following the deployment sequence of the nodes. Nodes are ranked by examining their in-degree, out-degree and edge weight. In-degree reveals the influence of the node whereas outdegree depicts how frequently the node responds to the nodes that approach it. The edge weight shows the interaction and trustworthiness between the two nodes. The ranking of nodes is between 0 and 1. When the algorithm identifies the credible influencer or information provider, the search comes to an end. Instead of traversing the entire network, the proposed algorithm uses a ranking parameter to find the most significant influencers. As a result, the approach takes much less time to compute than an exhaustive search mechanism.

The contributions of the paper are as below:

- i. The research aims to present an approach that eliminates the isolated/outlier nodes and traverses the network to find influencers in an efficient way.
- ii. The network pruning is performed to remove the irrelevant and non-influential nodes to avoid counterfeits. It directly enhances trustworthiness as well as traversing efficiency.
- iii. Identification of the meticulous influencers for the decisive information or service retrieval is performed. As a result, the algorithm traverses the network efficiently and accomplishes optimal accuracy than the unprocessed original network.
- iv. The proposed approach identifies the influencers in optimum computational time as compared to the existing approaches.

The remainder of the paper is organized as follows. Section 2 discusses the related work whereas Section 3 presents the proposed methodology in detail. Section 4 presents the results and discussion on the performance of the approach in comparison to the existing approaches whereas Section 5 concludes the paper.

2 Related Work

The growing use of social networks has attracted the attention of researchers to understand the fundamentals of connection patterns and the importance of each node in the network. Each network has nodes with a differently expressed interest in identifying the OSN influencers. In this section, the authors discuss some research works that are related to the aforementioned problem of identifying influencers. Reference [7] defined an influential user as someone who has the ability to influence other community members based on certain distinctive attributes, such as knowledge, personality and other distinguishing characteristics. Likewise, information-seeker, information-provider and strong social contacts are three key functions of influencers [8]. According to Rogers [9], the influential user has three characteristics, namely: (i) prominent social status, (ii) significant social responsibility and (iii) significant social involvement. According to Pei et al. [10], the approaches are divided into two categories: (i) for finding individual influencers (which rank the individual nodes in a network based on their spreading potential) and (ii) for finding multiple influencers (that find the optimal node collection to optimize the final collective influence). For finding individual influencers the steps include centrality methods, such as degree centrality [11] and the k-shell decomposition process [12,13]. Degree Centrality (DC) is a centrality measure, which simply counts the number of connections of an

individual node in the social network whereas the k-shell indicates the location of a node. On the other hand, centrality methods for example betweenness centrality [14] and page rank [15], are employed in dynamic-based steps. In the proposed work, the authors used a topological-based approach that is degree centrality as a metric to identify the Influencer. As it is known that the influencer has an impact on the community, therefore, the authors choose the highest degree node which shows the significant social status in the community.

In literature, there are several methods for identifying influential spreaders. According to Pei et al. [16], PageRank fails in ranking users' influence. Authors reveal that the top spreaders are consistently found in the k-core across a variety of social media sites. Namtirtha et al. [17], identify influential spreaders in an unweighted network using a k-shell hybrid method based on the network structure. According to the authors, not every node in the core area is the most powerful spreader. Even a lower shell node placed appropriately can be the most influential spreader. However, as the networks differ from each other in terms of structure in the case of the collaboration network, purchase network, communication network, social network and so on; therefore, the single indexing method is not suitable for all types of networks. Similarly, the scheme proposed in [17] only considers unweighted social networks. Reference [4] presents a study that uses a community-detection-based approach and various centrality measures to classify the most influential participants in a particular scientific discovery trend on Twitter. Authors identify influencers called opinion leaders in a community or group in online discussion networks like Twitter based on the interaction of people like replies, mentions and retweets. Reference [4] just considers a discussion network while in this research the authors use a directed and Weighted Signed Network (WSN) because communication-based features have been shown to be very important in quantifying the edge weights between users [18].

Structural analysis of the Facebook network is performed by Peng et al. [1]. The influential nodes of the network are recognized based on the high value of both degree centrality and betweenness centrality measures. As well community detection is performed in the overall network specifically around the influential nodes. The top-ranked users are filtered and upon this filtration, four top-ranked nodes of the network are selected based on the high value of degree centrality and betweenness centrality measures. In another research work, the authors utilize some characterization of the soft-set theory to deal with decision-making problems [19–22]. A weighted-based trustworthiness ranking model is presented using the soft-set theory to identify the trusted nodes. The risk of fraudulent transactions is reduced through the identification of the most trusted nodes. Rahmani et al. [23,24] presented an appropriate and optimal area coverage method that includes four phases. In the first phase, the overlap is calculated between the sensing range of sensor nodes in the network. In the next phase, an ON/OFF scheduling mechanism is designed based on fuzzy logic. In the third phase, a method is designed to predict the node replacement time to avoid possible holes in the network. In the last phase, the best replacement strategy is defined to reconstruct and cover the holes in the network. NS2 simulator is used for the simulation to compare the performance of the presented technique.

Alqurashi et al. [25] explored the public reaction and information dissemination by identifying influencers during the period of COVID-19, whereas the COVID-19 outbreak prediction is done in [26]. The authors used both HITS and PageRank algorithms to analyze the influence of information and found that both show the same results. Influencers can now determine the synthesized centrality (SC) in a new approach, according to [27]. The authors estimated the synthesized centrality by multiplying betweenness centrality with normalized degree centrality, then dividing the result by closeness centrality and comparing PageRank, HITS and SC. The final data revealed that the SC outperforms both the PageRank and the HITS when it comes to identifying influencers. In [28], the authors argued that a user can have three categories of influence (in-degree influence, retweet influence

and mention influence) and concluded that individuals can be categorized depending on their type of the influence. In the proposed method, the nodes have been ranked based on the popularity of nodes in the network by examining the in-degree.

Various methods for detecting Influencers have been proposed in the literature. Principal component analysis (PCA) was used in [29] to select and combine user attributes to locate influential people. Canali et al. [29] systematically compared several methodologies, such as data mining and learning methods, hybrid content mining, descriptive, statistical, stochastic and topological measures. In [30], the authors performed a comparative examination of centrality measures such as degree centrality, closeness centrality, betweenness centrality and eigenvector centrality. To identify Influencers in this proposed research, the authors have used a topological technique including the degree centrality statistic, which includes both in-degree and the overall degree of a node as metrics.

Reference [5], proposed a Class-Based technique for searching information in drone networks. It traverses up to six hops and halts in either way regardless of whether the desired information is found or not. It returns the path instead of the actual node that provides accurate information. In addition, the algorithm terminates if any node or path has already been traversed which causes irregular stoppage and may create a hindrance in information retrieval. According to [6], if the information is not found after traversing a specific network path (set of nodes), the path must be re-traversed from the beginning which consumes more time. As compared to the above-discussed techniques, our proposed approach identifies the influencers efficiently by pruning the complex network based on node ranking and finding influencers in specific communities based on high rank. The ranking of nodes is between 0 and 1. Nodes are ranked by examining their in-degree with the out-degree. The pruning mechanism has achieved less computational time while traversing the fewer nodes and reaching the target. During the node traversal, if the node is already traversed then the traversing will be performed on the other nodes excluding the current node. Thus, the proposed approach identifies the influencers in significantly less time than the state-of-the-art approaches.

3 Proposed Methodology

The social network is a network of people tied with each other by weak and strong ties. These ties have some pattern of contacts or interactions between them [31]. The patterns of friendships between individuals [32], business relationships between companies [33] and intermarriages between families [34] are all examples of networks. It is widely assumed that almost all social networks show a community structure where the vertices with a higher density of edges within them form a global network [35]. Moreover, people are divided into groups along the lines of interest, occupation, age and so on. Therefore, the detection of communities in a network is a significant factor in Social Network Analysis and it has drawn a large number of researchers from sociology, biology, computer science and other disciplines [36]. The interlinked social entities exploit the OSNs for fulfilling their needs in explicit disciplines, such as information retrieval or dissemination in a particular community.

Finding legitimate information/influencers in a large and complex network, such as a social network remains a big challenge. With billions of users connected to share information, the Internet is the best example of social networking. It is noted that an information seeker gets massive responses against a query. However, locating specific information or a provider of information among the search results remains a difficult task. The proposed approach takes into account, the participation of the nodes in a social network in terms of their interaction with other nodes to efficiently identify authentic influencer nodes that disperse information. In view of this research work there must be an influencer in a community who provides proper and accurate information or services. The exemplary scenario of

authentic influencer search has been depicted in Fig. 1a, whereas, the three major modules of proposed methodology named, (i) the pruning module, (ii) pruning technique, and (iii) the authentic influencer identification module has been depicted in Fig. 1b.

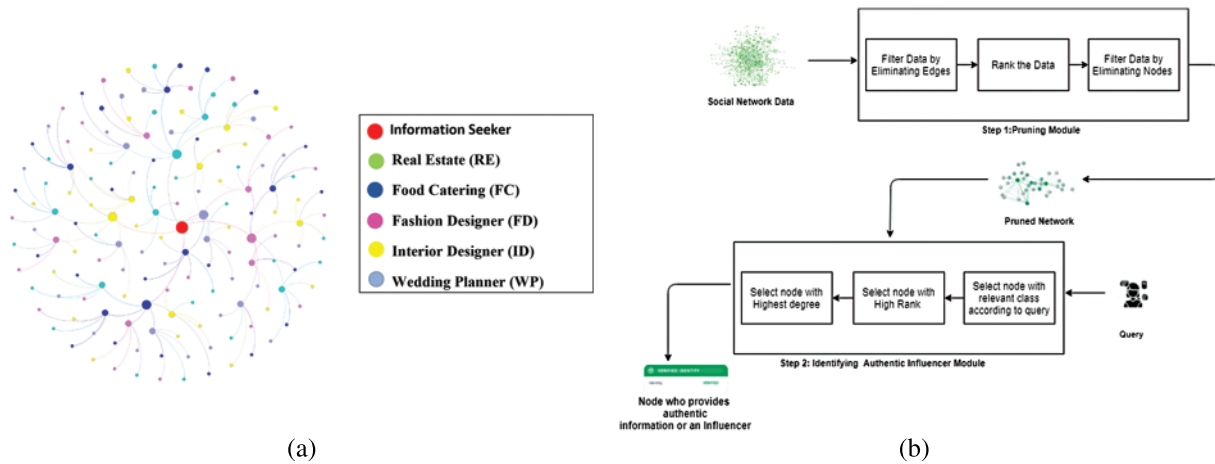


Figure 1: (a) Illustration of influencer identification, (b) block diagram of the proposed algorithm

Let us consider an individual's network where people from different backgrounds, such as medical, fashion industry, real estate, etc. are the members of the network. Assume that node x (information seeker) wants to approach any node y (information provider) that has significant information about real estate. Node y possesses in-depth knowledge about property dealing, such as buying and selling rates of residential and commercial properties and Return on Investment (RoI) projects, etc. There may be numerous people in the network who may have know-how about property dealing but the purpose of this research is to efficiently identify the ones who can provide authentic and reliable information in response to a user query. Fig. 1a represents an exemplary social network for illustrating the information searching mechanism where the node (Red in color) is searching for reliable information about the property or real estate. It requests for searching influencer nodes in the community that provides reliable information and services. The authors have considered communities as the classes of people who share mutual goals. In Fig. 1a, we have considered five communities named as classes which include: (i) Property or Real Estate (RE), (ii) Food Catering (FC), (iii) Fashion Designer (FD), (iv) Interior Designer (ID) and (v) Wedding Planner (WP).

The Red node is a requestor node and it is surrounded by five friend nodes. As our proposed system is a sequentially increasing network, each of these five nodes has five nodes and the depth of the network can vary. In this case, the network depth has been visualized up to three hops. To elaborate, suppose the requestor is always interested in finding information or seeking help to be served in the mentioned fields or communities, where Green nodes contain real estate information, Blue represents food catering information, Pink belongs to fashion designers and Yellow nodes have interior designer and Light Grey nodes are good in wedding planning, or they can provide any service to the requestor related to their fields. Consider an example of real estate information or a real estate service; there are many nodes in a network, instead of traversing all nodes or obtaining information from all nodes, our requestor node knows which node might have the information about the real estate field. Therefore, instead of contacting the other four nodes, the requestor will contact the Green node (real estate class node) directly. If that node has the information, it will provide the requisite information; otherwise, the Green node will take the same step, but instead of asking the other four nodes, it will ask its Green

child node. If the requestor is unable to obtain the information, then it will seek information from the other nodes.

For the practical evaluation, the authors considered an online social network dataset called Bitcoin Alpha that is intermittently connected, directed and weighted. Moreover, the closeness of two nodes defines their professional relationship whereas, the weights depict their transaction frequency and/or trustworthiness. The reason for using this dataset is that there is no network available that exactly matches our scenario. There are some social networks, for example, Facebook but it cannot be accessed due to their protocols. Bitcoin Alpha network is a bidirectional (in-degree (λ), out-degree (μ)) and weighted network (ω) network. In social network analysis, the terms network and graph (G) are used interchangeably. More detail about the Bitcoin Alpha network is presented in Section 4.

In the pruning module, the proposed algorithm revamps the network by evaluating the strength of the relationship among the nodes, prunes the insignificant nodes and outliers and determines their ranks. To refine the network, the graph is pruned based on ranking. The nodes are ranked based on their in-degree. The in-degree states how many nodes influence that node. On the other hand, the out-degree indicates how strongly a node influences the other nodes. The ranking of nodes of a specific class is between 0 and 1. In the pruning method, the nodes having a rank ($\alpha < 0.5$) are removed to reduce the network complexity.

In the authentic influencer identification module, the algorithm in response to a query from a requestor identifies the authentic information provider. As the proposed algorithm is an informed search greedy algorithm, therefore, it prioritizes the traversing in descending order. The algorithm compares the ranks of the successor nodes and traverses the highest rank node. If in some cases a situation arises where two or more nodes have the same rank, then the algorithm selects any one of them randomly. The algorithm in the proposed methodology inherits the characteristics of two existing algorithms namely, (i) Depth First Search (DFS) and (ii) Breadth-First Search (BFS). Therefore, it simultaneously travels in both directions based on the highest rank of the successor nodes. The details of the steps of the proposed method are presented in Section 3.1 and Section 3.3. The properties, advantages and shortcomings of the proposed algorithm are shown in Table 1.

Table 1: Complexity analysis

	Completeness	Optimality	Memory space	Complexity
Proposed algorithm	Not complete	Not optimal	Memory efficient $O(V)$	$O(n+nC_i)$

3.1 Pruning Module

As the size of the network increases, it is harder to identify the influencers from the network. Moreover, it is computationally expensive to manipulate and maintain the network. Our goal in this research is to create a pruned network that models only the most informative affiliations and relationships and efficiently identifies the authentic influencer in significantly less time. For the pruning network, the nodes removed based on the pruning algorithm (Section 3.2) and select the nodes with useful associations and connections. Deleting nodes from the network leads to deleting associated links. However, based on the existing studies [37] on link prediction and social network evolution, deletion or addition of the nodes does not affect the network state and can be easily identified the authentic influencer/information provider.

3.2 Pruning Technique

For the last decade, the size of OSNs is increasing vastly and by July 2021, the number of social media users was expected to be around 4.48 billion [31]. These numbers show that more than half of the world's population is connected through the well-known 17 OSNs. However, this number will be quite bigger if we consider all the other OSNs operating at smaller scales. These networks generate gigantic data and different researchers utilize this data for constructing several models including a human behavioral model. Some real-world applications compute Internet-based human social behavior and use them for different purposes, such as in business marketing [35], politics [32], knowledge management and dissemination of information [33] and other areas. The majority of researchers used OSN-based data over the previous approaches by using sampling or questionnaires [34] because the subsequent is cheaper and faster. However, here a problem of duplicate connections and information arises. To cater to this issue, the authors derived the following proposition of Graph Pruning methodology based on the suggested conditions.

Suppose set $X = \{x_1, x_2, x_3, \dots, x_n\}$ is a topological graph property of $G (N_G, E_G)$ [36]. Let's prune the insignificant elements of a set of nodes $\{n_1, n_2, n_3, \dots, n_n\}$ and/or an ineffectual links/set of edges $\{e_1, e_2, e_3, \dots, e_n\}$ in G . Graph $G' (N'_G, E'_G)$ is a product after the pruning process having the topological graph property $X' = \{x'_1, x'_2, x'_3, \dots, x'_n\}$ of graph G' . To intact the connectivity of the graph as well as to acquire the best nodes, authors have followed the algorithm steps depicted in Algorithm 1. The Graph Pruning method is used to eliminate extraneous nodes from a graph context, reducing the complexity of a large-scale social network while maintaining the original graph features. The pseudocode of pruning Algorithm 1 and Query-Based Algorithm 2 is given below.

Algorithm 1: Pruning Algorithm

Input: Social Internet of Things Network "G"

Output: Desired network "G': Prunednetwork;"

Function prune();

Flag: FALSE

Step1: Elimination of unwanted network connections:

IF ($\lambda t > \lambda s$)
 KEEP Connection(); *ELSE*
 REMOVE Connection()

Step2: Ranking Mechanism:

IF (*node's Indegree* > *node's outdegree*)
 RANK = 1 *ELSE*
 RANK = $\left(\frac{\lambda}{\mu}\right) * \omega$

Step3: Select fruitful nodes in network:

IF (*RANK* >= 0.5)
 KEEP node()
 ELSE
 REMOVE node()
 Return G'

Algorithm 2: Query-Based Algorithm*Input:* Desired network connections “Graph(G)”, Query “Q”, Classes “C_i”*Output:* Authentic Influencer Node “N_k”*Function* graph-tracking();

IF (G! empty)

{ FOR (i=1; i ≤ number of classes;)

 { IF (Q ∈ C_i) { FOR(j=1; j ≤ size. C_i; { IF(Rank_{max} (Node_j, Node_{j+1})&& DegreeN_{maxRank} == max)

{ Message (“Authentic Information Provider found”)

RETURN node;

Break; } ELSE

j++; }

} ELSE

i++ } } ELSE

Error(“Graph is empty, Influencer not found”)

End();

To illustrate the graph pruning of social network, the authors have simulated data under a variety of conditions to determine the cause of selecting explicit vertices and their connections that belong to certain groups or communities and to get the authentic influencer from the community who provide accurate information/services. The algorithm selects the connections and vertices from a social network based on conditions and continues this process until it reaches the threshold. Here, the threshold is required to avoid unnecessary connections or communities. In this paper the authors chose threshold alpha ($\alpha < 0.5$) because of fulfilling the desired condition of removing outlier nodes (isolated nodes, nodes with very less connections, distrustful edges with nodes) from the network. Although, experiments were conducted for different values of α , such as (0.1, 0.2, 0.3 and 0.4) (shown in [Table 2b](#)); however, the results were not found encouraging for these values.

Table 2: (a) Comparison of original and pruned data, (b) network pruning values

	(a)		(b)		
	Original data	Pruned data		Nodes	Edges
Nodes	3783	3302	Original Network	3783	24186
Edges	24186	11590	$\alpha \leq 0.1$	3496	23899
Average Degree	6.393	3.51	$\alpha \leq 0.2$	3464	23612
Average Path Length	3.679	3.351	$\alpha \leq 0.3$	3427	22636
			$\alpha \leq 0.4$	3343	20316
			$\alpha = 0.5$	3302	11590

First, the algorithm traverses the network and selects the nodes that belong to classes and prunes the edges between nodes based on a pruning model. In the pruning module, the algorithm removes the edges or connections whose in-degree of the target node (λ_T) is less than the in-degree of the source node (λ_S). After the edge pruning, the algorithm ranks the nodes by examining their in-degree with the out-degree. If the in-degree of a node is greater than the out-degree, rank is set equal to 1, otherwise, rank is set by dividing λ (in-degree) with μ (out-degree) of the node and multiplying the result with ω (edge weight). After spotting the node's rank, if the rank of a node is less than alpha (which is in this case, 0.5), the node is removed and the algorithm moves to the next iteration. Eventually, the resultant network (graph) contains all the best rank nodes of desired classes.

3.3 Authentic Influencer Identification Module

The algorithm begins by receiving a query from a requestor node. At first, the algorithm checks the requestor's connections whether the requestor node is connected to the network or is an isolated node. If so, the algorithm generates a message "*Graph is Empty*" and halts the process. Otherwise, the algorithm compares the query with the defined classes. In the worst-case scenario, if a query does not belong to any class, the algorithm generates a message "*Query does not belong to any of the classes*" and breaks the process. On the other hand, the algorithm compares the rank and the overall degree of the nodes connected to the requestor and chooses the node with the highest values of both. The process remains to continue until either the algorithm successfully discovers the information provider or the path gets over. Throughout the process, if a situation arises where assorted nodes may have the same rank, the algorithm picks any of the nodes randomly and accomplishes the task. As we know, the social networks dataset does not show the profile of any node and assigns an anonymous ID to each node. Similarly, the dataset used in this approach lacks the actual profile of the nodes, therefore, the algorithm returns the node ID. The pseudocode of the proposed Query-Based algorithm is given below.

In Algorithm 2, the input is pruned network, query and the classes and the authentic node's address/ID is output. In the algorithm, the tracking function first checks whether the graph is empty or not. If the graph is not empty, then there is a loop that checks the number of classes. In every iteration, the condition checks whether the query belongs to any class. If the query belongs to a specific class, the next loop will traverse to the size of the class i.e., all the nodes that belong to class C_i will be traversed. The algorithm then compares the rank and the overall degree of the nodes connected to the requestor and chooses the node with the highest values of both and that particular node is an influencer in class.

4 Results and Discussion

For the experiments, Bitcoin Alpha dataset [38] has been used, which is an open-source social network data repository. The dataset includes the true values of bitcoin traders' confidence ratings on the Bitcoin Alpha network. The rating scale runs from -10 (complete distrust) to $+10$ (complete trust). These ratings are the weights of edges from the source node to the target node, which indicate the transactions' trustworthiness and influence on each other. It is a bidirectional weighted network where each direction has an independent weight. Individuals are represented by nodes and the relations between individuals are represented by the edges. There are 3,783 vertices (nodes) and 24,186 edges (connections) in this network. It is observed that there are enormous connections available in this network that neither trust each other nor make any transactions. Such connections have a negative rating which shows the distrust and insignificance of the links. Therefore, the presence of this nothingness may cause network complexity and hindrance in identifying the right node for the required

service. The proposed algorithm analyzes the network connections and shears the unwanted and/or insignificant connections to reduce the complexity of the network. Consequently, after pruning, the network has 3,302 nodes and 11,590 edges. Fig. 2a depicts the original Bitcoin Alpha network whereas the pruned network is depicted in Fig. 2b.

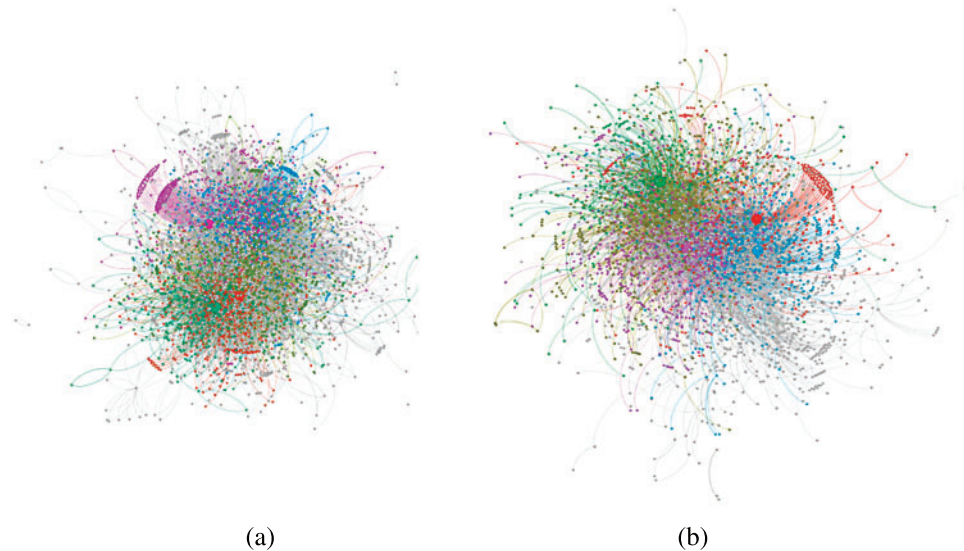


Figure 2: Bitcoin alpha (a) original network (b) pruned network

Table 2a shows a comparison of the original and pruned networks. For network visualization an open-source social network simulator named Gephi [39] is used. Gephi has the capability of visualizing complex structures by using different built-in algorithms. For the network visualization, the authors have used the Yifan-Hu algorithm [40]. The remaining network nodes are ranked with respect to their peculiarity and the nodes having a rank ($\alpha < 0.5$) are filtered out. For the network measures corresponding to the value of α refer to Table 2b

The key concept behind ranking is the goal of preserving the value of closeness bonds in social networks; however, there are few real-world datasets with edge ratings in social networks, therefore, this dataset is ideal for our case. This rating is used to rank the nodes in a conferred exemplary scenario. To minimize computational cost and rapid identification of an important/influencer node, the proposed approach just keeps the appropriate links. The process of network pruning with respect to the nodes and edges is explained and presented in Section 3.1.

Although the degree distribution of a network reveals only a small amount of knowledge about the network, however, it does reveal the network structure information. In a social network, the majority of the nodes have comparatively lower degrees, whereas a few nodes have a high degree. Therefore, to observe the network structure, researchers have calculated the average degree distribution of both the original and pruned networks. Figs. 3a and 3b show the average degree distribution of the original and the pruned networks. The average degree of the original network has been measured at 6.393, while after pruning it is reduced to 3.51. The in-degree and out-degree of both networks are shown in Figs. 3c, 3d and Figs. 3e, 3f respectively.

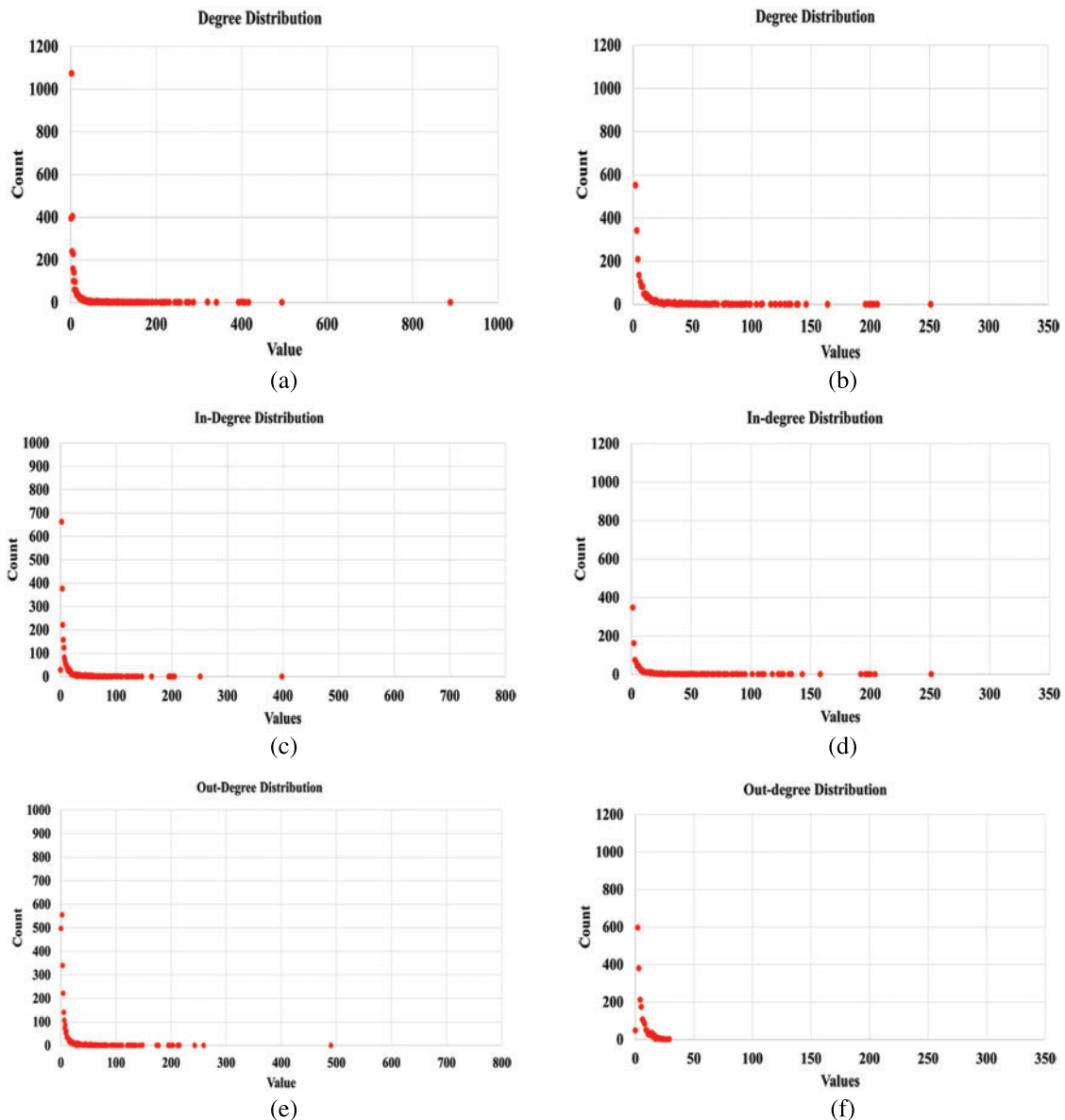


Figure 3: Degree distribution (a) original (b) pruned; in-degree (c) original (d) pruned; out-degree (e) original (f) pruned

The experiments were performed on a personal computer with an Intel(R) Core™ i7-3520 M CPU running at 2.90 GHz, 8 GB RAM and a 64-bit Windows 10 operating system for the evaluation of the proposed methodology. Python programming language was used to implement the algorithm. The authors have investigated the effectiveness of the proposed approach in terms of execution time and the number of hops traversed by comparing it with two existing works [5,6]. The execution time of these algorithms is shown in Figs. 4a and 4b. The X-axis displays the 100 runs of all three algorithms whereas the Y-axis represents the execution time taken by the existing and proposed approach against

each run. Figs. 5a and 5b show the comparison of the proposed algorithm with existing algorithms in terms of traversing a number of hops in the network. The X-axis displays the 100 runs of all three algorithms whereas the Y-axis represents the number of hops traversed in a network. These graphs show that by network pruning the proposed algorithm traverses a smaller number of hops/nodes and get the desired result easily and effectively.

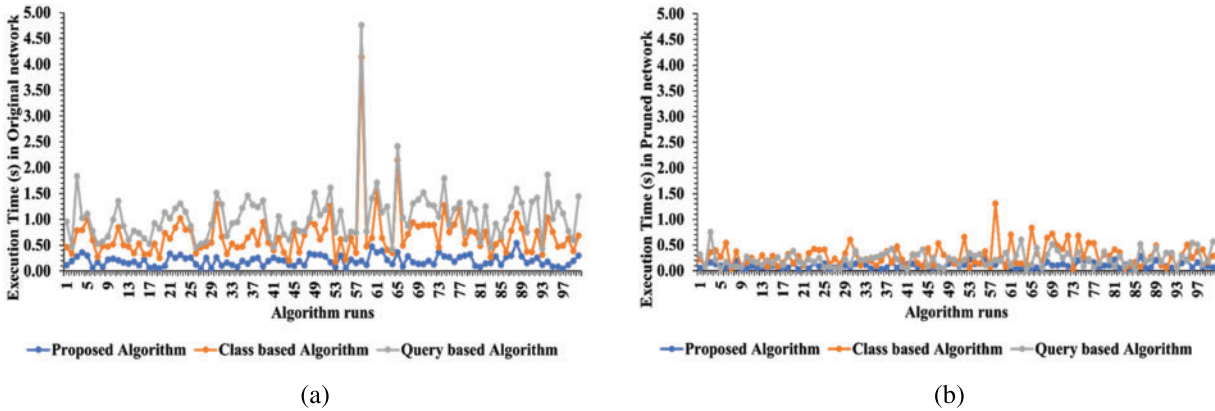


Figure 4: Computation time comparison with existing algorithms (a) original network (b) pruned network

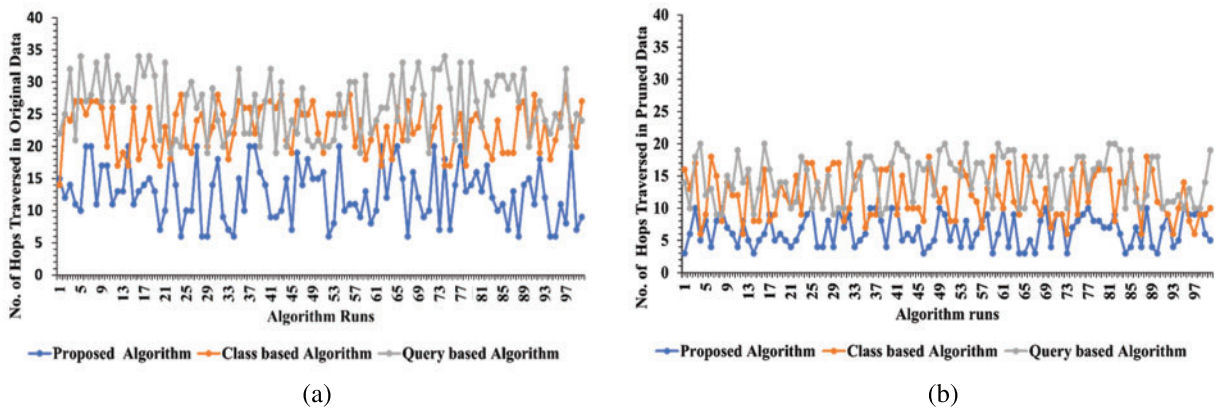


Figure 5: Number of hops traversed in comparison with existing algorithms (a) original network (b) pruned network

The mean execution time has also been calculated which is 0.49 s in the Query-Based algorithm [6], 0.39 s in the class-based algorithm [5] and 0.20 s in the proposed approach in the original network. In pruned network, the mean execution time is 0.24 s in the Query-Based algorithm [6], 0.29 s in the Class-Based algorithm [5] and 0.10 s in the proposed approach, respectively. The average execution time of these algorithms for a hundred algorithm runs against both networks can be seen in Table 3. The execution time of the proposed algorithm is quite low when considering either the overall network attributes of the original network or the pruned network. These comparison graphs illustrate that our proposed approach performs better in terms of time consumption.

Table 3: Average time of hundred algorithm runs against both (original/pruned) networks in (s)

	Algorithms		
	Query-based algorithm	Class-based algorithm	Proposed algorithm
Original network	0.49	0.39	0.20
Pruned network	0.24	0.29	0.10

For the evaluation of our proposed approach, two networks “original Bitcoin Alpha and pruned network” are traversed under the same conditions to identify the authentic influencer node. For example, in the Real Estate class “node 2” is the authentic influencer. The proposed approach is greedy in nature because it chooses the highest-ranked node in each layer and traverses that node rather than following the deployment sequence of the nodes. During the 100 algorithm runs, the authors observed that the algorithm failed eight times in identifying the authentic influencer which is “2” while traversing the original Bitcoin Alpha network. On the other hand, it failed only twice for the pruned network. [Table 4](#) shows the results pattern of 100 runs for both the original and the pruned networks. It clearly shows that there are more nodes in the original network to be traversed to reach the desired node whereas in the case of pruned network, due to the removal of unwanted and insignificant nodes, the algorithm traverses the fewer nodes to reach the target node.

Table 4: Nodes traversed in original and pruned networks to identify the influencer

Original network	Pruned network
52-3-6-7-26-30-33-39-43-47-48-58-61-63-65-68-73-80-83-85-95-100-111-116-130-134-136-143-150-154-156-157-168-170-172-177-179-182-184-188-196-198-216-218-227-228-245-249-252-259-260-301-312-318-327-329-330-333-358-360-369-491-585-708-801-1159-1691-2336-7334-7514-7595-7598-7601-7602-2	14-3-6-7-26-30-33-585-708-801-1691-2336-7334-7514-7595-7598-7601-7602-2
19-61-63-65-68-73-80-83-85-95-100-111-116-21-3-6-7-26-30-33-39-43-47-48-58-130-134-136-143-150-154-156-157-168-170-172-177-179-182-184-188-196-198-216-218-227-228-245-249-252-259-260-301-312-318-327-329-330-333-358-360-369-491-585-708-801-1159-1691-2336-7334-7514-2	67-3-6-7-26-33-39-43-47-48-58-61-63-65-68-73-80-83-85-95-100-111-116-130-134--136-2
59-66-70-104-113-139-146-148-166-176-197-202-211-255-272-274-278-1792-1811-1841-2319-2322-3079-3138-5679-7461-937-3002-7551-25-639-687-51-69-256-5-32-804-1020-106-129-534-815-967-35-267-689-38-57-160-933-208-20-2	70-3-6-7-26-30-21-22-42-50-66-70-104-113-139-146-7599-34-36-49-118-128-759-1568-4

(Continued)

Table 4: Continued

Original network	Pruned network
9-24-16-102-112-117-145-161-190-214-428-536-620-988-813-1008-1153-2244-34-36-49-118-128-759-1568-7564-10-29-53-56-25-639-687-51-69-256-5-82-87-141-222-375-96-103-107-123-124-174-207-244-288-430-433-886-969-1045-1157-1171-4	5-3-6-7-26-30-33-39-43-47-48-58-61-63-111-116-130-134-136-7601-7602-2
113-139-146-9-24-112-117-145-161-190-214-148-166-176-197-202-211-255-272-274-278-286-304-305-316-357-409-452-474-482-505-517-546-548-549-553-587-588-681-685-740-796-851-854-871-939-1022-1045-1157-1171-4-16-102-536-620-988-7564-10-29-53-56-96-11	8-3-6-7-26-30-33-585-708-801-1691-2336-7334-7514-7595-7598-7601-7602-2

The results show that the deletion of unnecessary and outlier nodes has a significant impact on the overall accuracy of the algorithm, which is 92% and 98%, for both the original and the pruned networks, respectively. The experiments reveal that the algorithm takes significantly lesser computational time for the pruned network than the original network and accurately identifies the intended node. The histogram and the line graph of execution time for both the original and the pruned networks are shown in Figs. 6a, 6b and 7. It is evident that the proposition of the pruning network and ranking process helps in reducing the time consumption. The execution time is quite low when considering the overall network attributes of the original and the pruned networks, indicating that our proposed approach is efficient.

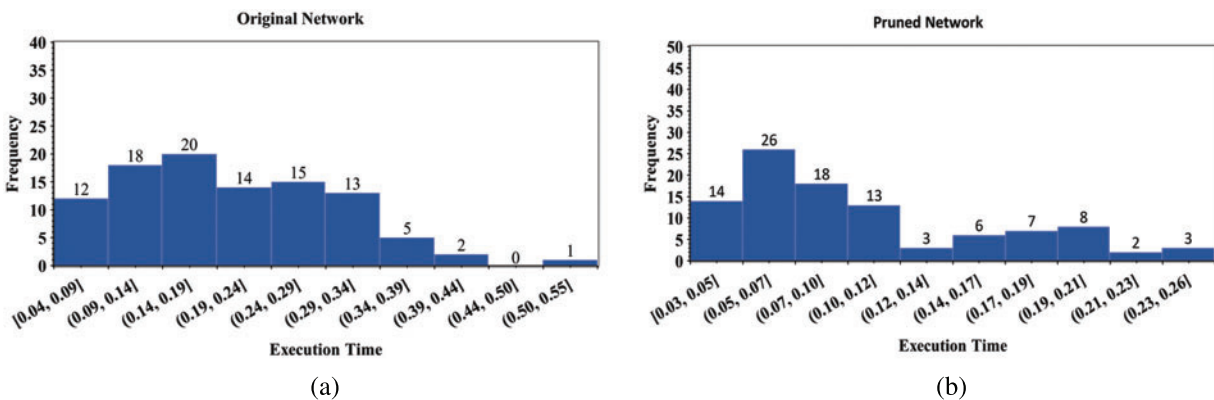


Figure 6: Execution time of (a) original network, (b) pruned network

The complexity of the pruning algorithm and the query-based approach has also been calculated. In the pruning algorithm, the unwanted network connections are eliminated and nodes are ranked. After ranking, the nodes are selected. Therefore, the complexity of these steps of the algorithm depends on the number of nodes n and it turns out to be: $O(n) + O(n) + O(n) = O(n)$

Hence, in the pruning algorithm, the overall complexity is $O(n)$.

Likewise, in a query-based algorithm, the complexity depends on the number of classes (C_i) and the size of nodes in class (nC_i).

Thus, the overall complexity would be: $O(n) + O(C_i + nC_i)$.

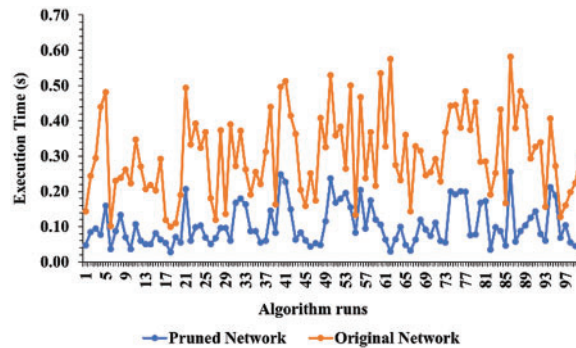


Figure 7: Execution time of original and pruned network

This research concludes that the proposed greedy approach not only reduces the network complexity but also minimizes the computational time. Moreover, the approach offers an efficient mechanism to enhance the traversal efficiency of the network. Despite its benefits, the proposed approach compares the rank at each level and consequently the approach may run in a loop when the network is too complex even after pruning. Therefore, the mentioned aspect may be explored in the future to address the issue.

5 Conclusion

In this paper, the authors have proposed an approach for identifying authentic influencer nodes in social networks in an efficient way. Numerous elements are required for a node to become an influential spreader including maximum connectivity among nodes and the magnitude of impact given by neighbors. The proposed approach ranks the nodes based on the in-degree, out-degree, and the link weights with their neighbor nodes. The proposed approach achieves the process of authentic influencer search based on a node ranking parameter in significantly less time. For the accomplishment of the desired results, initially the network has been pruned to remove the outlier and unnecessary nodes. Whereas, the authentic influencer has been identified based on the highest rank in that particular class. After performing the complete experiment the authors have observed that the proposed algorithm significantly reduces the network's complexity of the network and eliminates unnecessary connections thereby enhancing the time efficiency. Later, the performance of the proposed approach is evaluated by comparing it with two existing approaches i.e., the Query-Based information search algorithm and a Class-Based Information searching algorithm. The mean execution time of all three algorithms has been calculated against original and pruned network. The mean execution time of the Query-Based, Class-Based, and proposed Query-Based Greedy algorithm against original network are 0.49, 0.39, and 0.20 s respectively. Whereas, it is 0.24, 0.29, and 0.10 s respectively against pruned network. In conclusion, the proposed Query-Based greedy approach can be effectively employed for the efficient and accurate identification of influencers in continuously growing social networks.

Acknowledgement: Farah Batool has done all the experiments and wrote the manuscript. Dr. Abdul Rehman has revised the manuscript in several meetings, chose the experimental tool and refined the idea and results. Dr. Assad Abbas has supervised this research work. Whereas, Dr. Dongsun Kim, Dr. Raheel Nawaz, and Dr. Tahir Mustafa Madni reviewed the manuscript and helped in improving the paper quality.

Funding Statement: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1A5A1021944 and 2021R1I1A3048013). Additionally, the research was supported by Kyungpook National University Research Fund, 2020.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] S. Peng, G. Wang and D. Xie, "Social influence analysis in social networking Big data: Opportunities and challenges," *IEEE Network*, vol. 31, no. 1, pp. 11–17, 2017.
- [2] A. Rehman, A. Paul, A. U. Rehman, F. Amin, R. M. Asif *et al.*, "An efficient friendship selection mechanism for an individual's small world in social internet of things," in *2020 Int. Conf. on Engineering and Emerging Technologies (ICEET)*, Lahore, Pakistan, pp. 1–6, 2020.
- [3] S. Kumar and B. S. Panda, "Identifying influential nodes in social networks: Neighborhood coreness based voting approach," *Physica A: Statistical Mechanics and its Applications*, vol. 553, pp. 13, 2020.
- [4] A. U. Rehman, A. Jiang, A. Rehman, A. Paul, S. Din *et al.*, "Identification and role of opinion leaders in information diffusion for online discussion network," *Journal of Ambient Intelligence and Humanized Computing*, vol. 19, no. 9, pp. 1–13, 2020.
- [5] A. Rehman, A. Paul, A. Ahmad and G. Jeon, "A novel class based searching algorithm in small world internet of drone network," *Computer Communications*, vol. 157, pp. 329–335, 2020.
- [6] A. Rehman, A. Paul and A. Ahmad, "A query based information search in an individual's small world of social internet of things," *Computer Communications*, vol. 163, pp. 176–185, 2020.
- [7] P. Kotler, "Marketing management," in *Marketing Management: A South Asian Perspective*, 13th ed., Dehli, India: Dorling Kindersley, pp. 1–90, 2007.
- [8] R. S. Burt, "The social capital of opinion leaders," *The Annals of the American Academy of Political and Social Science*, vol. 566, no. 1, pp. 37–54, 1999.
- [9] E. M. Rogers, "Diffusion of Innovation," in *Simon & Schuster*, 5th ed., New York, NY, USA: Free Press, pp. 576, 2003.
- [10] S. Pei, F. Morone and H. A. Makse, "Theories for Influencer Identification in Complex Networks," In: S. Lehmann and Y.Y. Ahn (Eds.), *Complex Spreading Phenomena in Social System*, Switzerland: Springer Cham, pp. 125–148, 2018.
- [11] P. Bonacich, "Factoring and weighting approaches to status scores and clique identification," *The Journal of Mathematical Sociology*, vol. 2, no. 1, pp. 113–120, 1972.
- [12] S. N. Dorogovtsev, A. V. Goltsev and J. F. F. Mendes, "K-Core organization of complex networks," *Physical Review Letters*, vol. 96, no. 4, pp. 4, 2006.
- [13] S. B. Seidman, "Network structure and minimum degree," *Social Networks*, vol. 5, no. 3, pp. 269–287, 1983.
- [14] G. Sabidussi, "The centrality index of a graph," *Psychometrika*, vol. 31, pp. 581–603, 1966.
- [15] L. Page, S. Brin, R. Motwani and T. Winograd, "The PageRank citation ranking: Bringing order to the Web," Stanford InfoLab, 1999. [Online]. Available: <http://ilpubs.stanford.edu:8090/422/>.
- [16] S. Pei, L. Muchnik, J. S. Andrade, Z. Zheng and H. A. Makse, "Searching for superspreaders of information in real-world social media," *Scientific Reports*, vol. 4, no. 5547, pp. 1–12, 2014.
- [17] A. Namtirtha, A. Dutta and B. Dutta, "Identifying influential spreaders in complex networks based on Kshell hybrid method," *Physica A: Statistical Mechanics and its Applications*, vol. 499, pp. 310–324, 2018.
- [18] E. Gilbert, "Predicting tie strength in a new medium," in *Proc. of the ACM 2012 Conf. on Computer Supported Cooperative Work (CSCW)*, Seattle, Washington, USA, pp. 1047–1056, 2012.
- [19] A. U. Rehman, A. Jiang, A. Rehman and A. Paul, "Weighted based trustworthiness ranking in social internet of things by using Soft Set theory," in *2019 IEEE 5th Int. Conf. on Computer and Communications (ICCC)*, Chengdu, China, pp. 1644–1648, 2019.

- [20] A. Rehman, M. M. Rathore, A. Paul, F. Saeed, R. W. Ahmad *et al.*, “Vehicular traffic optimisation and even distribution using ant colony in smart city environment,” *IET Intelligent Transport Systems*, vol. 12, no. 7, pp. 594–601, 2018.
- [21] A. U. Rehman, R. A. Naqvi, A. Rehman, A. Paul, M. T. Sadiq *et al.*, “A trustworthy SIoT aware mechanism as an enabler for citizen services in smart cities,” *Electronics*, vol. 9, no. 6, pp. 918, 2020.
- [22] M. Raza, A. R. Barkat, A. U. Rehman, A. Rehman and I. Ullah, “Mobile crowdsensing based architecture for intelligent traffic prediction and quickest path selection,” in *2020 Int. Conf. on UK-China Emerging Technologies, UCET 2020*, Glasgow, UK, pp. 1–4, 2020.
- [23] A. M. Rahmani, R. Ali Naqvi, M. Hussain Malik, T. S. Malik, M. Sadrishojaei *et al.*, “E-Learning development based on internet of things and blockchain technology during COVID-19 pandemic,” *Mathematics*, vol. 9, no. 24, pp. 13, 2021.
- [24] A. M. Rahmani, S. Ali, M. S. Yousefpoor, E. Yousefpoor, R. A. Naqvi *et al.*, “An area coverage scheme based on fuzzy logic and shuffled frog-leaping algorithm (SFLA) in heterogeneous wireless sensor networks,” *Mathematics*, vol. 9, no. 18, pp. 41, 2021.
- [25] S. Alqurashi, A. Alashaikh, N. Mahmoud, E. Alanazi, S. Ahmed *et al.*, “Identifying information superspreaders of COVID-19 from arabic tweets,” in *2021 3rd IEEE Middle East and North Africa Communications Conf. (MENACOMM)*, Agadir, Morocco, pp. 109–114, 2021.
- [26] T. Sher, A. Rehman and D. Kim, “COVID-19 outbreak prediction by using machine learning algorithms,” *Computers, Materials & Continua*, vol. 74, no. 1, pp. 1561–1574, 2023.
- [27] A. Namtirtha, S. Gupta, A. Dutta, B. Dutta and F. Coenen, “Algorithm for finding influential user: Based on user’s information diffusion region,” in *2016 IEEE Region 10 Conf. (TENCON)*, Singapore, pp. 2734–2738, 2017.
- [28] M. Cha, H. Haddadi, F. Benevenuto and K. P. Gummadi, “Measuring user influence in twitter: The million follower fallacy,” in *Fourth Int. AAAI Conf. on Weblogs and Social Media (ICWSM)*, Washington, DC, pp. 10–17, 2010.
- [29] C. Canali and R. Lancellotti, “A quantitative methodology based on component analysis to identify key users in social networks,” *International Journal of Social Network Mining*, vol. 1, no. 1, pp. 27–50, 2012.
- [30] S. Arrami, W. Oueslati and J. Akaichi, “Detection of opinion leaders in social networks: A survey,” In: G. D. Pietro, L. Gallo, R. J. Howlett and L. C. Jain (Eds.), *Intelligent Interactive Multimedia Systems and Services*, Switzerland: Springer Cham, pp. 362–370, 2018.
- [31] S. Kemp, “DIGITAL 2022: GLOBAL OVERVIEW REPORT,” Datareportal, 2022. [Online]. Available: <https://datareportal.com/reports/digital-2022-global-overview-report>.
- [32] M. D. Ward, K. Stovel and A. Sacks, “Network analysis and political science,” *Annual Review of Political Science*, vol. 14, pp. 245–264, 2011.
- [33] F. Zhang and L. Zong, “Dissemination of word of mouth based on SNA centrality modeling and power of actors-An empirical analysis of internet word of mouth,” *International Journal of Business Administration*, vol. 5, no. 5, pp. 65, 2014.
- [34] A. Alamsyah, M. Paryasto, F. J. Putra and R. Himmawan, “Network text analysis to summarize online conversations for marketing intelligence efforts in telecommunication industry,” in *2016 4th Int. Conf. on Information and Communication Technology (ICoICT)*, Bandung, Indonesia, pp. 1–5, 2016.
- [35] A. Alamsyah and Y. Peranginangin, “Network market analysis using large scale social network conversation of Indonesia’s fast food industry,” in *2015 3rd Int. Conf. on Information and Communication Technology (ICoICT)*, Nusa Dua, Bali, Indonesia, pp. 327–331, 2015.
- [36] A. Alamsyah, Y. Priyana and B. Rahardjo, “Fast summarization of large-scale social network using graph pruning based on K-core property,” *Journal of Theoretical and Applied Information Technology*, vol. 95, no. 16, pp. 3749–3757, 2017.
- [37] R. Missaoui, E. Negre, D. Anggraini and J. Vaillancourt, “Social network restructuring after a node removal,” *International Journal of Web Engineering and Technology*, vol. 8, no. 1, pp. 4–26, 2013.

- [38] J. Leskovec and A. Krevl, "SNAP datasets: Stanford large network dataset collection," SNAP, 2014. [Online]. Available: <http://snap.stanford.edu/data>.
- [39] M. Bastian, S. Heymann and M. Jacomy, "Gephi: An open source software for exploring and manipulating networks," in *Third Int. AAAI Conf. on Weblogs and Social Media (ICWSM-09)*, San Jose, California, pp. 361–362, 2009.
- [40] Y. Hu, "Efficient, high-quality force-directed graph drawing," *The Mathematica Journal*, vol. 10, no. 1, pp. 37–71, 2006.