Journal of Information Systems Engineering and Management
2025, 10(21s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

HSF-Mobile V2 Net based Influence Maximization model for large scale networks

Vishakha Shelke[1,2] and Dr. Ashish Jadhav[3]

1. Department of Computer Engineering, Ramrao Adik Institute of Technology, D. Y. Patil deemed to be University, Nerul, Navi Mumbai, 400706, Maharashtra, India.

2. Department of Computer Science and Engineering (ICB), Dwarkadas J Sanghvi College of Engineering, Vile Parle(W), 400056, Maharashtra, India.

3. Department of Information Technology, Ramrao Adik Institute of Technology, D. Y. Patil deemed to be University, Nerul, Navi Mumbai, 400706, Maharashtra, India.

Corresponding author(s). E-mail(s): vishakhashelke21@gmail.com;
Contributing authors: ashish.jadhav@rait.ac.in;

ARTICLE INFO

Received: 25 Dec 2024

Revised: 30 Jan 2025

Accepted: 20 Feb 2025

ABSTRACT

Influence Maximization (IM) in online social networks deals with selection of seed nodes to maximize their influence spread underlying any influence diffusion models. In this paper, Heaviside Step Function (HSF)-Mobile V2-based IM model in social networks is proposed. At first, large-scale networks, such as Facebook, Instagram, etc. are established. Then, the graph is constructed for the large-scale network using the Sociogram technique. After that, graph analytics is performed for evaluating the influence score. On the other hand, popular Messages/event's propagation distance is calculated for evaluating the influence score of the most influenced node. The influence score is given as input into the HSF-Mobile V2 Net model for selecting the seed node. For these selected seed nodes, a new graph is constructed. Then, the baseline strength and requests are measured for the nodes presented in the newly constructed graph. Finally, the child node from the constructed graph is considered for evaluating the most influenced user using Naive-GCA algorithm. Lastly, experimentation analysis is done to verify the superiority of the proposed methodology.

Keywords: Sociogram Model(SM), Linear Threshold (LT), Heaviside Step Function activated Mobile Net V2 model (HSF-Mobile V2 Net), Naive sharding- Growing Clustering Algorithm (Naive-GCA), Naive Sharding (NS).

1. INTRODUCTION

Social networks significantly increased relevance in our daily life due to the various social media applications. Social media data is inherently streamed, gathered and analysed in real-time from a variety of sources which consist of huge volume of data. These large social networks require scalable solutions while performing social network analysis.

Nowadays viral marketing is an effective marketing strategy to capture large audience for newly introduced products. Large-scale online viral marketing campaigns carried out on online social networks like Facebook, Instagram and Twitter, because of their rising popularity among people. The influence maximization (IM) target to find limited group of seed nodes in a social network who can maximize their influence spread in network. According to survey influence diffusion models are used to model the influence spread of the node. Modelling influence distribution and influence maximization are two crucial technological components that would make it possible to conduct such extensive online viral marketing.

For modeling influence spread, the independent cascade (IC) diffusion model and its extensions are used to study IM [1] [2]. Most current studies on the impact of the maximization problem use the IC model and assume that the dynamics of information dispersion among people are independent [3]. For large networks greedy approximation approach with IC diffusion model select the node with the largest influence spread and add to the seed set in each round. These computationally intensive algorithms require more resources as the network size grows. It is difficult to scale the algorithm for the network with million or billion edges and nodes. The greedy algorithm is very slow to produce an accurate estimate of influence spread in large networks so Monte-Carlo simulations with numerous simulation applied. However, it has been demonstrated that it is NP-hard to calculate influence spread given a seed set [4, 5].

Diffusion models simulate the influence spread in large scale networks. Comparative influence spread analysis with different diffusion models with IM algorithms also highlight the need of scalable IM algorithm in large scale networks [6]. Different measures like centrality, k-core applied while searching the top(k) seed users within community. But the community structure evolves over the period of time so this is challenging as network structure grows. Community structure in dynamic nature of network can be efficiently identified by Wilcoxon Hypothesized K-Means (WH-KMA) algorithm. But implementation of existing IM algorithms is time consuming for dynamic snapshots. So scalable IM algorithms are required [7,8]. Influence maximization field requires the research to address various challenges [9, 10].

Another side of viral marketing is propagation of fake news. Irrespective of correctness of information, news goes viral in the network and wrong information gets spread within network. We can estimate the impact of viral information by influence spread to overcome the impact of news but in this case existing algorithms are time consuming and not efficient. So scalability is major aspect of IM algorithm [11].

As the size of the social network grows determining the initial diffusion seed nodes becomes increasingly challenging. A scalable influence maximization algorithm is essential for large social networks to efficiently identify initial diffusion seed nodes who can maximize the spread of information. Traditional algorithms become impractical as the size of social networks grows due to their high computational complexity and resource demands. A scalable algorithm ensures that influence maximization remains feasible and effective, enabling timely and cost-efficient marketing.

1.1 Problem Statement

The existing research methodologies have some drawbacks, which are listed below:

- Any social connection that exists between any two individual's changes with time and can occasionally be obscured, even in expansive networks.
- Due to distributed hashing schemes, the scalability of a network becomes a challenge.
- Seed node selection considers only direct neighbors and not the indirect node.

To overcome these problems, we proposed HSF-Mobile V2 Net for IM. The contributions of the proposed work are,

- A most influenced user is evaluated using Naive-GCA in a periodical manner.
- The influence score is evaluated using DVR-Fat Tree by evaluating the message propagation distance.
- A seed node is selected using HSF-Mobile V2 Net for considering direct and indirect nodes.

The remaining paper is organized as follows: section 2 elaborate literature survey, proposed methodology is explained in section 3, the experimental evaluation of the proposed model is given in section 4, and section 5 highlight conclusion of the paper with future planned work.

2. RELATED LITERATURE REVIEW

Saxena et al. developed hurst exponent for IM. Hurst exponent had been computed for assessing the self-similarity trend in nodes. Experimental results were achieved, which showed that the developed algorithm had been found to perform better. However, the accuracy of identifying influenced nodes was low due to considering the few amounts of node connections [12].

To identify seed set in influence spread Li et al. introduced influence user discovery algorithm. They implemented two-factor information propagation model considering the complexity of real networks. Experimental evaluation demonstrated the effectiveness of the model. However, the time consumption of the developed model was high due to the random position updation in the developed heuristic model on diversity networks [13]. Lozano et al presented fairness in IM through randomization. A heuristic approach was utilized for selecting seed nodes from the network. For the node-based problem, the approximation was achieved by the polynomial-sized linear program. Through the experiments, the multiplicative-weight routine yielded a better approximation result. Since the set-based features were not considered, the efficiency was still low [14].

Purba et al. demonstrated IM diffusion models. These models incorporated the user's activeness and engagement factors. Experiments demonstrated the higher efficiency of the developed model. Yet, the runtime was unbearable due to inappropriate tuning of parameters in the greedy algorithm [15]. Chen et al. propounded a multi-feature budgeted profit maximization (MBPM) oracle and a noise model was considered for the MBPM problem. Experimental results showed efficiencies of the model. But, the efficiency was not stable due to considering the features from the different datasets [16].

Zuo et al. exhibited an Online Competitive IM (OCIM). The proposed framework was adopted for the OCIM problem. Experimental evaluations demonstrated the effectiveness of the algorithms. The presence of a harder offline problem was solved by a new oracle with less feedback and easier offline computation that yielded a worse frequency regret bound [17]. Chatterjee et al. demonstrated a meta-heuristic approach for IM in social networks. A problem-specific local search strategy had been incorporated to increase exploitation in the search space. Experiments performed markedly well. However, tuning the parameters needed some degree of experimentation, which made the model more complex [18]. Yufeng Wang et al. implemented ITO IM algorithm with new index node key degree to model influence spread in large social networks. They compared the work with the state-of-the-art algorithms and proved significant contribution of their proposed work. However, their work was limited to disjoint communities [19].

Mehdi et al. applied combination of different measures and filtering approach for the effective seed node identification. Their work effectively identified seed node with less time complexity. This paper suggested reduction in run time of IM algorithm which is considered while designing proposed algorithm [20]. Jyothimon et al. introduced seed node selection in dynamic social network and performed experimental evaluation on various IM algorithm for the runtime evaluation. Research gap of reduced computational time in large social network is considered for proposed algorithm implementation [21]. Zareie et al. presented the challenges and research directions in the field of influence maximization and discussed the need of scalable algorithm to reduce the time complexity of existing IM algorithms [22]. Sumit Kumar et.al. implemented four-tier framework for influence maximization and applied the concept of clustering of similar nodes. Convolution Neural Network is applied on the proposed framework. The performance of CNN for accuracy is considered for our proposed model comparison [23].

Dizaji et al. implemented deep Q-learning (DQL) approach for influence maximization on dynamic networks and their proposed system evaluated for large size test network. Experimental work shows improved performance. The proposed work can be extended for the extensible large scale networks [24]. Shuxin Yang et al. proposed a Balanced IM framework based on deep reinforcement learning based on single entity network. Their proposed work outperforms the state of art algorithms for large scale networks. The research is open for applying diffusion models for IM [25]. Farzaneh Kazemzadeh et al. identified seed set based on optimal pruning approach on communities for large scale networks. Proposed algorithm shows marginal performance for efficiency and run time. So proposed work also implemented optimization concept to optimize the efficiency or higher propagation range of message passing in the large-scale network for scalability problem solution [26]. Yonggang Liu et al. provided one hop adaptation methodology for solving the IM problem in large scalable networks and presented results in term of computational cost of selecting seed set [27].

Chang Guo et al. implemented multi-scale propagation based on meta-path. Their experimental work provided feasible solution with good performance. Future direction is given in this paper is scalable algorithm for large scale networks [28]. Evren Guney et al. applied branch and cut approach for IM problem based on probabilistic IC model. The branch and cut method gave good result due to consideration of maximal covering [29]. Maitreyee Ganguly et al. implemented centrality based strategy for influence maximization instead of random node selection in cluster. Their strategy worked so well due to consideration of seed nodes in the large clusters. Proposed system implemented the clustering approach for efficiency [30].

M. Venunath et al. proposed community based influential maximization. The spread of influence is effective due to selection of seed nodes within communities and consideration of their spread in local communities [31]. Mani Saketh et al. provided detail review of existing IM algorithms for finding seed nodes in complex networks [32]. Banerjee S et al. presented the detail survey of work done in social influence maximization and discussed future trends [33].

Srinivas S et al proposed novel method based on mathematical model to identify seed nodes in communities and suggested to apply meta heuristic methods to find seed nodes in large scale networks [34].

Sara Ahajjam et al. proposed scalable solution for seed node detection using leader community detection approach. Here the seed nodes are selected and the spread of influence is identified using similarity metric [35]. Jinfang Sheng proposed the novel community detection algorithm inspired by human social behavior. Applied the closeness metric to identify user interest and then community structures are identified [36]. Krista Rizman Zalik et al. proposed community detection using node entropy to speed up the convergence and effective identification of communities [37] Another effective use of Graph Neural Network (GNN) framework for representing node embeddings for link prediction is presented in [38] [39] for large-scale network graphs. The experimental result exhibit the reduce complexity in the link prediction due to adaptation of Graph Convolutional Networks.

3. PROPOSED METHODOLOGY FOR SCALABLE INFLUENCE MAXIMIZATION FOR LARGE-SCALE NETWORKS

The proposed model efficiently identifies the social IM in a continuous timing manner. Figure 1 represents the system architecture of the proposed methodology.

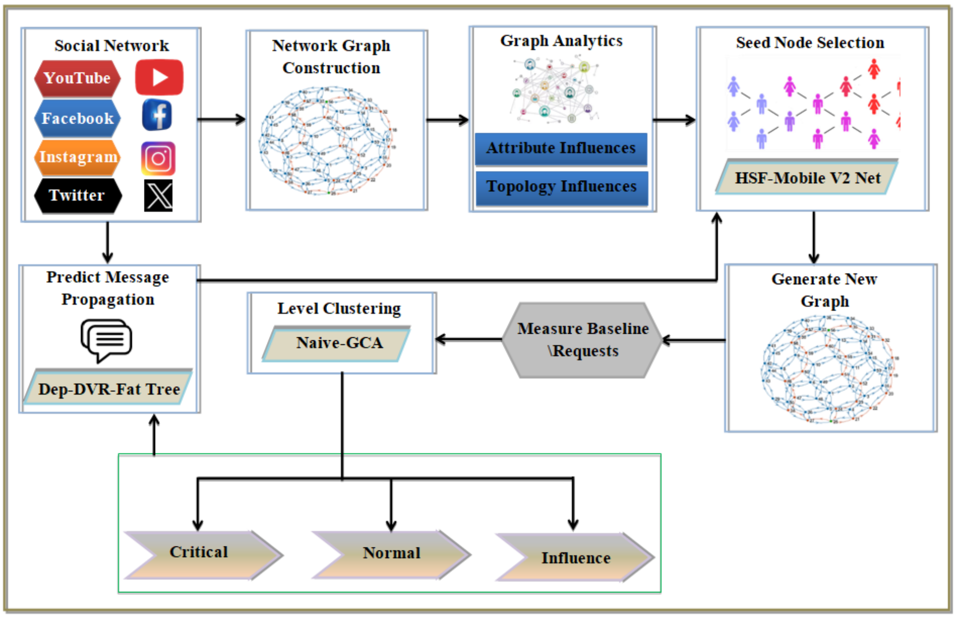

Figure 1: System architecture of the proposed framework

3.1. Social Network

A large-scale network $\Re$ comprising social connections from Facebook, Twitter, Instagram, and YouTube is considered.

3.2. Network Graph Construction

Here, a graph model is constructed for $\Re$ using SM. Sociogram quickly monitors the network activity. The network consists of a number of nodes $(\Im^n)$ (users) as shown in Eq. (1).

$$\Im^n = \left[\Im^1, \Im^2, \Im^3, ...\Im^N\right] \tag{1}$$

Here, the N^{th} number of nodes is depicted as $\Im^N$. The graph $\left(G^{GRAPH}\right)$ is drawn manually using the sociogram tool.

$$G^{GRAPH} = (\Im, \kappa) \tag{2}$$

Here in Eq. (2), the link between the two nodes is considered an edge point (κ). The graph is constructed by considering 5 parent nodes, 14 child nodes, and 36 edges. Constructed network graph using sociogram is shown in below figure 2.

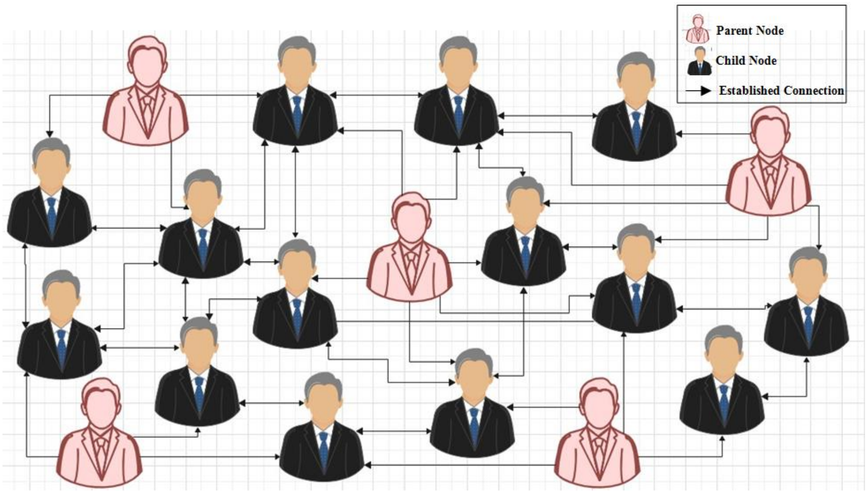

Figure 2: Network Graph

3.3. Graph Analytics

Here, the graph analytics is performed over the constructed sociogram-based network graph (G^{GRAPH}) using LT. The incoming influence is jointly decided by both incoming topology influence and incoming attributes influence. Here, the activation node $(\Im_i)$ depends on its neighbors $(\Im_j)$. If $\Im_j$ is active, it has an influence on $\Im_i$. The incoming topology influence is computed as shown in Eq. (3),

$$\Im_i^{TI}(\Im_i, \Im_j) = \frac{1}{\Im_i^{in-de}} \tag{3}$$

Here, $\Im_i^{in-de}$ denotes the in-degree of the node $\Im_i$, i and j are the node's number and number of views. The similarity of the attributes $\Im_{att}^{sim}(\Im_i, \Im_j)$ is calculated using Eq. (4),

$$\Im_{att}^{sim}(\Im_i, \Im_j) = \frac{a^{\Im_i} . a^{\Im_j}}{\left\| a^{\Im_i} \right\| \cdot \left\| a^{\Im_j} \right\|} \tag{4}$$

Here, $a^{\Im_i}$ and $a^{\Im_j}$ denotes the attributes of $\Im_i$ and $\Im_j$, respectively. Then, the normalization (A^{nor}) using edge-softmax is done using Eq. (5),

$$A^{nor} = \frac{\Im_{att}^{sim}(\Im_i, \Im_j)}{\sum_{\Im_l} \Im_{att}^{sim}(\Im_i, \Im_j)} \tag{5}$$

Here, $\Im_l$ denotes another in-neighbor of $\Im_i$. For identifying the influence score of $\Im_i$ (I^{score}), a linear combination of both influences is calculated using Eq. (6),

$$I^{score} = \beta^1 . \Im_i^{TI}(\Im_i, \Im_j) + \beta^2 . A^{nor} \tag{6}$$

Here, β^1 and β^2 denotes weight coefficients of $\Im_i^{TI}\left(\Im_i, \Im_j\right)$ and $\Im_{att}^{sim}\left(\Im_i, \Im_j\right)$. Likewise, the influence score for all the distributed nodes is identified. From that, the best influence score $\left(I^{best\ score}\right)$ and the best-influenced node's ID $\left(I^{best\ node}\right)^{ID}$ are carried out in further steps.

3.4. Message Propagation prediction

Following the identification of the influence score $\left(I^{best\ score}\right)$, the most popular events/messages' (M) propagation length over the network from the starting to end node is calculated using Dep-DVR-Fat Tree. The fat tree has a low diameter and fault tolerance, which help to connect any type of node. Therefore, this fat tree model has the ability to optimize the efficiency or higher propagation range of message passing in the large-scale network. The optimization is performed by adjusting the node's bandwidth of links in a tree. However, the fat tree takes a longer time to predict the message propagation due to its slow routing protocol. Hence, Depth First Distance Vector (DFDV) routing is used, which is easy to implement. However, still, Bellman Ford's search isn't precise, thus Depth First Search (DFS) is used.

Then, the flow of the event/message (M) is identified by the construction of the message fat flow tree, which comprises core switches $\left(F^{CORE}\right)$, aggregation switches $\left(F^{AGG}\right)$, and edge switches $\left(F^{EDGE}\right)$. The core switch (parent node) is the user who sends the M, aggregation switch is the child node and edge switches are the end node. The structure of the tree $\left(F^{TREE}\right)$ is mathematically expressed as shown in Eq. (7),

$$F^{TREE} = F^{CORE}\left(2 * F^{AGG}\left(4 * F^{EDGE}\right)\right) \tag{7}$$

For each message M_i, assuming M_j is the view of M. If there exists a tree which has j, make i a child of j. Otherwise, create a new tree with j as its root, and attach i to j. For identifying the distance of the message propagation without repetition of any nodes, the routing over the tree is performed using DFDV, which helps to search the nodes in the tree. The steps of DFDV are,

DFDV is an asynchronous algorithm in which a node $\Re_k$ sends a copy of its distance vector to all its neighbors. When the node $\Re_k$ receives the new distance vector $\left(d_{\Re_{NEI}}\right)$ from one of its neighboring vectors $\Re_{NEI}$, it saves the distance vector of $\Re_{NEI}$ and uses the DFS equation to update its own distance vector using Eq. (8).

$$D\left(\Re_k, \Re_{NEI}\right) = \min\left(c\left(\Re_k, \Re_{NEI}\right)\right) + d_{\Re_{NEI}} \tag{8}$$

Here, $c\left(\Re_k, \Re_{NEI}\right)$ represents the cost of $\left(\Re_k, \Re_{NEI}\right)$. The parameters of DFS are calculated below.

- Initially, create the stack of nodes. The stacks $\left(\Re^{F^{TREE}}\right)$ are expressed as shown in Eq. (9).

$$\Re^{F^{TREE}} = \left[\Re^1, \Re^2, \Re^3, ...\Re^{\Delta}\right] \tag{9}$$

Here, $\Re^M$ denotes the Δ^{th} node in the fat tree.

- Insert the root $\left(\Re^{ROOT}\right)$ through the unmarked or unvisited nodes $\left(U\left(\Re^{F^{TREE}}\right)\right)$ in the stack.

$$\Re^{ROOT} = \wp * \left(U\left(\Re^{F^{TREE}}\right) \oplus F^{TREE}\right) \tag{10}$$

In Eq. (10), $\wp$ denotes the routing function. The routing function avoids flooding the network with path search and updates messages through the unmarked nodes by using only parent–child links for message forwarding. This path selection helps to propagate the message in a wide range by consuming lower energy and bandwidth. Moreover, by this optimal path selection, the system allowed the model to efficiently find the propagation ratio of message through the network very efficiently.

- Calculate cost $c(\Re_k, \Re_{NEI})$ and distance $d_{\Re_{NEI}}$.

From the updated distance vector, the shortest routing path is evaluated manually. Then, views are assigned over the path. The messages with the large propagation distance indicate the maximum influenced nodes $(\Im^{par})$ (influenced node's ID $(\Im^{par})^{ID}$), and their influence score (I_{MESS}^{SCORE}) is computed using Eq. (11),

$$I_{MESS}^{SCORE} = \alpha^1.\Im^{par} + \alpha^2.(\Im^{par})^* \tag{11}$$

Here, α^1 and α^2 denotes the weight coefficients, and $(\Im^{par})^*$ represents the normalized $\Im^{par}$.

3.5. Seed Node Selection

Here, the $(\Im^{par})^{ID}$ and (I_{MESS}^{SCORE}) from section 3.4 and the $(I^{best\ score})$ and $(I^{best\ node})^{ID}$ from section 3.3 is input into the HSF-Mobile V2 Net to select the seed node based on the rule. The existing Mobile Net V2 model has a faster inference time. Yet, it learns slowly and causes pre-optimal convergence. Hence, Heaviside Step Function (HSF) is used instead of ReLu. The architecture of the proposed HSF-Mobile V2 Net is shown in Figure 3,

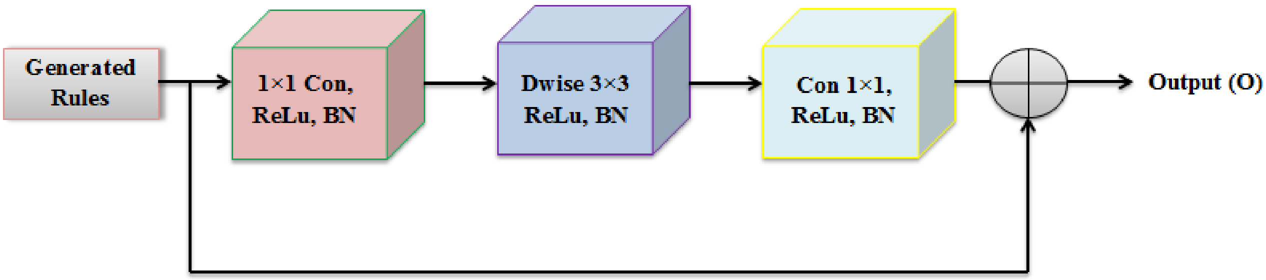

Figure 3: Architecture of proposed HSF-Mobile V2 Net

Initially, the rule (R) is generated to train the HSF-Mobile V2, which is expressed as shown in Eq. (12),

$$R = \begin{cases} \psi, & if\,(\Im^{par})^{ID} = (I^{best\ node})^{ID}\ \&\ I_{mess}^{score} = I^{best\ score} \\ \psi^{higher}, & if\,(\Im^{par})^{ID} \neq (I^{best\ node})^{ID}\ \&\ I_{mess}^{score} \neq I^{best\ score} \\ \psi^{bet}, & if\,(\Im^{par})^{ID} \neq (I^{best\ node})^{ID}\ \&\ I_{mess}^{score} = I^{best\ score} \end{cases} \tag{12}$$

If two same scores are from the same node ID, then it is classed as an influenced seed node (ψ). If scores from different node ID has different scores, then the higher score node (ψ^{higher}) is selected as an influenced seed node. If scores from different node ID has the same scores, then the between ness among the nodes (ψ^{bet}) are computed and selected as the influenced node. The generated rules are given into the first layer (1×1 convolution), which performs lightweight filtering by applying a single convolutional filter per input channel. The convolution operation (C) is given by using Eq. (13),

$$C(u,v)=\int_{-\infty}^{\infty} R(u,v)W(\varsigma,\varsigma)d(u,v) \tag{13}$$

Here, $R(u,v)$ and $W(\varsigma,\varsigma)$ denotes the rules with the size (u,v) and kernels with size (ς,ς). Following this, the HSF activation function (H) is performed to activate the appropriate neuron and is calculated as shown in Eq. (14),

$$H=\begin{cases}0, & if\ C>\eta \\ 1, & if\ C<\eta\end{cases} \tag{14}$$

Here, η denotes the neuron. Then, Batch normalization (B) is performed for normalizing C, and it is calculated using Eq. (15),

$$B=W\bullet\frac{\mu-\sigma}{\sigma}+C \tag{15}$$

Here, μ and σ denote the mean and standard deviation of C. The first block's output is concatenated with R and given as the input into the next residual block. The concatenation (Φ) is expressed as shown in Eq. (16),

$$\Phi=B\oplus\begin{cases}\psi, & if\left(\Im^{par}\right)^{ID}=\left(I^{best\,node}\right)^{ID}\ \&\ I_{mess}^{score}=I^{best\,score} \\ \psi^{higher}, & if\left(\Im^{par}\right)^{ID}\neq\left(I^{best\,node}\right)^{ID}\ \&\ I_{mess}^{score}\neq I^{best\,score} \\ \psi^{bet}, & if\left(\Im^{par}\right)^{ID}\neq\left(I^{best\,node}\right)^{ID}\ \&\ I_{mess}^{score}=I^{best\,score}\end{cases} \tag{16}$$

Then, the second and third layer is performed as same as the above steps by considering the output of the first layer as inputs. The first block's output is concatenated with R and given as the input into the next residual block. The procedure is continued till the last residual block. At the output layer (O), the classifier provides the best seed nodes (Γ_Z) by performing training till the last residual block. The pseudocode of the proposed model is,

Input: Generate Rules (R)

Output: Influenced Nodes (Γ_Z)

Begin

Initialize parameters u,v, H,

Compute weight value

For $\hbar=1 to \psi$ // $\hbar$ and ψ denotes residual block

For $\hbar=1$, **do**

Compute convolution operation C

Evaluate activation function

$$H=\begin{cases}0, & if\ C>\eta \\ 1, & if\ C<\eta\end{cases}$$

Compute BN (B)

Perform concatenation (Φ)

End for

Find Loss

If $Loss = 0$ **{**

Stop criteria

}**else**{

$iter = iter + 1$

}

End if

Return classified nodes

End begin

3.6. New graph generation (G) and Baseline/requests measuring (κ)

Here, using the obtained best seed nodes, a new graph is generated for Γ_Z using the sociogram technique. After the generation of G, the baseline strength of each node (δ) and requests (ϑ^{REQ}) are measured and it is calculated using Eq. (17),

$$\kappa \leftarrow (\delta, \vartheta^{REQ}) \tag{17}$$

The new sociogram graph is represented in Figure 4.

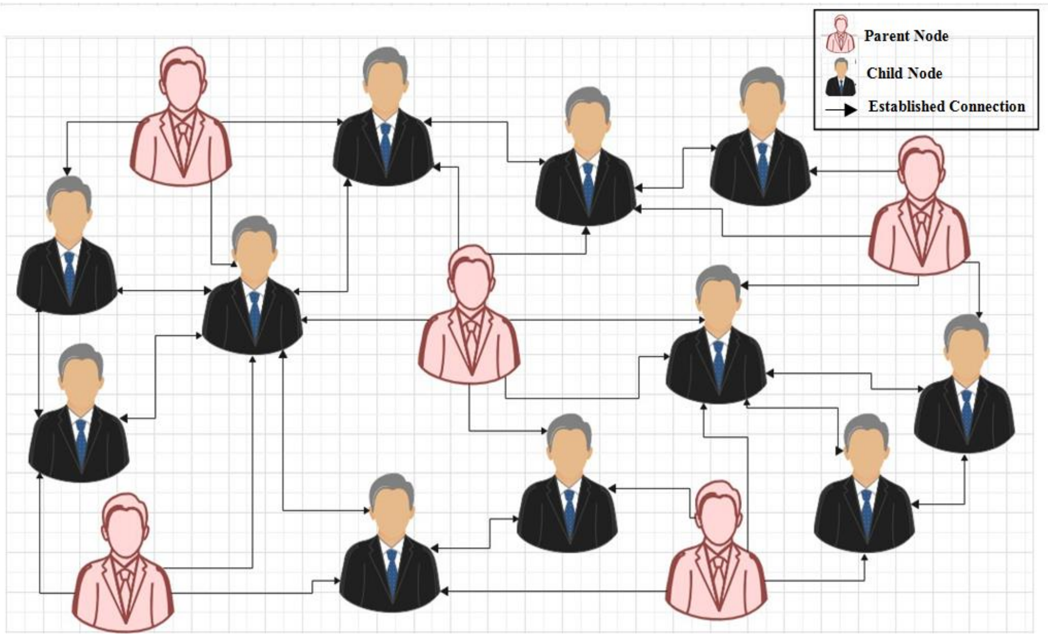

Figure 4: New graph using Sociogram

The new graph is the minimized version of the network graph, which is represented in section 3.3. This graph is constructed over the seed nodes to reduce the complexity of the accumulation of a large number of nodes. By constructing this new graph, the baseline request can be efficiently analyzed to predict the levels of influence. This graph is constructed with 5 parent nodes, 11 child nodes, and 28 edges.

3.7. Level Clustering

In this section, the influenced nodes are clustered using Naive-GCA over the child nodes (X) (established connection) from G. Based on message propagating distance, the nodes are clustered as: critical nodes, normal nodes, and maximum influenced nodes. Therefore, it is known as level clustering. From this level-based clustering, the highly influenced nodes are obtained, and these influenced nodes are further carried to the filter to find out the maximum influencer in the loop basis. The conventional Growing clustering algorithm is computationally efficient. But, it has the disadvantage of random center point selection, which may lead to wrong clustering. Hence, Naive Sharding (NS) is used for the seed selection. X is expressed as shown in Eq. (18),

$$X = \left[X^1, X^2, X^3, ...X^v\right] \tag{18}$$

Here, X^v denotes v^{th} child node. On the other hand, the message propagation distance (γ) over X is calculated using Eq. (19), which has been explained in section 3.4.

$$\gamma = b^1.X - b^2.(X)^* \tag{19}$$

Here, b^1 and b^2 denotes the weight coefficients, and $(X)^*$ represents the normalized X. Then, the seed points (centroids) (Ψ) calculated on X using naïve sharding are expressed as shown in Eq. (20),

$$\Psi = [\varpi / Q\kappa] \;\; where, Q\, represents\; the\; sharding\; over\; head \tag{20}$$

Here, ϖ denotes the randomly selected nodes. The regions start to grow from each seed by searching adjacent points based on the connection's distance. By this process, the lower distance nodes from Ψ are clustered as critical (N^{cri}), medium distance nodes as normal nodes (N^{nor}), and the high distance nodes are influential nodes $(N^{\inf l})$. The clustering result $(C(\gamma))$ is mathematically expressed as shown in Eq. (21),

$$C(\gamma) = \begin{cases} N^{cri}, \; if \; \Psi \Leftrightarrow X = \min imum \\ N^{nor}, \; if \; \Psi \Leftrightarrow X = \min imum < medium > \max imum \\ N^{\inf l}, \; if \; \Psi \Leftrightarrow X = \max imum \end{cases} \tag{21}$$

Here, $\Psi \Leftrightarrow X$ represents the distance between Ψ and X. Finally, the highly influenced cluster nodes are again given to the loop for predicting influence maximization.

4. RESULT AND DISCUSSION

4.1. Dataset description

The dataset used in the proposed framework is Amazon Products Sales Dataset 2023 and Social networks of trending social media. Amazon Products Sales Dataset 2023 is product data separated into 142 categories in CSV format [40]. Data preprocessing is done for data cleaning and conversion to structured .csv files. This standardized data was subsequently used for the experimental work. Social networks of Instagram [43], Twitter [44], Facebook [45] and YouTube [46], by gathering the information of node participation to influence the message.

4.2. Performance analysis of the overall proposed model

Here, the performance of the overall proposed model is analyzed and compared with the existing models.

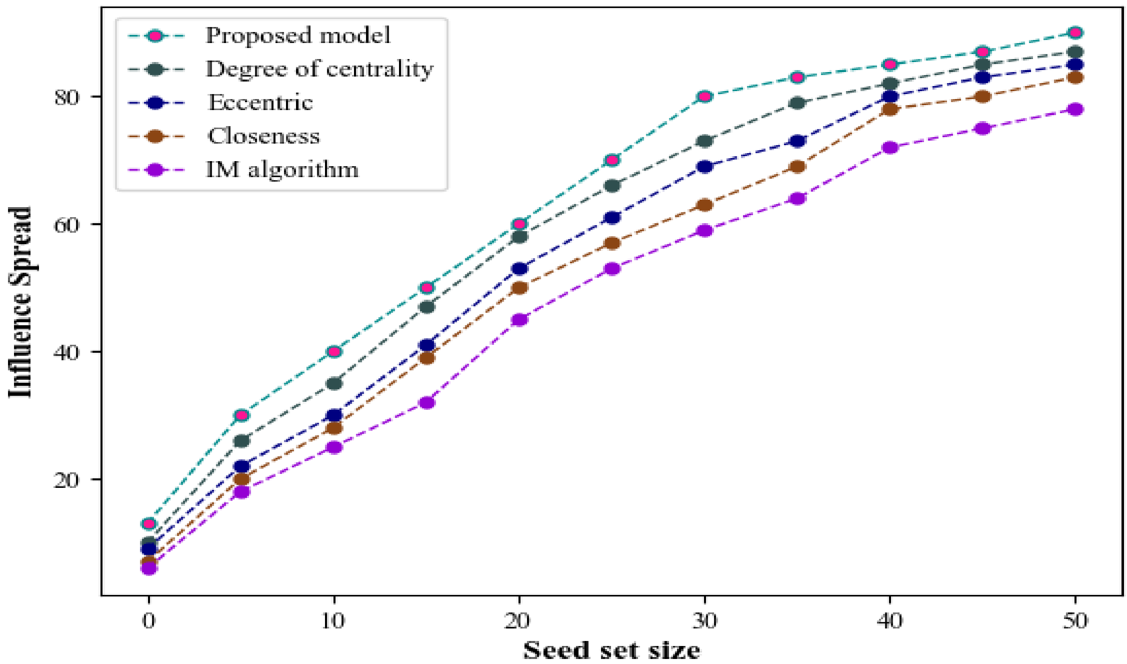

Figure 5: Performance analysis of proposed and existing models

Figure 5, shows the influence spread of algorithm at every iteration. Influence spread is measure to identify the influenced nodes at every iteration. Along with proposed model existing techniques, such as degree of centrality, eccentric, closeness, and IM algorithm is implemented to simulate the influence spread in every iteration at interval of 10. The result shows that at every iteration proposed model's influence spread is more as compared to existing models.

Table 1 represent the quantitative analysis of seed selection time and Figure 6 represents the Time analysis of influence over large-scale networks. In Figure 6, the seed selection time of the proposed model using four social networks is less as compared to existing models since considering the appropriate selection of neurons using HSF and also the generation of rules by comparing two influence scores. The seed selection time of the proposed model over the networks, such as YouTube, Twitter, Facebook, and Instagram are 3.214s, 3.214s, 3.21s, and 1.91s, respectively.

So proposed model proves the efficiency in seed selection with less computational complexity in large scale network which address the scalability issue discussed in the influence maximization.

Table 1: Time analysis of proposed HSF-Mobile V2 Net across social media platform

Social Media Platform	**Proposed HSF-Mobile V2 Net Model**	**Degree Centrality**	**Eccentric**	**Closeness**	**IM algorithm**
YouTube	3.214	6.098	9.562	12.459	13.89
Twitter	3.214	5.769	8.99	11.097	12.65
Facebook	3.21	5.345	7.423	10.673	11.35
Instagram	1.91	4.123	6.456	9.12	9.987

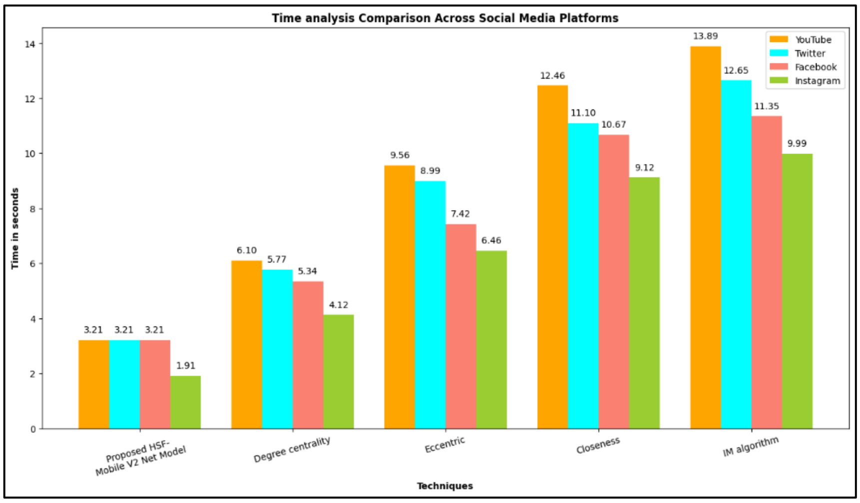

Figure 6: Time analysis of identifying the influence over the large-scale networks

4.3. Performance analysis of HSF-Mobile V2 Net

Table 2 provides the performance analysis and Figure 7 shows the performance of the proposed HSF-Mobile V2 and the existing methods like Mobile-V2 Net, Zeiler and Fergus Net (ZF Net), DenseNet, and Convolutional Neural Network (CNN). The proposed HSF-Mobile V2 Net is modified with the activation function. This reduces the Backpropagation error and prevents the neurons from dying, respectively. So, the proposed HSF-Mobile V2 achieves 98.77% of accuracy, 98.35% of precision, 97.33% of recall, 97.83% of F-measure, and the lower training time of the proposed model is 52136 ms.

Table 2: Performance analysis of proposed HSF-Mobile V2 Net with existing models

Social Media Platform	Proposed HSF-Mobile V2 Net Model	Mobile V2 Net	ZF-Net	Dense Net	CNN
Accuracy	98.77	94	92.11	90.09	87.94
Precision	98.35	95.9	93.73	89	87.02
Recall	97.33	95.2	92.9	88.6	87
F-measure	97.83	95.2	93	88	86

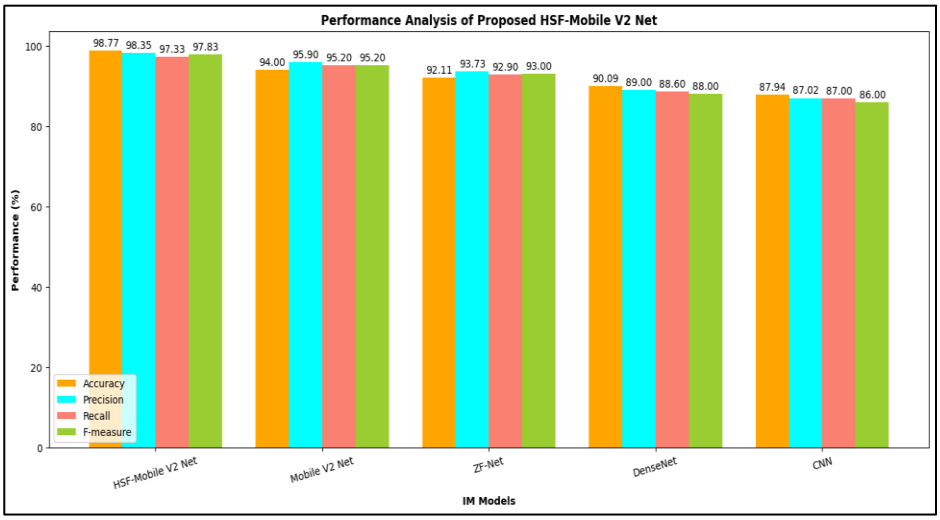

Figure 7 : Performance analysis of proposed HSF-Mobile V2 Net and existing models

Figure 8 shows the training time analysis. Proposed model is faster subject to computational time because it takes less computation compared to other algorithms due to Heaviside Step Function (HSF)-Mobile V2-based IM model.

The performance of the proposed HSF-Mobile V2 Net is analyzed and compared with the existing techniques.

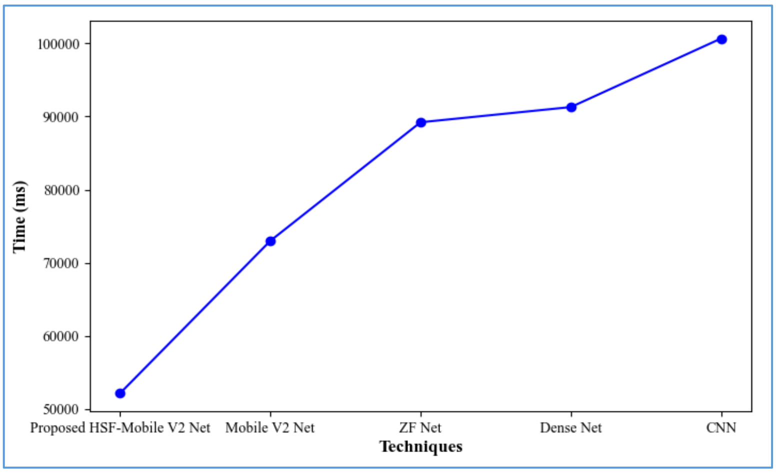

Figure 8: Training Time analysis

Table 3: Performance assessment of proposed HSF-Mobile V2 Net and existing models

Techniques	False Positive Rate (FPR)	False negative Rate(FNR)	Negative Predictive Value (NPV) (%)	Matthews Correlation Coefficient (MCC) (%)
HSF-Mobile V2 Net	11.03	4.92	97.97	97.91
Mobile V2 Net	19.1	11.45	95	94.34
ZF-Net	21.6	19.9	93.2	91
DenseNEt	23.9	24	90.1	89.22
CNN	30	29.06	87.6	86

Table 3 gives the performance assessment of proposed HSF-Mobile V2 Net and existing models. In HSF-Mobile V2 Net, the neurons remain active for a maximum number of iterations; thus, it significantly learns the inputs. Because of this, the proposed HSF-Mobile V2 Net withstands lower FPR (11.03) and FNR (4.92) rates, and it is shown in Table 1. Likewise, the NPV and MCC of the proposed model attain 97.97% and 97.91%, respectively. But, the prevailing models attain lower performance.

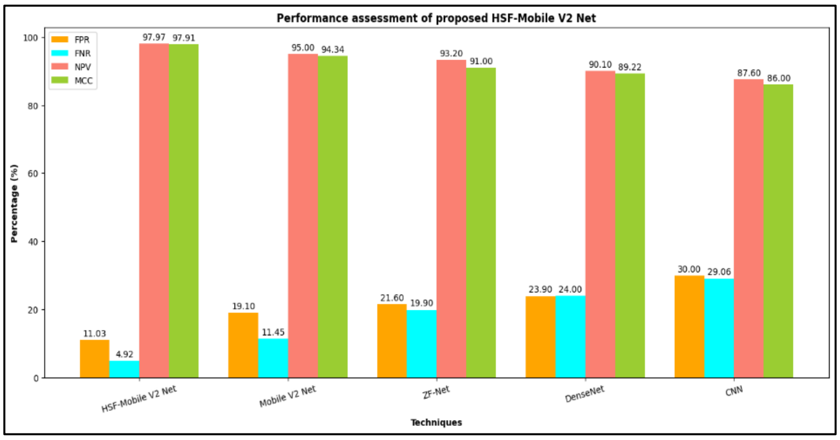

Figure 9: Performance assessment of proposed HSF-Mobile V2 Net

4.4. Performance of proposed Naive-GCA

Here, the Clustering Time (CT) and Clustering Accuracy (CA) of the proposed **Naive-GCA** and existing model are analyzed.

Table 4: CT and CA analysis

Techniques	Clustering Time (CT) (ms)	Clustering Accuracy (CA) (%)
Naive-GCA	10317	98.70
GCA	14583	94.50
K-means	17654	85.40
Fuzzy C-means	18906	80.26
D-means	21434	75.66

From Table 4, it is known that the proposed clustering remains to be more accurate and faster by using naive sharding. Thus, the iteration for clustering gets minimum with very accurate clustering results. The proposed method withstands better CT (10317), which is improved by 4,266 ms than GCA, 7,337 ms than K-means, 8,589 ms than Fuzzy C-means, and 11,117 ms than D-means. Likewise, the CA of the proposed model is 0.987, which is also increased than the existing models.

4.5. Comparative analysis with literature papers

Here, the comparative analysis of the proposed framework with the existing works is shown in Table 5 which includes Hurst exponent for IM [12], influence user discovery algorithm [13] and IM through Randomization [14] is performed based on efficiency.

IM Models	Efficiency (%)
Proposed HSF-Mobile V2 Net Model	97
Hurst exponent for IM [12]	95
Influence user discovery algorithm [13]	91
IM through Randomization [14]	87

Figure 9 compares the efficiency of the proposed and the existing research methodologies. The proposed method utilizes classifiers, such as HSF-Mobile V2 Net for seed node selection and Naive-GCA for influenced node identification. These algorithms improve efficiency by reducing training time, and appropriate selection of centroids.

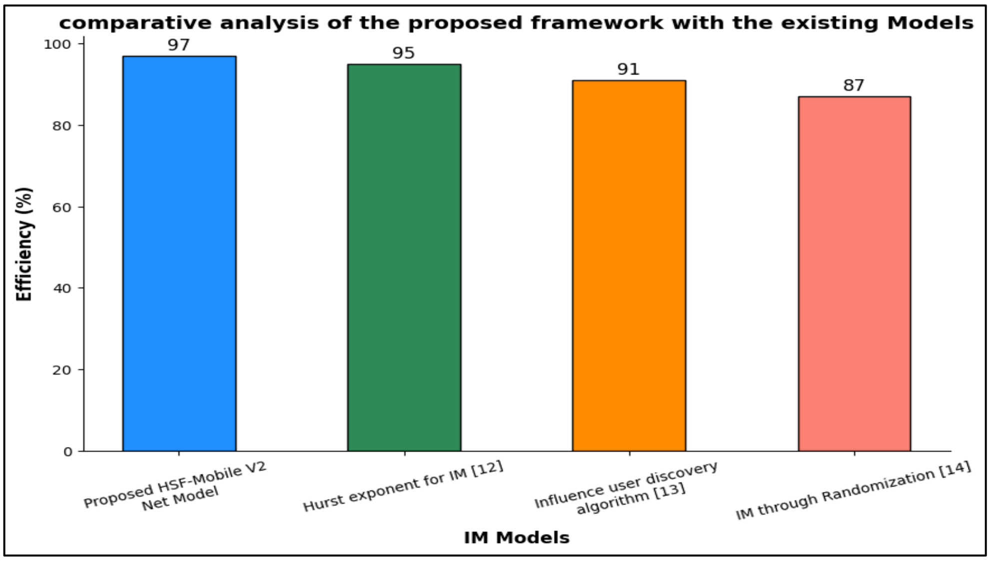

Figure 9: Comparative analysis of the proposed framework with the existing works

5. CONCLUSION

This work has proposed an efficient HSF-Mobile V2 Net-based IM with higher scalability. Initially, the graph structure was constructed for the distributed large-scale network. Then, the graph analytics was performed for evaluating the influence score by considering the attribute and topology influences. On the other hand, the influence score was calculated using DVR-Fat Tree for the node, which transmits the messages/events for a longer distance. Then, the seed node was selected using HSF-Mobile V2 Net with an accuracy of 98.77% and precision of 98.35%. For the obtained seed node, a new graph was constructed, and then baseline strength and requests were measured. Naive-GCA algorithm improve influenced node identification efficiency by reducing training time, and appropriate selection of centroids. This work focused on the scalability of the network for IM. Still, the elasticity of a social network needs to be considered for IM. In the future, an investigation of the hypothesis of the network that remains elastic to the level of interaction among network nodes needs to be done.

ACKNOWLEDGEMENT

I express deep gratitude to my guide Dr. Ashish Jadhav for his continuous guidance throughout the work and management of Ramrao Adik Institute of technology for providing research support.

REFERENCES

[1] Chen T, Guo J, Wu W. Adaptive multi-feature budgeted profit maximization in social networks. *Social Network Analysis and Mining*, *12*(1), 1–12, (2022). https://doi.org/ 10.1007/s13278-022-00989-3

[2] Fu B., Zhang J., Bai H., Yang, Y., & He, Y.. An IM Algorithm for Dynamic Social Networks Based on Effective Links. *Entropy*, *24*(7), 1–18, (2022). https://doi.org/10.3390/e24070904

[3] Li, W., Hu, Y., Jiang, C., Wu, S., Bai, Q., & Lai, E. ABEM: An adaptive agent-based evolutionary approach for IM in dynamic social networks. *Applied Soft Computing*, *136*, 1–14, (2023). https://doi.org/10.1016/j.asoc.2023.110062,

[4] Khatri, I., Choudhry, A., Rao, A., Tyagi, A., Vishwakarma, D. K., & Prasad, M. IM in Social Networks using Discretized Harris Hawks Optimization Algorithm and Neighbour Scout Strategy. *Arxiv*, (2022). 1–24. http://arxiv.org/abs/2211.09683

[5] Jinliang Y, Ruisheng Z Jianxin T, Rongjing H, Zepeng W, Huan Li. Efficient and effective influence maximization in large-scale social networks via two frameworks, Physica A: Statistical Mechanics and its Applications, Volume 526, 120966, (2019), https://doi.org/10.1016/j .physa.2019.04.202.

[6] V. Shelke and A. Jadhav, "Influence spread analysis a dimension to improve performance for epidemic diffusion models for scalable social networks," (2023) *14th International Conference on Computing Communication and Networking Technologies (ICCCNT),* Delhi, India, 2023, pp. 1-6, doi: 10.1109/ICCCNT56998.2023.10306466.

[7] Shelke Vishakha and Jadhav Ashish. 'Context Propagation Based Influence Maximization Model for Dynamic Link Prediction'. Intelligent-decision-technologies/idt230804, IOS Press, 1 Jan. 2024 : 1 – 17.doi: 10.3233/IDT-230804

[8] Ahmed MS, Sherine R, Tarek FG LKG: A fast scalable community-based approach for influence maximization problem in social networks, Physica A: Statistical Mechanics and its Applications, Volume 582, 126258, (2021). https://doi.org/10.1016/j.physa.2021.126258.

[9] Singh, S. S., Srivastva, D., Verma, M., & Singh, J. IM frameworks, performance, challenges and directions on social network: A theoretical study. *Journal of King Saud University - Computer and Information Sciences*, *34*(9), 7570–7603, (2021). https://doi.org/10.1016/j.jksuci.2021.08.009

[10]Ye, Y., Chen, Y., & Han, W. IM in social networks: Theories, methods and challenges. *Array*, *16*(October), 100264. (2022) https://doi.org/10.1016/j.array.2022.100264

[11] Peng Wu, Li Pan, Scalable influence blocking maximization in social networks under competitive independent cascade models, Computer Networks, Volume 123, Pages 38-50, 2017. https://doi.org/10.1016/j.comnet.2017.05.004.

[12] Saxena, B., & Saxena, V. Hurst exponent based approach for IM in social networks. *Journal of King Saud University - Computer and Information Sciences*, *34*(5), 2218–2230. (2022) https://doi.org/10.1016/j.jksuci.2019.12.010

[13] Li, W., Li, Y., Liu, W., & Wang, C. An IM method based on crowd emotion under an emotion-based attribute social network. *Information Processing and Management*, *59*(2), 1–18. (2022) https://doi.org/10.1016/j.ipm.2021.102818

[14] Lozano-Osorio, I., Sánchez-Oro, J., Duarte, A., & Cordón, Ó.. A quick GRASP-based method for IM in social networks. *Journal of Ambient Intelligence and Humanized Computing*, *14*(4), 3767–3779. (2023) https://doi.org/10.1007/s12652-021-03510-4
[15] Purba, K. R., Asirvatham, D., & Murugesan, R. K.. IM diffusion models based on engagement and activeness on instagram. *Journal of King Saud University - Computer and Information Sciences*, *34*(6), 2831–2839. (2022) https://doi.org/10.1016/j.jksuci.2020.09.012
[16] Ye, Y., Chen, Y., & Han, W.. IM in social networks: Theories, methods and challenges. *Array*, *16*(October), 100264. (2022) https://doi.org/10.1016/j.array.2022.100264
[17] Zuo, J., Liu, X., Joe-Wong, C., Lui, J. C. S., & Chen, W. Online Competitive IM. *Proceedings of Machine Learning Research*, *151*, 11472–11502, (2022). https://proceedings .mlr.press/v151/zuo22a/zuo22a.pdf
[18] Chatterjee, B., Bhattacharyya, T., Ghosh, K. K., Chatterjee, A., & Sarkar, R. A novel meta-heuristic approach for influence maximization in social networks. Expert Systems, 40(4). (2021) https://doi.org/10.1111/exsy.12676
[19] Yufeng W, Wenyong D, Xueshi D. A novel ITÖ Algorithm for influence maximization in the large-scale social networks, Future Generation Computer Systems, Volume 88, Pages 755-763, (2018) https://doi.org/10.1016/j.future.2018.04.026
[20] Mehdi H, Masoud A, Hesham F. SMG: Fast scalable greedy algorithm for influence maximization in social networks,Physica A: Statistical Mechanics and its Applications, Volume 420,Pages 124-133, (2015) https://doi.org/10.1016/j.physa.2014.10.088
[21] Yothimon Chandran, V. Madhu Viswanatham, Dynamic node influence tracking based influence maximization on dynamic social networks,Microprocessors and Microsystems, Volume 95, 2022,104689, (2022) https://doi.org/10.1016/j.micpro.2022.104689
[22] Zareie, A., Sakellariou, R. Influence maximization in social networks: a survey of behaviour-aware methods. *Soc. Netw. Anal. Min.* 13, 78 (2023). https://doi.org/10.1007/s13278-023-01078-9
[23] Gupta, S.K., Singh, D.P. Seed Community Identification Framework for Community Detection over Social Media. *Arab J Sci Eng* **48**, 1829–1843 (2023). https://doi.org/10.1007/s13369-022-07020-z
[24] Dizaji, S.H.S., Patil, K. & Avrachenkov, K. Influence Maximization in Dynamic Networks Using Reinforcement Learning. *SN COMPUT. SCI.* **5**, 169 (2024). https://doi.org/10.1007/s42979-023-02453-1
[25] Shuxin Yang, Quanming Du, Guixiang Zhu, Jie Cao, Lei Chen, Weiping Qin, Youquan Wang, Balanced influence maximization in social networks based on deep reinforcement learning, Neural Networks, Volume 169,2024, Pages 334-351, https://doi.org/10.1016/j.neunet.2023.10.030.
[26] Farzaneh Kazemzadeh, Ali Asghar Safaei, Mitra Mirzarezaee, Sanaz Afsharian, Houman Kosarirad,Determination of influential nodes based on the Communities' structure to maximize influence in social networks, Neurocomputing, Volume 534, 2023,Pages 18-28, https://doi.org/10.1016/j.neucom.2023.02.059.
[27] Yonggang Liu, Yikun Hu, Siyang Yu, Xu Zhou, Keqin Li, Conformity-aware adoption maximization in competitive social networks, Neurocomputing, Volume 573, 2024, 127224, https://doi.org/10.1016/j.neucom.2023.127224.
[28] Chang Guo, Weimin Li, Jingchao Wang, Xiao Yu, Xiao Liu, Alex Munyole Luvembe, Can Wang, Qun Jin, Heterogeneous network influence maximization algorithm based on multi-scale propagation strength and repulsive force of propagation field, Knowledge-Based Systems, Volume 291, 2024, 111580,https://doi.org/10.1016/j.knosys.2024.111580.
[29] Evren Güney, Markus Leitner, Mario Ruthmair, Markus Sinnl, Large-scale influence maximization via maximal covering location, European Journal of Operational Research, Volume 289, Issue 1, 2021, Pages 144-164, https://doi.org/10.1016/j.ejor.2020.06.028.
[30] Ganguly, M., Dey, P. & Roy, S. Influence maximization in community-structured social networks: a centrality-based approach. *J Supercomput* **80**, 19898–19941 (2024). https://doi.org/10.1007/s11227-024-06217-3
[31] Venunath, M., Sujatha, P., Koti, P. *et al.* Efficient community-based influence maximization in large-scale social networks. *Multimed Tools Appl* **83**, 44397–44424 (2024). https://doi.org/10.1007/s11042-023-17025-x
[32] Mani Saketh, C.V.S.S., Pranay, K., Susarla, A., Ravi Ram Karthik, D., Jaya Lakshmi, T., Nandini, Y.V. (2023). A Study on Influence Maximization in Complex Networks. In: Bhateja, V., Carroll, F., Tavares, J.M.R.S., Sengar, S.S., Peer, P. (eds) Intelligent Data Engineering and Analytics. FICTA 2023. Smart Innovation, Systems and Technologies, vol 371. Springer, Singapore. https://doi.org/10.1007/978-981-99-6706-3_10

[33] Banerjee, S., Jenamani, M. & Pratihar, D.K. A survey on influence maximization in a social network. *Knowl Inf Syst* **62**, 3417–3455 (2020). https://doi.org/10.1007/s10115-020-01461-4

[34] Sharan Srinivas, Chandrasekharan Rajendran, Community detection and influential node identification in complex networks using mathematical programming, Expert Systems with Applications, Volume 135, 2019, Pages 296-312, https://doi.org/10.1016/j.eswa.2019.05.059.

[35] Sara Ahajjam, Mohamed El Haddad, Hassan Badir, A new scalable leader-community detection approach for community detection in social networks, Social Networks, Volume 54, 2018, Pages 41-49, https://doi.org/10.1016/j.socnet.2017.11.004.

[36] Jinfang Sheng, Jie Hu, Zejun Sun, Bin Wang, Aman Ullah, Kai Wang, Junkai Zhang, Community detection based on human social behavior, Physica A: Statistical Mechanics and its Applications, Volume 531, 2019, 121765, https://doi.org/10.1016/j.physa.2019.121765.

[37] Krista Rizman Zalik, Borut Zalik, Memetic algorithm using node entropy and partition entropy for community detection in networks, Information Sciences, Volumes 445–446, 2018, Pages 38-49, https://doi.org/10.1016/j.ins.2018.02.063.

[38] V. Shelke and A. Jadhav, "Influence spread analysis a dimension to improve performance for epidemic diffusion models for scalable social networks," (2023) *14th International Conference on Computing Communication and Networking Technologies (ICCCNT),* Delhi, India, 2023, pp. 1-6, doi: 10.1109/ICCCNT56998.2023.10306466.

[39] Shelke Vishakha and Jadhav Ashish. 'Context Propagation Based Influence Maximization Model for Dynamic Link Prediction'. Intelligent-decision-technologies/idt230804, IOS Press, 1 Jan. 2024 : 1 – 17.doi: 10.3233/IDT-230804

[40] S. Bhatkar, P. Gosavi, V. Shelke and J. Kenny, "Link Prediction using GraphSAGE," *2023 International Conference on Advanced Computing Technologies and Applications (ICACTA)*, Mumbai, India, 2023, pp. 1-5, doi: 10.1109/ICACTA58201.2023.10393573.

[41] V. Shelke, J. Kenny, J. Gopani, S. Prasad and D. Sonani, "Node Degree Classification using Graph Modelling," *2023 International Conference on Advanced Computing Technologies and Applications (ICACTA)*, Mumbai, India, 2023, pp. 1-6, doi: 10.1109/ICACTA58201.2023.10393394.

[42] Dataset Link Available from: **Amazon-Products.csv**

Link: https://www.kaggle.com/datasets/lokeshparab/amazon-products-dataset

[43] Dataset Link Available from: : Instagram: https://www.kaggle.com/krpurba/im-instagram-70k-eg.

[44] Dataset Link Available from: Facebook: https://www.kaggle.com/code/boldy717/simple-network-analysis-of-facebook-data.

[45] Dataset Available from:

Twitter: https://www.kaggle.com/datasets/goyaladi/twitter-dataset?select=twitter_dataset.csv .

[46] Dataset Link Available from: Youtube: https://snap.stanford.edu/data/com-Youtube.html