An Efficient Algorithm for Influence Maximization in Signed Social Networks

Mozhgan Namiranian*, Mohammad Reza Meybodi
Department of Computer Engineering & Information Technology
Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran
na.mozhgan@aut.ac.ir, mmeybodi@aut.ac.ir

***Abstract*— Nowadays, much attention has been devoted to the issues of social networks and social influence. Social influence examines the user's behavioral changes under the influence of their neighbors. The issue of influence maximization is to find a subset of influential nodes that can maximize propagation in the network. The selection of people is very important and is the major aim of the studies. Hence, the current study aims to investigate the maximization of influence in signed social networks since in the psychology of society, negative opinions are superior to positive ones. The criteria considered for measuring influence and methods to increase it by identifying influential people are examined. The proposed solution of this paper is based on the label propagation algorithm. The algorithms used for maximizing influence in signed social networks namely a greedy algorithm and an innovative algorithm are outlined in the second section. To implement the algorithms and simulate the transfer of users' opinions in the graph network, the independent cascade propagation model is used. The proposed algorithm shows better performance and results compared to other algorithms and has less computational overhead since it finds primary nodes by detecting dense parts and not randomly. The significant novelty of the paper lies in the heart of the accuracy and authenticity of the proposed model in maximizing influence in signed social networks.**

Keywords*— *Social Network; Influence Maximization; Negative Opinions; Marginal Increase; Independent Cascade Model; Label Propagation

I. INTRODUCTION

Recent years have witnessed remarkable growth in the use of social networks which are regarded as a web-based platform of interaction through which people can easily communicate among themselves [13, 22]. The information is exchanged and the content generated in the meantime is modified using social networks [12]. The popularity of networks such as Facebook, Instagram, Telegram, and LinkedIn is obvious in the world. In general, a social network is defined as a network of interactions and communications where nodes are considered as users or actors and edges are known as the connections between them [15]. One of the significant applications of social networks is observed in viral marketing. A business strategy that employs a social network for promoting a product on different social media platforms is viral marketing [23]. Considering viral marketing or viral advertising, how the consumers share information regarding a product with other people is specified [27]. In social network, a certain number of vertices that are influential users are selected and the product is given to them [21]. Then. the products are distributed on the network. This method is introduced as a word-of-mouth technique whose role in pushing and directing people toward a new entity has been ignored in the literature [16].

Moreover, the problem of influence maximization in social networks confronts people with the challenge of how to choose a limited number of users in order to have the greatest impact on the entire network. Identifying influential users in social networks is an important issue that has received a lot of attention in recent years [38]. Influence spread indicates the number of people affected at the end of the diffusion process [40]. To solve the problem, the social network is modeled as a graph and the connections between people are indicated by the edges. Several heuristics such as node degree, betweenness and closeness have been proposed to maximize the impact on social networks [39]. FIP algorithm is proposed based on the overlapping communities for an influence maximization problem with the analysis of the node's emotional relationships [41]. Previous studies mainly failed to address the issue of influence maximization in signed networks. The importance of this topic is highlighted in people's relationships in real world that are not formed based on positive interactions only. Therefore, it is often important to explicitly consider negative relationships as well, especially when studying social media interactions. This is a very important parameter mainly neglected in previous studies [4, 25, 29].

It is noteworthy that discussion lists and disagreements have filled the sites. Social networking sites are a suitable place for conflict alongside friendship. Notably, negative comments have a remarkable psychological impact on people. Hence, the issue of influence maximization requires more consideration in the signed networks. As an innovation, present paper comprehensively examines the capability of several well-known algorithms in addressing the problem of influence maximization in signed networks. Evaluation criteria such as execution time, scalability, and accuracy are considered to specify the best one. The most important elements of social network analysis are introduced as they give a suitable tool for modeling. The clustering technique based on edge tokens has been rarely examined in related studies according to which the similarity of the edges can be determined. For this purpose, clustering in this research is conducted on the label propagation algorithm.

The rest of this paper is organized as follows: more related studies are reviewed in the second section to highlight the novelty of the present paper. The algorithms used for maximizing influence in signed social networks namely a greedy algorithm and an innovative algorithm are outlined in the third section. In this section, data selected for reaching the final results is also given. The obtained results are given in the form of a graph and table in the fourth section. In the fifth section, the obtained results are discussed and analyzed comprehensively. Finally, the conclusions and suggestions for future study are given in the sixth section.

II. RELATED WORK

Overall, modeling social networks as graphs have received much attention in related works [18, 37]. Also, in order to

979-8-3503-9969-1/23/$31.00 ©2023 IEEE

examine these networks, the functions with special characteristics are needed for proper modeling [5]. The concept of maximizing influence considers different criteria to find influential leaders that must be taken into consideration [31]. The influence maximization problem specifies a small set of nodes to maximize the amount of information dissemination in a social network, which has many applications in viral marketing. Accordingly, big companies use word of mouth to promote their products. Social networks are usually on a large scale and have a complex connection structure and are dynamic. Hence the proposed algorithm must be efficient and scalable. The challenge of maximizing influence in social networks states how to choose a limited number of users in such a way that they ultimately have the greatest impact on the entire network [6]. In order to solve the considered problem, the social network is modeled as a graph and the connections between people are indicated with edges [28]. The graph can be diverse in terms of direction, weight, sign, homogeneity, and static or dynamic. Some users are present on several social networks. The considered problem is modeled using multiple graphs based on which the optimization is conducted in various forms including subject-aware, distance-aware, location-based, and time-limited under the various spread models.

It is noteworthy that many studies in the broader literature have examined the issue of maximizing influence in signed social networks [10, 24, 32]. To mention a, a heuristic-greedy algorithm so-called the HEDVGreedy algorithm was presented by Aghaee and Kianian to solve the influence maximization problem in social networks [3]. The authors computed the expected diffusion value of the graph node based on the proposed algorithm and used the greedy technique for specifying the optimal nodes. The running time of implementation in this study was minimized using the Independent Cascade and Weighted Cascade models in eight real-world data sets. In 2020, Trivedi and Singh addressed the issue of influence maximization in social networks by employing an independent cascade model and obtained important results [30]. In another study, an innovative framework based on a meta-heuristic technique was presented for influence maximization in social networks [9]. The Shuffled Frog Leaping algorithm used in this research maximizes the two-hop spread of influence under the independent cascade model. Besides, the authors used local search strategies such as late acceptance-based hill climbing for enhancing the performance of the proposed solution. In 2022, Wu et al. utilized Boost Simulated Annealing (BSA) in social networks for budgeted influence maximization [35]. Additionally, three heuristic strategies were considered to improve the performance and speed up the presented algorithm. The empirical results confirmed the superiority of the proposed algorithm over the other state-of-art ones. In a major advance of 2022, an optimal heuristic algorithm was presented for influence maximization in social networks according to the redundancy weakening and two types of seeds into degree discount namely RWTDD [32]. The authors used the independent cascade model for comparing the proposed model with the six real social networks. The obtained results emphasize the better performance of RWTDD in terms of influence coverage and low time complexity. Additionally, Kumar et al. proposed another innovative model namely Communities-based Spreader Ranking (CSR) for solving the influence maximization in the social network according to the notions of communities and bridge nodes [19]. Li et al. introduced attributed influence maximization based on the crowd emotion for applying user's emotion and group features for examining the influence of multi-dimensional specifications on information spread and achieved many important results. In 2023 Bouyer and Ahmadi [42] proposed LMP which uses a special position among the important nodes for finding seed nodes. Nodes with gatekeeper and bridge property are important among important nodes and play a significant role in influence spreading. In the same year TSIFIM: A three-stage iterative framework for influence maximization in complex networks [43] was proposed which uses an adaptive search strategy for obtaining the optimal solution among the candidates.

Despite many state-of-art models presented so far, many gaps and shortcomings need to be tackled in the related studies. For example, the use of the independent cascade method can bring better results. Furthermore, the local tree structure increases the accuracy as it overcomes the limitations of the greedy algorithm in terms of speed and scalability. Concerning the studies reviewed above, the lack of a novel model for maximizing influence for the signed social networks is observed. In order to illuminate this uncharted area, the present paper proposes an efficient model for solving the problem of influence maximization in signed social networks.

III. METHODOLOGY

In this research, a greedy algorithm and an innovative algorithm with remarkable applications are used for maximizing influence in signed social networks. In the beginning, the structure of the greedy algorithm and innovative algorithm is explained. Then, the way of implementing the selected algorithms is specified.

A. Greedy Algorithm

The greedy algorithm is based on calculating the infiltration distributed by a set of nodes. Accordingly, Monte Carlo simulation is employed to estimate the infiltration rate, which causes the relatively slow performance of the algorithm [14]. The greedy algorithm adds each node to the set-in one-time step and mainly greedily selects the node with the largest marginal increase. However, the greedy algorithm is computationally expensive. The number of candidate nodes that must be tested to calculate the margin increase is large. To implement the algorithms and simulate the transfer of users' opinions in the graph network, the independent cascade propagation model is used in this research whose structure is highlighted in Fig.1.

As regarded in Fig.1, the main parameter used here is q, which includes the quality of the desired product. In other words, the quality of the product is also included in the opinions of users, which is simulated by the q parameter. The active nodes in the graph network can be positive with probability q and negative with probability 1-q. When a node becomes positive, it means that it is satisfied with the product. In the independent cascade propagation model, the node tries to make other nodes (users) positive. In contrast, nodes that are activated negatively attempt to convince other users. Since the opinion of each node is transferred to other neighboring nodes in this model, an independent coefficient namely the independent penetration coefficient, p, is considered. Notably, this coefficient is multiplied by the product quality coefficient or the same parameter q. Due to the obtained probability, if the node is positive, it will make the desired neighbor try to convince positive ones and the opposite is also true for the negative node. Besides, each node has only one chance to convince its neighbor, and otherwise, it has no more chances.

B. Heuristic Algorithm

The heuristic algorithm used here examines the area of influence in each node which is actually a Multi-channel Imager Algorithm (MIA) [7]. As shown in (1), μ (S, q) represents the positive propagation of seed group S in graph G with quality factor q under the MIA model. Also, pap (v, S, A, q) denotes the probability of node v becoming positive through the input branch A originating from S with quality factor q. In this algorithm, the aim is to find the influence areas considering that the nodes that cover a larger influence area, are selected for the initial dissemination of ideas. The MIIA parameter checks the area of influence, which is determined based on θ, and a number which means the limit of the distance from node v.(Eq.(1) and (2))

$$\mu(S,q) = \sum_{v \epsilon V} pap(v,s,MIIA(v,q,\theta),q) \quad (1)$$

$$MIIA(v,q,\theta) = \bigcup_{u \epsilon V, ppp(MIP(u,v)) \geq \theta} MIP(u,v) \quad (2)$$

In the beginning, the neighborhood area required for solving Equation is specified using (3).

$$\mathrm{MIP(u,v)} = \arg max_p \{ppp(P) | P \in \mathcal{P}(\mathrm{G,u,v})\} C \quad (3)$$

Considering $P = \langle u = p_1, p_2, \dots, p_m = v \rangle$, the probability of the positive propagation is obtained using (4).

$$ppp(P) = \prod_{i=1}^{m-1} p(p_i, p_{i+1}).q^m \quad (4)$$

Accordingly, the MIP parameter is the maximum penetration path between the two nodes, which is clearly the closest distance between the two nodes. Also, PPP denotes the probability that node U activates node V positively through the path P. MIP is a path like P initiating from u to v and has the highest PPP (P) value. The aim here is to find the set of seed nodes of size k that maximizes μ (S, q). After the initial cascading propagation at each node, the algorithm selects the path with the highest MIIA. In other words, the probability of positive activation of a node like v depends on two factors. The first one is the number of neighboring nodes of v that have been positively activated in the previous step. Second, the graph distance v from the seed node from which the propagation started. This parameter is the lowest number of edges from one of the s nodes to the v node.

In the following, the opinions are propagated for both algorithms concerning the nodes calculated as the initial seed. This algorithm is implemented in such a way that the initial seed nodes try to activate their neighbors (either positive or negative). Then, the ID names of the nodes are stored in the active nodes (saved nodes) and removed from the network list since it has affected their neighbors. Hence, the other one is useless since it is active itself. This action is taken for the next activated nodes and their ID is saved. The operation is repeated until no more nodes are activated.

C. Proposed Method

In this research, a method is presented for identifying influential people, considering the structure of society in social networks. For this purpose, society is identified in the beginning.

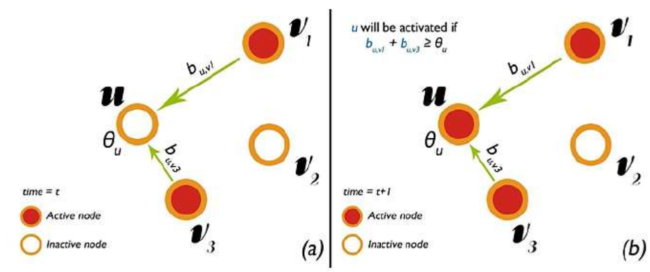


Fig. 1. The structure of the independent cascade propagation model [35]

In order to find the communities using clustering methods, communities are identified and then influential nodes among them are determined.

Recent studies investigated algorithmic aspects to maximize penetration in social networks for viral marketing based on two main infiltration cascade models, i.e., the independent cascade model and the linear threshold model. Such models are significant for a social network designed as a graph, starting from an initial set of vertices in the graph or a set of seeds. A random process determines the effect of a message from the seed on its neighbors until the process is over.

Fig.2(a) demonstrates the network can be segmented and balanced. Whereas, Fig.2(b) indicates that the network is partitionable but not balanced. The lines represent the positive edges and the dotted lines mean the negative edges. The network in Fig.2(a) can be segmented since it can be divided into three parts including (6, 7, 22, 23, 24, 25, 13, 14, 15, 16, 4, 5), (8, 9, 26, 27, 17, 18) and (20, 21, 10, 11, 12, 1, 2, 3, 19, 28. In this tripartite segmentation, the whole edge weights are positive and are sometimes negative between the groups. Also, the network is balanced since it can be divided into two parts including (6, 7, 22, 23, 24, 25, 13, 14, 15, 16, 4, 5) and (8, 9, 26, 27, 17, 18, 20, 21, 10, 11, 12, 1, 2, 3, 19, 28). Hence, the network can be segmented since it can be divided into three parts. However, it is not balanced because this network cannot be divided into two parts. Thus, two distinct parts are considered, in one of which the whole edges are positive and in the other, all the edges are negative.

As regards (5), the negative edges in the networks can be eliminated by specifying their clusters. The obtained subgraphs will contain only positive edges and subsequently, the subgraphs with separated nodes are achieved. In general, there are three different views regarding balance, as indicated in Table I.

A network is represented by the graph *G=(V,E)* where V is the set of nodes and *E={(u,v)|u,v∈V}* is the set of the edges between the nodes. The topology of the network is defined by (5)

$$\begin{cases} \omega_{ij} > 0 \quad < v_i, v_j > \in E \wedge (v_i \in P_k) \wedge (v_j \in P_k) \\ \omega_{ij} < 0 \quad < v_i, v_j > \in E \wedge (v_i \in P_k) \wedge (v_j \in P_l) \wedge (k \neq 1) \end{cases} \quad (5)$$

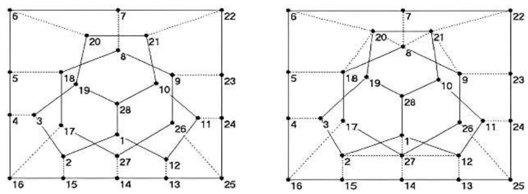
The adjacency matrix A. Accordingly, if two nodes h, v is directly connected, $A_{uv}=1$ and otherwise $A_{uv}=0$. The sublattice C⊂G in which the degree of the node u is divided into two parts is considered.

According to (6), $k_u^{in}(C) = \sum_{v \in C} A_{uv}$ is the number of edges that connect node u to the nodes related to the subnet C.

$$k_u = k_u^{in}(C) + k_u^{out}(C) \quad (6)$$

TABLE I. THE VARIOUS VIEWS AND FEATURES OF BALANCE

Features	***Views***
All cycles involving this node are positive	The balance of a node
The sign of multiplying three edges is positive	The balance of a triangle
Each triangle forming this network has a balance	The balance of a complete network
Either the complete network is balanced or people in each cluster have positive relationships within it and negative relationships between the two clusters.	The balance of an arbitrary network



(a) (b)

Fig. 2. Schematic representation of two signed networks

Also, $k_u^{out}(C) = \sum_{v \notin C} A_{uv}$ indicates the number of edges that connect node u to the rest of the network. A strong community is a subnet that satisfies the following condition (Eq.(7)):

$$\sum_{u \in C} k_u^{in}(C) > \sum_{u \in C} k_u^{out}(C), \forall u \in C \tag{7}$$

In a strong community, each node has more connections with its members than its connections with the rest of the network. Compared to the strong society, the weak society must be applied to (8):

$$\sum_{u \in C} k_u^{in}(C) > \sum_{u \in C} k_u^{out}(C) \tag{8}$$

As shown in (8), the sum of the degrees of the nodes in the community is greater than the sum of the degrees related to the rest of the network. It can be concluded that a strong society is also weak, but the opposite is not true. Although the use of clustering methods will not have the computational overhead of previous methods, the limitations of clustering can reduce the accuracy of communities. Therefore, clustering based on the tag spread algorithm in signed social networks is considered, which does not have the limitations of conventional clustering algorithms. Time complexity of this method is O(Tm) with T as time steps and m as graph edges.

The pseudo-code of the proposed method is presented in Algorithm 1.

D. Data Selection

The network assessment is made based on the data set presented in the study of Xie et al. [36] namely NetHEPT. Notably, the NetHEPT dataset is widely employed in many studies related to maximizing influence in social networks [1, 10, 32]. The characteristics of this data set are highlighted in Fig. 2.

Algorithm 1 .Proposed Method Algorithm.

1: Calculate similarity of two nodes I and j:

$$A_i^- = \{n_1^-, n_2^-, n_3^-, \dots, n_{m1}^-\}$$
$$A_i^+ = \{n_1^+, n_2^+, n_3^+, \dots, n_{m2}^+\}$$
$$A_j^- = \{n_1^-, n_2^-, n_3^-, \dots, n_{k1}^-\}$$
$$A_j^+ = \{n_1^+, n_2^+, n_3^+, \dots, n_{k2}^+\}$$
$$count\ Subscribe\ neighbor\ trust = A_i^+ \cap A_j^+$$
$$count\ Subscribe\ neighbor\ ditrust = A_i^- \cap A_j^-$$
$$sim(i,j) = count\ Subscribe\ neighbor\ trust + count\ Subscribe\ neighbor\ ditrust$$

for i=1:count_node
for j=1:count_node
if(i~=j)
sim_data(i,j)=sim_(i,j);

2: Data clustering based on LPA

Count new label based on current label and cluster number:

$$l'_u = arg_l max \sum_{v \in \sigma(u)} \delta(l, l_v)$$

[c1 c2]=find_two_max_cluster();
[m1 n1]=find(cluster==c1);
[m2 n2]=find(cluster==c2);
node=union(n1,n2);
for i=1:count_node
score(i,1)=i;
score(i,2)=score_node(i,node);
end when all the nodes are labeled.

3: Identify common nodes between two dense clusters and sort by calculated score and choose the influential one.

find_two_max_cluster
[s_sorted,index]= sort(s,'descend');
Display influential nodes by ('%d \n',score(i,1)

This information includes the number of network nodes and edges between nodes and other characteristics.

It is noteworthy that MATLAB 2017 b version is used for programming. The characteristics of the dataset used in the research is given in Table II. Accordingly, a computer with the specifications given in Table III is required to conduct the experiments.

E. Evaluation

To evaluate the proposed method, the algorithms given by [36] are used which include two methods:

- MIA-N Algorithm that uses several local tree structures. In this method, each tree is used to simulate local infiltration emissions.
- GREEDY Algorithm is employed to detect influential nodes in the network.

In this research, the quality factor parameter is considered to evaluate the network. This parameter specifies the influence of the selected nodes, which is a value between zero and one. Using this parameter, it is possible to determine the influence of selected nodes. The evaluation parameter that is used as the main parameter for calculation in this research is the Positive influence

TABLE II. THE CHARACTERISTICS OF THE DATASET USED IN THE RESEARCH [36]

Dataset	***NetHEPT***
Number of nodes	15K
Number of edges	31K
Average degree	4.12
Maximal degree	64
Number of connected components	1781
Largest component size	6794
Average component size	8.55

TABLE III. THE CHARACTERISTICS OF THE REQUIRED SYSTEM

Component name	***Characteristics***
Cpu	Intel ci7,12 core, 15 meg cach
Ram	16 Giga Byte DDR4

spread parameter based on [36]. This parameter indicates the number of nodes in the network that are influenced by influential nodes. It should be noted that the parameters are based on the value of the selected quality factor. The value of

the SEED parameter, which indicates the number of selected influential nodes is specified finally. The steps that must be taken for reaching the final results are as follows:

1.Calculating the similarity criteria between the nodes.

2. Clustering the dataset based on the label propagation algorithm (LPA) whose flowchart is given in Fig.3.

3. Detecting common nodes between two clusters with high density, ranking these nodes based on the calculated score value, and selecting the most influential node.

IV. RESULTS

The proposed method was applied to the considered dataset. This section gives the obtained results based on the various values of q from 0.1 to 1. The graphs of the whole three methods are drawn based on the seed value between zero and fifty and the comparison of the three methods is discussed. The obtained graphs are demonstrated in Fig.4 based on various values of q. The supplementary information regarding the results of various q values is also presented in Tables IV-XIII.

It should be noted that the best values obtained in each test were at a seed value of 50, as shown in Fig.5 and Table XIV. Based on the experiments, the average values of positive influence spread in each q are highlighted in Fig.6 and Table XV.

V. DISCUSSION AND FUTURE WORKS

Concerning Tables IV-XIII, the difference between the whole three methods is obvious. They moved exponentially and upward, but the proposed algorithm of this research had better results than other algorithms. As indicated in Table IV, the proposed algorithm has a high and better value, and the two graphs related to the other two methods have values close to each other. Nevertheless, they are far from the proposed algorithm. According to Table V, the algorithm was implemented based on q=0.3. The proposed algorithm has a high and better value. The distance between the two algorithms of [36] has become closer than the proposed algorithm. Due to the results presented in Table VII, the increase of the value of q led to increasing the

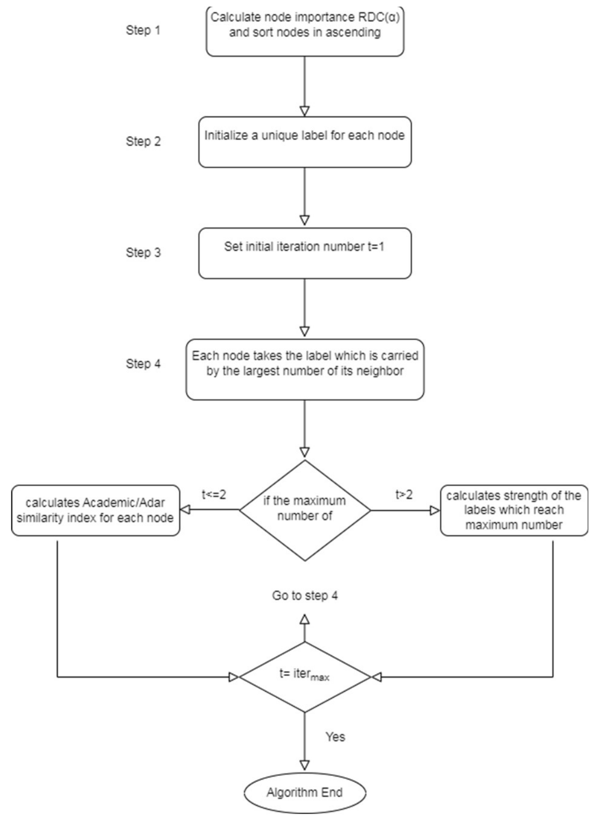


Fig. 3. The flowchart of LPA I detail [2]

spread of the positive influence in each execution. The values obtained based on q=0.4 are higher than q=0.3. The results obtained for q=0.5 were different since after the value of q=0.5, the other two algorithms are closer to the proposed algorithm, and the algorithms of the studies of Xie et al indicate much better results. In the following (Tables VII and VIII), the results did not change while the results of the proposed algorithm are better than the results of the other two algorithms. Notably, as the value of q approaches the final value, this superiority is still maintained. In the next two graphs (Tables IX and X), the experiments were carried out at their limit. In this test, the superiority of the algorithm has been maintained and the value of positive influence spread has reached a value higher than 500. As shown in Tables XI and XII, the results of the proposed algorithm have increased and had better values compared to the other two algorithms, which means that the proposed system has performed much better based on clustering. In the final test, the value of the positive influence spread parameter has reached its highest possible value. Thus, the best values obtained in each test were at a seed value of 50 according to Fig.5 and Table XIV.

According to the tests and obtained results, several points can be mentioned:

- The clustering algorithm based on density can find influential points and search for influential points, which is one of the prominent features of this algorithm.
- The proposed algorithm has less computational overhead than the other two algorithms since the MIA-N algorithm is performed using a tree structure in the graph, which brings a high overhead, and the second algorithm is Greedy.

- The proposed algorithm has a top-down view of the system while it does not choose primary nodes based on coincidence. However, search for influential nodes through the detection of dense parts in those points is necessary.

TABLE IV. THE OBTAINED RESULTS BASED ON Q = 0.1

Seed size set	*Label clustering*	*Greedy*	*MIA-N*
5	10.09	8.51	8.36
10	20	14.175	13.6
15	30	20.25	18.8
20	37.5	25.92	23.8
25	45	31.3875	29
30	50	37.4625	34
35	55	42.525	38
40	60	46.575	43
45	66.25	50.625	47
50	70	53.6625	50

TABLE V. THE OBTAINED RESULTS BASED ON Q = 0.2

Seed size set	*Label clustering*	*Greedy*	*MIA-N*
5	20.69	18.3	18.1
10	40	31.675	30.6
15	60	45.25	42.3
20	75	57.92	53.55
25	90	70.1375	65.25
30	100	83.7125	76.5
35	110	95.025	85.5
40	120	104.075	96.75
45	132.5	113.125	105.75
50	140	119.9125	112.5

TABLE VI. THE OBTAINED RESULTS BASED ON Q = 0.3

Seed size set	*Label clustering*	*Greedy*	*MIA-N*
5	30.53	28.6	28.3
10	60	49.175	47.6
15	90	70.25	65.8
20	112.5	89.92	83.3
25	135	108.8875	101.5
30	150	129.9625	119
35	165	147.525	133
40	180	161.575	150.5
45	198.75	175.625	164.5
50	210	186.1625	175

The results obtained in this study have not been presented in the literature so far. For example, Cai et al. proposed a new holistic influence diffusion model for considering the cyber and physical user relationships effectively [8]. The authors formulated a new issue of holistic influence maximization while

TABLE VII. THE OBTAINED RESULTS BASED ON Q = 0.4

Seed size set	*Label clustering*	*Greedy*	*MIA-N*
5	40.19	38.7	38.3
10	80	66.675	64.6
15	120	95.25	89.3
20	150	121.92	113.05
25	180	147.6375	137.75
30	200	176.2125	161.5
35	220	200.025	180.5
40	240	219.075	204.25
45	265	238.125	223.25
50	280	252.4125	237.5

TABLE VIII. THE OBTAINED RESULTS BASED ON Q = 0.5

Seed size set	*Label clustering*	*Greedy*	*MIA-N*
5	50.39	48.2	48
10	100	84.175	81.6
15	150	120.25	112.8
20	187.5	153.92	142.8
25	225	186.3875	174
30	250	222.4625	204
35	275	252.525	228
40	300	276.575	258
45	331.25	300.625	282
50	350	318.6625	300

TABLE IX. THE OBTAINED RESULTS BASED ON Q = 0.6

Seed size set	*Label clustering*	*Greedy*	*MIA-N*
5	60.3	58.8	48.6
10	120	101.675	81.6
15	180	145.25	112.8
20	225	185.92	142.8
25	270	225.1375	174
30	300	268.7125	204
35	330	305.025	228
40	360	334.075	258
45	397.5	363.125	282
50	420	384.9125	300

TABLE X. THE OBTAINED RESULTS BASED ON Q = 0.7

Seed size set	*Label clustering*	*Greedy*	*MIA-N*
5	70.6	68.32	68.01
10	140	119.175	115.6
15	210	170.25	159.8
20	262.5	217.92	202.3
25	315	263.8875	246.5
30	350	314.9625	289
35	385	357.525	323
40	420	391.575	365.5
45	463.75	425.625	399.5
50	490	451.1625	425

TABLE XI. THE OBTAINED RESULTS BASED ON Q = 0.8

Seed size set	*Label clustering*	*Greedy*	*MIA-N*
5	80.21	78.39	78.03
10	160	136.675	132.6
15	240	195.25	183.3
20	300	249.92	232.05
25	360	302.6375	282.75
30	400	361.2125	331.5
35	440	410.025	370.5
40	480	449.075	419.25
45	530	488.125	458.25
50	560	517.4125	487.5

TABLE XII. THE OBTAINED RESULTS BASED ON Q = 0.9

Seed size set	*Label clustering*	*Greedy*	*MIA-N*
5	90.21	88. 69	88.29
10	180	154.175	149.6
15	270	220.25	206.8
20	337.5	281.92	261.8
25	405	341.3875	319
30	450	407.4625	374
35	495	462.525	418
40	540	506.575	473
45	596.25	550.625	517
50	630	583.6625	550

TABLE XIII. THE OBTAINED RESULTS BASED ON Q = 1

Seed size set	*Label clustering*	*Greedy*	*MIA-N*
5	100.25	98.19	98.036
10	200	171.675	166.6
15	300	245.25	230.3
20	375	313.92	291.55
25	450	380.1375	355.25
30	500	453.7125	416.5
35	550	515.025	465.5
40	600	564.075	526.75
45	662.5	613.125	575.75
50	700	649.9125	612.5

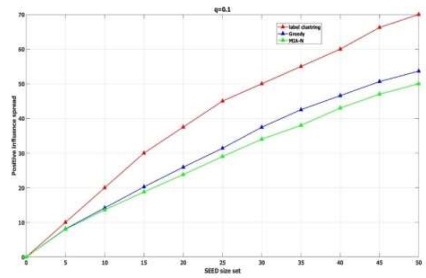
the spatial social network was not targeted at the signed social network. In the study, the greedy baseline algorithm was also developed for minimizing computing cost and time. In another research, Li et al. solved the Community-diversified Influence Maximization (CDIM) problem for social networks and obtained important results [20]. In a cutting-edge paper of 2022, soft computing techniques were used in social networks considering independent cascade models to obtain a dynamic seed set [17]. The authors modeled the graphs in a characteristic timestamp in which the edges and the nodes changed based on the various time intervals. The high scalability and accuracy of the proposed method to identify influential nodes in the snapshot graphs were also proved. Despite such interest and the benefits of the models, lack of an innovative model for solving the problem of maximizing influence in signed social networks in the related research is obvious. The issue of signed social networks considered in the presented model of this paper distinguishes it from the other state-of-art ones given in the literature.

As a matter of fact, the results obtained in the current study can be extended and enhanced using other heuristic algorithms or deep learning (DL) techniques. Looking forward, further attempts could prove quite beneficial to the literature. The following suggestions are proposed for future works:

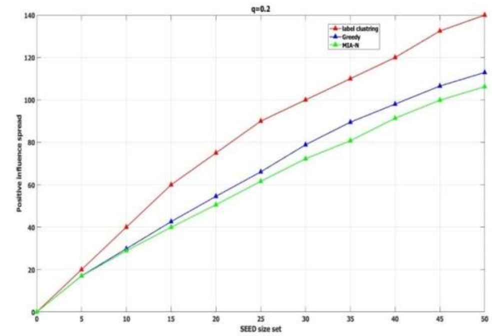
- Improving the proposed algorithm by combining it with other algorithms. For instance, the label propagation algorithm can be combined with meta-heuristic algorithms such as Genetic Algorithm (GA) or Particle Swarm Optimization (PSO).
- Using other clustering approaches such as clustering algorithms based on fuzzy logic.
- Identifying influential nodes based on the degree of the first and second neighborhood nodes. This method should be used along with the presented method to determine influential nodes.
- The use of routing algorithms to specify influential nodes. Using this technique, the nodes that are in high-traffic routes among other nodes can be recognized as influential nodes.

VI. CONCLUDING REMARKS

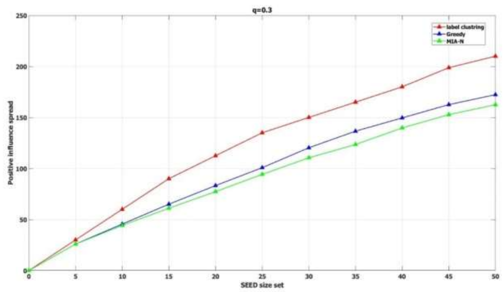
In summary, the current paper gives an efficient algorithm for maximizing influence in signed social networks. A novel solution was presented to determine influential nodes to maximize influence in social networks. For this purpose, the



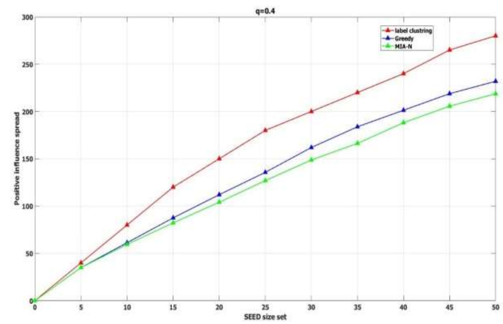
q = 0.1



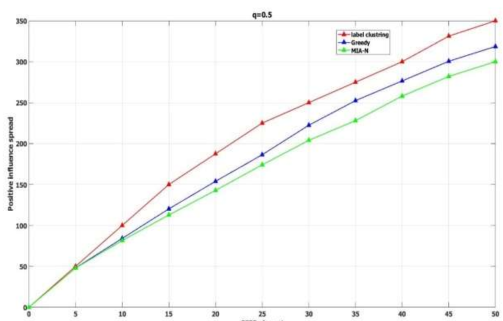
q = 0.2



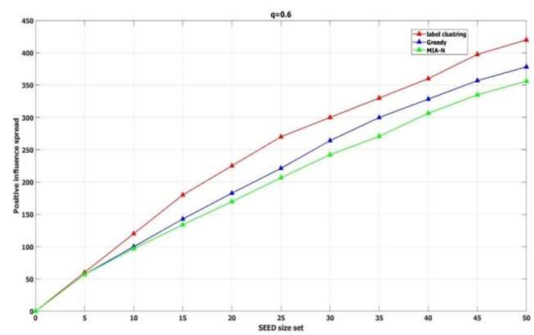
q = 0.3



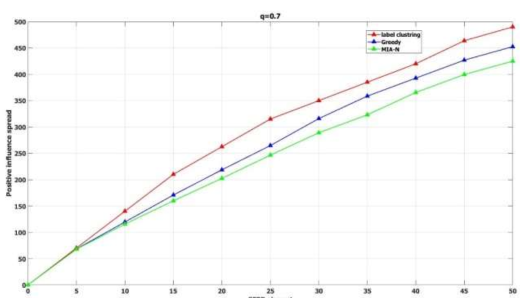
q = 0.4



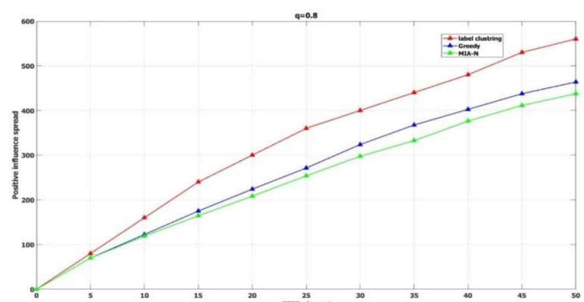
q = 0.5



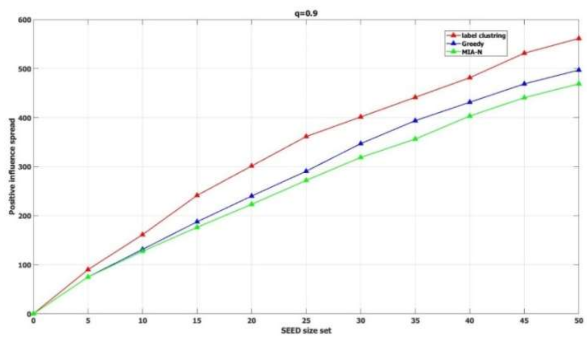
q = 0.6



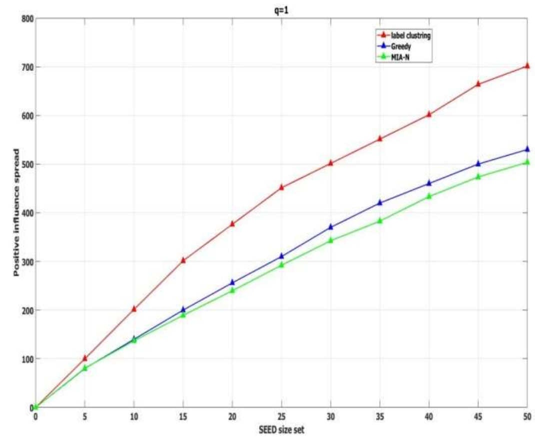
q = 0.7



q = 0.8



q = 0.9



q = 1

Fig. 4. The obtained results based on the different values of q

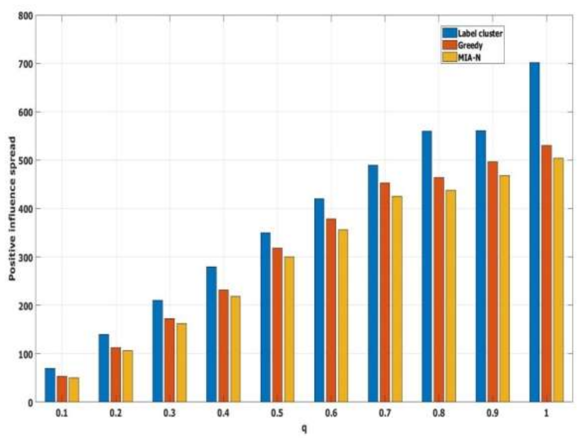


Fig. 5. The obtained results based on the best values of q

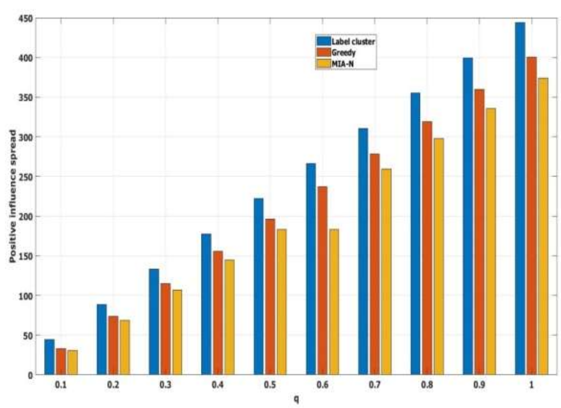


Fig. 6. Average value per q

similarity in the signed social network was calculated in the beginning. Then, the data were categorized and labeled using the label propagation algorithm, so that the data with high similarity was placed in a cluster. Then, the conventional nodes between two densely populated clusters were determined and influential nodes were calculated based on these nodes. After that, an

TABLE XIV. BEST VALUES IN SEED FIFTY BASED ON DIFFERENT Q

q	*Label cluster*	*Greedy*	*MIA-N*
0.1	70	53.6625	50
0.2	140	112.9563	106.25
0.3	210	172.3162	162.5
0.4	280	231.9412	218.75
0.5	350	318.5962	300
0.6	420	378.2875	356.25
0.7	490	452.4875	425
0.8	560	463.75	437.5
0.9	561.3	496.875	468.75
1	701.3	530	503.75

TABLE XV. AVERAGE VALUE PER Q BASED ON LABEL CLUSTER AND GREEDY

q	*Label cluster*	*Greedy*	*MIA-N*
0.1	44.384	33.1093	30.556
0.2	88.819	73.9133	68.68
0.3	133.178	114.7682	106.85
0.4	177.519	155.6032	145
0.5	221.914	196.3782	183.12
0.6	266.28	237.2633	183.12
0.7	310.685	278.0402	259.421
0.8	355.021	318.8723	297.573
0.9	399.3960	359.7273	335.7490
1	443.775	400.5023	373.8736

assessment of the clustering algorithm was applied in order to detect the most influential nodes. Based on the experimental results, the positive influence spread parameter and the proposed algorithm could obtain the best results. The different q values were used for the experiments based on which the various results were obtained. In the final test, the value of the positive influence spread parameter reached its highest possible value. Thus, the best values obtained in each test were obtained at a seed value of 50.

REFERENCES

[1] Abolghasemi, M., & Bagheri, E. A method to improve influence maximization in social networks based on community detection.

[2] Aghaalizadeh, S., Taghavi Afshord, S., Bouyer, A., & Anari, B. (2022). Improving the stability of the label propagation algorithm by propagating from low-significance nodes for community detection in social networks. Computing, 104. https://doi.org/10.1007/s00607-021-01027-4

[3] Aghaee, Z., & Kianian, S. (2020). The efficient influence spread estimation for influence maximization. Social Network Analysis and Mining, 10(1), 81. https://doi.org/10.1007/s13278-020-00694-z

[4] Bahutair, M., Al Aghbari, Z., & Kamel, I. (2022). NodeRank: Finding influential nodes in social networks based on interests. The Journal of Supercomputing, 78(2), 2098-2124.

[5] Balasubramaniam, K., Vidhya, S., Jayapandian, N., Ramya, K., Poongodi, M., Hamdi, M., & Tunze, G. B. (2022). Social network user profiling with multilayer semantic modeling using ego network. International Journal of Information Technology and Web Engineering (IJITWE), 17(1), 1-14.

[6] Banerjee, S., Jenamani, M., & Pratihar, D. K. (2020). A survey on influence maximization in a social network. Knowledge and Information Systems, 62(9), 3417-3455.

[7] Beheshti, Z., & Shamsuddin, S. M. H. (2013). A review of population-based meta-heuristic algorithms. Int. J. Adv. Soft Comput. Appl, 5(1), 1-35.

[8] Cai, T., Li, J., Mian, A., Li, R. H., Sellis, T., & Yu, J. X. (2022). Target-Aware Holistic Influence Maximization in Spatial Social Networks. IEEE Transactions on Knowledge and Data Engineering, 34(4), 1993-2007. https://doi.org/10.1109/TKDE.2020.3003047

[9] Chatterjee, B., Bhattacharyya, T., Ghosh, K. K., Chatterjee, A., & Sarkar, R. (2021). A novel metaheuristic approach for influence maximization in social networks. Expert Systems, e12676.

[10] Chen, W., Wang, C., & Wang, Y. (2010). Scalable influence maximization for prevalent viral marketing in large-scale social networks. Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining,

[11] Chen, Y., Qu, Q., Ying, Y., Li, H., & Shen, J. (2020). Semantics-aware influence maximization in social networks. Information Sciences, 513, 442-464.

[12] Choi, Y.-J., Jeon, B.-J., & Kim, H.-W. (2021). Identification of key cyberbullies: A text mining and social network analysis approach. Telematics and Informatics, 56, 101504.

[13] Dahesh, M. B., Tabarsa, G., Zandieh, M., & Hamidizadeh, M. (2020). Reviewing the intellectual structure and evolution of the innovation systems approach: A social network analysis. Technology in Society, 63, 101399.

[14] DeVore, R. A., & Temlyakov, V. N. (1996). Some remarks on greedy algorithms. Advances in computational Mathematics, 5(1), 173-187.

[15] Doostmohammadian, M., Rabiee, H. R., & Khan, U. A. (2020). Centrality-based epidemic control in complex social networks. Social Network Analysis and Mining, 10(1), 1-11.

[16] Hendrayati, H., & Pamungkas, P. (2020). Viral marketing and e-word of mouth communication in social media marketing. 3rd Global Conference On Business, Management, and Entrepreneurship (GCBME 2018),

[17] Jabari Lotf, J., Abdollahi Azgomi, M., & Ebrahimi Dishabi, M. R. (2022). An improved influence maximization method for social networks based on genetic algorithm. Physica A: Statistical Mechanics and its Applications, 586, 126480. https://doi.org/https://doi.org/10.1016/j.physa.2021.126480

[18] Kong, X., Shi, Y., Yu, S., Liu, J., & Xia, F. (2019). Academic social networks: Modeling, analysis, mining and applications. Journal of Network and Computer Applications, 132, 86-103.

[19] Kumar, S., Gupta, A., & Khatri, I. (2022). CSR: A community based spreaders ranking algorithm for influence maximization in social networks. World Wide Web, 25(6), 2303-2322. https://doi.org/10.1007/s11280-021-00996-y

[20] Li, J., Cai, T., Deng, K., Wang, X., Sellis, T., & Xia, F. (2020). Community-diversified influence maximization in social networks. Information Systems, 92, 101522. https://doi.org/https://doi.org/10.1016/j.is.2020.101522

[21] Motoki, K., Suzuki, S., Kawashima, R., & Sugiura, M. (2020). A combination of self-reported data and social-related neural measures forecasts viral marketing success on social media. Journal of Interactive Marketing, 52, 99-117.

[22] Muller, E., & Peres, R. (2019). The effect of social networks structure on innovation performance: A review and directions for research. International Journal of Research in Marketing, 36(1), 3-19.

[23] Muninger, M.-I., Hammedi, W., & Mahr, D. (2019). The value of social media for innovation: A capability perspective. Journal of Business Research, 95, 116-127.

[24] Panagopoulos, G., Malliaros, F. D., & Vazirgiannis, M. (2020). Multi-task learning for influence estimation and maximization. IEEE Transactions on Knowledge and Data Engineering, 34(9), 4398-4409.

[25] Raghavan, S., & Zhang, R. (2022). Influence maximization with latency requirements on social networks. INFORMS Journal on Computing, 34(2), 710-728.

[26] Rezvanian, A., Moradabadi, B., Ghavipour, M., Khomami, M. M. D., & Meybodi, M. R. (2019). Learning automata approach for social networks (Vol. 820). Springer.

[27] Sawaftah, D., Calıcıoglu, C., & Awadallah, R. (2020). The relationship between viral marketing and consumer purchase intention, the moderator role of brand image and age: Evidence from smartphone users in North Cyprus. Management Science Letters, 10(6), 1307-1320.

[28] Tang, J., Zhang, R., Wang, P., Zhao, Z., Fan, L., & Liu, X. (2020). A discrete shuffled frog-leaping algorithm to identify influential nodes for influence maximization in social networks. Knowledge-Based Systems, 187, 104833.

[29] Tong, J., Shi, L., Liu, L., Panneerselvam, J., & Han, Z. (2022). A novel influence maximization algorithm for a competitive environment based on social media data analytics. Big Data Mining and Analytics, 5(2), 130-139.

[30] Trivedi, N., & Singh, A. (2020). Efficient Influence Maximization in Social-Networks Under Independent Cascade Model. Procedia Computer Science, 173, 315-324. https://doi.org/https://doi.org/10.1016/j.procs.2020.06.037

[31] Urena, R., Kou, G., Dong, Y., Chiclana, F., & Herrera-Viedma, E. (2019). A review on trust propagation and opinion dynamics in social networks and group decision making frameworks. Information Sciences, 478, 461-475.

[32] Wang, F., She, J., Ohyama, Y., Jiang, W., Min, G., Wang, G., & Wu, M. (2021). Maximizing positive influence in competitive social networks: A trust-based solution. Information Sciences, 546, 559-572.

[33] Wang, J., Ma, X.-J., Xiang, B.-B., Bao, Z.-K., & Zhang, H.-F. (2022). Maximizing influence in social networks by distinguishing the roles of seeds. Physica A: Statistical Mechanics and its Applications, 604, 127881. https://doi.org/https://doi.org/10.1016/j.physa.2022.127881

[34] Wu, J., Gao, J., Zhu, H., & Zhang, Z. (2022). Budgeted Influence Maximization via Boost Simulated Annealing in Social Networks. arXiv preprint arXiv:2203.11594.

[35] Wu, P., & Pan, L. (2017). Scalable influence blocking maximization in social networks under competitive independent cascade models. Computer Networks, 123, 38-50.

[36] Xie, J., & Szymanski, B. K. (2011). Community detection using a neighborhood strength driven label propagation algorithm. 2011 IEEE Network Science Workshop.

[37] Zhang, B., Zhang, L., Mu, C., Zhao, Q., Song, Q., & Hong, X. (2019). A most influential node group discovery method for influence maximization in social networks: a trust-based perspective. Data & Knowledge Engineering, 121, 71-87.

[38] Arazkhani, N., Meybodi, M. R., & Rezvanian, A. (2019, April). An efficient algorithm for influence blocking maximization based on community detection. In 2019 5th International Conference on Web Research (ICWR) (pp. 258-263). IEEE.

[39] Mashayekhi, Y., Meybodi, M. R., & Rezvanian, A. (2018, April). Weighted estimation of information diffusion probabilities for independent cascade model. In 2018 4th International Conference on Web Research (ICWR) (pp. 63-69). IEEE.

[40] Aghaee, Z., Beni, H. A., Kianian, S., & Vahidipour, M. (2020, April). A heuristic algorithm focusing on the rich-club phenomenon for the

influence maximization problem in social networks. In 2020 6th International Conference on Web Research (ICWR) (pp. 119-125). IEEE.

[41] Bouyer, A., Beni, H. A., Arasteh, B., Aghaee, Z., & Ghanbarzadeh, R. (2023). FIP: A fast overlapping community-based Influence Maximization Algorithm using probability coefficient of global diffusion in social networks. Expert systems with applications, 213, 118869.

[42] Bouyer, A., & Beni, H. A. (2022). Influence maximization problem by leveraging the local traveling and node labeling method for discovering most influential nodes in social networks. Physica A: Statistical Mechanics and its Applications, 592, 126841.

[43] Dong, C., Xu, G., Yang, P., & Meng, L. (2023). TSIFIM: A three-stage iterative framework for influence maximization in complex networks. Expert Systems with Applications, 212, 118702.

.