International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

An Improved Gazelle Optimization Algorithm for Influence Maximization to Identify Influential Nodes in Social Networks

Srinu Dharavath [1], Natarajasivan Deivarajan[2], C. Nalini [3]

Submitted:14/03/2024 **Revised**: 29/04/2024 **Accepted**: 06/05/2024

Abstract: The goal of the Influence Maximization (IM) issue is to choose a component of the k-most influential nodes in a system so that the amount of influence spread by the seed set is maximized.When the transmission probability is high, greedy algorithms have a difficult time effectively approximating the predicted spread of influence of a particular node set and are not readily scalable to large-scale systems.Low solution accuracy or high memory costs are common issues with traditional heuristics based on constrained diffusion channelsor network topology. To address the IM issue more effectively, an Improved Gazelle-Based Optimization Algorithm for Influence Maximization (IGOA-IM) is proposed in this research.A unique local exploitation technique that combines random walk and deterministic procedures is proposed to enhance the suboptimal meme of everymemeplex to facilitate the global exploratory solution.The study findings on the spread of influence in twelve real-world networks demonstrate that IGOA-IM outperforms numerous state-of-the-art alternativesfor IMin choosing targeted influential seed nodes.

***Keywords**: Influence maximization, seed selection, improved gazelle-based optimization algorithm, real-world datasets, and accuracy*

1. Introduction

Influence Maximization (IM) is the process to recognize a group of people or network nodes that, if they are given a message or intervention, will have the biggest impact on the attitudes or behaviors of the general public. Finding the smallest group of people to target to have the greatest impact is the aim of IM. IM is a significant issue in a variety of fields, including marketing, public health, and social media [1, 2].Finding the most influential nodes in a network using network analysis and graph theory is a well-liked method for influence maximization. Utilizing algorithms, this method analyses the network's structure to pinpoint the nodes that are the most central or most interconnected.

By adding billions of loyal consumers, social networks have developed into potent platforms for the dissemination of knowledge and viral marketing. The social impact, which tracks the connections between people in the networks and can be assessed based on reputation and trust, is an underlying factor supporting the abilities [3-5].Viral advertisement, which recognizes the significant impact "word-of-mouth" lives in the connections and influence connection of customers and can change user’s behaviors and views is the typical application encouraged by social networks. Domingos and Richardson initially defined the issue in terms of the perspective of networks, which identifies the greatest number of possible customers to increase the anticipated profit of a product promotion operation. Online social networks have become a viable means of data transfer as mobile Internet connection has become more common [6, 7].Because of the relatively low average degree of user separation, rumors, advertisements, and news spread quickly on these networks. Communication networks, where people share files linked to various contents, containing video, pictures, and audio ,are another place where data is transmitted. A relatively heterogeneous framework, in which the majority of users are weakly connected but a small subset of them have many connections, is another characteristic of interactions and social networks [8-10].

Furthermore, in some social networks, a disassortative wiring pattern is defined by the propensity of high-level degree vertices to interconnect with low-level degree vertices. The data transmission is impacted by the networks' complicated structure, which creates a hierarchy among the nodes. In other words, nodes that increase the medium size of outbreaks are present in networks as special nodes that are the most effective spreaders during the transmission process [11].To comprehend and regulate the spreading process in social networks, it is crucial to identify these powerful nodes. The IM issue is particularly concerned with choosing a collection of η spreaders that, by a spreading dynamic, because the greatest cascade of new adopters. For the majority of spreading models, the difficulty of locating this collection of starting spreaders is NP-hard, which makes IMP challenging for network scientists [12, 13]. Heuristic algorithms are used to tackle the IMP because it is not possible to produce the ideal outcomes for the majority of networks.

[1]*Research Scholar, Department of Computer Science and Engineering, Annamalai University, Annamalainagar, Tamil Nadu, 608002.*

[2]*Department of Computer Science and Engineering, Faculty of Computer Science and Engineering, Annamalai University, Annamalai Nagar, Tamil Nadu, 608002.*

[3]*Department of Computer Science and Engineering, BIHER, Chennai-600073.*

The challenge of influence maximization presents two difficulties. Accurately estimating the spread of influence of a provided node-set is the initial issue and has been proven to be P-hard. Then offer powerful and effective approaches for choosing a small subset of prominent nodes that can increase the spread of influence throughout the system. Influence maximization was first stated as an optimization issue, and the greedy strategy with assured solution accuracy was suggested [14, 15].However, the outcomes of the experiments demonstrated how time-consuming the greedy method is, particularly in the biggest networks. It is due to the approaches needed to execute k-rounds to choose the desired seed nodes. We presented a unique technique called Improved Gazelle Based Optimization Algorithm (IGOA) to anticipate and resolve the influence problems to circumvent the issues that arise in influence maximization. The key contribution is,

- We provide a unique approach that makes use of the network's community spread and seeding phase forIM dissemination.
- The evaluation findings on four real-world datasets of varying sizes and applications show that the proposed approach performs better than many of the other IM algorithms.
- We carry out testing with real datasets. According to the testing findings, algorithm IGOA-IM performs much higher than state-of-the-art algorithms in terms of the time of efficiency and spread of influence.

The outline of this essay is as follows. In Section 2, the study problem is discussed, along with our findings and a review of some relevant literature. In Section 3, the IGOA-IM technique and associated algorithms are explained. Section 4 presents the tests on three actual datasets. Section 5 comes to an end.

2. Literature Survey

In this section, we refer to a few papers on influence maximization and discuss them in detail. For the IM challenge on interconnected networks, Keikhaet al. [16] suggested a deep learning-based technique called "DeepIM" by using network embedding. The diverse structural characteristics, cross-linkages, and bridge nodes of the given networks make it extremely difficult to maximize influence across interconnected networks. Additionally, because of an increase in issue size brought on by the expansion of network nodes, IM issues on linked networks are more complicated than IM challenges on traditional networks. To the best of our understanding, the suggested approach is the first approach to use network embedding to an issue of instant messaging.

It is suggested by Tang et al. [17] to use a discrete shuffled frog-leaping method to discover influential nodes for IM. A local degree-based replacement method is offered to work in conjunction with the local exploitation to enhance the suboptimal meme of every memeplex in the suggested framework, which is based on network topology for discrete encoding mechanisms and evolutionary conditions. In the meantime, the DSFLA's parameters-setting process is optimized using the orthogonal experimental design approach to ensure that the technique evolves successfully.

Bagheriet al. [18] suggested FAIMCS, a quick and precise algorithm, for IM in social networks built on community frameworks. The amount of nodes that must be looked at to find seeds is decreased by FAIMCS without sacrificing quality. Utilizing the CoDA method, FAIMCS first extracts communities from the input network. Then, based on the community structure, it determines the allotment of seed nodes for each community. Cohesive and overlapping communities' quotas are constrained, while 2-mode communities' limits are presumptively nil. As a result, it speeds up the selection of the seed nodes. In the end, seed nodes are chosen from candidate nodes utilizing the very accurate SimPath algorithm.

Dynamic Node Strength Decomposition (DNSD), based on dynamic network decomposition was a technique Li & Sun [19] developed to detect and rank node influence: both the influence of decomposition and the distinction of edges order on the node ranking are taken into consideration. They use the SIR framework to simulate the propagation process in 4 real networks and assess the Kendall's between spreading capacity and node ranking to assess the efficacy of the suggested technique. The outcomes of the experiments demonstrate that the approach outperforms other approaches for recognizing influential spreaders and has a good advantage in resolution ratio.

An approach to IM utilizing the concepts of graph neural networks and graph embedding was suggested by Kumar et al.[20]. In this research, the issue of influence maximization in intricate networks will be reduced to a pseudo-regression issue. To build an embedding for each node in the network using the struc2vec node embedding, they first use this method in the approach. The resulting embedding then serves as a feature for every node. Then, a GNN-based regress or receives the nodes and their characteristics. Calculating each node's effect under the SIR and IC information diffusion models yields the labels needed to train the GNN for the regression problem. Then, using parametric analysis on artificial test networks, they choose the best training network. The trained model can be employed to forecast how likely it is that nodes will influence the target network. (SGNN).

3. Preliminaries

3.1. IMIssues

Assume that $G=(V,E)$ is a network, with set E being the edge set and V being the node. The goal of the IM issue is to choose $k(1\leq k\leq|V|)$ targeted influential nodes as the seed set S in a way that ensures the highest influence spread σ(S) possible for the given transmission method.

$$S^{*}=\underset{S\subseteq V,|S|=k}{\arg\max}\,\sigma(S) \tag{1}$$

where S is a potential seed set, S* is the ideal seed set to increase the influence spread, and σ(S) is the influenced node's anticipated amount that S is projected to trigger. An optimization problem is the IM depicted in Eq. (1).

3.2. Influence estimator system

The second problem of influence maximization is to build efficient methods to reliably predict the anticipated influence of a particular node set. This difficulty is related to the topic of establishing efficient techniques to choose anaimed seed set that can increase the influence spread.

According to studies on the mechanics of influence spreading in social networks, impact declines as one's circle of neighbor's narrows. The LIE can be written as Eq. (2) based on the recommendation.

$$LIE(S)=\sigma_0(S)+\sigma_1^{*}(S)+\tilde{\sigma}_2(S) \tag{2}$$

where $\sigma_1^{*}(S)$ and $\tilde{\sigma}_2(S)$ are the estimated influence spreads of the set S's one-hop and two-hop areas, respectively, and $\sigma_0(S)$ is the size of the seed set. When the adjacency matrix of the nodes in S is used to express the local influencer estimator of a one-hop region and a two-hop area, the LIE can then be determined using Eq. (3).

$$LIE(S)=k+\sigma_1^{*}(S)+\frac{\sigma_1^{*}(S)}{\left|N_S^{(1)}\setminus S\right|}\sum_{u\in N_S^{(2)}\setminus S}p_u^{*}d_u^{*}=k+\left(1+\frac{1}{\left|N_S^{(1)}\setminus S\right|}\sum_{u\in N_S^{(2)}\setminus S}p_u^{*}d_u^{*}\right)\times\sum_{i\in N_S^{(1)}\setminus S}\left(1-\prod_{(i,j)\in E,j\in S}\left(1-p_{i,j}\right)\right) \tag{3}$$

where $N_S^{(2)}$ and $N_S^{(1)}$ stand for the candidate set S's two-hop and one-hop areas, respectively. A propagation cascade model's modest constant probability is called p_u^{*}. The number of edges that node u has within $N_S^{(2)}$ and $N_S^{(1)}$ is d_u^{*}.

As a result, choosing the k most influential nodes becomes an optimization issue to choose a seed set that maximizes the fitness value of Equation (2). We optimize the local influencer estimator method and investigate the influential nodes for IM in this paper to offer an efficient improved gazelle-based optimization technique.

3.3. Influence propagation model

We use the conventional IC model, based on the influence estimator, to evaluate the influence spread in specific networks. A probability system called the IC system imitates how data spreads through social networks. Every node in the IC model can be in one of just two states of inactive or active and can flip between the two, but not the other way around. The cascade model's propagation probability (p) explains how likely it is for inactive individuals to be impacted by their nearby active neighbors. An active node u has one chance to activate each of its neighboring inactive neighbors v at step t, with a success edge $(u,v)\in E$ and probability of p_{uv}. Whether the activation was successful or not, v won't be activated in the phases that follow again. If the v node is triggered by nodev, u will continue to be active and will have one opportunity to trigger every of its nearby inactive neighbors in step t + 1. If at step T, no node is activated, the diffusion process ends, and the spread of influence σ(S) made up of all the active nodes is returned.

4. Proposed Methodology

In this paper, we utilize Improved Gazelle Optimization Algorithm (IGOA-IM) for influence maximization. Optimization methods can be used to increase the LIE function's fitness value since, as was said before, the envisioned influence spread of a set of candidate nodes can be assessed by the local influence estimator. The usefulness of the Gazelle-Based Optimization Algorithm, a sophisticated meta-heuristic algorithm, on optimization problems, has been confirmed in numerous investigations. In this research, we aim to perform more studies on the method and propose an Improved Gazelle Based Optimization method specifically for the IMP.The basic concept of mimetic evolution is first introduced in the following subsections, after which evolutionary rules and discrete encoding mechanisms are developed for a gazelle based on network topology characteristics, and finally, the design of the Improved Gazelle-Based Optimization Algorithm (IGOA) for influence maximization is presented [21].The newly created GOA algorithm imitates how gazelles manage to survive. The optimal strategy involves grazing in the absence of a predator and running for cover when one is spotted. As a result, the described GOA algorithm process having of two parts.

4.1 Exploitation

At this point, it is believed that either a predator is not there or is just pursuing the gazelles as they peacefully graze. In this stage, the Brownian movement, which is attributed to controlled steps and uniform, is effectively utilized to cover neighborhood portions of the domain. Equation (4) shows how to calculate this occurrence mathematically.

$$gazele_{i+1} = gazele_i + s.M * M_B * (Elite - M_B * gazele_i) \quad (4)$$

Where s is the rate at which the gazelles graze, $gazele_{i+1}$ denotes the answer for the next iteration, $gazele_i$ denotes the answer for the present iteration, M_B denotes a constant random integers vector [0, 1], and the M denotes a vector of various random amounts denoting the Brownian movement.

4.2 Exploration

The exploration phase starts when a predator is spotted. Scaling the 2 m height to an amount between 0 and 1 mimics the 2 m height. When faced with danger, gazelles can turn their tail and stamp their four feet up to two meters into the air. The Lévy flight, which consists the periodic huge jumps and a series of little steps, is used in this algorithmic phase. This tactic has improved search functionality in the optimization literature. Both runs show a sharp turn in the guidance of travel, which is denoted by the mu. This research assumed that the gazelle shifts its direction on each iteration, moving in one way when the iteration number is odd and in the opposite direction when the iteration amount is even. We proposed that the gazelle utilizes the Lévy flight to migrate because it reacts first.

Equation (5) shows the mathematical formula for the gazelle's actions once it spots the predator.

$$\overrightarrow{gazele_{i+1}} = \overrightarrow{gazele_i} + S.\mu.\overrightarrow{M_L} * (\overrightarrow{Elite_i} - \overrightarrow{M_L} * \overrightarrow{gazele_i})$$

(5)

Where,

$$CF = \left(1 - \frac{iter}{\max - iter}\right)^{\left(2\frac{iter}{\max\_iter}\right)}$$

(6)

Even though Mongolian gazelles are not endangered, research on them also indicated that they had 0.66 an annual survival rate, which equates to just 0.34 cases where predators are effective. Predator success rates, or PSRs, affect a gazelle's ability to escape. The effect of PSRs is modeled in equation (8).

$$\overrightarrow{gazele_{i+1}} = \begin{cases} \overrightarrow{gazele_i} + \mathrm{CF}\left[\overrightarrow{\mathrm{LB}} + \vec{M} * (\overrightarrow{UB} - \overrightarrow{LB}) * \vec{U}\right] & \text{if } r \le \mathrm{PSRs} \\ \overrightarrow{\mathrm{gazele_i}} + [PSRS(1-r) + r](\overrightarrow{gazele_{r1}} - \overrightarrow{gazelee_{r2}}), & \text{else} \end{cases}$$

(7)

$$\vec{U} = \begin{cases} 0 & \text{if } r < 0.34 \\ 1 & \text{otherwise} \end{cases} \quad (8)$$

4.3 Improved Gazelle-Based Optimization Algorithm for Influence Maximization

The proposed Improved Gazelle Optimization Algorithm is to highlight its main functions and organizational framework. The proposed strategy uses the 3 main methods OL, RDR, and GOA, subject to 3stages that come after a transition mechanism. According to an assumption (IF *rand*< 0.2), the proposed IGOA changes the locations of the solutions. At the conclusion, the search process is examined to see whether it should be stopped or continued. If this is the case, the OL's search operations will be conducted; IF *rand*< 0.5, the RDR's search operations will be conducted. If this is not the case, the search processes will be waived in line with the Gazelle Optimization Algorithm's exploitation and exploration.

The proposed method solves the drawbacks of the standard approaches (i.e., OL, RDR, and GOA) by developing a new arrangement and employing3 integrated techniques. The fact that there may not be a wide range of potential answers is one of the GOA's major problems [22]. The recommended solution has a suitable arrangement among previous approaches to address these problems to address clustering obstacles more accurately.

Finally, we show how the problems were fixed. During an exploration examination of RDR and OL for half of the iteration and an exploitation search of the gazelle optimization algorithm for the other half, the imbalance between the search processes is first addressed. By choosing one exploitation or exploration process out of three techniques in every iteration, the recommended force configuration may equalize the search operations. and promote variation in the candidate solutions. Second, the speed of convergence would be controlled by modifications to the search strategy performed by the proposed transition mechanism. Due to this, the optimization process avoids the local search area and instead looks for the best solution. Then, applying numerous update strategies by the advised technique will maintain the variety of the used solutions.

4.4 The proposed IGOA's time complexity

The starting point of the candidate solutions, the aiming procedure of the previous solutions, and the modification of the candidate solutions are used to offer information about the proposed method’stotal time complexity.

Assume that N represents the total amount of solutions that have been used and that O(N) represents the complexity of initializing those solutions. The updating of the solutions has a temporal complexity $X(Y \times Z) + X(Y \times Z \times Dim)$, where a total amount of iterations is T employed. The issue's location size is Dim. As a result, the following is a description of the IGOA's time complexity.

$$X(IGOA) = (Z) \times X(GOA) + X(OL) + X(RDR)$$

(9)

Three major search operators such as GOA, OL, and RDR determine how time-consuming the proposed approach is. The following table lists the complexity times for these techniques.

$$X(OL) = X(Z \times (\max\_iter \times Dim + 1)) \tag{10}$$

$$X(GOA) = X(Z \times (\max\_iter \times Dim + 1)) \tag{11}$$

$$X(RDR) = X(Z \times Dim) \tag{12}$$

As a result, the IGOA's overall time-based complexity is provided as follows.

$$X(IGOA) = X(\max\_iter \times Z \times (Dim + 1) + 1(Z \times Dim) + (Z \times Dim)) \tag{13}$$

$$X(IGOA) = X(Max\_iter \times Z \times (Dim + Z)) \tag{14}$$

5. RESULTS AND DISCUSSION

The effectiveness and efficacy of the IGOA-IM algorithm have been evaluated on several databases, such as the NetHept network and interconnected networks, in this section. We contrast our findings with an amount of industry-standard IM methods, which are described below.

5.1 Dataset Description

The evaluation of performance is conducted on twelve real social networks to verify the performance of the proposed IGOA-IM on the IM challenge. The NetInfective network describes how visitors responded when interacting face-to-face at the INFECTIOUS exhibition. Peer-to-peer file-sharing network Gnutella30. The additional networks, which describe the relationships between authors across different research disciplines, are the co-author networks NetHEPT, NetPHY, and NetScience.Two undirected cooperation networks, AstroPh and CondMat, respectively. Technology-related news social network Slashdot is regarded as an undirected network and is well-known for its user base. Epinions is a generic consumer review website called Epinions.com's who-trust-whom online social network. The trust relationships between site users are illustrated by directed edges and can be determined by the members of the site. The Eu-Email network was created by a significant European research organization, and each node represents an email address, while each email received or sent represents aconnected edge. A Stanford is a sizable web graph that was taken directly from Stanford University, with nodes standing in for web pages and directed edges for relationships between them.

5.2 Algorithms for Comparing with seeding strategies

The details of the compared existing seeding approachesare given below:

- DeepIM – This algorithm was diverse structural characteristics, cross-linkages, and bridge nodes of the given networks making it extremely difficult to maximize influence across interconnected networks.
- DSFLA - A discrete shuffled frog-leaping method to discover influential nodes for IM.
- FAIMCS –This algorithm works on the number of nodes that must be looked at to find seeds is decreased by FAIMCS without sacrificing quality.
- DNSD - To detect and rank node influence: both the influence of decomposition and the distinction of edges order on the node ranking are taken into consideration.
- SGNN – This algorithm is utilized to solve the issue of IM in intricate networks and will be reduced to a pseudo-regression issue
- IGOA-IM- Our proposed algorithm. It uses gazelle optimization to correct the solution.

5.3 Parameter setting

The proposed approach has taken into account the parameters β, ρ, andα. The significance of the pheromone and the caliber of the node are related by the parameters α and β. We tried both and, with values ranging from 0 to 1, discovered that the fitness value rises as the values of both parameters rise. We set the values of the other parameters ρ = 0.2,and β = 0.5 to give the parameter value ofα.As a result, we run experiments on various datasets and discover that the fitness value rises as grows from 0.1 to 0.5, but that it becomes approximate at 0.5 and above. Similarly, we test the parameter value of β with the settings of ρ = 0.2, α = 0.5. IGOA-IM performs at its best when α = 0.5, and β = 0.5.The value 0.2 is given to the evaporation rate parameter rρ. This happens as a result of the pheromone value evaporating too quickly. IGOA-IM, therefore, checks every potential node. We set α = 0.5 and β = 0.5 to the parameter value ofρ.

5.4 Performance analysis

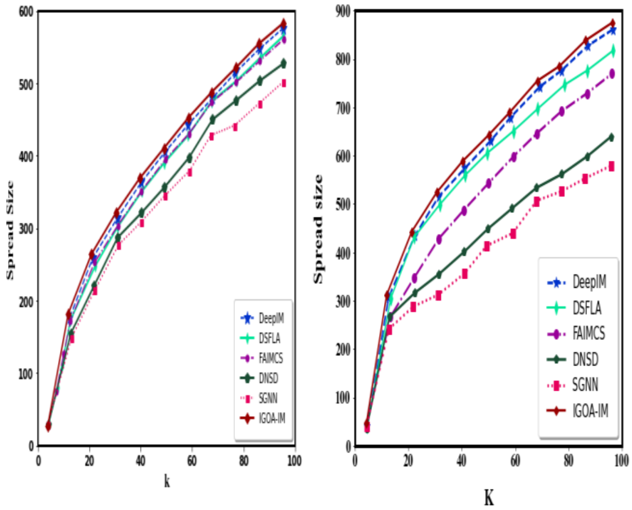
The evaluation of IGOA-IM and the other five baseline methods on the spread of influence under the IC model at transmission probabilityp = 0.01 on the 12 networks is demonstrated in Figure. 1. In the twelve large-scale networks, as demonstrated in Figure. 1 (a)–(l), IGOA-IM obtains a satisfactory spread of influence at the specified seed size. The following Table 1 gives the data values of the proposed scheme influence spread comparison with existing systems in several datasets for the IC model.

Table 1: Influence spread comparison of the proposed algorithm over existing algorithms in various datasets for the IC model

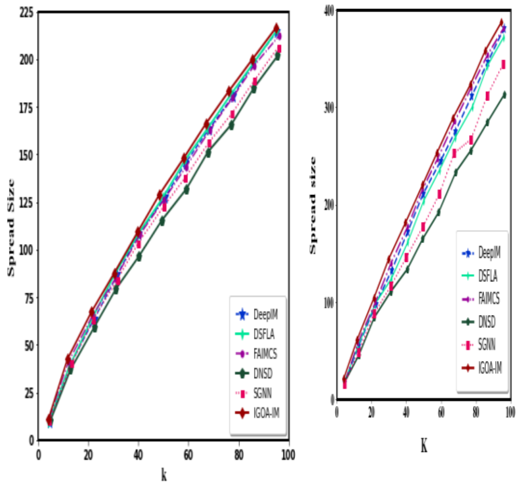
Datasets	Seed Size	Algorithms					
		SGNN	DNSD	DSFLA	FAIMCS	DeepIM	IGOA-IM (Proposed)
NetInfective [23]	20	56.78102	61.89567	79.78488	77.22029	82.33172	85.61822
	40	89.12242	99.34689	118.31816	113.94476	119.77972	124.8976
	60	102.48449	117.43929	141.89826	138.64401	148.46802	159.79068
	80	122.78132	132.25969	157.45186	148.32557	164.02163	177.166
	100	146.72318	150.37628	178.11207	171.90245	186.8653	194.16182
Slashdot [24]	20	426.6923	1118.7177	1141.5316	1165.8412	1156.7468	1199.9558
	40	552.88165	1317.1207	1400.8016	1426.4863	1416.0047	1449.2672
	60	667.63991	1511.7636	1591.6301	1622.7678	1614.4077	1641.7489
	80	752.00392	1679.7541	1717.759	1777.3768	1751.9979	1792.5701
	100	786.91983	1821.1406	1881.9835	1916.8464	1900.989	1939.5855
Eu-Email [25]	20	172.34313	176.92666	234.28786	191.84594	240.02398	243.45134
	40	276.94992	293.00927	344.63257	313.6593	353.81573	360.68297
	60	345.99635	371.2317	435.47446	405.65379	441.21774	449.23757
	80	426.51144	442.57257	505.66095	478.13831	518.28579	520.57486
	100	456.54687	473.75702	571.27644	555.21352	578.14905	581.58535
NetScience [26]	20	38.6416	39.7105	46.5029	43.2829	43.2855	48.9988
	40	48.7141	65.2811	74.7853	69.4276	74.426	83.3573
	60	77.3583	98.5707	106.9993	101.6442	112.7141	116.6427
	80	93.5018	111.9295	120.9994	115.9978	124.9264	129.9285
	100	109.6437	135.3572	137.1419	135.3535	141.072	147.4996
Gnutella30 [27]	20	78.6406	89.9606	145.673	150.9598	150.9771	155.4761
	40	108.3209	142.2189	185.1633	177.6235	200.9824	212.2679
	60	130.4649	174.9261	221.6129	208.0605	231.3953	251.741
	80	145.0794	203.8497	247.5439	236.9875	261.8495	285.9375
	100	161.2211	247.0926	291.5367	280.258	294.5636	327.6877
NetPHY [28]	20	336.028	377.9781	384.0749	405.05	411.0255	434.9277
	40	515.9882	650.9352	824.8568	863.9769	846.026	881.8792
	60	711.045	833.9682	944.9403	971.9637	944.9888	1058.9851
	80	773.9118	975.051	1158.1907	1155.1422	1151.9482	1254.0419
	100	903.1164	1089.1832	1320.1515	1302.1035	1377.0283	1464.0012
NetHEPT [29]	20	155.2405	194.4283	217.9264	235.1086	250.7528	255.5398
	40	188.5037	312.1555	368.4891	390.3855	404.4916	426.4153
	60	207.3695	418.8617	487.7725	501.8787	506.5291	537.8994
	80	248.5777	514.6925	563.2177	566.3665	594.5151	619.5786
	100	253.5649	615.1829	655.8905	673.1364	693.4402	724.8013

LiveJournal [30]	20	264.7111	282.4531	315.974	296.2721	315.9896	341.5992
	40	381.4604	407.1198	478.1188	452.4905	491.9223	511.6273
	60	496.2559	531.7617	598.816	573.1783	608.6809	634.3155
	80	609.0698	640.6277	762.9021	757.0098	778.6966	808.2945
	100	688.3472	743.5672	891.4949	863.8817	907.2738	964.4818
Stanford [31]	20	83.76582	87.65324	97.78496	97.79606	99.34511	103.99595
	40	133.23426	145.67475	161.25774	172.15288	177.60045	181.46198
	60	192.02289	210.70768	235.61456	244.96806	247.26203	253.48043
	80	255.48827	267.16534	300.64749	311.53154	320.86653	323.1864
	100	313.52828	344.63876	371.88032	381.22641	381.21901	387.41521
CondMat [32]	20	65.67797	63.55932	59.32203	65.25424	63.55932	67.37288
	40	108.05085	108.47458	96.18644	108.05085	103.38983	109.32203
	60	147.0339	145.76271	131.77966	143.64407	137.71186	147.88136
	80	181.77966	180.50847	165.67797	179.66102	171.61017	183.05085
	100	215.25424	214.40678	201.69492	212.28814	206.35593	216.52542
Epinons [33]	20	287.46144	315.65013	431.89108	347.35018	433.65515	442.91339
	40	356.4892	400.51533	559.03512	486.81471	573.12248	588.18898
	60	439.58475	490.65596	650.93981	598.10174	679.10054	689.17323
	80	526.20843	561.4338	746.38381	691.7649	776.30581	784.84252
	100	579.37358	639.25393	817.15327	769.59342	861.18499	874.88189
AstroPh [34]	20	214.04175	219.73435	112.71347	126.37571	314.2315	264.27221
	40	308.5389	286.90702	248.19734	252.75142	405.31309	368.62004
	60	379.12713	356.35674	350.66414	350.66414	479.31689	452.55198
	80	441.74573	449.71537	428.08349	430.36053	514.61101	521.73913
	100	502.08729	527.13472	564.70588	561.29032	576.09108	582.98677

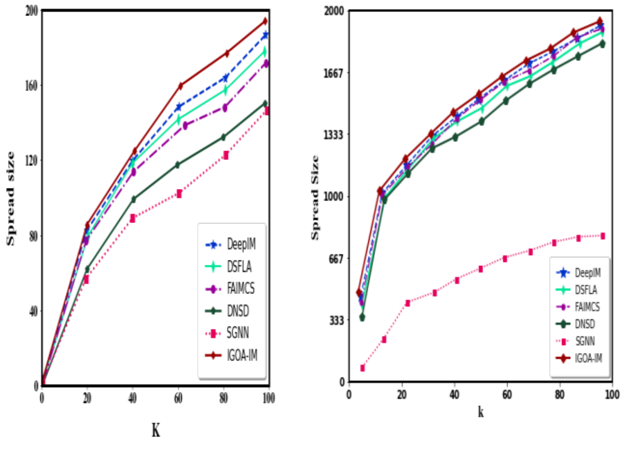
In comparison to the other five state-of-the-art methods, IGOA-IM obtains an even greater influence spread than DNSD on the CondMat, as demonstrated in Figure. 1(c). As can be seen in Fig. 1(b), IGOA-IM beatsSGNN, DNSD, and DSFLA in all cases except the Epinions network.In other words, because of its memetic evolutionary rules, the proposed IGOA-IM is successful in recognizing prominent nodes. In most cases, except the one in Figure. 1(h), the SGNN spreads its influence less than DNSD and IGOA-IM.



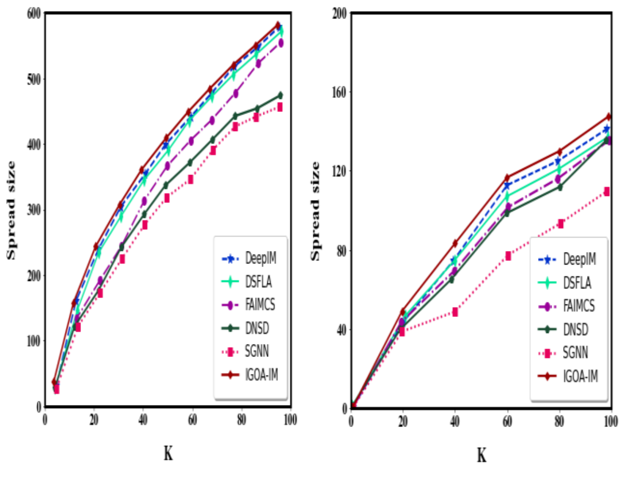
(a) AstroPh **(b) Epinons**



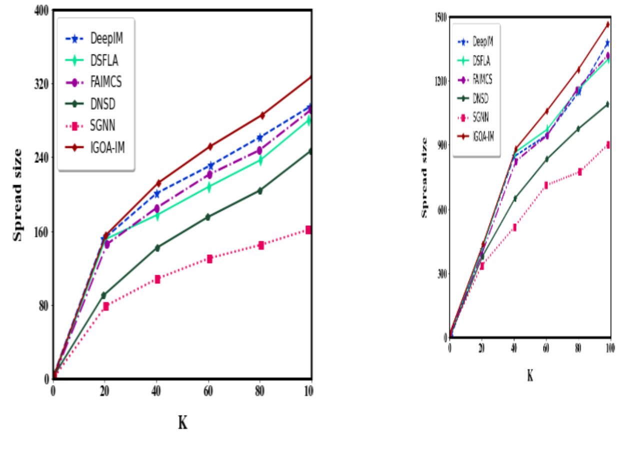
(c) CondMat (d) Stanford



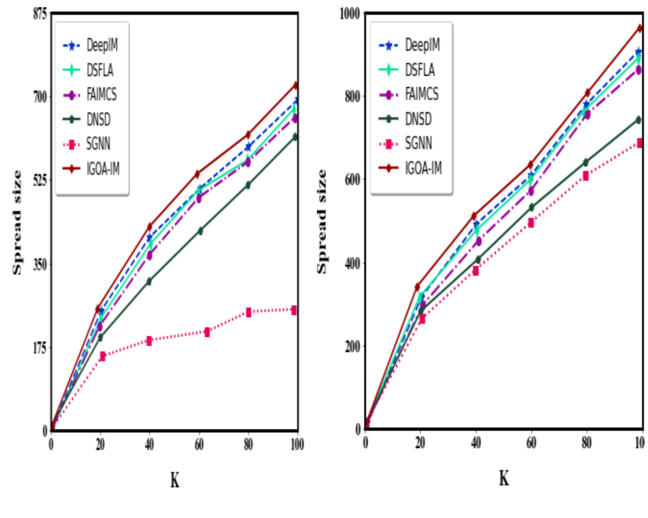
(e) NetInfective (f) Slashdot



(g) Eu-Email (h) NetScience



(i) Gnutella30 (j) NetPHY



(k) NetHEPT (l) LiveJournal

Fig. 1. Analysis of the five algorithms' influence spread over the twelve social networks using the IC model with propagation probability p = 0.01.

Although the IGOA-IM evolutionary rules are effective, the local search strategy used in IGOA-IM tends to make it easy for the approach to become trapped in a locally optimal solution. The issue of imprecise assessment of the node's spread of influence, which assesses the anticipated node'sdiffusion value inside its one-hop area, makes DNSD act as the worst of the 3 meta-heuristic methods, as illustrated in Fig. 1. SGNN performs the worst of the five methods in determining the most powerful nodes for IM.Then, when the targeted seed set size k increases, the algorithm frequently fails to identify the most significant nodes.

5.5 Comparison of run time with other algorithms

To demonstrate the effectiveness of IGOA-IM to recognize influential nodes for impact maximization, a comparison of the five algorithms' running times on the six networks at the specified seed set size of 100 is provided in Figure. 2. The bar graphs in Figure. 2 demonstrate that both DNSDand SSA can locate the intended seed nodes on the twelverealtime networks in just a few seconds.Contrarily, the bar graph shows that the DNSD is the more time-consuming algorithm, even taking 31 hours to choose the desired seed set in the Slashdot network, even though it outperforms all other algorithms in terms of effectiveness.

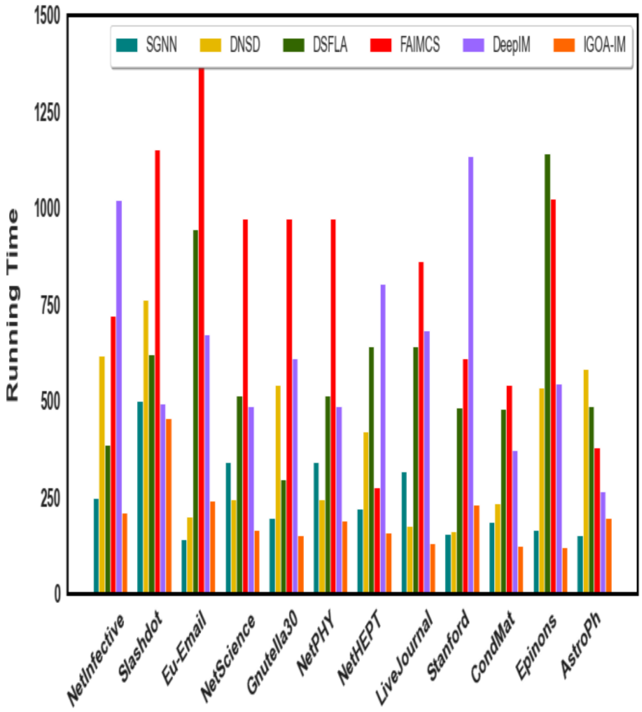


Fig. 2. Comparison of the five methods' execution times on the twelve networks when the desired seed set size is 100.

As shown in Fig. 2, where IGOA-IM's processing time is only half that of DSFLAis 2432 times faster than SGNN on the Slashdot network, the proposed IGOA-IM outperforms DSFLA and SGNN at recognizing influential nodes and is scaleable to large-scale networks.

6. Conclusion

On a variety of challenging optimization tasks, the Improved Gazelle Optimization Algorithm, which combines random search and deterministic techniques, performs exceptionally well. To find prominent nodes for influence maximization, an Improved Gazelle Optimization Algorithm is proposed in this research. A local degree-based replacement method is developed to work in conjunction with local exploitation to enhance the suboptimal meme of everymemeplex in the proposed structure. The Gazelle Optimization and evolutionary rules are conceptualized based on network topology.In the meantime, the IGOA-IM's parameters-setting process is optimized using the orthogonal evaluation modelapproach to ensure that the algorithm evolves successfully. The results of the experiments conducted on various situations

show that the proposed approach may successfully recognize the key nodes in networks.Further work on the IM challenge will primarily focus on the development of efficient spread of influence estimators and more sophisticated evolutionary procedures that are scalable to the biggest networks.

References

[1] Gong, Y., Liu, S., & Bai, Y. (2021). A probability-driven structure-aware algorithm for influence maximization under independent cascade model. Physica A: Statistical Mechanics and its Applications, 583, 126318.

[2] Peng, Z., Huang, W., Luo, M., Zheng, Q., Rong, Y., Xu, T., & Huang, J. (2020, April). Graph representation learning via graphical mutual information maximization. In Proceedings of The Web Conference 2020 (pp. 259-270).

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

[4] Zareie, A., Sheikhahmadi, A., &Khamforoosh, K. (2018). Influence maximization in social networks based on TOPSIS. Expert Systems with Applications, 108, 96-107.

[5] Cui, L., Hu, H., Yu, S., Yan, Q., Ming, Z., Wen, Z., & Lu, N. (2018). DDSE: A novel evolutionary algorithm based on degree-descending search strategy for influence maximization in social networks. Journal of Network and Computer Applications, 103, 119-130.

[6] Aral, S., &Dhillon, P. S. (2018). Social influence maximization under empirical influence models. Nature human behaviour, 2(6), 375-382.

[7] Huang, H., Shen, H., Meng, Z., Chang, H., & He, H. (2019). Community-based influence maximization for viral marketing. Applied Intelligence, 49, 2137-2150.

[8] 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.

[9] Zareie, A., &Sheikhahmadi, A. (2018). A hierarchical approach for influential node ranking in complex social networks. Expert Systems with Applications, 93, 200-211.

[10] Manchanda, S., Mittal, A., Dhawan, A., Medya, S., Ranu, S., & Singh, A. (2019). Learning heuristics over large graphs via deep reinforcement learning. arXiv preprint arXiv:1903.03332.

[11] Wang, Y., Cai, Z., Zhan, Z. H., Gong, Y. J., & Tong, X. (2019). An optimization and auction-based incentive mechanism to maximize social welfare for mobile crowdsourcing. IEEE Transactions on Computational Social Systems, 6(3), 414-429.

[12] Tschiatschek, S., Singla, A., Gomez Rodriguez, M., Merchant, A., & Krause, A. (2018, April). Fake news

detection in social networks via crowd signals. In Companion proceedings of the the web conference 2018 (pp. 517-524).
[13] Tong, G., Wu, W., & Du, D. Z. (2018). Distributed rumor blocking with multiple positive cascades. IEEE Transactions on Computational Social Systems, 5(2), 468-480.
[14] Wu, L., Sun, P., Fu, Y., Hong, R., Wang, X., & Wang, M. (2019, July). A neural influence diffusion model for social recommendation. In Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval (pp. 235-244).
[15] Wen, T., Pelusi, D., & Deng, Y. (2020). Vital spreaders identification in complex networks with multi-local dimension. Knowledge-Based Systems, 195, 105717.
[16] Keikha, M. M., Rahgozar, M., Asadpour, M., &Abdollahi, M. F. (2020). Influence maximization across heterogeneous interconnected networks based on deep learning. Expert Systems with Applications, 140, 112905.
[17] 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.
[18] Bagheri, E., Dastghaibyfard, G., &Hamzeh, A. (2021). FAIMCS: A fast and accurate influence maximization algorithm in social networks based on community structures. Computational Intelligence, 37(4), 1779-1802.
[19] Li, X., & Sun, Q. (2021). Identifying and ranking influential nodes in complex networks based on dynamic node strength. Algorithms, 14(3), 82.
[20] Kumar, S., Mallik, A., Khetarpal, A., & Panda, B. S. (2022). Influence maximization in social networks using graph embedding and graph neural network. Information Sciences, 607, 1617-1636.
[21] Abualigah, L., Diabat, A., &Zitar, R. A. (2022). Orthogonal Learning Rosenbrock's Direct Rotation with the Gazelle Optimization Algorithm for Global Optimization. Mathematics, 10(23), 4509.
[22] Altbawi, S. M. A., Khalid, S. B. A., Mokhtar, A. S. B., Shareef, H., Husain, N., Yahya, A., ... &Alsisi, R. H. (2023). An Improved Gradient-Based Optimization Algorithm for Solving Complex Optimization Problems. Processes, 11(2), 498.
[23] Singh, S. S., Kumar, A., Singh, K., & Biswas, B. (2019). LAPSO-IM: A learning-based influence maximization approach for social networks. Applied Soft Computing, 82, 105554.
[24] Li, Y., Chen, W., Wang, Y., & Zhang, Z. L. (2013, February). Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships. In Proceedings of the sixth ACM international conference on Web search and data mining (pp. 657-666).
[25] Qin, Y., Ma, J., &Gao, S. (2017). Efficient influence maximization under TSCM: a suitable diffusion model in online social networks. Soft Computing, 21, 827-838.
[26] Nguyen, D. L., Nguyen, T. H., Do, T. H., &Yoo, M. (2017). Probability-based multi-hop diffusion method for influence maximization in social networks. Wireless Personal Communications, 93(4), 903-916.
[27] Singh, S. S., Kumar, A., Singh, K., & Biswas, B. (2019). LAPSO-IM: A learning-based influence maximization approach for social networks. Applied Soft Computing, 82, 105554.
[28] Aghaee, Z., &Kianian, S. (2020). Efficient influence spread estimation for influence maximization. Social Network Analysis and Mining, 10, 1-21.
[29] Goyal, A., Lu, W., &Lakshmanan, L. V. (2011, March). Celf++ optimizing the greedy algorithm for influence maximization in social networks. In Proceedings of the 20th international conference companion on World wide web (pp. 47-48).
[30] Kim, J., Kim, S. K., & Yu, H. (2013, April). Scalable and parallelizable processing of influence maximization for large-scale social networks?. In 2013 IEEE 29th international conference on data engineering (ICDE) (pp. 266-277). IEEE.
[31] Kianian, S., &Rostamnia, M. (2021). An efficient path-based approach for influence maximization in social networks. Expert Systems with Applications, 167, 114168.
[32] Banerjee, S., Jenamani, M., &Pratihar, D. K. (2019). ComBIM: A community-based solution approach for the Budgeted Influence Maximization Problem. Expert Systems with Applications, 125, 1-13.
[33] Chen, W., Wang, C., & Wang, Y. (2010, July). Scalable influence maximization for prevalent viral marketing in large-scale social networks. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1029-1038).
[34] Kianian, S., &Rostamnia, M. (2021). An efficient path-based approach for influence maximization in social networks. Expert Systems with Applications, 167, 114168.