EFSLIM : INFLUENCE MAXIMIZATION USING ENHANCED SHUFFLED FROG LEAPING APPROACH IN SOCIAL NETWORKS

[1]K. GEETHA, [2]A. R. NASEER and [3]M. DHANALAKSHMI

[1] CS, SRR Govt. Degree College, SATAVAHANA UNIVERSITY Telangana, India

[2]SGCS, INHA University, Incheon, South Korea & SOCIE INHA University, Tashkent

[3]IT, JNTUH University College of Engineering, Jagitial, Karimnagar, Telangana, India

E-mail: [1]kandukurigeetha01@gmail.com, [2]dr_arnaseer@hotmail.com, [3]dhana.miryala@gmail.com

ABSTRACT

Social networks are influencing people to make choices and decisions on others. With the advertisement of the business products on the web and other sources, the development of social network has increased tremendously. Many social networking organizations develop their network nodes by using a popular concept known as Influence Maximization, which is a greedy approach. The objective of this approach is to maximize the nodes by identifying minimum subset nodes formed at the base level, which has the capability to influence other nodes. The existing algorithm, Independent Cascade Model, in which the activation probability of every node is computed and an influential set is generated based on the behaviour of other nodes due to the influence of the parent nodes. The major disadvantage of this mechanism is the potential creation of vulnerable nodes which spread the information without knowing the adverse effect on the individual. For example, advertising the junk food attractively may have impact on the obese person. The issue with this approach is influencing the entire population using vulnerable nodes. The proposed model tries to influence the targeted audience by maximizing the non vulnerable nodes in the graph. Since the interaction is associated with the behavioural patterns of the individuals, the model uses the genetic algorithm termed as Enhanced Shuffled Frog-Leaping. It searches the local space by encrypting the cumulative responses from other nodes and it updates the fitness function based on the utility. It is evident from the obtained experimental results that the proposed Enhanced Shuffled Frog Leaping Approach for Influence Maximization (EFSLIM) in social network showed the influence spread and statistical tests as an effective and advanced model for overcoming the influence maximization problems. The proposed model showed better performance and reached 1400 of spreading size for 100th node but the existing DFLA obtained 800 of spreading size.

Keywords: *Influence Maximization, Seed Nodes, Shuffled Frog Leaping Algorithm, Social networks, Spreading Size*

1. INTRODUCTION

Social networks have shown a booming development in media such as Facebook, Google+, Twitter for network analysis which is the research hotspot. The social networks showed an impact in information propagation that has become a good platform for exchanging and propagating information among people [1]. Social

networks are the most powerful platforms for performing information diffusion resulting in the expansion of viral marketing with billions of loyal users. The fostering was caused by the capabilities that influenced socially mapped interactions among the network individuals [2]. This is evaluated on the basis of reputation and trust. The

typical applications promoted the social networks based on viral marketing appreciate the word of mouth effect that involved with indwelled interpersonal relationships among the consumers. This reshaped the behaviours and the consumer's attitudes. Thus, the popularity will maximize the influence thereby increasing the social network popularity [3].

With the unprecedented rise of Social networks, more and more people are able to share information to become popular by creating a new platform in social media [4]. The new platform is created by making the marketing viral that will promote the innovations, opinions, and the products [5]. It will be motivated by preceding with the applications, firstly making the selection of the seed that formalizes the viral marketing as a discrete optimization approach which is known as Influence Maximization. The companies will target users in smaller numbers where they use seed sets for advertising and recommending the new products for their friends [6]. The number of people adopting the product influences the surges with the growing popularity in Online Social Networks (OSNs) which is having the information that spreads quickly among the people. There are lots of works that were done to overcome the problem and to put forward using an appropriate algorithm having information of the model certainly.

The challenge of influence maximization is to identify a small selection of nodes (seed nodes) in a social network that can extend influence as widely as possible. It is a fundamental computational difficulty in social influence analysis. The problem requires a very effective and scalable solution since social networks are huge, have intricate link structures, and are highly dynamic. The study of how information spreads on social networking sites has received a great deal of attention recently due to its substantial practical utility in viral or word-of-mouth marketing. To maximize the number of customers who purchased the product, the challenge selects a small number of initial nodes to disseminate the product's information in a social network. The influence Maximization (IM) problem, often known as the process of identifying important nodes in social networks, focuses on identifying a small selection of nodes that have the most influence across the network.

The most important property for any influence maximization is selection of “seed” value. The system needs to identify the minimum seed value using either canonical setting or dynamic control. Usage of these techniques sometimes may overlap the spreading nodes. So the proposed model has introduced the concept of memetic algorithm which identifies the behavioural patterns of the individuals based on the search strategy applied by frogs for hunting their food. The main advantage of this approach is its hybrid mode where it combines the local search and global search. This is a meta heuristic algorithm because the population of frogs are enhanced by scattering their eggs among different birds population.

The SFLA algorithm is a metaheuristic memetic-based algorithm. A population-based technique called the memetic algorithm is utilized to solve challenging and important optimization problems. The way frogs look for food served as the model for the SFLA algorithm. This algorithm searches locally among frog subgroups using the nomometric approach. The frog hybrid jump method takes advantage of the hybrid approach and permits message exchange in local search. The benefits of particle group optimization and the nomometric algorithm are combined in this algorithm. Words are exchanged in both local and global searches in the frog hybrid jump algorithm. Numerous nonlinear, undetected, and multi-state issues can be resolved with the frog hybrid jump technique. The speed of convergence and precision of this algorithm in looking for global solutions are its key benefits. The management of water resources, computers, the production of electricity and energy, and engineering are the principal fields in which SFLA is used.

There were various researches that focused on single social networks in the real world. The users joined multiple social networks that influenced the spread among the users that were common in the multiple networks [7]. The existing works used included the simulation-based proxy approach for sketching-based approaches. These approaches suffered from distinct issues such as feasibility, efficiency, and scalability- based approaches. Because of these approaches, the explored networks showed influence diffusion in terms of computation. The previous algorithms showed information loss during the exploration of network because of the strategies used for pruning model [8-9]. The proposed approach will employ deep learning techniques for learning the feature vectors present among the network nodes that are

Journal of Theoretical and Applied Information Technology
15th December 2022. Vol.100. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

preserved with local and global information. The proposed method obtains the best for network embedding thereby solving the IM problem.

In this paper, we propose an Enhanced Shuffled Frog Leaping Approach for Influence Maximization(EFSLIM) in Social Networks . The proposed model tries to influence the targeted audience by maximizing the non vulnerable nodes in the graph. Since the interaction is associated with the behavioural patterns of the individuals, the model uses the genetic algorithm termed as Enhanced Shuffled Frog-Leaping. The remainder of the paper is organized as follows. Section 2 provides a brief review on the related work carried out in the area of Influence Maximization in social networks. The proposed EFSLIM Model, Data sets used for evaluation, Network Input Data & pre-processing, Influence node detection & segregation steps are presented in Section 3. Section 4 provides a detailed discussion on the proposed Enhanced Shuffled Frog Leaping Algorithm. Section 5 provides performance evaluation and experimental results and discussion along with comparative analysis of the approaches. Finally, Conclusions and future directions are presented in Section 5.

2. RELATED WORK

This section briefly discusses the related work carried out in the area of Influence Maximization in Social Networks.

Jianxin Tang et al [9] developed a Discrete Shuffled Frog Leaping Algorithm (DSFLA) that identified the influence nodes for finding the maximization in the social networks. They carried out experimental evaluation of 6 real- world networks for studying influence Maximization in social networks. An advantage of this study was that the statistical tests showed that the DSFLA performed effectively for selecting target influencing seed nodes for influence maximization. However, in spite of remarkable performances, the model failed to develop an effective influence spread estimator that needed evolutionary rules for scaling more advanced large scale networks.

Weijia Ju et al. [10] developed an algorithm for performing positive influence maximization for the signed network based on the Empirical results. The empirical results obtained for those social networks were because of using an efficient algorithm that gained a positive IM in signed networks. The developed independent cascaded model showed that the algorithm used has shown more influence in spreading positively compared with other methods. The obtained results were better but the developed model failed to investigate the problem for detection of influence nodes in the multiple networks. These nodes joined various trade and social networks influenced spread across distinct networks.

Shan Tian et al. [11] developed a Deep Reinforcement learning-based model for tackling up the Topic aware Influence Maximization(TIM). The results obtained by the meta-learning approach showed that the learned heuristics were generalized and solved distinct kinds of TIM problems. The TIM with larger graph size on distinct graph had instances having same topics which showed excellent performances. However, the developed model is still required to succeed to some extent as the time consumption is seen by the increase in the seed set size.

Mohammad Mehdi Keikha et al [12] evaluated IM across interconnected networks which is heterogeneous in nature using a deep learning approach. The experimental results were interconnected with two networks that include Twitter Foursquare and DBLP that illustrated DeepIM guaranteed an optimal solution in terms of exponential power approximation. However, the developed model found the relevant nodes at the most using heuristics that still remained as a concern.

Jingyi Ding et al. [13] evaluated IM based on the realistic independent cascade model. The experiments are evaluated for the real-world networks and demonstrated that D-greedy was combined with M greedy and R greedy algorithm advantages. The developed model showed better performances as it consumed less time when compared with the existing algorithms. However, the developed model failed for improving the performances, further lowered the running time to select the seeds which were under the uniform setting distribution.

Zahra Aghaee and Sahar Kianian [14] evaluated IM in social networks using Group of Influential

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Nodes (GIN) algorithm. The developed model reduced the search space in the network and considered to find an optimal solution with low running time and acceptable accuracy. The GIN algorithm created distinct groups of graphs having more connection compared with other groups. However, specific nodes from each of the group were required to reduce the search space for finding the influential nodes.

Bhawna Saxena and Padam Kumar [15] determined the node activity and connectivity in social networks for IM. The developed model considered influence spread and two activity based diffusion model including activity based independent cascade model and linear threshold model that influenced the propagation that was actually performed from the past. However, still the model was required to be extended with user activity levels for processing.

Sanjay Kumar et al [16] developed a Modified Degree with Exclusion Ratio algorithm (MDER) for Influence Maximization in social networks. The developed model was based on the notion that showed maximum coverage in the information. The minimum interference was showed in the novel semi-local algorithm, spreading interference that used modified degree centrality. The model used the exclusion ratio that determined the influential nodes in the network with various diverse locations. Yet, the developed model required similar core that required nodes with various numbers that failed for not considering the nodes that had the same core value for influence maximization.

Neda Binesh and Mehdi Ghatee et al [17] developed a Distance Aware Optimization model for the identification of nodes in Social networks. The model used a new Distance Aware Spreader Finding (DASF) algorithm for the detection of problem in the community. The DASF selected the anchor nodes using a novel threshold. The social distance is evaluated between random walk processes and anchor nodes. The distance is regularized with the neighbourhood degree. The model finds an influential spreader that are coming

under the Independent Cascade (IC) as a diffusion model. However, DEIM failed to determine the spreaders who had large networks limited in terms of hardware.

Salim Bouamama and Christian Blum et al [18] developed an Improved Greedy Heuristic model for finding the Influence dominating set problem in the social networks. The results obtained by the greedy algorithm showed that the best quality solutions are obtained especially with respect to the SNAP networks. The developed model showed better performances even for small and medium sized problem instances that tackled the problem occurring in the complex networks. However, the model faced difficulty in finding and identifying the network characteristics showing dominating problem.

In [19], Sanjay Kumar et al suggested IM-ELPR (Influence Maximization using Extended h-index, and Label Propagation with Relationship Matrix). Information from chosen influential nodes is disseminated using Independent Cascade. The chosen nodes identify numerous labeled nodes and are not neighboring. In real-time, the algorithm is applied to 8 datasets of varying sizes. The algorithm is divided into four stages: the seeding phase, the label propagation phase, the merging of communities using a relationship matrix, and the finding of influential nodes. To determine how related two communities are and to potentially join them, a Relationship Matrix is employed. To assess the system's performance, four measures are employed. The suggested model is effective for use on large-scale networks because the time complexity of the IM- ELPR is linear with respect to network size. The proposed algorithm performs better since its density is lower. The node has extremely few connections to other nodes.

In [20], Tarun K. Biswas et al presented a meta-heuristic strategy based on multi-criteria decision-making (MCDM) to address the IM issue in social networks. To cut down on computing expenses, the suggested model chooses potential nodes by removing less important ones during the preliminary stage. To determine the best answer, a modified form of Simulated Annealing is utilized using an improved search method. With three steps—node ranking, candidate pool selection, solution development, and evaluation—the suggested model is a metaheuristic approach. The model features a self-adaptive mechanism and is user-controlled. The model employs a greedy hill-climbing technique as well. It quickens the probability-based convergence process.

ASA(Adaptive Simulated Annealing) and SAW (Simple Additive Weighting) are the methods employed. The work fell short of resolving issues with real-world limitations like time, place, competition, etc.

In [21], Weimin Li et al suggested a dynamic approach for Influence Maximization based on cohesive entropy. To reduce the selection range, the ODP algorithm is used. The model separates the community that overlaps. To examine the proximity of users and separate their effects, the impact propagation algorithm is proposed. By overlooking user associations and choosing the wrong threshold, randomness that results from these factors is reduced. To assess if a user can be a propagatable leader and have an impact on others, a cohesive force that combines cohesive entropy and self-information is developed. Internal community members have a better chance of sharing information than users from other groups. The findings of the fuzzy clustering partition provided by CeCOPRA are utilized to exclude unimportant nodes to streamline the seed selection procedure. The identification of several connections between users and various sorts of impacts is lacking inthe model.

In [22], Sahar Kiranian et al suggested a powerful path-based algorithm HIPA(Heuristic Independent Path algorithm). To assess the performance, extensive empirical tests are also carried out. HIPA is scalable and effective. It effectively eliminates pointless nodes and lowers computational costs. HIPA performs better since it delivered precise IS results. One-by-one diffusion is the mode, and there are only two possible states for nodes: active or inactive. HIPA is a method for resolving issues with influence maximization that falls within the IC model. To streamline the correlation and quicken the computation procedure, HIPA uses the potent heuristic grade function to determine each node's value. Preprocessing with HIPA and VCP decreased the time and cost of computation. HIPA performs better in networks that are sparse and have low average degrees. HIPA uses a greedy method to determine influence spread. As it relies on determining a path between the nodes, HIPA is dependent on the number of edges. HIPA additionally presupposes that influence can only reach a node through pathways that are inside of threshold theta. The paper lacked a framework for modifying the path length criterion to identify legitimate paths.

In [23], Liqing Qiu et al devised an approach, Local-Influence Descending search approach to produce a node-set (LIDDE). The node set has quite a significant impact. The swarm intelligence methodology is the model's foundation. To arrive at the world's best solution, the model models the biological evolutionary process. To hasten differential evolution's convergence, LIDDE employs the LFV technique. EDIV is added to increase precision and effectiveness. The diffusion factor within two-hop is combined with the local impacts computed by LEV in EDIV. Diffusion values over four hops are used by EDIV to calculate the effect of a node-set. To improve parameter quality while lowering variability, a zoom factor is introduced. All local effects are combined with diffusion value in the fitness function. According to a set probability, the suggested algorithm chooses nodes to produce the crossover individual. These algorithms take a long time and employ greedy tactics.

In [24], Michael Kahr et al centered on situations of competition where the original members of one unit are indeed known. A Benders decomposition-based algorithmic framework is created. The framework also makes use of early heuristics and preprocessing. The method is evaluated using recently acquired data and real- world examples. The number of network vertices that an influence cascade can cover was maximized by the model. The authors demonstrated how the maximum coverage location issue can have a stochastic variation in CIMP. BEN is utilized inside the branch and slice structure is used to resolve the CIMP problem. The suggested algorithm uses a greedy strategy. the effects of changing the seed set producing assets on various options for selecting the leader seed collection are studied. Each experiment is conducted on ten SAA repetitions. One significant advantage of the technique is that the tight estimate ratio of MAR just wouldn't apply to problem variations. BEN is a simple average estimation framework that is built on a set covering formulation.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 1: Analytical Report on Existing Works

S. N.	Author	Method / Algorithm	Merits	Demerits
1.	Sanjay Kumar	IM-ELPR	uses a relationship matrix and determines the selective nodes, lower density	works efficient with fewer nodes
2.	Tarun K. Biswas	ASA and SAW	removes all the unnecessary nodes and less computational cost	can't resolve real world issues like time, place,etc.,
3.	Weimin Li	ODP, CeCOPRA		can't identify various sorts of users and impacts
4.	Sahar Kiranian	HIPA	less computational costs	can't modify path length criterion
5.	Liqing Qiu	LIDDE	Swarms methodology, LFV technique is used	takes more computational time
6.	Michael Kahr	Bender's Decomposition	tight estimation ratio by MAR	greedy strategy and increases time complexity
7.	Feng Wang	TrCIM	great gain influence,tie-breaking methods are used	need of trimming on large datasets
8.	Mohammad Mehdi Daliri Khomami	CFIN	dynamic algorithm produces fuzzy results	and cannot deal with multilayer social networks
9.	Pei Li	duplicate forwarding model	user influence both above and below are analyzed, the defined ranking system	only works for symmetric connections, and can't work when a network has a million users.

In [25], Feng Wang et al developed a new model of belief competitive influence diffusion to mimic the transmission of both positive and negative influence. The authors calculated influence probabilities using the predicted trust levels they obtained from generalized network flows. The dynamic dissemination of competitive influence is simulated by the model TrCID. The approach uses influence estimation and TrCIM (Trust-based Competitive Influence Maximization) to iteratively identify the seeds with the greatest gains in influence. The tie-breaking method used by the model was random. The LT model is expanded upon by the TrCIM. The authors also made comparisons using actual datasets without using a greedy strategy. On large-scale datasets, the model was constructed by trimming the MC simulations, which resulted in high time efficiency. The robustness algorithm has to be enhanced because the results for seed detection are unstable.

In [26], Mohammad Mehdi Daliri Khomami et al suggested a quick and expandable algorithm. In order to disseminate influence as widely as possible across networks, the suggested algorithm CFIN (Community Finding Influential Node) chooses k-users based on community structures. The model is divided into two parts: local community spreading and seed selection. The meaningful nodes are chosen in the first section, which involves the collection of seed nodes from networks due to computational complexity. The impact grows inside independent communities in the second section. The dynamic and agent-based algorithms blended competition and cooperation among particles and produced fuzzily structured results. The confidence level is derived by an overlapping measure. Any community detection approach can be replaced by the suggested model. Multilayer social networks can't be supported by the suggested model. To make the research of these networks more realistic, overlapping needs to be taken into account.

In [27], Pei Li et al developed a duplicate forwarding model to describe the social network diffusion process. The user influence both above and below the diffusion threshold was also examined by their approach. The correlation between two rankings is calculated using a model that uses a Spearman-like correlation coefficient. The identical forwarding model analysis findings are also found, and the accuracy was greatly improved. Rankings of users and influences are

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

correlated. The model assigns users with a similar rank value, which itself are equal to the average of their locations in ascending order, who have the same user influence. To make the model simpler, only symmetric connections and uniform user behavior are taken into account. Asymmetric relationships, where different users pass messages with differing probabilities, were not discussed in this research work. If the network contains millions of users, then this approach is not appropriate.

In the next section, we present our proposed EFSLIM Model, Data sets used for evaluation, Network Input Data & pre-processing, Influence node detection & segregation steps

3. PROPOSED EFSLIM MODEL

The Block Diagram of the Proposed Model on Influence Maximization in Social Network EFSLIM is shown in figure 1. The block diagram consists of Dataset, network input data, pre-processing, influence node detection and segregation, influence maximization. The dataset includes HepPh, Graph 30, Phy, Epinion, ama, Amazon, Epi, Enron email dataset, and Stanford Dataset. At the next step, the network input data is given to the pre-processing step. As the social networks consisted of huge data, an effective model is required for data pre-processing. Therefore, it is important to select the raw input data for selecting among the pre-processed data. The pre-processed data processes for influence node detection and the model determines the influence node in the network using the proposed ESFLA. The ESFLA algorithm has mainly two main stages such as Local Exploration and Global Exploration. The local exploitation strategy has been at its worst meme having each submemeplex able to exploit the influential nodes. Mostly it will help the collective evolution. The global explorations have memeplexes that were contributed and combined for obtaining solutions.

The experimental results showed that the influence spread and the statistical tests showed advanced and effective model overcoming the IM problem. Thus, an effective influence estimator showed an easier identification of accuracy and the IM is performed for the data.

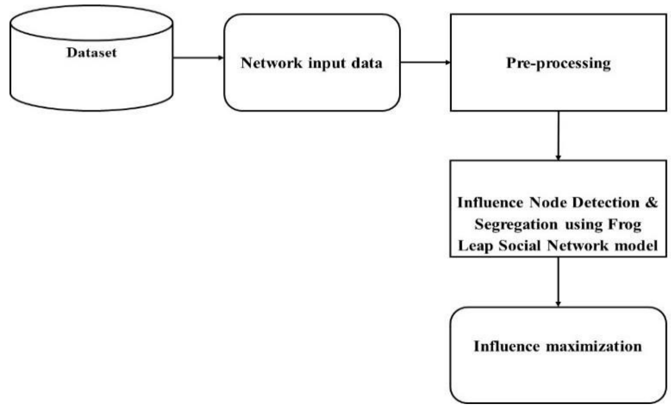

Fig. 1: Block Diagram Of Influence Maximization In Social Networks

3.1 DATASET

The proposed model evaluates results for real-world dynamic datasets that include the following datasets:

Enron email dataset

The Enron email consists of mainly 500,000 emails that are generated by the Enron Corporation employees. These emails were obtained by the Federal Energy Regulatory Commission that will be investigated for Enron's collapse [19].

Epinions

The Epinions dataset is a who-trust- whom online social network that is having a general consumer review site which is having Epinions.com. The members from the site decide to trust each other or not. All the trust relationships are managed and interacted from the Web of Trust that is then combined with the review rating, will find the reviews which are shown to the user [20].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 2: Epinions Dataset Description

Dataset statistics	
Number of triangles	1624481
Edges	508837
Nodes	75879
Average clustering coefficient	0.1378
Diameter (longest shortest path)	14
Fraction of closed triangles	0.0229
90-percentile effective diameter	5

Stanford Dataset Collection:
Stanford dataset consists of a total 20 number of datasets. Among that the following datasets are only used [21]

Social networks: Social networks are the online social networks having the edges that represent the interaction among the people.

Networks with ground-truth communities: Networks with ground-truth communities are having social networks and information networks.

Communication networks: The email communication networks are having edges that represent the communication.

Citation networks: The citation networks provide a representation of networks and edges representing the citations.

Collaboration networks:
The nodes are represented by scientists and the edges are representing the collaborations.

HepPh
High Energy Physics Phenomenology is the citation graph that is obtained from the e-print arXiv that covers the citations present overall. Within the dataset, totally of 34,546 papers and with 421,578 of edges [22].

Graph 30
The Amazon has made the graph challenge datasets which is available in a community with free of charge as a part of AWS public dataset. The data present in several formats, files and there is wide variety of paths to access them [23].

Amazon
This dataset was created in house teams at Prompt Cloud and the dataset contains 30K records in it. Total Records Count: 697053 Domain Name: amazon.com and Date Range: 01st Jan 2020 - 31st Jan 2020.

BHOSLIB (Multi Level Graph Visualization): Benchmark graphs are used to test NP-hard graph algorithms including the minimum vertex cover, maximum independent set, maximum clique problem, and vertex colouring.

Table 3: Statistics of BHOSLIB

1	Nodes	450
2	Edges	83.2K
3	Density	0.823539
4	Maximum degree	407
5	Minimum degree	327
6	Average degree	369
7	Assortativity	-0.0291078
8	Number of triangles	25.2M
9	Average number of triangles	56.1K
10	Maximum number of triangles	67.5K
11	Average clustering coefficient	0.821076
12	Fraction of closed triangles	0.820998
13	Maximum k-core	340
14	Lower bound of Maximum Clique	24

Gnutella: A series of images taken in August 2002 showing the Gnutella peer-to-peer file sharing network. Nine snapshots of the Gnutella network were taken overall in August 2002. In the Gnutella network structure, nodes stand in for hosts, and edges represent the relationships between Gnutella hosts.

AstroPh: Scientific collaborations between authors of articles submitted to the Astro Physics category are covered by the Arxiv ASTRO-PH (Astro Physics) collaboration network, which is part of the e-print arXiv. The graph has an undirected edge from author I to author j if they both co-authored a paper. The co-authorship of k authors results in a fully linked (sub)graph on k nodes for the publication.

Table 4: Statistics Of Gnutella

1	Nodes	6301
2	Edges	20777
3	Nodes in largest WCC	6299 (1.000)
4	Edges in largest WCC	20776 (1.000)
5	Nodes in largest SCC	2068 (0.328)
6	Edges in largest SCC	9313 (0.448)
7	Average clustering coefficient	0.0109
8	Number of triangles	2383
9	Fraction of closed triangles	0.006983
10	Diameter (longest shortest path)	9
11	90-percentile effective diameter	5.5

Table 5: Statistics Of Astroph

1	Nodes	18772
2	Edges	198110
3	Nodes in largest WCC	17903 (0.954)
4	Edges in largest WCC	197031 (0.995)
5	Nodes in largest SCC	17903 (0.954)
6	Edges in largest SCC	197031 (0.995)
7	Average clustering coefficient	0.6306
8	Number of triangles	1351441
9	Fraction of closed triangles	0.1345
10	Diameter (longest shortest path)	14
11	90-percentile effective diameter	5

Table 6: Statistics Of AS-22JULY06

1	Nodes	23K
2	Edges	48.4K
3	Density	-
4	Maximum degree	2.4K
5	Minimum degree	1
6	Average degree	2.10930627531
7	Nodes	23K
8	Edges	48.4K

AS-22JULY06: A symmetrized snapshot of the Internet's organization at the level of autonomous systems is depicted in the graph as-22july06 and was created using BGP tables that were uploaded to archive.routeviews.org. This snapshot, which Mark Newman developed using data for July 22, 2006, was not previously released.

3.2 Network Input Data & Pre- Processing

The model considers the social networks that correspond to the semantics as an input where those networks are having the ability to model as the graph. The datasets are applied for community detection that requires local information that is suitable to analyze the large weighted network. The community detection is significant as it uses the local information to analyze the large weighted network. The input data is used for performing network construction and also community construction that reads the data and constructs the network graph that is constructed on the basis of given data. Thus, the input parameters are analyzed lastly for finding the influence values among the nodes in terms of the probability function. Social networks are usually having huge data and thus an effective method is needed for data processing. Thus, the most appropriate raw input data is selected should be pre-processed. If it is not pre-processed, the network will not produce the forecast accurately for IM. The variables of input are modified in the pre-processing phase that will match better and generate an output. The data quality is enhanced and is promoted for the meaningful extraction from the data. The data will be prepared, cleaned, and organized makes it suitable for further steps to evaluate the influence maximization.

3.3 Influence Node Detection & Segregation

There are various methods involved with influence maximization and influence node detection techniques that were developed for the discovery of influential users having certain social network conditions. Each of the nodes will affect other nodes having some degree through the influence of propagation and influential nodes are detected on the basis of the characteristics. The proposed research finds the influence node in the social network. As far as IM problem is concerned, the node propagation is processed on the node to influence the network. It uses analogous for meme evolution for the frog population. The social individuals have shown profit from the promotional information sharing. It is the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

generalized behaviour among the other interactive members that will resize the behaviour when the influence starts to propagate through the network. Thus, the evolutionary mechanism for the SFLA requires itself for tackling up the problem of IM. An efficient method is developed which is subjected to selecting the seed set as the target will maximize the influence of spread. It effectively constructs and estimates the needed influence spread accurately which sets the given node-set which was another challenge for IM.

3.4 Proposed Approach EFSLIM Using Enhanced Shuffled Frog Leaping Algorithm

The Enhanced Shuffled Frog-leaping algorithm (ESFLA) works on the basis of a metaheuristic memetic algorithm which is based on the population but showed complexity significantly. The main aim of the algorithm is to search for the local leader within the genetic algorithm structure that shows improvement in terms of aqueous performances. The memetic algorithm will encrypt the initial answers and then the algorithm evaluates the utility of each response on the basis of generated fitness functions showing the new solution. The inspiration of the ESFLA algorithm is based on searching for food for frogs. The algorithm mainly uses the Nomometric method for searching the food locally among the frogs' subgroups. The proposed research utilizes a hybrid jump algorithm that utilizes mainly the hybrid strategy that allows for message exchanging in the local search. The advantages of particle group optimization and the Nomometric algorithm are because of their combined optimization approach. The words are exchanged not with the local search but with the global search also. Therefore, the local and global searches are combined with the algorithm. The hybrid jump algorithm is searchable at a higher rate and will be easy to implement. The hybrid jump algorithm will solve

undetectable, non-linear, and multi-state problems.

The steps involved in the SFLA frog jump algorithm is as follows:

- The SFLA algorithm's meta-exploration strategy is summarized in the two main stages of Global exploration and Local exploration according to the following steps.

Global exploration stages :

Step 1: Select the terms M, N that represents the memeplexes number and the number of frogs respectively. Therefore, the population size of the pond is totally obtained with the relation as $F = M \times N$.

Step 2: Production of the virtual population

From the space available, the virtual frogs $(1), U(2), \ldots., (F)$ are calculated which is having the competency value as $f(i)$ for each $U(i)$. $(i) = (U^1, U^2, \ldots, U^d)$. Thus, the number of decision variables is represented as d.

Step 3: Sorting and Grading of the frogs

The frogs are stored in descending order based on the metrics according to the array $X = \{(i), f(i), \text{where } i = 1, \ldots, F\}$, The positions are recorded at their best as $P_x$for frog present in the population which is provided in Eq. (1)

$$U = P_x \tag{1}$$

Step 4: The frogs are divided into the memeplexes

The array X and Y each of them consisting of N number of frogs.

Step 5: Each of the memeplexes shows the evolution in memetic.

Each of the memeplexes Y_k where $k = 1,2, 3, \ldots., M$ will be evolved with the local search using the frog jump algorithm which is described as given below.

Step 6: The memeplexes are Combined

Once after reaching a certain number, memetic evolution will take place that has $Y_1, \ldots Y_M$ in X

that sets a relation as $X = Y_k$, where $k = 1,2, \ldots, M$. The best position $P_x$is updated.

Step 7: The Convergence study is performed

If the convergence conditions are met, then stop the process or else perform the fourth step to perform a global search.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Local exploration steps:

At the fifth stage in the global search, the evolution for each of the memeplexes is performed N times independently. Once the evolving of the memeplexes, the algorithms will return with the global search for performing a complete combination. The details of the local search are described in detail at each memeplex are given below:

Step 1: Initialization

The value of i_M and i_N is set to $\overline{\text{zero}}$ and the number of memeplexes is counted as i_M, i_N counts with the evolution steps which are numbered as follows :

Step 2: Evaluate $i_M + 1 = i_M$ (2)

Step 3: Evaluate $i_N + 1 = i_N$ (3)

Step 4: Creation of sub-memeplexes

The goal of the frog is to move the positions optimally and improve memes. The selection of submemeplex is done by assigning more weights of frogs thereby showing higher performances. The lower weights of the frogs are generated which showed lowering the values in terms of performances. The weights obtained are assigned

Where the value of S is the step size of the frog which is obtained as shown below :

If the position generated is best than the past one, then the new value (q) is having the former
(Q), Then step 8 is stepped for the local search or else step 6 is performed for the local search.

Step 6: The step size having the size P_x fails if there is no better result which is as shown in step
5. Then the step size of the frog will be calculated as shown in the below equation:

The new position (q) is calculated by setting the relation as $U(q) = S + PW$. If the equation is within the possible space, then the efficiency of (q) is calculated. If the value of (q) shows better when compared to the previous ones. It is then replaced with (q) as the new one which is former $U(q)$ goes for the 8th step in the local search. Else, the seventh step of the local search is obtained.

Step 7: The process of Censorship is performed if the new position updated is not at the achievable area which is not better compared to the previous position, the newly generated frog(r) are generated randomly that is available for the location. This replaces the frog whose newly generated positions are not suitable for advancement. The value of
(r) and the value of $U(q)$ are set to r which is equal to the value of $f(r)$.

Step 8: The memeplexes are updated

Once the mimetic change occurs then the worst frogs are in the submemeplex places the frogs at Z which is their original position represented as Y_i^m.

with probability distribution $2(j-1+n)\ P_i = \{ {}_{(n+1)}, j =$
The order of performances is sorted in the
descending order as Y_i^m
$1, \ldots, n\}$. Thus, in order to build the submemeplex, the arrays Z, the randomly selected frogs q are selected from the n number of frogs for each of the memeplex. The submemeplex are denoted with PW and PB respectively.

Step 5: Correction of the worst frog position.

The worst frog and the new position of the frog are considered with the submemeplex that means the frog showing the worse performances are evaluated with the relation which is provided in Eq. (4)

$$(q) = S + PW \quad (4)$$

Step 9: If $>$ i_N, perform a local search using step 3.

Step 10: If $m > i_m$, then the local search is performed using step 1, or else the global search is returned for combining the memeplexes.

The results obtained from the experiment showed that the influence spread and statistical tests showed that the ESFLA is an effective and advanced model for overcoming the influence maximization problems. There are always various factors and parameters which are to be adjusted for a meta-heuristic algorithm that has reasonable parameters for setting up the strategy. The strategy involves the ability of convergence and improves the algorithm performances. Therefore, the current proposed model performs shuffling to partition the frog population. The model is divided to memeplexes that contributed for obtaining solutions.

The Sub node from Z_i or P_X having the entire
population is initialized which is represented as

$F = M \times N$
The virtual frogs that are generated as
$(1), U(2), \dots, U(F)$
new_node $\leftarrow \Phi$
for each node$\in P_b(P_X)$ do
$N(1) \leftarrow One\ ho(node)$
$SN(1) \leftarrow Sort(N(1))$
for each $node \in SN(1)$ do if $node \ \in/ \ new\_node$ $then$
$new\_node \leftarrow new\_node \cup \{node\}$
break end if end for end for
return new_node

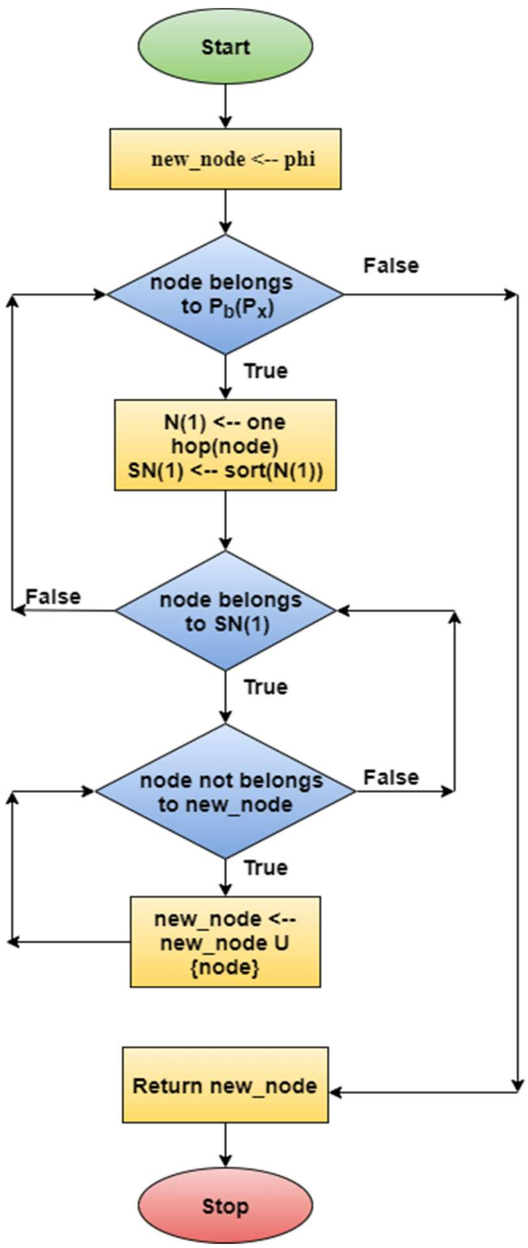

Figure 2: Flow Chart for Enhanced SFL Algorithm

The local exploitation strategy considers frogs with worst positions for corrections that had submemeplex which are suitable to exploit the influential nodes. The model performed the collective evolution and the model showed easier influence on the nodes improved accuracy.

4 PERFORMANCE EVALUATION & EXPERIMENTAL RESULTS

The experimental setup used consisted of Intel i7 processor which is having 8GB of RAM and 500 GB hard disk. The proposed method EFSLIM was implemented using python 3.6 having the networks and community. The proposed method results are analysed in terms of memory usage, execution time, and average influence spread.

4.1 Performance Metrics

Influence Spread

combinatorial optimization problems yet the algorithm showed slow convergence when it comes under local optimal solution and premature convergence. The proposed EFLA-based Social network model for overcoming the problem of IM showed better results when compared to the GA and FL optimization approaches. The results are evaluated with respect to three datasets as Epinions, Stanford, and Enron Email dataset. It is evident from figure 3 that as the number of seed nodes increase from 10 to 100, the spread size also increases. The spread size for GA varies from 100 to 810 with respect to the Epinions dataset. Similarly, the FL algorithm obtains results ranging from 120 to 980 and the proposed EFSLIM method produces better results ranging from 150 to 1400.

Influence spread is defined as the process of choosing the initial people to maximize the number of people who will receive the information of the product in the social network. It is shown in equation (5)

$$S = \arg\max(S) \; \{ \subseteq V, |S| = k \qquad (5)$$

Where, S is a seed set, (S) denotes the influence spread of S, V is the nodes set and k is the number of seed nodes to be selected.

Journal of Theoretical and Applied Information Technology
15th December 2022. Vol.100. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Memory requirement

The Memory performances are evaluated using the relationship among the latency and speed which are related closely to use the information for optimizing the performance in the memory.

4.2 Quantitative Analysis

Table 7 shows the results obtained for Epinions Dataset in terms of spread size for the existing GA, FL algorithms which are compared with the proposed EFSLIM algorithm. The existing GA algorithm used for overcoming the problem of IM showed repeated fitness function evaluation indicating complexity problems limited the performances of the model. The existing SFLA was only suitable for solving various

Table 8 shows the spread size obtained for the Stanford dataset. It is clearly seen from the figure 4, as the number of the seed nodes increase from 10 to 100, the spread size also increases. The spread size for GA varies from 190 to 880 with respect to the Stanford dataset. Similarly, the FL algorithm obtained results ranging from 140 to 1040 and the proposed EFSLIM method produced better results ranging from 200 to 1480.

Table 7: The Spread Size (K) Obtained For Epinions Dataset

Seed Nodes	GA	FL	Proposed
10	100	120	150
20	140	200	300
30	220	280	500
40	300	380	600
50	370	460	750
60	450	560	900
70	570	640	1050
80	620	700	1100
90	720	850	1200
100	810	980	1400

Table 9 shows the spread size obtained for the Enron Email dataset. From figure 5, it is observed that as the number of the seed nodes increase from 10 to 100, the spread size also tend to increase. The spread size for GA varies from 220 to 900 with respect to the Enron email dataset. Similarly, the FL algorithm generated results ranging from 220 to 1110 and the proposed EFSLIM method produced better results ranging from 220 to 1530.

Table 8: The Spread Size (K) Obtained For The Stanford Dataset

Seed Nodes	GA	FL	Proposed
10	190	140	200
20	240	290	330
30	310	310	560
40	340	470	630
50	440	550	840
60	520	590	960
70	660	730	1080
80	700	750	1160
90	740	890	1240
100	880	1040	1480

Table 9: The Spread Size(K) Obtained For Enron Email Dataset

Seed Nodes	GA	FL	Proposed
10	220	220	220
20	280	320	410
30	410	400	660
40	360	520	730
50	500	640	930
60	540	660	1010
70	690	820	1180
80	790	770	1230
90	810	930	1330
100	900	1110	1530

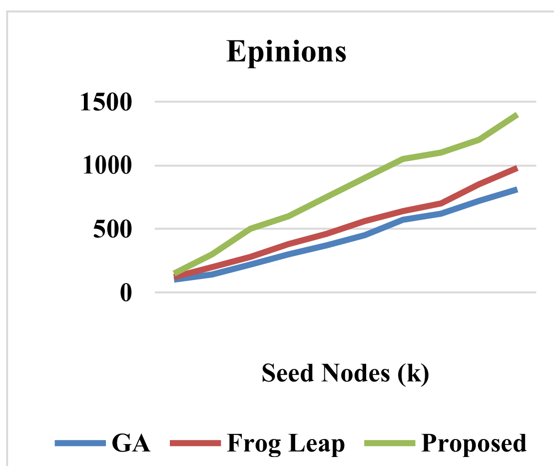

Figure 3: The Spread Size Obtained For The Epinions Dataset

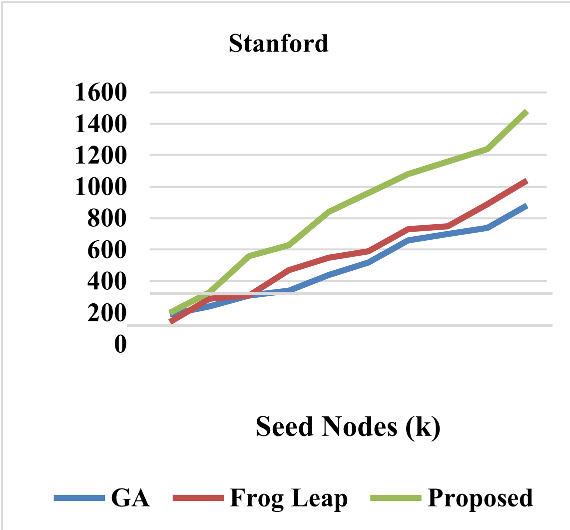

Figure 4: The spread size obtained for Stanford

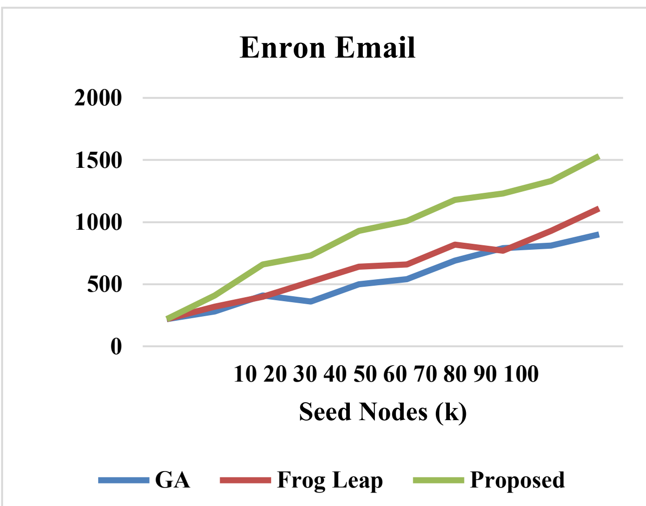

Figure 5: The spread size obtained for the Enron Email dataset

Table 10: Average Execution time(s)

Seed Nodes (k)	GA	Frog Leap	Proposed
10	6	4	2
20	8	4	2
30	8	5	3
40	9	5	3
50	9	5	4
60	10	9	4
70	14	9	8
80	14	11	9
90	16	15	10
100	16	15	10

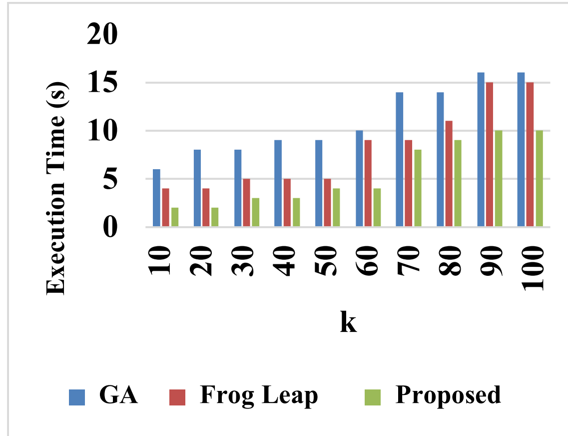

Figure 6: The average execution time obtained for the proposed method and the existing methods

Table 10 shows the average execution time obtained by the proposed FSLIM when compared with the existing GA and Frog leap algorithm. It is evident that as the number of seed nodes increases the execution time also increases. When the seed node reaches 100, the execution time is also high but compared with the existing frog leap and GA, the proposed model utilized lesser execution time. Figure 5 shows the average execution time obtained for the proposed method and the existing Frog leap, GA methods. In this proposed model, the memory requirements are evaluated for distinct node numbers. The memory requirement of algorithm also shows that the proposed model utilized lesser memory and was consistent in the estimation.

The Table 11 shows performance evaluation for different datasets such as HepPh, Graph 30, Phy, Epinion, ama, Amazon, Epi, BHOSLIB, gnutella08, AstroPh and AS- 22JULY06. From the table results, HepPh dataset has 40 number of targeted nodes of 15233 that has the execution time of 2.52s whereas the graph 30 dataset has less dataset but has high performance even with 30 number of nodes consuming execution time of 0.0417s. Table 12 shows performance evaluation of datasets - BHOSLIB, gnutella08, AstroPh and AS-22JULY06. This Table shows the Execution time, Memory usage & Influence Spread for variation in seed size K from 10 to 50.

Journal of Theoretical and Applied Information Technology
15th December 2022. Vol.100. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 11: Metrics Evaluation Obtained For Datasets

Dataset	Targeted Nodes (k)	Spread Size	Execution time (s)	Total Number of Nodes
HepPh	40	1523	2.522	15233
Graph 30	11	3	0.0417	30
Phy	121	3715	8.529	37149
Epinion	2931	7588	4087.87	75879
ama	41	23276	17.511	232761
Amazon	41	23278	13.9868	232780
Epi	2925	7588	2281.13	75879
BHOSLIB	1000	3050	2009	3064
Gnutella08	500	1503	750	1003
AstroPh	275	401	100.93	562
AS-22JULY06	843	2964	1890	1285

Table 12: Performance Evaluation For Different Datasets

Dataset	Seed Size	Best Affinity Value	Execution Time	Memory Usage	Influence spread
BHOSLIB	K =10	9.9	5.98 ms	3.2	15.43
	K = 20	9.1	5.5 ms	3.1	14.78
	K = 30	8.6	4.9 ms	2.95	14.54
	K = 40	8.1	5.3 ms	2.83	13.91
	K = 50	7.4	4.8 ms	2.5	12.10
Gnutella08	K =10	4.36	1.08 ms	1.75	10.99
	K = 20	3.0	0.97 ms	1.43	10.71
	K = 30	3.0	1.04 ms	1.45	10.71
	K = 40	3.0	0.28 ms	1.06	8.92
	K = 50	3.0	0.35 ms	1.47	9.01
AstroPh	K =10	4.1	0.17ms	1.40	10.3
	K = 20	3.2	0.22ms	1.36	9.8
	K = 30	3.5	0.56 ms	1.38	8.9
	K = 40	3.0	0.31ms	1.39	9.01
	K = 50	3.0	0.34ms	1.40	9.01
AS-22JULY06	K =10	5.5	2.35 ms	2.98	8.95
	K = 20	5.3	2.35 ms	2.45	8.44
	K = 30	4.8	2.04 ms	1.8	8.03
	K = 40	4.1	2.04 ms	1.8	8.01
	K = 50	3.5	1.98 ms	1.34	7.55

4.3 COMPARATIVE ANALYSIS

Table 13 shows the comparative analysis for the proposed method in terms of spreading size evaluated for three datasets such as Epinions, Stanford, and Email. The results are compared with the 100th and 50th node with the existing DSFLA [11], Group of Influential Nodes [14], and activity-based Independent Cascade model, and the Activity-based Linear Threshold model [15] for the evaluation of results.

Table 13: Comparative Analysis

Methods	Seed Node (k)	Data set		
		Epinions	Stanford	Email
Deep Reinforcement Learning-Based Approach [11]	100	800	400	600
Connectivity- based model [14]	50	668.16	NA	67-69
Influence maximization algorithm [15]		NA	NA	47
Proposed EFSLIM method	50	750	NA	930
	100	1400	1480	1530

NA – Not Available

The DSFLA model showed the problem of developing an effective influence spread estimator that required more advanced evolutionary rules which is scalable for large-scale networks and thus results in lower performances when compared to the existing models. The GIN model showed better performance for the Epinions dataset when compared to the email dataset. Yet, the GIN model failed to reach better performance when compared to the proposed model due to the selection of the specific nodes from each group the model reduced the search space for finding the most influential nodes.

5 CONCLUSION & FUTURE WORK

The shuffled frog-leaping algorithm combining the deterministic and random search strategies showed excellent performance. The model combined the random search strategies and deterministic search strategies thereby showing greater improvement in IM. The proposed model performed discrete encoding mechanism that constructed evolutionary rules conceived using the network topology. The local degree based replacement strategy with local exploitation showed improvement for each memeplex. The orthogonal experimental results

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

showed that the optimized parameters evaluated the EFSLIM effectively. The experimental study on various datasets demonstrated that the proposed method EFSLIM successfully identified in large number the influential nodes in networks. The proposed model EFSLIM showed better performances and reached 1400 of spreading size for 100th node but the existing DFLA obtained 800 of spreading size. The major limitation of the work it needs many computation resources to run the processors parallel. As a future work, it is proposed to handle BigData sets in social networks and to design & implement parallel algorithms for Influence Maximization using Hadoop & Map Reduce framework or SPARK framework.

REFERENCES:

[1] Li, J., Cai, T., Deng, K., Wang, X., Sellis, T. and Xia, F., 2020. Community-diversified influence maximization in social networks. *Information Systems*, *92*, p.101522.

[2] Biswas, T.K., Abbasi, A. and Chakrabortty, R.K., 2021. An MCDM integrated adaptive simulated annealing approach for influence maximization in social networks. *Information Sciences*, *556*, pp.27-48.

[3] Qiu, L., Tian, X., Zhang, J., Gu, C. and Sai, S., 2021. LIDDE: A differential evolution algorithm based on local-influence-descending search strategy for influence maximization in social networks. *Journal of Network and Computer Applications*, *178*, p.102973.

[4] Kumar, S., Singhla, L., Jindal, K., Grover, K. and Panda, B.S., 2021. IM-ELPR: Influence maximization in social networks using label propagation based community structure. *Applied Intelligence*, pp.1-19.

[5] Kahr, M., Leitner, M., Ruthmair, M. and Sinnl, M., 2021. Benders decomposition for competitive influence maximization in (social) networks. *Omega*, *100*, p.102264.

[6] Ghayour-Baghbani, F., Asadpour, M. and Faili, H., 2021. MLPR: Efficient influencemaximization in linear threshold propagation model using linear programming. *Social Network Analysis and Mining*, *11*(1), pp.1-10.

[7] Caliò, A. and Tagarelli, A., 2021. Attribute based diversification of seeds for targeted influence maximization. *Information Sciences*, *546*, pp.1273-1305.

[8] Sivaganesan, D., 2021. Novel Influence Maximization Algorithm for Social Network Behavior Management. *Journal of ISMAC*, *3*(01), pp.60-68.

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

[10] Ju, W., Chen, L., Li, B., Liu, W., Sheng, J. and Wang, Y., 2020. A new algorithm for positive influence maximization in signed networks. *Information Sciences*, *512*, pp.1571- 1591.

[11] Tian, S., Mo, S., Wang, L. and Peng, Z., 2020. Deep reinforcement learning-based approach to tackle topic-aware influence maximization. *Data Science and Engineering*, *5*(1), pp.1-11.

[12] Keikha, M.M., Rahgozar, M., Asadpour, M. and Abdollahi, M.F., 2020. Influence maximization across heterogeneous interconnected networks based on deep learning. *Expert Systems with Applications*, *140*, p.112905.

[13] Ding, J., Sun, W., Wu, J. and Guo, Y., 2020. Influence maximization based on the realistic independent cascade model. *Knowledge-Based Systems*, *191*, p.105265.

[14] Saxena, B. and Kumar, P., 2019. A node activity and connectivity-based model for influence maximization in social networks. *Social Network Analysis and Mining*, *9*(1), pp.1-16.

[15] Aghaee, Z. and Kianian, S., 2020. Influence maximization algorithm based on reducing search space in the social networks. *SN Applied Sciences*, *2*(12), pp.1-14.

[16] Kumar, S., Lohia, D., Pratap, D., Krishna, A. and Panda, B.S., 2022. MDER: modified degree with exclusion ratio algorithm for influence maximisation in social networks. *Computing*, *104*(2), pp.359-382.

[17] Binesh, N. and Ghatee, M., 2021. Distance-aware optimization model for influential nodes identification in social networks with independent cascade diffusion. *Information Sciences*, *581*, pp.88-105.

[18] Bouamama, S. and Blum, C., 2021. An improved greedy heuristic for the minimum positive influence dominating set problem in social networks. *Algorithms*, *14*(3), p.79.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[19] Kumar, S., Singhla, L., Jindal, K., Grover, K., & Panda, B. S. (2021). IM-ELPR: Influence maximization in social networks using label propagation based community structure. Applied Intelligence. doi:10.1007/s10489-021-02266-w

[20] Biswas, T. K., Abbasi, A., & Chakrabortty, R. K. (2021). An MCDM integrated adaptive simulated annealing approach for influence maximization in social networks. Information Sciences, 556, 27–48. doi:10.1016/j.ins.2020.12.048

[21] Li, W., Zhong, K., Wang, J., & Chen, D. (2020). A Dynamic Algorithm based on Cohesive Entropy for Influence Maximization in Social Networks. Expert Systems with Applications, 114207. doi:10.1016/j.eswa.2020.114207

[22] Kianian, S., & Rostamnia, M. (2020). An Efficient Path-Based Approach for Influence Maximization in Social Networks. Expert Systems with Applications, 114168. doi:10.1016/j.eswa.2020.114168

[23] Qiu, L., Tian, X., Zhang, J., Gu, C., & Sai, S. (2021). LIDDE: A differential evolution algorithm based on local-influence-descending search strategy for influence maximization in social networks. Journal of Network and Computer Applications, 178, 102973. doi:10.1016/j.jnca.2020.102973

[24] Kahr, M., Leitner, M., Ruthmair, M., & Sinnl, M. (2020). Benders decomposition for competitive influence maximization in (social) networks. Omega, 102264. doi:10.1016/j.omega.2020.102264

[25] 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. doi:10.1016/j.ins.2020.09.002

[26] Khomami, M. M. D., Rezvanian, A., Meybodi, M. R., & Bagheri, A. (2020). CFIN: A community-based algorithm for finding influential nodes in complex social networks. The Journal of Supercomputing, 77(3), 2207– 2236. doi:10.1007/s11227-020-03355-2

[27] Li, P., Liu, K., Li, K., Liu, J., & Zhou, D. (2021). Estimating user influence ranking in independent cascade model. Physica A: Statistical Mechanics and Its Applications, 565,125584. doi:10.1016/j.physa.2020.125584

[28] Rossi, R. A., & Ahmed, N. K. (2015). The Network Data Repository with Interactive Graph Analytics and Visualization. AAAI. Ahttps://networkrepository.com

[31] Shetty, J. and Adibi, J., 2004. The Enron email dataset database schema and brief statistical report. *Information sciences institute technical report, University of Southern California*, *4*(1), pp.120-128.

[32] Boroujeni, G.A.A., 2013, April. A trust model for epinion dataset. In *7th International Conference on e-Commerce in Developing Countries: with focus on e-Security* (pp. 1-7). IEEE.

[33] https://snap.stanford.edu/data/

[34] Chekanov, S.V. and Kuvshinov, V.I., 1996. Bunching parameter and intermittency in high- energy collisions. *arXiv preprint hep-ph/9605379*.

[35] Koone, M.E. and Elmasri, R., 2021, April. Public Datasets: Access, Download and Cleaning (AWS). In International Conference on Data Science, E-learning and Information Systems 2021 (pp. 70-75).

[29] M. Ripeanu and I. Foster and A. Iamnitchi. Mapping the Gnutella Network: Properties of Large-Scale Peer-to-Peer Systems and Implications for System Design. IEEE Internet Computing Journal, 2002.

[30] J. Leskovec, J. Kleinberg and C. Faloutsos. Graph Evolution: Densification and Shrinking Diameters. ACM Transactions on Knowledge Discovery from Data (ACM TKDD), 1(1), 2007.