Applied Computing and Informatics 15 (2019) 102–108

Contents lists available at ScienceDirect

Applied Computing and Informatics

journal homepage: www.sciencedirect.com

Original Article

A SI model for social media influencer maximization

Jyoti Sunil More [a,*,1], Chelpa Lingam [b,1]

[a] Department of Computer Engineering, Ramrao Adik Institute of Technology, Nerul, Navi Mumbai, India
[b] Pillai's HOC College of Engineering and Technology, Rasayani, India

ARTICLE INFO

Article history:
Received 11 May 2017
Revised 11 November 2017
Accepted 14 November 2017
Available online 15 November 2017

Keywords:
Influencers
Social network analysis
Diffusion model
SI model
Multithreading
Marketing strategies

ABSTRACT

Social network mining can be divided into two categories, namely, the study of structural characteristics and content analysis. One of the most significant problem in the context of a social network is finding the most influential entities within the network. This task has significance in viral marketing, since the most influential entities can be targeted for endorsing new products in the market. However, the problem of discovering the most persuasive node in a social network has proved to be NP-hard and also the exact algorithms cannot be designed. This creates a wide scope for developing approximation methods and algorithms that are able to produce solutions with proven approximation guarantees. Greedy algorithm serves as a base for most of the existing algorithms designed for dealing with these problems. Greedy algorithm can achieve a good approximation, but it is found to be computationally expensive. Therefore, in this paper we propose a two level approach, designed based on Suspected-Infected (SI) epidemic model for maximizing the influence spread. We further propose that, multithreading approach for implementation of algorithm for the proposed SI model aids to further elevate the performance of proposed algorithm in terms of influence spread per second.

1. Introduction

Viral marketing has been acknowledged as an effective marketing strategy. Eventually a large number of people get connected through social networks, such as Facebook, Flickr, and Twitter. The impact of social network on their lives has increased significantly. The social influence acts as a motivating force, governing the diffusion of the information in the network. Although there are millions of users on social platforms, the activities of a selected number of users are acknowledged and spread through the network. These dominant users generate trends and play a significant role to shape or manipulate opinions in social networks. These opinions are crucial in areas such as marketing or opinion mining. Many companies have started targeting the key individuals called influencers, who are in contextual alignment with their brand and operate for the companies indirectly for finding the potential customers. This is an indirect form of marketing also called influencer marketing.

Social media influencers are the entities in the social network, who help potential customers make a buying decision by influencing his opinion, through social networking. An influencer can be any person who reviews product, posts a blog about a new product, any industry expert or any person who has a potential to influence people. The problem of influencer identification can be presented as, given a group of individuals which are to be motivated to adopt a new product or information, find the optimum target subset of individuals (seed set), which can further influence the nodes. The ultimate goal is to maximize the spread the information to a large population.

Recently, there are large advances in the social networks field. It has focused on the study of relationships that includes quantitative measures of social networks like influence, authority, centrality, modularity, connectedness, etc. [1]. Influence maximization can be defined as the problem of forming an objective function for selecting appropriate target nodes in a social network such that it maximizes the influence spread. These target nodes in turn will propagate the influence to their connected nodes. This will be helpful to design marketing strategies or diffuse a new idea in a network related work in influencer detection in social networks.

* Corresponding author.
E-mail addresses: Jyotis8582@gmail.com (J.S. More), Chelpa.lingam@gmail.com (C. Lingam).
[1] Affiliated to Mumbai University.
Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.aci.2017.11.001

The major issue of concern here is: how to improve the diffusion for the given seed selection? Greedy algorithm proposed by Kemp et al. [2], is a significant algorithm for selecting the seed nodes. This is a NP-Hard problem. Greedy algorithms require optimal local choices. The solutions provided by greedy are suboptimal. Hence there is a need to find a solution to get a better algorithm which provides maximum influence spread. The problem of finding the minimal set of activated nodes to spread information to the whole network or to optimally immunize a network against epidemics can be exactly mapped onto optimal percolation. The most influential nodes are the ones forming the minimal set that guarantees a global connection of the network. At a general level, the optimal influence problem can be stated as follows: find the minimal set of nodes which, if removed, would break down the network into many disconnected pieces. The natural measure of influence is, therefore, the size of the largest (giant) connected component as the influencers are removed from the network. Hence an epidemic model, Susceptible-Infected (SI) is used. It is best fitted to deal with the progressive nature of the model in context.

2. Related work

Kemp et al. [2] proposed a greedy approach for finding K influential nodes out of all existing nodes. They provided for the first time an approximation guaranteed solution for the greedy algorithm. They proposed an analysis framework for finding the seed nodes. This framework is based on submodular functions. The framework also showed that a feasible solution can be obtained using a greedy strategy. They also proposed triggering model and showed that their proposed approximation algorithm worked better as compared to other known node selection strategies in social networks.

The performance of greedy algorithm for influence maximization can be improved by exploiting the submodularity, by an approach called Cost-Effective Lazy Forward selection (CELF) [3]. Eventually, an improvised algorithm called CELF++ [4] was proposed, which exploits the property of submodularity of the spread function.

On the other hand, the social network diffusion was also modelled [5] using various theories like bond percolation, resulting in the proposal of Susceptible Infected Recovered (SIR) model. Meanwhile, Graph evolution parameters, such as densification and shrinking diameters, were analysed [6] for modelling social networks. Based on global social network metrics, such as betweenness centrality and closeness centrality, a semi-local centrality measure was proposed to design an effective ranking method. This design along with the SIR epidemic model was used to evaluate the performance of the diffusion model by considering the parameters such as the rate of influence spread and the number of infected nodes i.e. influenced nodes [7,8]. Later a new heuristic and scalable solution based on maximum influence path was proposed [9].

Social networks were analysed for quantifying user influence and they dealt with web semantics to learn about influence in heterogeneous social networks [10–12]. Social influence was further exploited for the study of human dynamics and human behaviour [13–17]. Social networks are evaluated for influence maximization [18,19] and a two phase model for information diffusion was employed [20] selecting the seed nodes and further activating it in multiple levels. Recently, linear-time implementation of Collective-Influence (CI) algorithm is used to find the minimal set of influencers in networks via optimal percolation [21,22]. Further, scalable algorithms were proposed for massively large social networks [23].

The bond percolation theory and epidemic models are studied and utilization of the SI epidemic model [24] for modelling the diffusion in social networks is proposed. There is majority of work based on bond percolation and on SIR model. The SI model is preferred since SI model is a progressive model and hence can be better exploited in the influence maximization problem.

3. Problem discussion

Different models and frameworks have been defined by different researchers to obtain an optimal solution for the above stated problem. Some of the approaches are discussed in this section.

Social network can be interpreted as a directed graph G = (V, E) where V denotes the nodes in the graph, which represent the users in the social network and E denotes the edges, that represent the relationship between the users. In this context the relationship would be that of the influencer and influenced node i.e. who influences whom. The influence maximization problem deals with optimally selecting the seed set of users such that they contribute to maximize the expected spread of influence or diffusion in the given social network, in the context of a given propagation model.

Let $S_t \subseteq V$ be defined as the active set, containing the active nodes at given time t. The active nodes which participate in spreading the influence to the next level of influence will be termed as seed nodes. Let S_0 be the seed set, containing the seed nodes. In other words, the seed nodes in this set are called the seeds of influence diffusion. These seed nodes are the initial nodes at the root level, which are selected to propagate the influence throughout the network. For example, as a marketing strategy, the initial users selected by the promotional campaign of a new product, designed as marketing strategy.

In progressive diffusion models, the active sets are monotonically non-decreasing and hence the superset, V is finite, for a finite number of steps and the set of active nodes, S_t remains unchanged. Eventually, the active nodes belonging to the active set leads to the final active set and is denoted as $\phi(S_0)$, where S_0 is the initial seed set.

There exists two classic progressive models, originally proposed in the mathematical sociology, are described as follows:

3.1. Independent cascade model (IC model)

Independent cascade (or IC) model was the first progressive model [25,26]. The key characteristic of this model is that diffusion events associated with every edge in the given social graph are mutually independent.

3.2. Linear threshold model (LT model)

The linear threshold (or LT) model is a progressive, stochastic information diffusion model, proposed by Granovetter [27].

3.3. Triggering model

Kempe [2] proposed the triggering model, which is mainly based on two basic propagation models, already discussed above, namely, the IC and the LT models. In IC and LT propagation models, this influence maximization problem is proved to be NP-hard by Kempe et al. [2]. He also showed that the maximization function $\sigma_m(S)$ follows the properties of monotonicity and submodularity.

3.4. Submodularity and monotonicity of a function

The propagation models i.e. IC, LT and triggering models satisfy two important properties in terms of their influence spread function, σ. These properties are submodularity and monotonicity.

Submodularity can be interpreted in this context as diminishing marginal return i.e. when more nodes are added to the seed set, there is no great effect on the performance of the model. In this context, monotonicity can be interpreted as if more elements are added to a seed set, it cannot reduce the size of the final set containing the active nodes (influenced nodes).

3.5. Greedy algorithm proposed by Kempe et al. [2]

Input: G, k, $\sigma_m(S)$
Output: seed set S, θ
1. $S \leftarrow \phi$
2. while $|S| < k$ do
3. $u \leftarrow \text{argmax}_{w \in V-S} (\sigma_m(S + w) - \sigma_m(S))$;
4. $S \leftarrow S \cup \{u\}$

The line 3 of the greedy algorithm is most important. Here, it selects the node that provides the maximum influence spread, in other words, the largest marginal gain $\sigma_m(S + w) - \sigma_m(S)$ with respect to the total expected influence spread of the seed set in context i.e. S. This step helps to ensure the submodularity and monotonicity.

Greedy algorithms require the optimal local choices. Greedy algorithm works only if locally optimal choices have potential to lead to a global optimum and the sub problems are optimal. But if it fails, then the greedy algorithm performs poorly. The same thing is observed here. The greedy algorithm only finds local minimum edge at every iteration and hence it fails to reach more nodes. There could be a possibility that in a large perception, the local optima is far weaker than global optima. Hence we tried to exploit the graphical structure of the graph.

4. Proposed model

This maximization problem can be expressed as a discrete optimization problem. It can be modelled as a graphical model for learning tree distribution. A discrete approach aims to choose the optimal set of nodes that constitute an optimal path in a spanning tree, emerging out of the seed node. In other words, finding a spanning tree of social graph G of best fit for the triggering nodes (seed nodes), such that when the nodes are traced along the given path length (also termed as threshold), it provides a subset of the solution i.e. subset of final active set.

4.1. SI epidemic model – (Susceptible infected model)

The SI model [24,28], categorizes the entire population in the context into two groups, namely, the susceptible individuals who may get infected by the given disease i.e. who are likely to get infected and the other group is that of the infected individuals, who get infected by the disease and further may carry or spread the disease to the next set of individuals i.e. susceptible group. Once a susceptible entity becomes infected, he or she gets added into the infected set, thereby increasing the size of the infected set and ultimately decreasing the size of the susceptible set of individuals. We utilize this characteristic of the epidemic model to model the influence spread across a social network. Fig. 1 shows the proposed two phase SI model for the influence maximization problem.

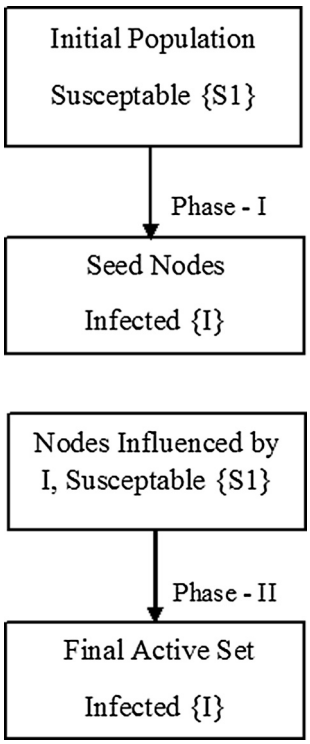

Fig. 1. SI-based two phase model.

4.2. Proposed method

Fig. 2 represents the proposed method. Here, in phase I, we propose that the initial population of the nodes will be the candidates i.e. susceptible nodes, represented as {S1}. The nodes which get triggered i.e. the active nodes will now act as seed set and will be categorized as infected nodes {I}. These infected nodes will now serve as the influence carriers. In phase II, the nodes other than the seed nodes are all susceptible. Once these susceptible nodes are influenced by the seed nodes, they become active nodes. The nodes once active do not become inactive. This is the progressive behaviour as stated earlier. Hence it fits in the framework of SI model.

Let G be a graph, with initial population assumed as susceptible for spread. Let {S} denote the set of seed nodes obtained from the greedy algorithm, {S1} denote the set of susceptible nodes and {I} denote the set of infected nodes, i.e., the nodes responsible to spread the influence. μ represents the threshold, considered as eccentricity, i.e., the maximum distance (using the spanning tree) from given node v to any other node in the graph. It is the diameter of sub graph, used to find diffusion for one infected node ∈ {I}. This is how we actually compute the reachability of the nodes. We assume that the nodes which are reachable are more likely to get infected. Hence, it is a significant parameter in the process of influence spread algorithm. The set {S1} represents graphically all the reachable nodes which are at a distance ≤ μ. The algorithm can be stated as follows:

4.2.1. Algorithm

Algorithm SI_Influence_Spread (G, S, S1, θ, I, Ψ)

Input: S, μ
1. Initialize set $S \leftarrow \phi$
2. Data preprocessing- Build the social network graph.
3. Assume initial population as susceptible (S) and identify the seed set from greedy algorithm depending on threshold θ
4. Phase I: Identified nodes become infected i.e. $\{I\} \leftarrow \{S\}$
5. Phase II: For Each node $v \in \{I\}$, find the set $\{S1\}_{vi} \subseteq \{S1\}$ such that $v_i \in \{S1\}$ only if vi is at distance ≤ μ

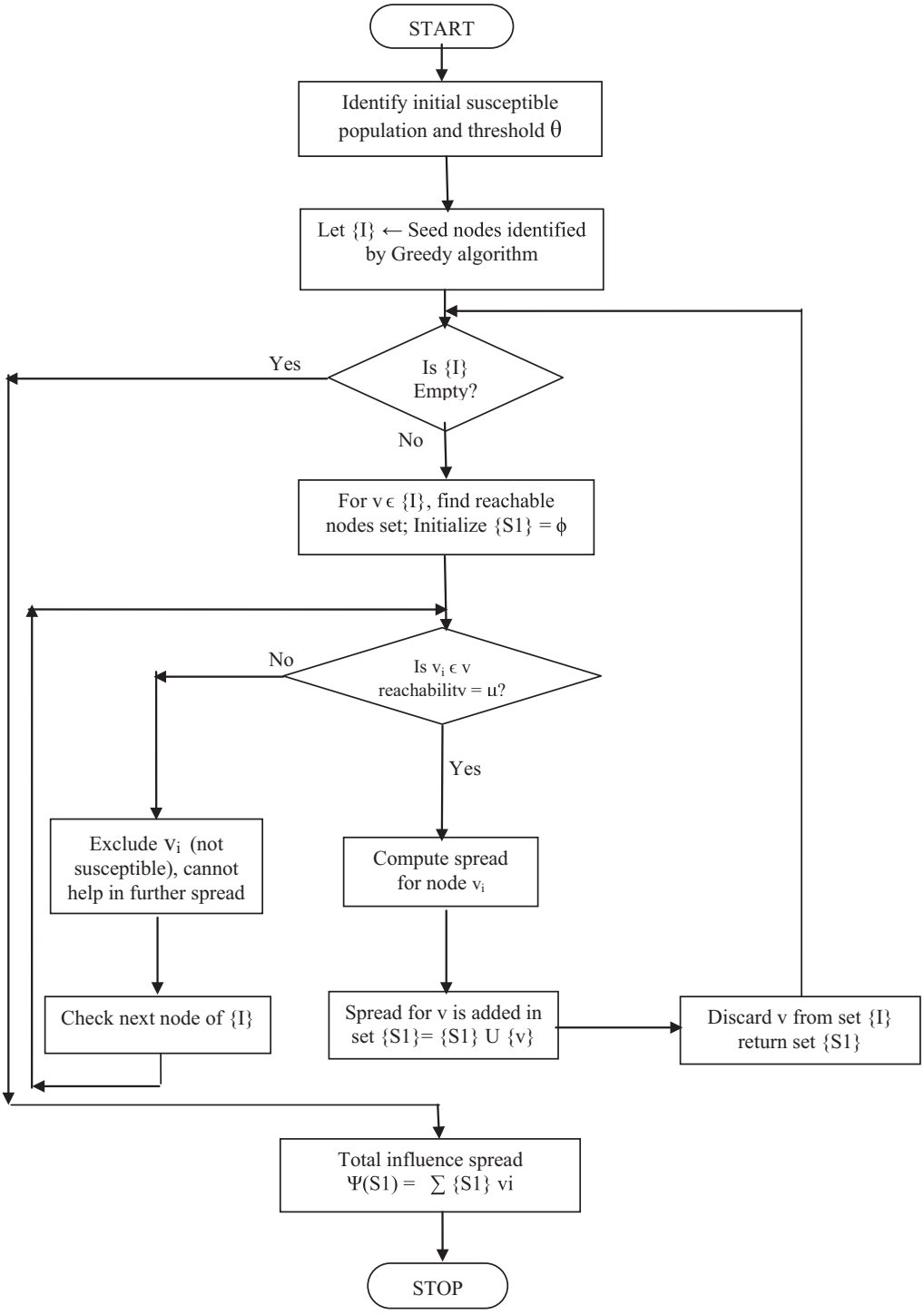

Fig. 2. Flowchart for the proposed SI model.

6. Total_influence_spread, $\Psi(S1) = \sum \{S1\}_{vi}$
7. return $\Psi(S1)$
8. end For

The shortest path is traversed using spanning tree which makes sure that the vertex with maximum influence spread is passed to the next iteration. This leads to an incremental influence spread. The nodes returned by the above algorithm represents the set of nodes influenced by the source node. It is observed that the spread is wider than the greedy algorithm.

Theorem 1. *In the SI model, the number of susceptible individuals decreases monotonically, that is* $S_{n+1} \leq S_n$, *for all n. We also have that the number of infected individuals increases monotonically, i.e.,* $I_{n+1} \geq I_n$, *for all n* [28].

Proof. The input for the algorithm is a set of infected nodes {I} which are derived from the population by using the greedy algorithm. It is clear that the target seed set obtained from the greedy algorithm is submodular as well as monotonous [2].

Table 1
Features of the datasets.

Datasets	CA-AstroPh	Cit-HepTh	Cit-HepPh	Soc-Eopinions
Nodes	18772	27770	34546	75879
Edges	198110	352807	421578	508837
Average clustering coefficient	0.6306	0.3120	0.2848	0.1378
Diameter	14	13	12	14

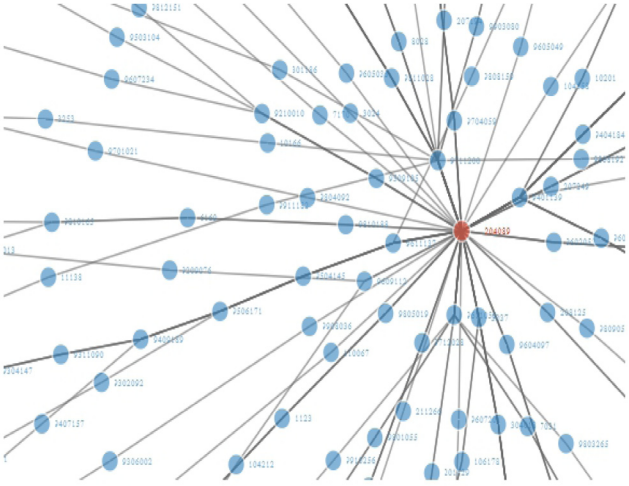

Fig. 3. Sample hybrid algorithm for μ = 7 (664 nodes) for seed node 204089 for dataset cit-HepTh (visualization using R).

Hence, if we are using incremental approach to find the influence spread, then the monotonicity is reserved in this approach. Further, experimentally, we have proved that the number of nodes influenced by a seed set at earlier step n is more than the next step, i.e., n + 1.

Theorem 2. *The number of susceptible individuals is never negative,* $S_n \geq 0$, *and the number of infected individuals is never more than the total population size,* $I_n \leq N$ [28].

Proof. In our approach, the initial population, i.e., a set of infected nodes is never empty. Hence, even for minimum one seed node, it is not possible to have susceptible individuals < 0. As we follow graphical model, we know that at least two nodes and one edge will be required. Hence, if one node is selected as the seed node, i.e., infected node, then the other is susceptible (as it is reachable from infected node). Therefore, the set of susceptible individuals can never be negative. In the worst case, where all the nodes are infected, the susceptible node can be 0 but not negative.

On the other hand, the seed nodes, i.e., infected node set {I} can contain the entire population in the best case. The greedy algorithm terminates when the seed node set contains all the nodes of the entire population N. In this case, the seed set becomes universal set. Hence, even if any infected node gets added later, it will be a subset of universal set and according to set theory, |U| = N, i.e., the total number of nodes. Hence, the number of infected nodes can be a maximum of N.

The computational time can be reduced substantially. Once we get the target nodes using the greedy algorithm, we can simultaneously execute the algorithm on all the seeds. This is the reason why the computational time gets reduced. For achieving further improvement in time efficiency, we propose to use the Multithreading approach.

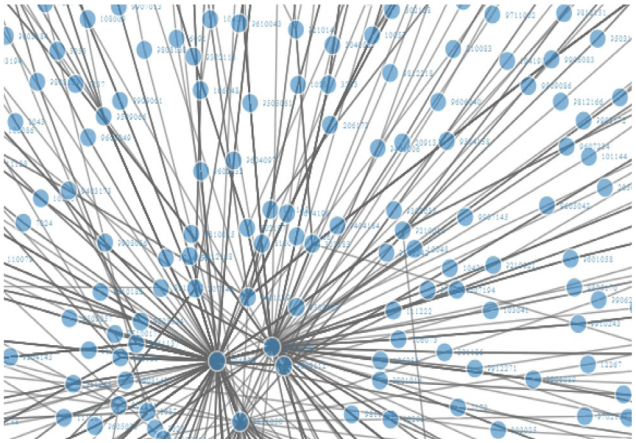

Fig. 4. Sample hybrid algorithm for μ = 6 (2650) for seed node 204089 for dataset cit-HepTh (visualization using R).

5. Dataset description

We consider three datasets, available on Stanford Large Network Dataset Collection (SNAP), published by Stanford University (Available at: http://snap.stanford.edu/data/com). Table 1 represents the features of datasets.

6. Working of algorithm

The algorithm will work as follows:
Step 1:

Input: Set of Susceptable nodes
After Processing using greedy Algorithm: Susceptible nodes → Infected nodes
P → {v1, v2, v3, v4, v5, v6, v7} → seed sets for influence spread (assuming threshold as 7)

Step 2: Input: Set of Susceptible nodes (Those found as infected in previous step) i.e. {v1, v2, v3, v4, v5, v6, v7}

After modelling: Infected nodes

{v1}- {v11, v12. . ., v1n} assuming the threshold as μ
e.g. for dataset citHepTh, the infected nodes for a seed node 204089, for (μ = 7), it influences 664 nodes (shown in Fig. 3).

Figs. 3 and 4 illustrates the visualization of the influence spread for a node. The central hub (node) is the seed node, selected during phase I whereas other nodes are the infected nodes during phase II. The influence is spread by the set of seed nodes. The set of all active nodes at this phase represents the final active set. It clearly shows the effect of threshold on the total spread. As the threshold μ, increases, the total spread shrinks.

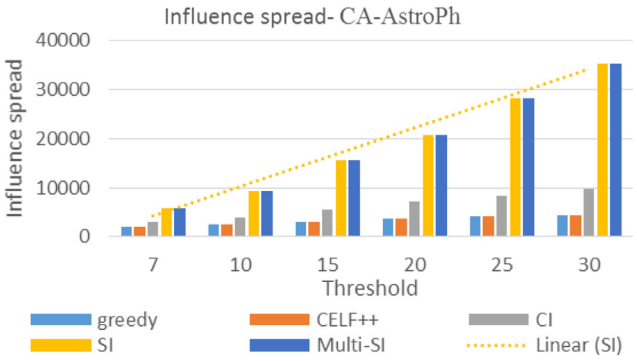

Fig. 5a. Influence spread for CA-AstroPh Dataset.

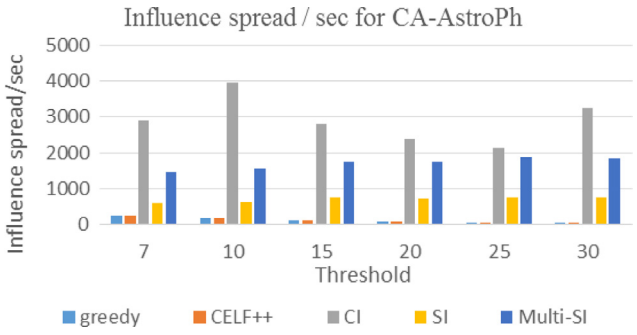

Fig. 5b. Influence spread per second for CA-AstroPh Dataset.

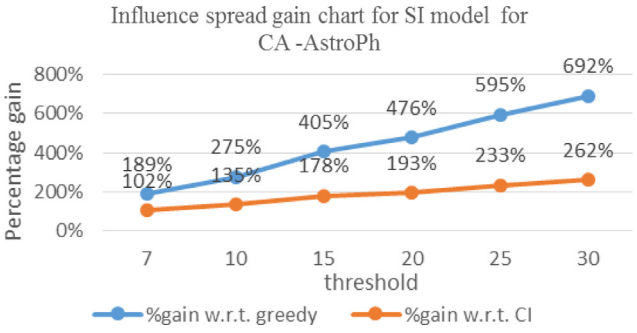

Fig. 5c. Influence spread gain chart for CA-AstroPh Dataset.

7. Results discussion

Initially, the seed set is computed using the greedy algorithm. To find the influence spread, the next cascade is found by considering that an edge connecting two nodes mean that one is influencing the other. Further, by considering the longest among shortest paths, the given nodes influenced (which is computed by cascaded operation) will be restricted by specified threshold (which specifies the path length). The SI model is based on the incremental approach where the spread is cumulative. It exploits the graph properties. The spanning tree enables the SI approach to find the best possible longest path which helps in increasing the influence spread. The greedy approach is restricted to the local search. CELF++ is based on the submodularity imposed on greedy algorithm and CI is based on the adaptive bottom up approach utilizing the finite radius of sphere of social networks. The performances of influence spread with different values of thresholds θ ($\mu = 6$ for SI model) are illustrated in Fig. 5a. Linear (SI) represents the trend of the performance of algorithm as the threshold varies. The performance elevates linearly with increase in the threshold.

Fig. 5b show the influence spread per second for different algorithms for the dataset CA-AstroPh Dataset. It is clear that the influence spread is elevated to a large extent by using SI model as compared to greedy algorithm. The influence spread per second shows the significant outperformance when a multithreaded approach is used. CI gives best influence spread per second but the total influence spread is best achieved by SI model as shown in following Fig. 5c. In SI based algorithm, spanning tree data structure is used where in, while pre-processing, the reachability of the nodes is checked for the seed nodes which are shortlisted by the greedy algorithm. This enables us to obtain the infected nodes. However, the time complexity of the CI-algorithm is better than the proposed algorithm because of use of max-heap data structure for storing and processing the CI values. The finite radius ℓ of the CI sphere, allows to process the CI values in a max-heap data structure [22,23]. The basic idea is that, after each node removal, there is a need to recompute CI just for a O(1) number of nodes, and find the new largest value. It follows bottom up approach. Whereas, in SI based approach, the computations are based on computations of the longest possible path among all the shortest paths, which is an incremental approach. Here, the execution time of CI is better than SI model, but still fails to achieve the influence spread as good as SI model. We further propose to use multithreaded approach to obtain a better performance of SI model in terms of execution time to some extent.

The comparison of performance gain for influence spread of different algorithms for dataset CA-AstroPh are as depicted in Fig. 5c. Table 2 shows the comparison in terms of performance i.e., influence spread of proposed SI model with other models for different datasets. Table 3 depicts the performance gain (Influence spread per second) by using multithreaded approach for SI model. In Table 3, the performance of multithreaded approach for SI model is compared with the basic SI model. It is proved that by multithreading though doesn't provide the best influence spread per second, still manages to give a significant boost to the speed of the influence spread. This performance improvement is reported in Table 3.

Table 2
Performance comparison (Influence Spread) of SI model with other models ($\mu = 6$).

Datasets→	Cit-HepTh		CA-AstroPh		Cit-HepPh		Soc-Eopinions	
Seed set size	%gain w.r.t. greedy	%gain w.r.t. CI	%gain w.r.t. greedy	%gain w.r.t. CI	%gain w.r.t. greedy	%gain w.r.t. CI	%gain w.r.t. greedy	%gain w.r.t. CI
7	685%	552%	189%	102%	444%	405%	56%	20%
10	966%	757%	275%	135%	553%	499%	107%	48%
15	1246%	922%	405%	178%	698%	581%	172%	79%
20	1716%	1222%	476%	193%	859%	686%	285%	143%
25	1880%	1283%	595%	233%	963%	735%	355%	179%
30	2257%	1489%	692%	262%	1039%	764%	415%	200%

Table 3
Performance gain (Influence spread per second) by using multithreaded approach for SI model.

Datasets→	CitHepTh	CA-AstroPh	Cit-HepPh	Soc-Eopinions
7	123%	150%	71%	129%
10	133%	150%	48%	105%
15	127%	134%	48%	114%
20	154%	142%	51%	104%
25	141%	147%	84%	116%
30	145%	142%	82%	185%

8. Conclusion and future scope

As discussed in the earlier part, greedy algorithms require optimal local choices at each stage with the hope of finding a global optimum. If locally optimal choices yield a global optimum and the sub-problems are optimal, then the algorithm works. If it fails, then the greedy algorithm performs poorly. This has been confirmed by this study too. The greedy algorithm only finds local minimum influence spread at every iteration, hence it fails to reach more nodes. It is observed that the influence spread observed in the greedy algorithm is limited and generally requires more run time. The proposed two phase SI based algorithm performs better than greedy algorithm in terms of time and the overall influence spread. Hence, we show that the graphical structure of the social network can be exploited to improve the reachability and hence improving the influence spread.

Here, a novel approach is proposed based on SI epidemic model for influence spread, the longest shortest path concept for reachability and implementation of multithreading for improving the time efficiency which iteratively improves the greedy cascaded model exponentially. The influence spread in this model is maximized as compared to the basic greedy model. The efficiency in terms of speed is an added benefit. In this study, we evaluated the algorithm for different seed sizes with different datasets against different approaches proposed earlier. We observed that our ultimate aim of maximizing the influence spread is achieved using SI Model, but at the cost of execution time. Hence we used multithreading to improve the total number of nodes influenced per second, i.e., indirectly decreasing the computational time.

This work provided an overview of the influencer identification and the influence maximization. This study concludes that by identifying the influential users in social media, different business strategies can be planned, e.g., efficient launching and marketing new products, targeting the potential consumers, etc. It is obvious that the influence maximization and social influence mining together will form the significant components to enable extensive viral marketing through online social networks.

Identifying influential users may be proposed through different models, algorithms and statistical techniques. Also parallel problems like link prediction, social network content analysis, etc. could be considered as potential problems for social network mining to deal with in future.

References

[1] Jiawei Han, Micheline Kamber, Jian Pei, Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, 2011.
[2] David Kemp, Jon Kleinberg, Eva Tardos, Maximizing the spread of influence through a social network, in: KDD 03, USA, 2003, pp. 137–146.
[3] Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen and Natalie Glance, Cost-effective Outbreak Detection in Networks, in: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD, 2007, pp. 420–429.
[4] Amit Goyal, Wei Lu, Laks V.S. Laxmanan, Celf++: optimizing the greedy algorithm for influence maximization in social networks, in: ACM Proceeding, 2011, pp. 47–48.
[5] Masahiro Kimura, Kazumi Saito, Ryohei Nakano, Hiroshi Motoda, Extracting influential nodes on a social network for information diffusion, Data Min Knowl Disc, Springer, 2009, pp. 70–97.
[6] Jure Leskovec, Jon Kleinberg, Christos Faloutsos, Graph evolution: densification and shrinking diameters, ACM Trans. Knowl. Disc. Data 1 (1) (2007).
[7] Duanbing Chen, LInyuan Lu, Ming Sheng Shang, Yi-Cheng Zhang, Tao Zhou, Identifying influential nodes in complex networks, Phys. A: Stat. Mech. Appl. (2011) 47–55.
[8] Christine Kiss, Martin Bichler, Identification of influencers-measuring influence in customer networks, Decis. Supp. Syst. (2008) 233–253.
[9] Chi Wang, Wei Chan, Yajun Wang, Scalable influence maximization for independent cascade model in large-scale social networks, Data Mining and Knowledge Discovery, Springer, 2012, pp. 1029–1038.
[10] Eytan Bakshy, Brian Karrer, Lada Adamic, Social influence and the diffusion of user-created content, in: Proceedings of the 10th ACM Conference on Electronic commerce, 2009, pp. 325–334.
[11] Eytan Bakshy, Jake M. Hofman, Winter A. Mason, Duncan J. Watts, Everyone's an influencer: quantifying influence on twitter, in: WSDM Proceedings of Fourth ACM International Conference on Web Search and Data Mining, 2011, pp. 65–74.
[12] Lei Tang, Huan Liu, Leveraging social media networks for classification, Data Mining and Knowledge Discovery, Springer, 2011, pp. 447-478.
[13] Lu Liu, Jie Tang, Jiawei Han, Shiqiang Yang, Learning influence from heterogeneous social networks, Data Mining and Knowledge Discovery, Springer, 2012, pp. 511–544.
[14] Bogart Yail Mrquez, Manuel Castaon-Puga, Juan R. Castro, Eugenio D. Suarez, Jos Sergio Magdaleno-Palencia, Fuzzy models applied to complex social systems: modeling poverty using distributed agencies, Int. J. New Comput. Arch. Appl. (IJNCAA) 1 (2) (2011) 292–303.
[15] Na Li, Denis Gillet, Identifying influential scholars in academic social media platforms, in: ASONAM Proceedings IEEE/ACM International Conference on Advances in Social Network Analysis and Mining, 2013, pp. 608–614.
[16] Wei Pan, Wen Dong, Manue Cebrian, Taemie Kim, James H. Fowler, Alex SandyPentland, Modeling dynamical influence in human interaction, ACM Trans. Web (ACM TWEB) (2012) 77–86.
[17] Symeon Papadopoulos, Yiannis Kompatsiaris, Athena Vakali, Ploutarchos Spyridonos, Community detection in social media-performance and application considerations, Data Mining and Knowledge Discovery, Springer, 2012, pp. 515–554.
[18] Kundu Suman, C.A. Murthy, S.K. Pal, A new centrality measure for influence maximization in social networks, in: 4th International Conference on Pattern Recognition and Machine Intelligence (PReMI11), Springer-Verlag, 2011, pp. 242–247.
[19] Sankar K. Pal, S.C.A. Murthy, Centrality measures, upper bound, and influence maximization in large scale directed social networks, Fundam. Inform. (2014) 317–342.
[20] Swapnil Dhamal, K.J. Prabuchandran, Y. Narahari, Information diffusion in social networks in two phases, IEEE Trans. Network Sci. Eng. (2016) 197–210.
[21] F. Morone, H. Makse, Influence maximization in complex networks through optimal percolation, Nature 524 (2015) 65–68.
[22] Flaviano Morone, Byungjoon Min, Lin Bo, Romain Mari, Hernan A. Makse, Collective Influence Algorithm to find influencers via optimal percolation in massively large social media, Sci. Rep. 6 (2016) 30062.
[23] Sen Pei, Xian Teng, Jeffrey Shaman, Flaviano Morone, Hernan A. Makse, Efficient Collective Influence maximization in cascading processes with first-order transitions, Sci. Rep. 7 (2017) 45240.
[24] Linda J.S. Allen, Some discrete-time SI, SIR, and SIS epidemic models, Math. Biosci. 124 (1) (1994) 83–105.
[25] Jacob Goldenberg, B. Libai, E. Muller, Talk of the network: a complex systems look at the underlying process of word-of-mouth, Market. Lett. (2001) 211–223.
[26] Jacob Goldenberg, B. Libai, E. Muller, Using complex systems analysis to advance marketing theory development: modeling heterogeneity effects on new product growth through stochastic cellular automata, Acad. Market. Sci. Rev. (2001) 1–18.
[27] M. Granovetter, Threshold models of collective behavior, Am. J. Sociol. (1978) 1420–1443.
[28] Kacie M. Sutton, Discretizing the SI epidemic model, Rose-Hulman Undergraduate Math. J. 15 (2014) 192–208.