



# Maximal Clique Based Influence Maximization in Networks

Nizar Mhadhbi<sup>1</sup>(✉) and Badran Raddaoui<sup>2</sup>(✉)

<sup>1</sup> CRIL - CNRS UMR 8188, University of Artois, Lens, France  
mhadhbi.nizar@gmail.com

<sup>2</sup> SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, Évry, France  
badran.raddaoui@telecom-sudparis.eu

**Abstract.** Influence maximization is a fundamental problem in several real life applications such as viral marketing, recommendation system, collaboration and social networks. Maximizing influence spreading in a given network aims to find the initially active vertex set of size  $k$  called seed nodes (or initial spreaders (In this paper, we use seed set and initial spreaders interchangeably.)) which maximizes the expected number of the infected vertices. The state-of-the-art local-based techniques developed to solve this problem are based on local structure information such as degree centrality, nodes clustering coefficient, and others utilize the whole network structure, such as  $k$ -core decomposition, and node betweenness. In this paper, we aim at solving the problem of influence maximization using maximal clique problem. Our intuition is based on the fact that the presence of a dense neighborhood around a node is fundamental to the maximization of influence. Our approach follows the following three steps: (1) discovering all the maximal cliques from the complex network; (2) filtering the set of maximal cliques; we then denote the vertices belonging to the rest of maximal cliques as superordinate vertices, and (3) ranking the superordinate nodes according to some indicators. We evaluate the proposed framework empirically against several high-performing methods on a number of real-life datasets. The experimental results show that our algorithms outperform existing state-of-the-art methods in finding the best initial spreaders in networks.

**Keywords:** Influence maximization · Maximal clique · Independent cascade model

## 1 Introduction

Influence maximization problem in social networks has become a hot topic in recent years due to the great deal of real-life applications concerned, such as viral marketing and disease spreading. One of its application in the field of viral marketing is to select a set of highly influential users to adopt a particular product and the goal here is to attract as much as possible of users for purchasing this product [4, 6, 10]. In order to model the process of the spread of an idea or

an information through a given network, Kempe et al. [15] proposed two models named **Independent cascade model (ICM)** and **Linear threshold model (LTM)**. The independent cascade model is the most common model for information diffusion. In this model (Algorithm 1), the input are a graph in which every edge  $(u, v)$  is associated with a propagation probability  $p_{uv}$  (represents the probability that node  $v$  can be influenced by node  $u$ ) and a set of initially activate nodes (seed set) and the diffusion probability called also activation threshold  $\theta$ . In this model, nodes can have two states, either active or inactive. Nodes are allowed to switch from inactive to active but not in the other. As shown in Algorithm 1, the diffusion model starts with an initial set of active nodes ( $t = 0$ ). In time  $t$ , an active node  $u$  will get chance to activate its inactive neighbor  $v$ .  $v$  will become active if  $p_{uv} \geq \theta$ , otherwise  $u$  will not get any further chance to activate  $v$ . The process of diffusion stops when no further activation is possible. This method is called independent because the activation of a node does not depend on the history of active nodes. In the linear threshold model, the idea is that a node becomes active if a large part of its neighbours is active. More formally, each node  $u$  has a threshold  $t_u$ . The threshold represents the fraction of neighbours of  $u$  that must become active in order to active the node  $u$ . Influence maximization problem is the problem of assigning a subset of  $k$  users as seed nodes in a graph that could maximize the spread of influence by maximizing the expected number of influenced users.

## 2 Related Works

Choosing the best  $k$  initially active nodes in order to maximize the number of activated nodes at the end of the diffusion process had made a prominent place in several works. Numerous techniques have been developed for both efficient and effective influence maximization. Related works can be classified in to four categories: *local-based approaches*, *global-based approaches*, *community-based approaches*, and *approximation-based approaches*. Algorithms of the local-based category use the local information of the network in order to select the best  $k$  influential nodes. The first solution proposed in this category is to select nodes with higher number of neighbors. That is, select the nodes based on their degree scores. Domingos and Richardson [10,17] were the first to study this as an algorithmic problem. Algorithms in the global-based approaches exploit the information of the whole network. In this category, a plethora of centrality measures such as betweenness centrality [11,12], M-centrality [13] and coreness centrality [16] are proposed in order to rank nodes according to their topological importance in the network. M-Centrality measure combines the information on the position of the node in the network with the local information on its nearest neighborhood. The position is measured by the  $K$ -shell decomposition, and the degree variation in the neighborhood of the node quantifies the influence of the local context. Coreness is a well-established centrality index that focuses on the structure of networks. Authors in [16] found that the most efficient spreaders are those located within the core of the network as identified by the  $k$ -shell

decomposition analysis. Betweenness centrality identifies key nodes in a network called bridges. A bridge is a node that has short paths to other nodes in the network. Despite the efficiency of the local-based and the global-based approaches in terms of time which is very fast, these approaches may result in less influence over the network. Studies in [3] showed that the degree-based and centrality-based approaches may result in less influence over the network. The reason behind it might be, these measures do not consider the effect of neighborhood. Indeed, a given group of connected nodes may have a high degree or a high centrality score, but if their adjacent nodes are overlapped then the information may not propagate through the rest of the network. Several approaches are proposed to deal with the problem of neighborhood overlapping. In [7], a faster method that considers the neighbors of each node is required in order to avoid overlapping. Algorithms of the category of community-based use the communities in the network as an intermediate step to select the most influential nodes [8]. Authors of [8] improve the efficiency of influence maximization by incorporating information on the community structure of the network into the optimization process. They detect the community structure of the input network using the concept of (maximal) cliques problem. Algorithms in the category of approximation-based give the worst case bound for influence spread [15]. However, most of them suffer from the scalability issues, which means, with the increase of the network size, running time grows heavily.

---

**Algorithm 1:** Independent Cascade Model

---

**Input:** A network  $G = (V, E)$  and  $A_0 = \{v_0, \dots, v_k\}$  denote the set of initially activated nodes;  $\theta$  denote the activation threshold

**Output:** A set of final activated nodes

```

1 while  $A_i \neq \emptyset$  do
2      $A_i \leftarrow$  Newly activated nodes;
3     for  $u \in A_i$  do
4         for  $v \in N_u$  do
5             if  $\theta \leq p_{uv}$  then  $A_{i+1} \leftarrow A_{i+1} \cup \{v\}$ ;
6 return  $A_0 \cup A_1 \dots \cup A_i$ 

```

---

In this paper, by utilizing maximal clique problem, we propose *IMSN (Influence Maximization using Superordinate Nodes)*, which is a novel algorithm for influence maximization in large networks. *IMSN* is based on superordinate nodes to look for the initial vertex set which maximizes the expected number of the infected vertices in the independent cascade model. *IMSN* starts by discovering all the maximal cliques from the complex network represented as a graph. We then denote the vertices belonging to the set of maximal cliques with size greater than or equal to  $\alpha$  as superordinate vertices. As a next step, we propose two indicators to rank influential individuals in the networks. We then simulate the information spread using the complete random simulation used in [15]. We

also compare the simulation results of our *IMSN* algorithm against two popular algorithms for influence maximization problem.

### 3 Background

Formally speaking, a complex network is generally abstracted as a graph with entities as the vertex set and the relationships (co-authorship, friendship, etc.) between them as the edge set. Graphs discussed in this paper are simple and undirected. Formally, an undirected graph is defined as a pair  $G = (V, E)$  where  $V$  is a set of nodes and  $E \subseteq V \times V$  is a set of edges. We denote by  $n$  (respectively  $m$ ) the number of nodes (respectively edges) in  $G$ . For a node  $u \in V$ , we denote by  $N_u$  the set of neighbors of  $u$ , i.e.,  $N_u = \{v \in V : (u, v) \in E\}$ . The *degree* of a node  $u \in V$ , denoted by  $d_u$ , is equal to  $|N_u|$ . The concept of maximal cliques is defined as follows:

**Definition 1 (Clique, Maximal Clique).** *Let  $G = (V, E)$  be an undirected graph. Then, a clique of  $G$  is a subset of nodes  $C \subseteq V$  such that whatever  $v_1$  and  $v_2$  belong to  $C$ , then the edge  $(v_1, v_2)$  belongs to  $E$ . A clique  $C$  of  $G$  is said maximal if for any  $x \in V \setminus C$ ,  $C \cup \{x\}$  is not a clique. The set of all maximal cliques of  $G$  will be denoted by  $\mathcal{C}$ .*

A clique represents a densely connected structure in the graph, as such it can be used to recover the locally most related elements, useful for several data mining tasks such as clustering, frequent patterns and community mining [14].

Given a network  $G$ , an integer  $k$  and an activation threshold  $\theta$ , in the problem of influence maximization we are looking for the initial active vertex set of size  $k$  which maximizes the expected number of the infected vertices. As shown by Kempe et al. in [15], the optimization problem is NP-complete. A plethora of work has emerged in order to solve the problem of influence maximization, for which we refer the reader to existing surveys [3].

### 4 Influence Maximization Based on Maximal Cliques

Maximal cliques are widely used in several real life applications. For instance, in anomaly detection, signals of rare events are defined as a set of large maximal cliques [5]. In data visualisation, maximal cliques are used to visualize a large graph where the cliques are grouped together in the display. In community detection problem, also known as graph clustering, a rigorous way to model communities is to consider maximal cliques, that is, maximal (with respect to set inclusion) subgraphs in which any pair of nodes is connected by an edge. The objective of this paper is to develop a solution for influence maximization in real networks based on maximal clique problem. To illustrate our method for influence maximization problem, we build our intuition from the following simple but very relevant principle: *a node can be a good infector in a network if multiple maximal cliques contain it.* This intuition is based on the fact that the

presence of a dense neighborhood around a node is fundamental to the maximization of influence, because in this way the node can spread the information between dense regions of the network. The proposed *SNIM (Superordinate Nodes for Influence Maximization)* follows three phases (Fig. 1). The first phase detects the maximal cliques of the input network, where a maximal clique is a clique that cannot be extended by including one more adjacent vertex. Such maximal cliques are really the ideal communities structures, that one would like to find. In complex networks, a community structure is a subset of individuals who interact with each other more frequently than other individuals outside the community. In the second phase, maximal cliques with size smaller than a fixed threshold  $\alpha$  are removed. This simple tactic may also find the most largest maximal cliques. We then denote the vertices belonging to the rest of maximal cliques with size greater than or equal to  $\alpha$  as superordinate vertices. The main idea here is to select from each set of intersecting maximal cliques the most  $k$  influential nodes. To do this, we use some indicators that can find a group of nodes of size  $k$  that by acting all together maximize the expected number of influenced nodes at the end of the spreading process, formally called *Influence Maximization*.

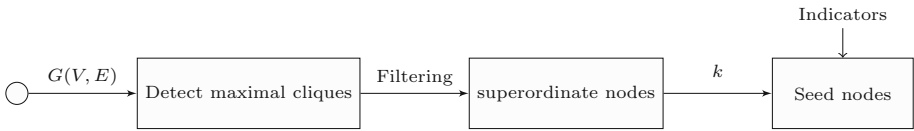


Fig. 1. Overview of SNIM algorithm

Let us now introduce two simple but very relevant indicators which scores the superordinate nodes based on their connectivity to other nodes. The first indicator called superordinate vertex frequency is defined as follows:

$$\mathcal{F}(u) = \sum_{c \in C} \sigma_{u,c} \tag{1}$$

where the sum is over maximal cliques  $C$  obtained after the procedure of filtering,  $\sigma_{u,c} = 1$  if  $u \in c$  and 0 otherwise. This first indicator quantifies the ability of a node to connect different maximal cliques. Then, a high score of frequency of a given node is obtained if the node belongs to many maximal cliques. Let us now denote the set of intersecting maximal cliques on a given node  $u$  by  $I_u$ , the second indicator quantify the set of vertices that could be influenced directly by a given node, i.e, the number of nodes in the union of maximal cliques containing the node  $u$ . More formally, this indicator is defined as follows:

$$\mathcal{W}(u) = \left| \bigcup_{c \in I_u} c \right| \tag{2}$$

where the union is over maximal cliques containing the node  $u$ . It is clear in this second indicator that the more nodes in the union of the set of intersecting maximal cliques on node  $u$ , the higher value of  $\mathcal{W}(u)$  is obtained.

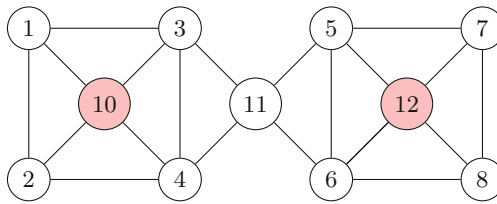
Next, we show how to select the seed nodes using the two indicators (1) and (2). Let  $f$  be a function which assigns a number to each node in the graph. Such number is computed using the indicators (1) and (2) as follows:

$$f: \left\{ v \mapsto \mathcal{F}(v) * \mathcal{W}(v) \right.$$

The function  $f$  scores the superordinate nodes based on the principle that a node can be a good infector if it is a member of many maximal cliques and has a strong connection with the other nodes. The nodes with large value  $f$  can work as an infection bridge between different maximal cliques, since in real life a person or a company can be a good infector if it appears in many different areas of life [1].

*Example 1.* Let us consider the undirected network depicted in Fig. 2. Using the two indicators  $\mathcal{F}$  and  $\mathcal{W}$ , for each node we have:  $\{\mathcal{F}(1) = \mathcal{F}(2) = \mathcal{F}(7) = \mathcal{F}(8) = \mathcal{F}(11) = 2; \mathcal{F}(3) = \mathcal{F}(4) = \mathcal{F}(5) = \mathcal{F}(6) = 3; \mathcal{F}(10) = \mathcal{F}(12) = 4\}$  and  $\{\mathcal{W}(1) = \mathcal{W}(2) = \mathcal{W}(7) = \mathcal{W}(8) = 4; \mathcal{W}(3) = \mathcal{W}(4) = \mathcal{W}(5) = \mathcal{W}(6) = \mathcal{W}(10) = \mathcal{W}(11) = \mathcal{W}(12) = 5\}$ .

As a result, the two nodes 10 and 12 have the best value of  $f$ .



**Fig. 2.** Example of undirected graph

### 4.1 Algorithm

Algorithm 2 describes the general feature of our superordinate nodes based procedure to determine the initial spreaders in the graph. It proceeds as follows: first the set of maximal cliques are identified. Then, a procedure of filtering is done using a parameter  $\alpha$ . As a next step, we rank all the superordinates nodes following the principle using the function  $f$ .

After that only  $k$  nodes are selected to be initial spreaders. The question now is how to select the  $k$  initial spreaders among all the superordinates nodes? The easiest solution would probably be a selection of the top  $k$  nodes with the

**Algorithm 2:** Superordinate Nodes for Influence Maximization

---

**Input:** A network  $\mathcal{N} = (V, E)$ ,  $\alpha, k > 1$   
**Output:** A set of initial spreaders

```

1  $C = \text{maximalCliques}(\mathcal{N});$ 
2  $\mathcal{NS} = \text{superordinatesNodes}(C, \alpha);$ 
3 for  $u \in \mathcal{NS}$  do
4      $\text{Compute}(\mathcal{F}(u));$ 
5      $\text{Compute}(\mathcal{W}(u));$ 
6      $f(u) = \mathcal{F}(u) * \mathcal{W}(u);$ 
7  $\mathcal{S} \leftarrow \emptyset;$ 
   // Top  $k$  not connected nodes
8 while  $|\mathcal{S}| \leq k - 1$  do
9     Let  $v$  the node with
10    the maximum score of  $f$ ;
11    if  $\forall u \in \mathcal{S} (u, v) \notin E$  then
12         $\mathcal{S} \leftarrow \mathcal{S} \cup \{v\};$ 
13         $\mathcal{NS} \leftarrow \mathcal{NS} \setminus \{v\};$ 
   // Link-discount
14 while  $|\mathcal{S}| \leq k - 1$  do
15     Let  $v$  the node with
16     the maximum score of  $f$ ;
17      $\mathcal{S} \leftarrow \mathcal{S} \cup \{v\};$ 
18     for  $u \in N_v$  do
19         if  $u \in \mathcal{NS}$  then
20              $\mathcal{W}(u) \leftarrow \mathcal{W}(u) - 1;$ 
21              $\text{Recompute}(f(u));$ 
22      $\mathcal{NS} \leftarrow \mathcal{NS} \setminus \{v\};$ 
23 return  $\mathcal{S}$ 

```

---

highest value of  $f$ . However, if two nodes having several friends in common are selected as initial spreaders then their influence will overlap and will cause negative effect for influence maximization. For example, suppose that the nodes 2 and 3 are selected as initial spreader in the graph presented in Fig. 2, since most of the neighbors of node 1 are also neighbors of node 2, then their influence will overlap and will cause negative effect for influence maximization. To overcome this drawback, we propose to select the initial spreader nodes following two methods. (1) The first method called  $IMSN_{nc}$  (line 8–13 in Algorithm 2) in which we select the top  $k$  not connected nodes with the highest value of  $f$ . (2) Our second method called  $IMSN_{ld}$  (line 14–22 in Algorithm 2) in which the main idea is that if one superordinate node is considered as seed, then the links connecting this node with the other superordinate nodes not yet chosen will be discounted, i.e., when considering the next node, the links connecting this node with the other superordinate nodes already in the seed set will be discounted.

*Example 2.* Let us consider the undirected depicted in Fig. 2 and let  $k = 2$ . Using the principle of 'top  $k$  not connected superordinate nodes' procedure, nodes 10 and 12 are considered as initial spreaders.

## 5 Experimental Evaluation

The proposed algorithm, referred to as  $SNIM_{nc/ld}$  was written in Python. Given an input network as a set of edges, our algorithm starts by generating the set of maximal cliques. To detect maximal cliques, we consider the state-of-the-art algorithm proposed in [9]. We compare our algorithm with two popular algorithms in influence maximization problem, namely degree-based algorithm and degree-discount algorithm. The degree-based heuristic is commonly used in the sociology literature as estimates of a node's influence [18]. The degree-based heuristic chooses nodes  $v$  in order of decreasing degrees  $d_v$  in a given graph  $G$ . Authors of [2] and [18] used high-degree nodes as influential nodes. The General idea of the degree discount algorithm proposed in [7] is that if one node is considered as seed then the links connecting this node with the other node will not be counted as a degree, i.e., when considering the next node, the links connecting with the nodes already in the seed set will be discounted.

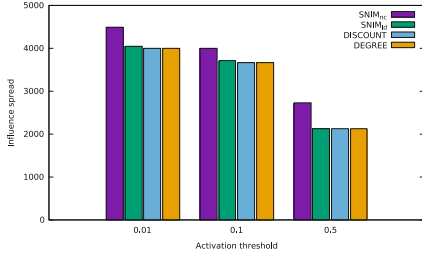
The comprehensive performance study conducts on two real world datasets, Amazon network and Dblp network. In each experiment, we vary parameters, of the diffusion model, to compare the influence spread (number of activated nodes) of four algorithms. For our experimental study, all algorithms have been run on a PC with an Intel Core i7 processor and 16 GB memory. We imposed 1 h time limit for all the methods.

**Complete Simulation:** To compute the expected number of infected vertices, we use the random simulation used in [15]. More specifically, for a given activation threshold  $\theta$  and for each seed set identified by an algorithm, we simulate the independent cascade model 10000 times (number of iteration). At each time, we choose randomly the propagation probability  $p_{uv}$  between each two connected nodes  $u$  and  $v$  in the graph. The expected number of infected nodes is the total of activated nodes throughout the simulation process divided by the number of iteration.

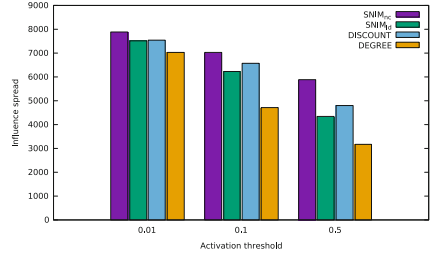
**Results on Amazon Instance:** We discuss the influence maximization on a large real-life dataset, Amazon network [19]. It is a product network, where nodes denote the products. If a product  $i$  is frequently co-purchased with product  $j$ , the graph contains an undirected edge from  $i$  to  $j$ . Amazon network contains 334 863 nodes and 925 872 edges. We analyze the efficiency and influence spread of our algorithms ( $IMSN_{nc}$  and  $IMSN_{ld}$ ) with respect to different numbers of seeds and values of parameters. For the filtering procedure, we set the minimum size of the maximal cliques  $\alpha$  to the value 3. Figure 3(c) shows the influence spread of different algorithms with different number of seeds on Amazon. The  $x$ -axis indicates the number of seeds and  $y$ -axis indicates influence spread. In most cases,  $IMSN_{nc}$ 's influence spread  $> IMSN_{ld}$ 's influence



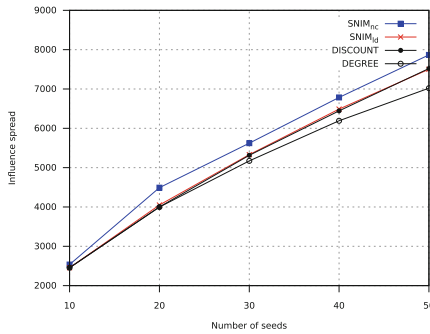
spread > Degree discount’s influence spread > degree’s influence spread. With the increasing number of seeds,  $IMSN_{nc}$  get better influence spread than the baselines.



(a) Influence spread of 20 seeds



(b) Influence spread of 50 seeds

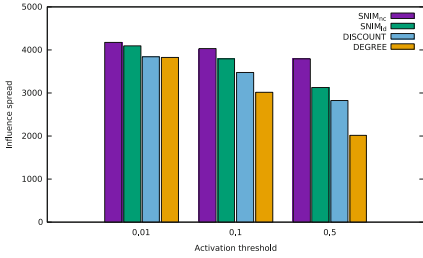


(c) Influence spread of different algorithms with  $\theta = 0.01$

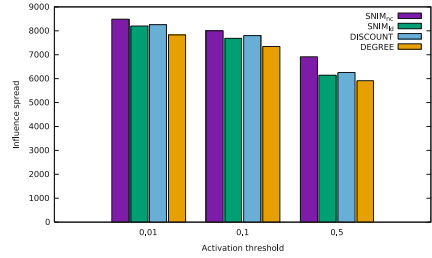
**Fig. 3.** Results on Amazon network

Figures 3(a) and 3(b) perform the influence spread of 20 seeds and 50 seeds with different  $\theta$  values (0.01, 0.1 and 0.5). The  $x$ -axis indicates activation threshold and  $y$ -axis indicates influence spread. The results reflected in the figures show that although total influence spread of the four algorithms will decrease as  $\theta$  increases. Notice that  $IMSN_{nc}$  improves its influence spread with the increasing of  $\theta$ .

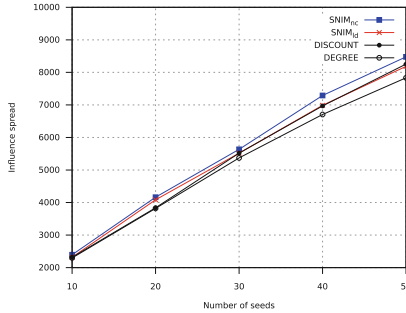
**Results on Dblp Instance:** Now, we discuss the influence maximization on a second dataset, named Dblp [19]. It is a large real-life academic collaboration dataset in Computer Science. Each node in the undirected network represents an author. If an author  $i$  co-authored a paper with author  $j$  (they publish at least one paper together), the graph contains an undirected edge between  $i$  and  $j$ . Dblp network contains 317 080 nodes and 1 049 866 edges. For the filtering procedure, we set the minimum size of the maximal cliques  $\alpha$  to the value 3. In Fig. 4(c), we report the influence spread of the four algorithms for different



(a) Influence spread of 20 seeds



(b) Influence spread of 50 seeds



(c) Influence spread of different algorithms with  $\theta = 0.01$

**Fig. 4.** Results on Db1p instance

values of  $k$  and for  $\theta = 0.01$ . Specifically, we simulated the information spread for  $10 \leq k \leq 50$ . Figure 4(c) clearly shows that our  $SNIM_{nc}$  algorithm outperforms  $SNIM_{Id}$ , degree-based algorithm and degree-discount algorithm in most cases of  $k$ . Figures 4(a) and 4(b) perform the influence spread of 20 seeds and 50 seeds, respectively with different  $\theta$  values (0.01, 0.1 and 0.5). The  $x$ -axis indicates activation threshold and  $y$ -axis indicates influence spread. The results reflected in the figures show that although total influence spread of the four algorithms will decrease as  $\theta$  increase. It is clear in these figures that  $SNIM_{nc}$  algorithm outperforms the other algorithms in all cases of  $k$  and  $\theta$ .

Overall, in terms of influence spread,  $SNIM_{nc} > SNIM_{Id} > degree - discount > degree - based$ . In terms of scalability, our algorithm is able to maintain the same efficiency when the number of nodes and edges increase. Indeed, several efficient parallel algorithms to solve the problem of maximal cliques are proposed in the last decade.

## 6 Conclusion

In this paper, we proposed an algorithm for influence maximization problem in networks based on maximal clique problem. In particular, we make an original use of a particular concept of nodes called *superordinate nodes*. Then, we

introduced two indicators in order to select the most influential nodes in the graph. There are many possible directions for future works. Possible improvements can be obtained by designing better indicator functions for superordinate nodes selection. Another direction is to extend our proposed framework in order to use other cohesive structures such as  $k$ -plex,  $k$ -truss, etc. We also plan to extend our method in order to deal with the problem of influence maximization in dynamic networks.

## References

1. Community based influence maximization in the Independent Cascade Model. Zenodo, September 2018
2. Albert, R., Jeong, H., Barabasi, A.: Error and attack tolerance of complex networks. *Nature* **406**(6794), 378–382 (2000)
3. Banerjee, S., Jenamani, M., Pratihari, D.K.: A survey on influence maximization in a social network. CoRR, abs/1808.05502 (2018)
4. Bass, F.M.: A new product growth for model consumer durables. *Manag. Sci.* **50**(12 Supplement), 1825–1832 (2004)
5. Berry, N., Ko, T., Moy, T., Smrcka, J., Turnley, J., Wu, B.: Emergent clique formation in terrorist recruitment, January 2016
6. Brown, J.J., Reingen, P.H.: Social ties and word-of-mouth referral behavior. *J. Consum. Res.* **14**(3), 350–362 (1987)
7. Chen, W., Wang, Y., Yang, S.: Efficient influence maximization in social networks. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2009, pp. 199–208. ACM (2009)
8. Chen, Y., Zhu, W., Peng, W., Lee, W., Lee, S.: CIM: community-based influence maximization in social networks. *ACM TIST* **5**(2), 25:1–25:31 (2014)
9. Cheng, J., Zhu, L., Ke, Y., Chu, S.: Fast algorithms for maximal clique enumeration with limited memory. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2012, pp. 1240–1248. ACM (2012)
10. Domingos, P., Richardson, M.: Mining the network value of customers. In: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2001, pp. 57–66. ACM, New York (2001)
11. Freeman, L.C.: A set of measures of centrality based on betweenness. *Sociometry* **40**(1), 35–41 (1977)
12. Freeman, L.C.: Centrality in social networks conceptual clarification. *Soc. Netw.* **1**, 215 (1978)
13. Ibnoulouafi, A., Haziti, M.E., Cherifi, H.: M-centrality: identifying key nodes based on global position and local degree variation. *J. Stat. Mech: Theory Exp.* **2018**(7), 073407 (2018)
14. Jabbour, S., Mhadhbi, N., Raddaoui, B., Sais, L.: Detecting highly overlapping community structure by model-based maximal clique expansion. In: IEEE International Conference on Big Data, Big Data 2018, Seattle, WA, USA, 10–13 December 2018, pp. 1031–1036 (2018)
15. Kempe, D., Kleinberg, J., Tardos, E.: Maximizing the spread of influence through a social network. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2003. ACM (2003)

16. Kitsak, M., et al.: Identification of influential spreaders in complex networks. *Nat. Phys.* **6**(11), 888–893 (2010)
17. Richardson, M., Domingos, P.: Mining knowledge-sharing sites for viral marketing. In: *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2002*, pp. 61–70 (2002)
18. Wasserman, S., Faust, K.: *Social Network Analysis: Methods and Applications*, vol. 8. Cambridge University Press, Cambridge (1994)
19. Yang, J., Leskovec, J.: Defining and evaluating network communities based on ground-truth. *Knowl. Inf. Syst.* **42**(1), 181–213 (2013). <https://doi.org/10.1007/s10115-013-0693-z>