Computer Science Review 40 (2021) 100393

Contents lists available at ScienceDirect

Computer Science Review

journal homepage: www.elsevier.com/locate/cosrev

Review article

New trends in influence maximization models

Mehdi Azaouzi [a,b,*], Wassim Mnasri [a], Lotfi Ben Romdhane [a]

[a] MARS Research Laboratory LR17ES05, University of Sousse, Tunisia
[b] L3i, La Rochelle University, France

ARTICLE INFO

Article history:
Received 16 October 2020
Received in revised form 12 January 2021
Accepted 4 March 2021
Available online 22 March 2021

Keywords:
Social networks
Social influence analysis
Influence maximization
Influential node
Influential group
Privacy protection

ABSTRACT

The growing popularity of social networks is providing a promising opportunity for different practical applications. The influence analysis is an essential technique supporting the understanding of real-life activities. Accordingly, certain reviews and surveys have been presented, focusing on models, methods, and evaluation aspects related to social influence analysis. However, the ultimate goal is that the background social influence analysis methods developed in research could be employed in real applications. In this context, social influence analysis still remains a number of challenges including the privacy of the massive networks that have been recently mentioned by researchers. Motivated from these facts, in this paper we provide a state-of-the-art survey on the influence analysis techniques in addressing these challenges. In this detailed survey, we divide the diffusion models into two categories, individual and group node-based models. Our primary focus is to investigate the research methods and techniques and compare them according to the above categories. In the sequel, we especially further provide an overview of the existing methods for influence maximization under privacy protection. The recent advanced applications of social influence are also surveyed. In the end, open issues are discussed to enable the researchers to a better understanding of the present scenario and suggest several potential future directions for research in influence maximization.

Contents

* Corresponding author at: L3i, La Rochelle University, France.
E-mail addresses: mehdi.azaouzi@univ-lr.fr (M. Azaouzi), mn.wacym@gmail.com (W. Mnasri), lotfi.BenRomdhane@isitc.u-sousse.tn (L. Ben Romdhane).

https://doi.org/10.1016/j.cosrev.2021.100393
1574-0137/

1. Introduction

Influencer, Influenced, Influence, has increased enormously in many fields such as in social media, business, and scientific research. The term influence analysis has been widely used to describe a some degree in his or her daily communications. In marketing, business firms to presentation their products in the marketplace and to succeed in competition with other competitors in selling products uses social media and useful users on social networks as their advertising platform. In political, the Trump case and how he got to the presidential election, he used the social network twitter to influence, attract, and convince people with his ideas and vote for him. In music, the influence of millions of young people through a genre of music. Likewise, the influence of millions of children and adolescents through a culture, the influence of customers through a product, etc. During the propagation information, nodes affect other nodes via their relationship-based behaviors, e.g., accepting, approving, labeling, forwarding, and agreeing to the opinions of others via friend relationships (acceptance behaviors) or ignoring and rejecting information from strangers (rejection behaviors). One of the severe challenges in social networks is the diffusion of phenomena and information in the system, which is of great interest to researchers. The influence analysis is constructed based on this issue. Influence analysis is generally accepted as one of the most significant problems for social network mining because the influence degree of a user or a group reveals its impact on communication or information propagation in terms of scale, level, or other factors. Influential user discovery based on users' complex relationships and interactions is a key challenge for influence analysis in social networks. Identification of influential users in social network this problem is known as the Influence Maximization problem (IM). The influence maximization, first introduced by Domingos and Richardson [1], is an important research issue which concerns influence spread process in the social networks. IM is the problem of addressing a small subset of influential users that can maximize the influence spread in the network. This set of nodes is named the target set. Influence spread is simulated based on information diffusion models. In these models, various states are considered for nodes where each state may change based on some defined transition conditions. States are defined based on the concept of influence acceptance. In most of the existing information diffusion models, the states are either active or inactive. Nodes enter active state when they accept influence. The models are known as progressive if nodes can only switch from inactive to active state and are known as symmetric if nodes can switch in both directions [2]. The solutions through the use of influence analysis can be obtained to help reduce economic and labor cost while achieving better profits in viral marketing [1], social recommendation [3], rumor control [4], network monitoring [5], healthcare [6], revenue maximization [7], etc. With the exponential growth of social networks, for example according to *Hootsuite* [8], 2.414B monthly active users (Q_4 2019) on Facebook, of which more than 1.62B users are active each day, Youtube 2B monthly active users (Q_4 2019), WhatsApp 1, 6B monthly active users (Q_4 2019). The challenges facing researchers in the problem of social influence, or rather the problem of influence maximization, have become more complicated and complex with this massive number of users and the amount of extraordinary data shared between users. In order to influence the greatest possible number of users in the network, hundreds and thousands of solutions, models, and algorithms are proposed to solve this problem.

In recent years, there have been numerous efforts on IM. According to the type of features used in existing IM methods, we divide them into two categories: topological networks-based methods and semantic-based methods. The topological networks-based detection methods mainly utilize the positions and topological locations of nodes such as centrality measures and community detection, extracted from the topology of a network since it is often difficult in many networks to access contextual information. The semantic-based methods generally rely on the interaction characteristics among abundant users, such as commenting, reposting, and following. The social context-based methods generally rely on the interaction characteristics among abundant users such as commenting, reposting, and following, as well as the topic features, such as opinion and topic specific.

According to knowledge, until the end of 2018, the models are only interested in the individual to maximize influence, ignoring the group effect. A new path in scientific research to solve this influence maximization problem is open, new models are interested in groups and select a k-top group instead of selecting k-top nodes to maximize influence, and this poses a new problem which is the maximization of influence based on groups (GIM).

To study the GIM Problem, a group is abstracted as a several users and this group will be activated if a certain number of users in this group are activated. Though topological and semantic techniques in both these influence maximization problems (individual or group) implemented for building an efficient solution, these techniques exhibit challenge. Privacy protection is one of the major challenges which limiting the dissemination of information on the network.

1.1. Motivation

Most of the previously presented literature surveys focused on reviewing the state-of-art in a certain direction in-depth. To the best of our knowledge, they do not show the GIM problem and the problem of protecting user's privacy. As far as we know, this is the first attempt to bring the group-based IM problem by taking into account the privacy preserving techniques. To summarize, the key contributions of this survey are four-fold: (1) We conduct a systematic review for IM models based on node topologic techniques; (2) We provide an overview and summary for group-based models. (3) We elaborate the impact of security and privacy preserving techniques on the different IM models (4) We discuss the challenges and open issues, and identify the new trends and future directions in this research field to share the vision and expand the horizons of IM research.

1.2. Related works

There are several survey articles in the literature that considers issues of the influence maximization in social networks, in particular, [9–14]. More specifically, the authors in Ref. [9] covers the types of social networks and categorizes the influence maximization algorithms according to adaptive strategy. In the same line, Li et al.'s survey [10] presents an overview of studies that enable IM tasks, which exploit how the solution methodologies are used to solve the context-aware IM. Microscopic models and macroscopic models were considered to predict user influence power. Based on these models, the survey described in Ref. [11] covers the influence maximization and also the influence minimization and flow of influence. They also discuss the influence evaluation metrics. The role of semantics in online social networks is presented in Ref. [12]. The survey summarizes the algorithms based on different topics such as influence metrics, information flow and influence, the properties of Network and the applications of IM. The work in Ref. [13] is focused on the hardness results of this problem in both traditional and parameterized complexity framework, followed by the solution methodologies. Finally, other work in [14] focus on the influential nodes detection problem in dynamic social networks.

One can see that the previous surveys mainly focused on influence maximization algorithms in social networks and lack the investigation on some important points. They have not covered two other emerging techniques, i.e., Group IM (GIM), the parallel techniques and the impact of security, and privacy preserving techniques of the users on IM, where the framework, methodology and features are discussed in this paper.

1.3. Contributions

Our paper provides a novel framework for the influence maximization in SNs which is missing in previous most recent reviews. To the best of our knowledge, there has not been a survey conducted on IM under privacy protection mechanisms — which our paper objectively illustrates, as new challenges of information diffusion in social networks. After that, we introduced a new IM problem based on group (GIM). Furthermore, we categorize and compare a number of relevant research works on GIM in social networks. The major contributions of this survey are as follows:

(1) Firstly, in order to make our survey comprehensive, we provide an overview on traditional relevant research works on influence maximization. These traditional works study in most survey is "node-level" study. They has only focused on the subject node and social tie in social influence phenomenon. To the best of our knowledge, this is the first work that addresses the parallel algorithms in node-level based IM. Additionally, we aim to conduct a comprehensive survey that takes into account the new problem of "group-level" influence maximization. Hence, we define the new extended concept, Group IM (GIM), to provide more comprehensive coverage of the field.

(2) The second major contribution of this survey paper uncovers important challenges of privacy considerations in the information diffusion model. As opposed to all existing surveys, our work reviews related work in the problem of maximizing influence with privacy considerations in social networks. A part of representative state-of-the-art approaches is then used to tackle real-life issues related to privacy and security of OSNs in social influence.

(3) Finally, we then identify open issues raised from social influence analysis under privacy protection mechanisms as new challenges. Our goal is to pave a solid foundation for novel promising research directions in social influence analysis.

1.4. Article organization

The entire paper is organized according to the series mentioned below. The diffusion of influence in social networks and some preliminary knowledge, fundamental concepts concerning the propagation of influence and social networks are specified in Section 2. Then, we will discuss the most recent and typical models used for influence analysis, of which we will talk about two main families which are the node-based models in Section 3, and the group-based models in Section 4. Thereafter, the user privacy barrier that limits the maximization of influence in social networks and the taxonomy of IM models under privacy protection are depicted in Section 5. At last, the recent and promising applications of influence analysis based on this study are shown in Section 6. Afterward, we will discuss our future directions and our open challenges in social networks in Section 7. Finally, the conclusions of the paper are given in Section 8.

Table 1
Table of notations.

Notations	Descriptions
$G = \langle V, E\rangle$	A graph G with V a set of vertex and E a set of edge.
n	The number of nodes in the graph G, i.e., $\|V\|$
e	The number of edges in the graph G, i.e., $\|E\|$
C_i	A community of the graph G
Grp_i	A group of the graph G
N_u	The neighborhood set of u
D_i	Vector characteristic of a vertex
I_{i_z}	An element of the vector D_i corresponds to a center of interest of a vertex V_i
$\sigma(u)$	The area or zone of influence of a vertex u
$w(u, v)$	The weighting of the edge (u, v)
p	The probability of propagation of the influence
$Cand$	The set of influential candidate vertices
A	The set of seed vertices in the social network
S	The set of neighboring nodes of a node v which have already tried to influence the node v
m	A published message
$Cntxt(m)$	The contexts of the published message m
$Cntxt(I_u)$	The contexts of the user's interest I_u

2. Preliminary knowledge in social networks and influence maximization

In Table 1 we find the main notations used in this article Survey.

2.1. Overview of social networks

A social network according to scientists and sociologists is generally assumed to be viewed as graphs. It designates an arrangement of links between individuals who are people or users, and/or organizations that are a set of users who form a social group. The edges represent interactions or the links among users. These users may have different interests, despite a collective objective, and they constitute a grouping which has a meaning: a family, friends, etc. The most appropriate structure for presenting a social network is a graph $G = \langle V, E\rangle$. V is the set of vertices in G; and E represents the set of edges connecting pairs of vertices and it represents the interactions between them, noted by $(u, v) \in E$. A subgroup of vertices with strong interactions between them and little interaction with the outside are called community (group). For any arbitrary vertex or group, a direct neighbor of the vertex or group v in G is a vertex or group u if the vertex or group v and u are well connected by a link. The boundary or neighborhood set $N(v)$ is the set of all vertices with direct links to the vertex u; noted by $N(v) = \{u \in V : (v, u) \in E\}$.

The semantics social network is a kind of social network in which the knowledge characterizing individuals are integrated. These knowledge can be represented by the profile or group information, the social tags, the topics, the conformity, the actions (*like*, *share*, *comment*), etc. The semantics social network stores the data in a structured form without compromising their rich, contextual character.

The vector characteristic of a user, noted as D-dimensional vector of a user v is the set of centers of interest of user v. D_i notes this set and it is expressed as $D_i = (d_{i_1}, \ldots, d_{i_D})$ where d_{i_d} corresponds to a center of interest d_d of the user U_i.

2.2. Overview of social influence

Social influence is the force or pressure exerted by a user (individual) or group on each of its members, whose objective is to impose by means of contact, dominant standards or to modify or change the attitudes, opinions or feelings of an individual or a group. Formally, the social influence is recognized as influence maximization (IM) problem and is defined as followings: "*given a social network modeled as a graph G and a nonnegative number k, find a set of k seed nodes, such that by activating them initially, the influence propagation scale of the seed node set is maximum under a certain diffusion model* [2]". Generally, all nodes in G have two states, namely activate and inactive. An active vertex who is "infected" by the information may keep sending it to its neighbors, and the inactive vertex can be influenced by this information but it cannot keep sending it to its neighbors. Social influence is not necessarily symmetric. The fact that vertex v influences an vertex u does not necessarily means that u also influences v, where v and u are two entities. In this case, the area or zone of influence $\sigma(v)$ of a vertex v is the set of vertices who are in the path of diffusion (propagation) of the information disseminated by the vertex v, and which are already influenced by the content propagated by the vertex v.

The level of social influence can be measured on the vertex or on the edge. The level of social influence of a vertex (nodes or groups) or what is also called the importance of a nodes or groups in a social network is a probabilistic value in the interval [0, 1] which presents the power that this node or group can exercise on each of its neighbors to influence it with an idea, product, concept, etc. The vertex social influence can be regarded as centrality measures, link topological ranking measures, entropy measure, and so on. These evaluation metrics can be deduced which are listed as follows: the degree of a vertex, betweenness, closeness, similarity, pageRank, etc. The level of social influence of edge completes the strength of the vertices, whose edges represent the interaction between the vertices and the influence exerted by one vertex on another. However, the links between the vertices can be direct links or indirect links. For this we can measure the power or the strength of edges as follows: direct relation, indirect relation.

2.3. Formulation of the influence maximization problem

The problem of influence maximization (IM), a problem which got along a lot with the extraordinary growth and the exponential one of the number of users and data shared in the social networks. This problem attracted the attention of the scientific researchers and sociologists, as it has attracted politicians, business people, etc. To convince a person or group of people to buy a product, companies advertise using banner ads on the street, or on TV channels despite being expensive. This system can influence people but it is so expensive, and sometimes not too effective. Using social networking big data in recent years, advertising has become easier, less expensive, and more efficient. Thus, one of the issues of primary interest in social network analysis is social influence. It is based on the fact that one's ideas can affect those of his friends. The existence of social influence may result in the spread of ideas or information like an epidemic. Inspired by the "viral marketing" problem which is one of the widespread applications of influence spread through a social network, the authors of Ref. [1] firstly collated and summarized the influence maximization work. On this basis, Ref. [2] showed that the IM issue is a non-deterministic problem with polynomial complexity, which is Non-deterministic Polynomial hard (NP-hard).

Formally, the IM in a social network can be regarded as finding a seed set of users (nodes) or groups (communities) who could maximize the influence spread over a social network. From the viewpoint of mathematics, given a directed social network $G = (V, E)$, in which V denote the set of nodes (individuals) and E represents the set of edges between them. Considering a seed set $S \in V$ is a set of initial active nodes (users) or communities (groups) at the beginning of the cascading process, and $\sigma(S)$ is the number of nodes expected to be influenced by S under a certain stochastic diffusion model. In particular, Figs. 1 and 2 gives an illustration of influence spread. According to the definition of influence spread, the IM problem can be formulated as follows.

Definition 1 (*Influence Maximization*)**.** For the given network $G = (V, E)$, an positive number $k < |V|$, propagation probability P_{uv} between each edge uv and the diffusion process M; the goal of the influence maximization problem is to find the seed set S^* containing k nodes where the expected number of active nodes using S^*, $\sigma(S^*)$, is maximized at the end of M, which is shown in Eq.

$$\mathfrak{S}_M = \underset{|S^*|=k}{\arg\max} \quad \sigma(S^*) \tag{1}$$

Regarding maximizing the spread of influence, it is important to know how the influence spreads in social networks. In fact, the diffusion model should be determined at first. Influence diffusion models are usually treated as specific cases of the epidemic models within the active nodes and they help to identify the final number of active users in the network. Several models have been presented to demonstrate how influence propagates in a social network. IM models can be regarded as two categories: node or user interactions based model and group interactions based model. With regard to the node based existing models, the most popular models for describing influence propagation are linear threshold model (LTM), the independent cascade model (ICM), and heat diffusion model (HDM).

(1) Linear threshold model The linear threshold model was proposed by Granovetter [15]. In this model, a individual v will be activated by its activated neighbors according to a weight metric b_{uv}. Otherwise, weight b_{uv} denotes the influence of node u on node v. In order to measure the probability to be activated any inactive node, a threshold θ_v is randomly assigned. In general, the threshold θ is selected uniformly at random from 0 to 1. The model expands in the form of discrete timestamps, i.e., once a node is active in step t, then it will remain active at the remaining of the lifecycle. At each step, if the total weight summary of its active neighbors satisfies the following threshold:

$$\sum_{u \in N(v)} b_{uv} \geq \theta_v$$

The status of nodes changes from "inactive" to "active". Since there is no clear definition to determine the threshold, the IC model is usually adopted to simulate the process of influence spread in most previous works.

(2) Independent cascade model Goldenberg et al. [16] proposed the independent cascade model, in which the diffusion process triggers a cascade of activations in discrete timestamps. In the model, each edge $e = (u, v)$ is related to the probability that node v is activated by node u successfully. At timestamp t, activated node u attempts to activate its inactive neighbor node v with a

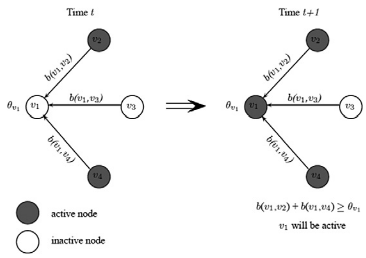

Fig. 1. An illustration of influence spread with LT model.

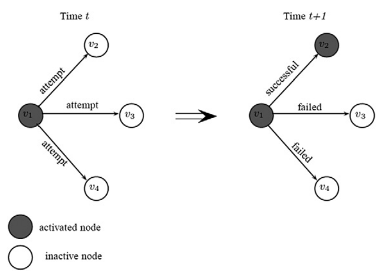

Fig. 2. An illustration of influence spread with IC model.

certain probability, then node v will become active at timestamp $t+1$, if it is activated successfully, otherwise, it will keep inactive. No matter the activation is successful or failed, node u has no chance to try it again on node v until the end of the process. This process will end until no more nodes are activated. We note that the weighted cascade model is a special case of the independent cascade model.

(3) Greedy algorithm In either the IC or the LT model, the exact calculation of the optimal solution for the IM problem is NP-hard [2]. The goal of the greedy algorithm is to iteratively choose the nodes with the maximal marginal influence increment as seed nodes. Let $f(v|S_i) = f(vS_i) - f(S_i)$ denote the marginal influence increment of node v, where $f(S)$ is the influence spread of seed nodes S. Although the Greedy algorithm is conceptually simple, it requires significant computational overheads because Monte Carlo simulations are performed over the entire network to compute the marginal influence increment. Therefore, the greedy solution is approximated to within a factor of $(1 - 1/e - \epsilon)$.

To solve the IM problem, there is a vast body of literature on IM algorithms. These algorithms can be classified in two groups, which are discussed in the following. The first category is based on the node or the person to influence the others, and other models based on the group and the strength that the group gives to maximize influence. However, the security and privacy of networks imposes limits on certain models in the propagation of information, others models and solutions have also been proposed, and is discussed in this paper.

3 Individual node-based influence analysis

The problem of influence maximization is well initialized since the definition of social influence by Ritzer [17], who consider social influence as an orientation and a change in the thoughts and attitudes of an individual through its interaction with another individual or group in a social network.

Many frameworks have been proposed to solve this problem. With over a decade of research, there have been a large number of publications on various kinds of algorithms and applications in this area of the identification of the top-k node. In these frameworks, a minimal set of k-top nodes that are capable of generating maximum influence and widest information spread to their connected nodes, are selected. The idea is a reflection of real life, that a person can influence a very large number of people, such as famous people like football players, whose millions of fans are obsessed with these famous people. For that they can easily get attached to ideas or brands of products, etc. that their heroes (famous people) use, superstars like music stars and movie stars, leaders in our lives like our teachers and parents, close friends, etc. Much more detail for identifying top-k nodes is provided in Ref. [18].

At present, much related research involves the problem of influence maximization based on individual node. These works are generally based on the two abovementioned fundamental propagation models. The node individual based solutions methodologies can be divided into four categories: (1) submodularity-based algorithms; (2) Heuristic-based heuristic algorithms; (3) community-based algorithms and (4) Parallel models. To sum up, we list recent representative works in this paper in the following: At first, we describe the variations of these models in Sections 3.1.1 and 3.1.2. Then, the available solution methodologies in the literature have been described in Sections 3.2 to 3.4. Finally, Table 2 shows a comparison of the individual influence models.

3.1 Models based on influence diffusion processes

In this subsection, we aim to examine the trend and progress of diffusion models to identify influential nodes. Two main models are formulated by Kempe and Kleinberg [2]: the independent cascade (IC) model and linear threshold (LT) model. For both the IC and LT models, a approximation algorithms is designed to provide the best solution. The approximation algorithms (the greedy algorithms) makes iteratively an optimum local choice at each stage, expecting it to converge to a global one in which it seeks to calculate the k influence elements. These algorithms rely on time-consuming Monte Carlo simulations to estimate accurate marginal influence spread. However, These is unsuitable for large-scale social networks due to its high time complexity. Meanwhile, a number of variations of these models have been proposed to estimate influence of nodes in effective information propagation. The variations of both the IC and LT models and models that differ from the IC or LT models can be seen throughout the research in this section.

3.1.1 Variations of the IC and LT models

In the IC model, the most influential nodes are selected according to their capacities for activating nodes. However, the IC model is not practical due to its high computational complexity. Another main limitation of this model is that only the activated nodes are considered and the inactivated nodes are disregarded in the algorithm. To overcome the drawbacks of the IC model, some researchers have considered time delay and time-critical constraints for influence diffusion. To overcome the basic specification of the IC model, whose activation of the nodes is independent and does not depend on a history of activation attempts, Kempe et al. [19] proposed a second model called Decreasing Cascading (DC), which takes into account the information saturation problem. Generally, an inactive node can receive more than one activation attempt of more than one neighbor, this can happen at different times as they can happen at the same time. Similarly, Kimura et al. [20] proposed a special case of IC Model called Shortest Path Model (SP Model). In this model, an inactive node will get a chance to become active only through the shortest path from the initially active nodes. In [21] have proposed a model called CTMC-ICM (Continuous-Time Markov Chain-Independent Cascade Model) which derived from the basic model the IC model. CTMC-ICM uses a classification metric called *SpreadRank* which takes advantage of the information propagation and the influence capacity of each vertex of the social network, and also incorporate a Markov chain theory in continuous time. However, it is able to efficiently estimate the number of vertices $\sigma(A)$ influenced by a set of vertices A which presents the seeds vertices in the social network. The authors have defined the network classification equation which based on the theory already given as follows:

$$SR(v) = \alpha + (1-\alpha) \sum_{w \in PA(v)} \pi_{wv} SR(w) \tag{2}$$

where α presents the decay factor, π_{wv} presents the transition matrix between the two vertices w and v, and $SR(v)$ is the *SpreadRank* value of the vertex v. In the CTMC-ICM model, the identification of the most influential individuals remains an estimate and any information on a fixed number is provided. Srivastava et al. [22] proposed an unified model in which the local neighborhood effects, aggregate social behavior, and external factors are combined. This framework does not require extensive simulations. Also, since in viral marketing companies and their products are in competition, Lin and Liu [23] proposed another model called (General Competitive Independent Cascade) GCIC which is a model based on the IC model and which allows to effectively define the diffusion of the two competing sources in the same social network. Until recently, Borg et al. [24] present a near-linear time IM algorithm which simultaneously reduce the time complexity and returns a (1-1/e-ϵ)-approximate solution with at least (1-1/n^l) probability. Retaining the approximation guarantee with the same high confidence, methods of TIM [25] and IMM [26] both decrease time complexity, while the latter one further reduce the unnecessary computational costs. Li et al. [27] proposed a novel conformity-aware cascade model, which used conformity theory to calculate influence. Mohamadi-Baghmolaei et al. [28] considered important time and trust factors, and proposed the trust-based latency-aware independent cascade (TLIC) model. This is the first time trust has been studied in a classic IC model. In the TLIC model, a node can change its state (i.e., as active or inactive) with different probabilities for a trusted neighbor node than for a distrusted neighbor. Social influence is divided into two groups: positive and negative. In this case, the social influence is considered as signed. In [29], Wang et al. modified ICM to show the process through which users build opinions. Tong et al. [30] have focused on adaptive vertex seed selection strategies to propose a variation of the IC, named Dynamic Independent Cascade (DIC). DIC capable of capturing the uncertainty of the diffusion process and the dynamic aspects of a real social network. To find a set of seed nodes that would provide a maximum-positive or a maximum-negative influence, Li et al. [31] presented a polarity-related IC (IC-P) as the diffusion model. For the IC model, Liu et al. [32] proposed maximum likelihood-based scheme based heuristic algorithm, named MLIM. In this algorithm, the thumbnails of the social network are constructed, then the L-value for each vertex using the maximum likelihood criterion is computed. A greedy algorithm is proposed to sequentially choose the seeds with the smallest L-value. The experiments were also conducted to show that the proposed method can provide a wider influence spreading while obtaining lower time consumption.

To outperform the drawbacks of the IC model, many different models have been proposed based on a LT model in the literature. Firstly, Mossel and Schoenebeck [33] proposed a new extension of the LT model, the Voter model whose idea is that a node u can be influenced, change and orient its opinions according to the influence exerted by most of its neighboring nodes $N(u)$, whose probability $p_u = \frac{|N_a(u)|}{|N(u)|}$ to influence and activate the node u is proportional to the number of neighbors $N_a(u)$ already activated and influenced by an idea, a concept, a product, etc. The main difference between the Voter model and the LT model or LT-based models is that the Voter model allows a user to change their opinions thanks to the influence exerted by their active neighbors. However, unlike the monotonous LT-based models, where the user can no longer change their opinions and state after a change, where a node has become active after being inactive, it will remain active without the possibility of changing its state the other times, the voting model is not monotonous. Based on the voter model, another model is proposed by Wang et al. [34] called PIDS whose objective is to efficiently select a set of seed nodes capable of maximizing influence in a network. The PIDS (Positive Influence Dominating Set) model is based on the principle that once at least half of the neighboring users $N(u)$ of a user u have a positive impact on this user, then its probability p_u of positive influence on the other inactive neighboring nodes increases and becomes high. By improving the voter model, Li et al. [35] have proposed a new model dedicated to signed networks which makes it possible to analyze the dynamics of propagation of the positive and negative influence which represent two opposing opinions. However, this model derived the exact and closed formulas separately, to ensure long-term and short-term dynamics. Similarly, in order to promote and popularize a new product from a company than another product already on the market and in competition, He et al. [36] proposed a new model called CLT (competitive linear threshold) in which they studies the problem of maximizing the influence of a product by limiting the propagation of the influence of a second competing product. In Ref. [37], a hybrid influence maximization algorithm was proposed based on a LT model. This algorithm is divided into two phases: a heuristic phase and a greedy phase. In the heuristic phase, some influential nodes with high influence are selected as seeds, whereas the greedy phase is similar to the IC model. One significant improvement is that the potential effects of seeds on their inactivated nodes are considered in this algorithm. Like the principle of the models already mentioned, the process of spreading the PIDS model is repeated several times until no individual becomes active. In [38], Liu et al. extend the linear threshold model to establish the diffusion-containment (D-C) model by incorporating the realistic specialties and characteristics of the containment of positive influence spreading. The traditional LT model is not applicable to that a situation concerns both the diffusion and the containment of the influence. In the D-C model, the state of a node is described by the activation probability; each node is only influenced by a neighbor with a higher probability, and the sum of the probabilities of possible node states is not greater than 1. They are designed to deal with non-competitive as well as competitive timevarying propagation scenarios. In Ref. [39] a Linear Threshold with multi-level Attitude (LT-MLA) model is proposed. In this model, an attitude weight is assigned to every node related to the entity which is propagating in the social network. The value of this parameter which can be updated, determines the current state of the node. The states consist of (1) negative and active, (2) negative and inactive, (3) positive and active, (4) positive and inactive, and (5) neutral. There exists another parameter, the influence weight for every edge in [−1,1] range indicating that it considers both trust and distrust. In the process of influence diffusion, there exists a direct relation between the attitude parameter value of a node and its state regarding change. This process ends when the state of no node can be changed to the state of positive and active. A very recent technique proposed by Sheng et al. [40], adopts an opposite influence propagating model and an influence propagation function is defined to estimate the positive influence propagating of a seed set. An algorithm is presented to select the seed nodes which can obtain the largest positive influence spreading in the signed network. The algorithm employs the greedy strategy to sequentially select the seed nodes according to their spreading increments, which are estimated by the influence propagation function.

In summary, all the existing approaches have their own weaknesses. Simulation-based approaches [2,5,19,41,42] rely on time-consuming Monte-Carlo (MC) simulations so that these approaches are computationally inefficient. Proxy-based approaches [20,43] compromise some precision in influence spread to improve computational efficiency. Sketch-based approaches [25,26,44,45] are theoretical efficient but these approaches are aligned with some specific diffusion models.

3.1.2 Other models

Most of the studies in the literature are based on IC and LT diffusion models, which are time-independent, i.e., in these models, the timing of the seed selection is not important. To improve such models, therefore, other studies on influence evaluation have been conducted.

Firstly, to study the problem of influential node detection, many approaches suggest a targeted IM. Early studies have focused on the special case of a single selected target-node, but more general targeted IM methods aim at maximizing the probability of activating a target set of arbitrary size by discovering a seed set which is neither fixed and singleton nor has constraints related to the topological closeness to a fixed initiator. Majority Threshold Model (MT Model) in Ref. [46] is the deterministic threshold model. In this model, the vertex will become active, when at least half of its neighbors are already active in nature. The most influence resistant model of diffusion called Unanimous Threshold Model (UT Model) [47] where vertex threshold value is set to its degree. Doo and Liu [48] proposed a new model called Probabilistic Social Influence model (PSI) with the aim of defining a probabilistic diffusion of influence in social networks. However, PSI defines for each user a probability of propagation of influence using its characteristic attributes in the social network, unlike the IC and LT models which use a uniform probability. Among the characteristics of each node, we cite the number of interactive activities $IA(u, v)$ and the number of non-interactive activities $NA(u)$. Then, the probability of diffusion $w(u, v)$ of each node which is based on the activities can be presented as follows:

$$w(u, v) = \alpha \frac{NA(u)}{MAX(NA)} + (1-\alpha)\frac{IA(u, v)}{\sum_{s:(u,v)\in E} IA(u, s)} \tag{3}$$

Of which, α has a damping factor.

Alternatively, Zeng Yifeng et al. [49] addressed an alternative influence maximization model that is based on constrained simulated annealing with a reliability constraint for nodes in social networks. This model can identify the top-k most influential nodes given a threshold of influence loss. Similarly, Zhang et al. [50] have proposed a trust-based most influential node discovery (TMID) method for discovering influential nodes in a social network. Four phases are performed to establish influence degrees for influential node discovery: (1) an influence propagation process, which reveals the influence diffusion records among nodes for obtaining the categories of nodes in the social network; (2) a trust evaluation method, which provides methods for calculating two types of trust relationships among users, namely, direct trust and indirect trust; (3) an influence evaluation phase, which calculates the explicit binary influence among users (named active influence), the potential binary influence among users (named inactive influence), and the unary influence of nodes (named node influence); and (4) a set of algorithms for discovering the most influential nodes, which comprise two phases: a heuristic phase and a greedy phase.

In a viral marketing network, the problem of identifying influential individuals is studied in Ref. [51]. The proposed algorithm takes into account not only the users' position in the network, but also their contextual information, and thus could effectively cover interested users. The users interest on the messages can be considered to measure the closeness between two entities. The interested users are determined and the most influential ones among them are selected. Contrariwise, some researchers have to focus on non-target users in a viral marketing network. In marketing, a non-target user could be one who "hates" the product that a campaign wishes to promote. Padmanabhan et al. [52] labeled the nodes in the network with two labels T (target users) and N (non-target users). Then, they define a constrained influence maximization problem and they presented the greedy algorithm to produce a (1-1/e) approximation. They also developed a multi-greedy algorithm that attempts to keep multiple seed sets and improves upon the greedy algorithm. However, naive implementations of this algorithm is not practically viable due to prohibitively high time overhead. To address this issue, we develop a two-phase heuristic framework to improve the run times.

A novel targeted IM problem in which the objective function is defined in terms of spreading capability and topology-based diversity w.r.t. the target users is addressed in Caliò et al. [53]. They proposed two alternative algorithms, L-DTIM and G-DTIM, to solve the problem under a proposed objective function. Their experiment results on real-world datasets indicated that the proposed method performs better than some algorithms. However, the major problem is that only considers specific notions of diversity that are driven by the topology of the information diffusion graph.

Recently, He et al. [54] proposed a two-stage iterative framework for the influence maximization in social networks (TIFIM) to ensure efficiency and accuracy of the proposed schemes at the same time. In the first stage, the nodes with less influence are excluded to generate a candidate seed set. In the second stage, vertex advantage and influence overlap are considered for final selection. Using the last iteration results and the two-hop measure the authors provided a provable performance guarantee for their solution based on an efficient FLAS. Additionally, they proposed Removal of the Apical Dominance (RAD) to determine the seed nodes from candidate nodes and prove that the influence spread of TIFIM according to RAD converges to a specific value within finite computations.

More recently, a new methods of influence maximization is proposed in [55]. In this method, Li et al. considered the proportion of influential nodes, which are called "elites", to ordinary nodes, which are called "grassroots", to achieve a greater diffusion performance. They attempted to select ordinary grassroots as seeds and showed that grassroots are better choices than elites in the influence maximization problem from the aspects of relationship strengths and polarities. They developed a grassroots-oriented seed user seeking algorithm which fully explores the community information of the network structure. Similarly, a novel targeted influence maximization problem which accounts for the diversification of the seeds according to side-information available at node level in the general form of categorical attribute values is proposed in [56]. They defined a class of nondecreasing monotone and submodular functions to determine diversity of the categorical profiles associated to seed nodes.

Pros: This category of individual level approaches is based on specific diffusion model. They easily incorporate any diffusion models with ease by plugging in the model-specific Monte Carlo simulation module to evaluate the influence. They has a good theoretical property that it usually returns a solution with a constant bounded ratio of approximation if the underlying influence

Table 2
Comparison of the individual influence methods.

Study	Year	Algorithm	Spread model	Complexity
Greedy [2]	2003	Greedy	IC and LT	$\mathcal{O}(kNMI)$
SP1M [20]	2006	Greedy	IC	$\mathcal{O}(knm)$
CELF [5]	2007	Greedy	IC	$\mathcal{O}(kNMI)$
CELF++ [41]	2011	celf	IC and LT	$\mathcal{O}(kNMI)$
SIMPATH [43]	2011	LDAG	LT	$\mathcal{O}(klNP_{\theta})$
IRIE [57]	2012	Greedy	IC	$\mathcal{O}(k(n_{o\theta}k + M))$
PMIA [29]	2012	Greedy	SP1M	$\mathcal{O}(nt_{i\theta} + kn_{o\theta}.n_{i\theta}(n_{i\theta} + logn))$
IPA [58]	2013	Greedy	IC	$\mathcal{O}(\frac{NO_v n_{nv}}{c} + k^2(\frac{O_v n_{nv}}{c} + (c-1)))$
SVIM-L [35]	2013	Dynamic	Voter model	$\mathcal{O}(k+l)(\|E\| + min(bn_z^3 + n_x^3, t_C m_B))$
CTMC-ICM [21]	2014	CTMC	IC	–
PSI [48]	2014	Greedy	IC and LT	–
Borg et al. [24]	2014	Greedy	IC	$\mathcal{O}(kl^2(n+m)log^2 n/\epsilon^3)$
TIM [25]	2014	Greedy	IC	$\mathcal{O}(\frac{k(n+m)log^2 n}{\epsilon^2})$
IMM [26]	2015	Greedy	IC	$\mathcal{O}(\frac{k(n+m)log^2 n}{\epsilon^2})$
C^2 [27]	2015	Greedy	IC and LT	$\mathcal{O}(kmn)$
DIC [30]	2016	Greedy and heuristic	IC	–
LT-MLA [39]	2016	Greedy	LT	–
CSA-Q [49]	2016	CSA	ISP	$\mathcal{O}(\|PATH(S)\|) + \mathcal{N}\mathcal{O}(\|PATH(v)\|)$
D-C [38]	2016	Greedy	Diffusion-Containment	$\mathcal{O}(K\|V\|R)$
DIN [59]	2016	Greedy	LT	$\mathcal{O}(n^2)$
DP [60]	2016	Greedy	SIR	$\mathcal{O}(ln)$
PIM-SN [31]	2017	Greedy	IC	$\mathcal{O}(m.n)$
BCT [61]	2017	Greedy	IC	$\mathcal{O}((k+l)(n+m)logn/\epsilon^2)$
BCT [61]	2017	Greedy	LT	$\mathcal{O}((k+l)nlogn/\epsilon^2)$
SA [31]	2017	Simulated annealing	Polarity-related IC	–
CoFIM [62]	2017	Heuristic	Diffusion	$\mathcal{O}(k^2 \cdot n \cdot k_{max})$
SAIM [63]	2018	Greedy	bfs-tree	$\mathcal{O}(n+m)$
MLIM [32]	2018	Greedy	Likelihood IC	$\mathcal{O}(m)$
DTIM [53]	2018	DIC	LT	$\mathcal{O}(\theta^k)$
IMUD [51]	2019	Heuristic	Other	$\mathcal{O}(ncc^{(2)})$
NATURAL GREEDY [52]	2019	Greedy	IC and LT	$\mathcal{O}(k \cdot \|V\| \cdot Inf)$
MULTI GREEDY [52]	2019	Greedy	IC and LT	–
TIFIM [54]	2019	Greedy	IC and LT	$\mathcal{O}(T_I * \sum_{i-1}^{N} \|V_i^2\|)$
ADITUM [56]	2021	Heuristic	RIS	$\mathcal{O}((k+l)(\|\mathcal{E}\| + \|\mathcal{V}\|)log\|\mathcal{V}\|/\epsilon^2)$

function is monotone and submodular. In addition, most models shows a significant reduction in the time complexity by lazy forward decrease in calculations with high efficiency.

Cons: The major problem of proposed models is computational efficiency. Mostly, the time consuming in large networks. They requires tens of thousands of Monte Carlo simulations, which seriously limits its application on large-scale networks. Then, when the network size increases, the computational time will increase dramatically, which prevents the greedy based algorithm to become a feasible solution. To solve the time efficiency problem of traditional greedy algorithms, a spectral of heuristic-based algorithms were proposed by researchers in recent years.

3.2 Heuristic-based models

Due to the high computational complexity of model-based algorithms and the high efficiency of meta-heuristic algorithms, many meta-heuristic algorithms have been proposed to reduce the solution-solving time and pursue higher algorithm efficiency in recent years. In these algorithms, by defining a fitness function, IM problem is modeled as an optimization problem, and methods like evolutionary optimization algorithms are applied to solve it. In recent years, several heuristics have been proposed based on different bio-inspired evolutionary techniques by many researchers. In these methods, a score is assigned to each user by using a recursive approach, and the users with the highest score are selected as seeds. Here, we will describe the more recently population-based heuristic solution methodologies from the literature. In [64], a change in the structure of the IM problem is suggested in order to tailor it to swarm intelligence algorithms and to achieve a general slope on the state-space surface of its objective function. More precisely, if a social network is envisioned as a graph and individuals as nodes, reshaping means sorting the nodes in descending order (from largest to smallest) according to the metrics under consideration (i.e., metrics that give an idea about the level of influence of an individual) and renumbering the nodes according to this order. Thus, the nodes those are close to each other in terms of level of influence become closer to each other in the state-space. This creates a general slope on the state-space surface of the objective function. This simple idea paves the way for applying all swarm intelligence algorithms to this kind of problem. Another heuristic is proposed in [65] in which the gray wolf optimization algorithm is used to solve the IM problem. In [66] a multi-objective function is first defined as an influentiality measure, and finding such an initial is framed as an optimization problem. Then, using artificial bee colony optimization two approaches are proposed to solve the problem. In paper [67], a novel algorithm, based on the evolutionary algorithm called ITÖ, is designed to solve the Influence Maximization (IM) problem (this new algorithm is denoted as ITÖ-IM). There are three properties and two operators in ITÖ-IM: the formers include particles radius, particles activeness and environmental temperature, the later ones are drift operator and fluctuate operator. The particle radius is mainly used to simulate the characteristics of particles in Browns motion, and it is inversely proportional to the particle's activeness. The environmental temperature controls the motion ability of particles. During the searching process, the particles in ITÖ-IM can cooperate with each other to effectively balance the contradictions between exploration and exploitation exited in most of meta-heuristic algorithms. Cui et al. [68] proposed a degree-descending search strategy (DDS). They designed a new DDSE-based evolutionary algorithm that eliminates unnecessary simulations of the greedy algorithms and thus significantly improves the time complexity of such algorithms. According to the framework of DDSE, the discrete operations including mutation, crossover and selection involved in the differential evolutionary (DE) can identify influential nodes efficiently. Moreover, extensive experimental results demonstrated that it is a promising way to solve influence maximization problem by taking the advantages of metaheuristic optimization algorithms.

Apart from network structure, both online and offline interactions of users are considered in [69]. In this work, an information maximization strategy in Online and Offline double-layer propagation scheme (IMOOP) is proposed, where first the topological graph for online social network and offline connection graph of probability, respectively, is formed. Then, the two layers are compressed into a single-layer communication graph. The influence maximization in double-layer propagation scheme is NP-hard and a practical greedy heuristics is proposed. A new heuristic-greedy algorithm named the HEDVGreedy algorithm is introduced in [70]. In this algorithm, the expected diffusion value of the graph nodes was calculated using the heuristic method, and then, the effective nodes were selected using the greedy method. Experimental results showed that the proposed algorithm has a high performance than the baseline algorithms while, it significantly reduces the running time of the computations in the eight real-world data sets. In [71], a novel encoding mechanism and discrete evolutionary rules are conceived based on network topology structure for virtual frog population. Then, an effective discrete shuffled frog-leaping algorithm (DSFLA) is proposed to solve influence maximization problem in a more efficient way. To facilitate the global exploratory solution, a novel local exploitation mechanism combining deterministic and random walk strategies is put forward to improve the suboptimal meme of each memeplex in the frog population. The experimental results of influence spread in six real-world networks and statistical tests show that DSFLA performs effectively in selecting targeted influential seed nodes for influence maximization and is superior than several state-of-the-art alternatives.

Finally, the strengths and weaknesses of these swam optimization based approaches are detailed in Table 3 and are listed as follows:

Pros: A common superiority of metaheuristic algorithms is that these methods avoid generating tens-of-thousands subgraphs or traversing Monte-Carlo simulations to compute the influence by optimizing the evaluation function to estimate the expected influence spread of given node set. Moreover, extensive experimental results demonstrated that it is a promising way to solve influence maximization problem by taking the advantages of metaheuristic optimization algorithms.

Cons: The overlap between the members of the seed has been neglected in all of the above mentioned meta-heuristic methods. In addition, some drawbacks of these methods is the low algorithm accuracy. This low stability is due to the near-optimal selection.

On a different track, an enormous amount of pioneering research for detecting influential nodes based on network topology has been achieved. The centrality measures are one of the most studied heuristics. Some of the widely-used measures are as follows: centrality measures [76], closeness centrality [77], degree centrality [78,79], DegreeDiscount [80,81], heuristic clustering [82], eigenvector centrality [83], betweenness [84], k-shell centrality [85], h-index centrality [86], KATZ centrality [87], neighborhood coreness [88], pagerank (PR) [63], leaderrank [89], Shell clustering coefficient [90], degree punishment [60], truss number [91,92], gravity formula [93], local-gravity [94] and INF [95]. Further we compared it in one Table 4. In [76], two algorithms, namely, "MaxCDegKatz d-hops" and "MinCDegKatz d-hops" are proposed. The algorithms attempt to use classic centrality metrics in combination to benefit from the power of local and global centralization of each node. Then, the selection of a seed set is separated by the radius and one-hop for all graph data, since each network exhibits different topology and characteristics. The algorithms have various advantages compared with existing approaches in terms of maintaining a balance between the spread of influence and time complexity. In [78], authors used a approach called SNDUpdate to combine the semantic and structural in dynamic social networks and then the influential nodes are detected. The degree of centrality for each node models the structural aspect. Since the degree-based methods suffer unsatisfactory accuracy because only covers a limited considered scale over the whole network of interest and also lacks discriminatory power. In [79], Rui et al. proposed a novel influence maximization method, named Fixed Neighbor Scale (FNS), which extracts useful information from multiple levels of neighbors for a target node to estimate its influence strength, rather than only considering directly connected neighbors as in degree-based methods. In [80], the characteristic of membership marketing are introduced into the study of influence maximization problem. On MBIC model, the DegreeDiscount algorithm is used to quantify the influence ability. Additionally, the IRR algorithm is proposed to quantifies the ability of each node at influence and reference stage. IRR ranks the node according to the weighted sum of the two stages' abilities, and iteratively selects the node with largest weighted sum as seed node. The main advantage of the IRR algorithm is the use of different states based on real-world criteria for nodes, but this algorithm is dependent on the number of nodes in the graph. In [82], the similarity of node pairs is calculated and the nodes are clustered into k clusters and the center of each cluster is considered to be the influential node. In [88], the centrality information of neighboring nodes is used to measure the influence degree of nodes. Based on the k-shell indices of the neighbors, the spreading influence of a node is estimated. Their experiment results showed that, the proposed method can quantify the node influence more accurately and provide a more monotonic ranking list than other ranking methods. A new model is proposed M. Azaouzi and Ben Romdhane named SAIM [63]. SAIM is mainly based on social actions, of which it has defined different weights which are friendship factors attributed to qualified the different social actions. It mainly consists of two phases, the first consists in calculating the power of influence of each peak in the social graph based on the calculation PageRank. Then, the second consists of selecting the seed vertices. This calculated influence power is used for distributed community detection in [96]. The authors proved that their SAIM model shows good efficiency in the identification of seed nodes which are capable of maximizing influence in the social network, but it has a significant temporal complexity. Zareie et al. [90] also presented an improved cluster rank approach that takes into account local clustering coefficient and uses similarity of connections between neighboring nodes. Two nodes having neighbors within the same parts of the networks are considered as correlated nodes in the proposed method. Therefore, a node having low correlation with its neighbors could be better candidate to be an influential node due to its ability in spreading messages to different parts of the network by the help of its neighbors. In the method of Ref. [60] with repetitive punishment process. The nodes with the most influentiality are included in a seed set and their first and second level neighbors are punished by reducing their influentiality level. By including node z as a new member to the seed set, the influentiality of first and second level neighbors are updated according to the degree punishment. The advantage of this algorithm is that complexity is low. The authors in [91] uses an improved K-truss technique which is an extension of K-core decomposition to select the k-top

Table 3
Comparison of the meta-heuristic algorithms.

Study	Base algorithm	Time complexity	Advantages	Drawback
Reshaping [64]	GW, WOA & PR	$\mathcal{O}(n^2 log(n) + n^2 + nm)$	Simple process	Swarm intelligence algorithms often become trapped in local optima
GWIM [65]	GW	$\mathcal{O}(\|V\| + max_t.n.\|V'\| + max_t.n.k.d.d^{(2)})$	Less computational time	Bad local searching capability
MOABC [66]	ABC	–	More effective, a near optimal solution in an acceptable time	Users' humanity not taken into account
ITÖ [67]	NKD	–	Fast convergence, robust	Less performance on small dimension database
DDSE [68]	Greedy	$\mathcal{O}(n \cdot k \cdot D(-) \cdot g_{max})$	High efficiency and incredibly	Poor global search ability
IMOOP [69]		$\mathcal{O}(n^2)$	Low time complexity	Low accuracy in selecting the near-optimal
DPSO [72]	PSO	$\mathcal{O}(k^2 logknD^2)$	Reliable estimation of influence spread	The local search strategy to lead to sub-optimal solution
LAPSO-IM [73]	DPSOLA	$\mathcal{O}(l_{max} \cdot n \cdot k(logk + D^2) + n \cdot k \cdot N)$	No premature convergence	Less solving accuracy
ACO-IM [74]	ACO	$\mathcal{O}(I_{max}\|R_G\|V\| + \|V\|log\|V\|)$	High efficiency and scalability, Low time complexity	Less solving accuracy, Poor global search ability
IM-SSO [75]	SSO	$\mathcal{O}(n(kD_{avg}(N)^2 + n) + nlogn)$	High global search ability, Time-efficient	Poor local search ability

seeds capable of maximizing influence. The k-truss is a concept closely related to the k-core. The k-truss or the truss of order k of G is the maximal subgraph of G where every edge is involved in at least (k-2) triangles (i.e., cycles of length three) within the subgraph. Ma et al. propose a gravity model (GR) [93] which considers the gravity between two nodes is proportional to the product of k-core, and inversely proportional to the power of the distance between them. Afterward, the node importance is measured by the gravity accumulation of the neighbors within three steps of a focal node.

In conclusion, the centrality measures are holding relatively high computational complexity, which becomes prohibitive in applying for large-scale networks. These measures are successful in many real-life scenarios, but are criticized by some certain limitations.

A group of research models the influence through diffusion paths in the network. For example, Tang et al. [25] proposed a Two-phase Influence Maximization (TIM) to form a bridge over theory and practice. On the practice side, TIM incorporates novel heuristics that significantly improve its empirical efficiency without compromising its asymptotic performance. In [97], a random walk algorithm is first proposed to prune uninfluential nodes, then a heuristic algorithm is applied to select the most influential nodes. The requirement to experimentally specify the number of nodes for calculating influence spread at the end of preprocessing in an experimental manner is considered as the drawback of this approach. Recently, an efficient heuristic independent path algorithm (HIPA) is proposed [98]. Three key ideas are involved in HIPA to accomplish its objective by improving time complexity while maintaining efficiency. The first feature of HIPA is in its applying a vertex cover algorithm as a preprocess on the network which enables the algorithm to reduce the number of computational operations by removing the useless nodes from the influential nodes' selection domain. The second feature of HIPA is to approximate the influence spread by applying the number of paths between nodes, which needs great volume of memory to save paths. HIPA applies an appropriate structure for saving paths to reduce space complexity. The third feature is to combine the degree heuristic and the influential paths to accomplish a more accurate approximation of the influence.

Based on evidence theory, Refs. [99,100] propose a new methods to identify the influential nodes. In [99] estimates influence in network of networks. This network can be treated as a special multilayer networks and their inter-layer edges only can be used to link the nodes which have the same label but in the different layers. In the new method, evidence theory is used to fuse the influence of node on different single layer networks. Then the result of fusion is used as a quantification to identify the influential nodes in the Network of Networks. Similarly, the approach in [100] considers the degree and the neighbor information of every node in a network. The main advantage of this algorithm is its low time complexity.

Pros: The most significant advantage of the heuristic approaches is to reduce the number of evaluations on the influence spread of nodes which leads to low time complexity than the greedy approaches. Furthermore, these algorithms are achieve a balance between efficiency and effectiveness and more scalable and faster on larger networks.

Cons: However, these algorithms also have some disadvantages, such as they cannot provide any theoretical guarantee to the optimal solutions. These approaches suffer unstable accuracy. As a result, on some datasets they perform much worse than greedy or submodularity-based algorithms. Moreover, since these algorithms have to maintain a large amount of influence centrality, they usually require huge memory.

3.3 Community-based models

Due to the huge size of the real-life social networks, solving the IM problem by considering the network as a single entity, incurs a huge computational burden. A community [101] is a group of people with some common properties. With take advantage of the community structure, researchers have tried to scale down the IM problem to the community level, which performs well in most of the real-world social networks. It is reported in the literature, this category is divided into three phases, as follows in Fig. 3 : (1) At first, the communities of the input graph are extracted; (2) Then, selecting candidates seed in each community and (3) Finally, the node which has the maximum influence spread among candidate nodes is selected as a seed node.

There are several community-based models used by researchers, for this in this subsection, discussion will be limited to only those more recently published works. In [102], first authors identified a limited number of communities. Indeed, the communities are selected according to their ability to influence based on the proximity centrality. Then, proportional to the size of the community, a number of nodes are selected from each community based on degree centrality to from candidates set. Later, ComPath algorithm is proposed to chosen the most influential nodes based on the path length of propagation of influence of candidates set. The INCIM algorithm [103] computes the spread value of each node as a combination of its local and global influences to track the effect of each node in its community and also, the effect of each community in the input graph. Another

Table 4
A brief comparison of node centrality measures.

Metric	Equation	Topology	Metric complexity	Algorithm complexity
Katz [a] [76]	$C_{\mathtt{katz}}(v) = \sum_{k=1}^{\infty} \sum_{u=1}^{n} \alpha^k (A^k)_{uv} + \beta$	Local + Global	$\mathcal{O}(n^3)$	$\mathcal{O}(n + m + k \cdot log(k))$
k-core [b] [85,88]	$k_s(i) = k, s.t.i \in S^k_{\mathtt{core}}, i \notin S^{k+1}_{\mathtt{core}}$	Global	$\mathcal{O}(m)$	$\mathcal{O}(m)$
K-truss [c] [91,92]	$K_s(i) = k, s.t.i \in T^{k-1}_{\mathtt{truss}}, i \notin T^k_{\mathtt{truss}}$	Global	$\mathcal{O}(m)$	$\mathcal{O}(m^{1.5})$
GR, GR+ [d] [93]	$G(i) = \sum_{j \in \psi_i} \frac{k_s(i) k_s(j)}{d^2_{ij}},\ G_{+}(i) = \sum_{j \in \Gamma_i} G(j)$	Global	$\mathcal{O}(n^3)$	$\mathcal{O}(n^3)$
H-index [86]	$H(i) = \arg\max_{h \in \mathcal{N}} \{\forall \lvert \Gamma(j) \rvert \geq h, 1 \leq h \leq \lvert \Gamma(i) \rvert, j \in \Gamma(i)\}$	Semi-local	$\mathcal{O}(n + m)$	$\mathcal{O}(n + m)$
EC [83]	$EC(i) = k_1^{-1} \sum_j A_{ij} x_j$	Global	$\mathcal{O}(n + m)$	$\mathcal{O}(n + m)$
CC [77]	$CC(i) = \frac{n-1}{\sum_{j \neq i} d_{ij}}$	Global	$\mathcal{O}(nm + n^2 logn)$	$\mathcal{O}(nm + n^2 logn)$
DC [78]	$DC(i) = \sum_j^n a_{ij}$	Local	$\mathcal{O}(n)$	$\mathcal{O}(n^2 \cdot log(M))$
PR [e] [63]	$PR_k(i) = \sum_{j=1}^{k} a_{ij} \frac{PR_{k-1}(j)}{k_j^{out}}$	Global	$\mathcal{O}(n + m)$	$\mathcal{O}(n + m)$
BC [84]	$BC(i) = \sum_{s \neq i, s \neq t, i \neq t} \frac{g_{st}(i)}{g_{st}}$	Global	$\mathcal{O}(nm + n^2 logn)$	$\mathcal{O}(g * n * \lvert maf(O^*) \rvert^3)$
SCC [90]	$SCC_i = \sum_j ((2 - C_{i,j}) + (2 \frac{d_j}{d_{max}} + 1))$	Local	$\mathcal{O}(n)$	$\mathcal{O}(n + m)$
LGR [94]	$LG_R(i) = \sum_{d_{ij} \leq R, j \neq i} \frac{k_i k_j}{d^2_{ij}}$	Semi-local	$\mathcal{O}(n^2)$	$\mathcal{O}(n^2)$
INF [95]	$INF(i) = \sum_{j \in \Gamma(i)} \frac{w_{ij}}{k_j}$	Local	$\mathcal{O}(nk)$	$\mathcal{O}(nk)$
DP [f] [60]	$c_i(n+1) = c_i(n) - d_z \cdot w, c_j(n+1) = c_j(n) - d_z \cdot w^2$	Local	$\mathcal{O}(n)$	$\mathcal{O}(l \cdot n)$
LR [g] [89]	$LR_i = LR_i(t_c) + \frac{LR_g(t_c)}{n}$	Global	$\mathcal{O}(n + m)$	$\mathcal{O}(n + m)$
HC [82]	$S_{ij} = A^{(2)} \cdot \lambda A^{(3)}$	Local	–	–

[a] A denotes the adjacency matrix, n denotes the number of nodes, m is the number of edges in the network, k denotes the averaging degree, α denotes the attenuation factor and β is considered a bias constant.
[b] $S^k_{\mathtt{core}}$ represents the node set after removing nodes with degree less or equal to k, $S^{k+1}_{\mathtt{core}}$ represents the node set after removing nodes with degree less or equal to k+1.
[c] Each K-truss is a subgraph of a (k–1)-core.
[d] Γ_i denotes the neighbors of node i, ψ_i denotes the neighbors whose distance to node i is less than or equal to 3, d_{ij} is the shortest path length from node i to node j.
[e] a_{ij} denotes the element at the ith row and j-column of the adjacency matrix, M and T denotes the number of communities and snapshots, respectively. k_j^{out} is the number of edges from node j to i.
[f] $c_i(n)$ denotes the influentiality of node i in the nth step, d_z is degree of node z, w is the punishment rate, l denotes the number of iterations.
[g] $LR_i(t_c)$ denotes the score of node i at time t, $LR_g(t_c)$ denotes the score of the ground node at steady state.

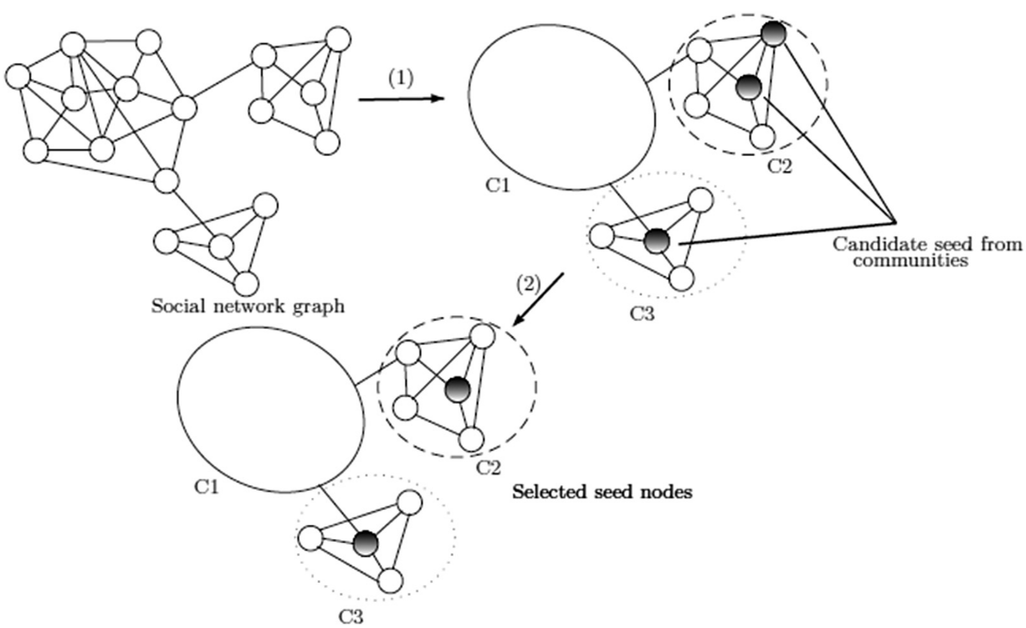

Fig. 3. Flowchart of the community-based solution framework.

earlier study was conducted by [59], called DIN. He described the IM problem in two phases: The first step in the graph partition into communities using a DOCNet algorithm [104]. The next step was weighing the nodes according to the structure and the semantic aspect. Thereafter, PageRank calculation is performed on each community to generate all of the candidate nodes in the community. Finally, the k-top seed nodes are generated based on farness centrality. Shang et al. [62] proposed CoFIM as another community-based solution framework. In this framework, a diffusion model is introduced and it works in two phases: firstly, the seed set was expanded to their neighbor nodes, which would be usually allocated into different communities. Secondly, influence propagation within the communities was computed. Based on this diffusion model, they developed an incremental greedy algorithm for selecting seed set. The advantage of CoFIM is its high speed, and its disadvantage is determining each community's budget without considering the features of the community. Li et al. [105] used a specific geographical location for each user and a spectral clustering technique to identify communities. Then, they adopted a seed selection methodology by considering community-based influence index. They further improved that this methodology is more efficient than many others methodologies, while achieving almost the same influence spread. The study in [106] used the multi-neighbor potential in community networks to solve the problem of maximizing influence. In this approach, the influence diffusion process is divided into two phases: a multi-neighbor potential-based seeds expansion and an intra-community influence propagation. The advantages of the CoFIM algorithm are: relatively high influence spread and low running time but, this algorithm only for non-overlapping communities is suitable. Another extension of community-based solution was proposed in [107]. A three-phases based on community structure are proposed: firstly, the partition phase in which key nodes are found in each community to construct a candidate set by detecting community structure. Secondly, the most potential influence nodes from a candidate set are detected by combining the influence weight of nodes and the community influence of nodes through the analysis of the community structure of the impact on nodes. Thirdly, the nodes with maximization marginal gain from remaining a candidate set are selected. C2im [108] focuses on both user's interest and community structure in order to identify influential nodes. The network is partitioned into communities. Then, authors proposed two new algorithms, a non-desirable nodes finder technique and seed selection algorithm. The first is proposed to identify non-desirable nodes. The second to compute most influential seed nodes based on diffusion degree of nodes. The advantage of this algorithm is that accuracy and effectiveness are relatively high. Bozorgi et al. [109] proposed a community-based solution approach for the same problem under an extended linear threshold model to solve the positive influence maximization problem. Their method was developed based on finding the influential communities by incorporating local and global influences. Huang et al. [110] developed a new community-based influence maximization method for viral marketing that integrates community detection into influence diffusion modeling, instead of performing community detection independently, to improve the performance. They build a comprehensive latent variable model which captures community-level topic interest, item-topic relevance and community membership distribution of each user, and we propose a collapsed Gibbs sampling algorithm to train the model. Later, they infer community-to-community influence strength using topic-irrelevant influence and community topic interest, and further infer user-to-user influence strength using community-to-community influence strength and community membership distribution of each user. Finally, they proposed an algorithm for selecting seed, which selects the influential nodes with a divide-and-conquer strategy, considering both topic-aware and community relevant to enhance quality and improve efficiency. Recently, Beni and Bouyer [111] have proposed TI-SC, another similar community detection heuristic based on scoring criteria. The TI-SC algorithm selects the influential nodes by examining the relationships between the core nodes and the scoring ability of other nodes. After selecting each seed node, the scores are updated to reduce the overlap in selecting the seed nodes. Furthermore, the discovered communities with low expansion are not considered in the seed node selection phase, and this is useful for reducing computational overhead. Recently, Cai et al. [112] takes the community property to accelerated the speed of the influence maximization algorithm. The community-based greedy algorithm, named RSRW, is proposed. The experimental result on four datasets shows the proposed method can reduce the time taken to find the high influential seed node while also ensuing that the selected node has a high influence. The existing community-based methods only consider the number of nodes in a community and ignore the density of edge connections in a community. For this reason, Wu et al. [113] propose community closeness-based influence maximization algorithm (CCIM) to select most influential nodes. CCIM considers the influence of point-to-point and point-to-community, reflecting the micro-level and meso-level influence. Similarly, Huang et al. [114] proposed a new algorithm to solve community-based influence maximization problem in attributed networks, which consists of three steps: community detection, candidate community generation and seed node selection. A similar approach is provided in [115] to address the opinion maximization problem. More recently, Li et al. [116] presents a Dynamic algorithm based on cohesive Entropy for Influence Maximization (DEIM), the goal of which is to find the most influential nodes in social networks. Firstly, the Community Overlap Propagation Algorithm based on Cohesive Entropy (CECOPA) is put forward for the discovery of overlapping communities in networks, and potential nodes in the gathering area are selected to construct the candidate seed set. Then, the Optional Dynamic influence Propagation algorithm (ODP) is designed based on narrowing down the selection range of seeds. It utilizes a variety of entropy calculations to obtain the cohesive power between neighboring nodes and then determines whether the node has the ability to become a propagable pioneer of another node; thus, information continues to diffuse effectively. Finally, via many times experiments on several data sets, it is confirmed that the proposed DEIM algorithm in this paper can successfully affect the ideal number of users in different scenarios.

The community-based solution methodologies use community to study the problem of identifying influential nodes.

Pros: The most significant advantage of the community-based solution methodologies is the good tradeoff between the influence spread and the running time. The efficiency of these algorithms outperforms the general greedy algorithms largely in scalability and fastness, as they assume the independence between communities and suits to run parallelized (shortly discussed in Section 3.4).

Cons: One of their major findings observed that the community structure can achieve a good tradeoff between the influence spread and the running time. On the contrary, the proposed algorithms demonstrate several major drawbacks: Firstly, they have no effective methods to reduce the search space the for selecting seed nodes in the large networks. Secondly, they suffer from the problems of the overlap of seed nodes. Thirdly, they do not consider the role of core nodes in the influence spread.

3.4 Parallel models

Parallel algorithms are algorithms that do not follow iterative execution with a simple loop, but they perform iterations in parallel while taking advantage of the machine's graphical architecture where they typically run on multiple CPU processors and sometimes on GPU. They are not heavy and not very complex and sometimes extremely fast. Despite this advantage, most of the algorithms for identifying influential nodes presented in the literature are serial in nature. The issue of scalability in influence maximization problem can be tackled by developing distributed and parallel algorithms. To the best of the authors' knowledge, except a few parallel algorithms have been developed recently, there are no distributed algorithms existing in the literature. So, this is an open area to study the influence maximization problem and its variants under parallel and distributed settings [117].

To increase the speed of the greedy algorithm based on an approximation of hope, Liu et al. presented a framework called IMGPU [118] that accelerates the maximization of influence by using the parallel processing capability of a graphics processing unit (GPU). The first step, in this application, is the use of a directed acyclic graph to efficiently convert the social graph and to avoid redundant calculations and the second one consists in mapping the inherent parallelism by using an ascending path algorithm. The IMGPU algorithm significantly reduces the execution time of the existing sequential influence maximization algorithm while maintaining good influence propagation. Zong et al. in [119] proposed an incremental updating method based on IRIE, called dIRIEr, to reduce the overhead of repeated computation in IM problem. Therefore, in [120] authors proposed a divide-and-conquer strategy with parallel computing mechanism to tackle the influence maximization in mobile social network. To coarsely estimate the influence spread to avoid massive estimation of heat diffusion process, Wu et al. [121] proposed a modified algorithm of greedy, called CSIM. Indeed, the k-shell decomposition method is employed to divide a social network and generate the candidate shells. Then, the heat diffusion model is used to model the influence spread. Finally, the seeds of candidate shells are selected in parallel by using the CSIM algorithm. In the same context, authors in [122] implemented a parallel algorithm for identifying influential nodes that is also capable of running on multi-GPU systems. Recently, Xiao et al. [123] developed an efficient parallel algorithm for detecting influential nodes for large biological networks by exploiting the massive computing capability of a modern GPU. The essential concept behind this work is that several computationally expensive procedures in detecting influential nodes are redesigned and transformed into quite efficient GPU-accelerated primitives such as parallel sort, scan, and reduction. To measure the nodal influence, four local metrics are used : Degree Centrality, Companion Behavior, Clustering Coefficient, and H-Index.

Pros: The work applies the parallelism to identify the influential nodes significantly reduced the time consumption compared to other algorithms. The results in most algorithms show a good acceleration effect with parallel cpu, which effectively improves the time performance and memory usage.

Cons: Despite the advantages of parallel algorithms, little work applies the parallelism to identify the influential nodes. This is due to the fact that the high degree of data dependency in social networks.

4 Group-based influence analysis

In the majority of early existing work, it is looked only the individual and its effect in the study of social influence analysis. However, one important aspect has been neglected, that is, the group of individuals despite the use of the community structure. In fact, understanding people's group is critical to understand people's personal behaviors. In addition, the human behavior is mostly easily influenced by their group's behavior and most of the world's decisions or works are done by groups or teams. This concept of social graph is a different from the classic community, which pays more attention to diffusion proximity properties of nodes besides cohesive properties in structure. In real-world online social network, group influence can be examined in two aspects: (1) The influence intra-group occurs when most people share the content which published by most of theirs friends, and their final decision is the decision made by the majority and automatically goes in the same direction as the group, even without knowing all the details. (2) In another way, the influence inter-group can be seen by a mutual influence is among the groups, or a group that has already decided, can offer advice to a group that has not yet.

As discussed earlier, the influence of a group is stronger than the influence of an individual thanks to the quantity, because usually humans follow most of the time the majority and do not risk to go towards the unknown. Since the group presents a dense interconnection, even if the individual has a different opinion from the others when he interacts with the strong dominant majority he will be weak in front of this dominant majority and head towards the same meaning. For this, government or company tries to influence group rather to care more about personal influence in order to obtain a maximum benefit. In fact, to be able to influence by group, we must first maximize the influence within this group itself by selecting certain seed node influencers locally in each group and then through the activated groups or influencers maximize the influence in the global social network. In this respect, detection of influential groups is a new variant of IM problem, and the group influence maximization (GIM) problem is introduced. Mathematically, the GIM problem [124] can be described as follows:

Definition 2 (*Group Influence Maximization*)**.** For the given network $G = (V, E)$, a group U is defined as a subset of V, an positive number $k < |V|$, propagation probability P_{uv} between each edge uv and the diffusion process M. Let $\mathcal{U}$ be the set of groups and l be the number of total groups. Given an activation threshold $0 < \beta < 1$, a group is said to be activated if β percent of nodes in this group are activated under M model. The objective of the GIM problem is to look for k initial seed users where the expected number of eventually activated groups is maximized.

$$\arg\max_{|S|=k} \quad \rho(S) \tag{4}$$

where S is the initial seed set and $\rho(S)$ is the expected number of groups eventually activated for given initial seed set S.

This section discusses the existing work on identifying influential group in social networks, which is a relatively new research area. Contrary to conventional models which solve the problem of Competitive Influence Maximization (CIM) by neglecting the importance and effect of groups, and which uses individual nodes as a unit, then calculate the number of influenced nodes globally to represent the propagation of influence, Dai et al. [125] introduced the concept of the group for measuring influence in a social network. Indeed, they used groups as a unit and a metric to maximize influence spread, instead of using them only to reduce the size and complexity of the social graph. On this basis, the authors underlined the group-based competitive influence maximization (GCIM) problem. The competitive group-linear threshold (CG-LT) model was proposed which takes advantage of the local effect and analyzes it under two fundamental aspects. The first aspect, the intra-group, means that the individuals or the nodes in the same community will influence each other in the form of "word-of mouth", and the final decision in this community is made by the majority. The group will be activated if the activated nodes reach a certain threshold. In CG-LT, the node in group have four different states (positively active, negatively active, neutral, inactive). The node will be activated when his activated intra-group neighbors reaches a certain threshold. The second aspect is the inter-group effect, whose influence is between the groups. The idea means that the group does not consider the node impact outside the group and integrates them into group influences. In addition, the authors proved that group-based influence computation in the LT model is NP-hard and then provided performance guarantees for a greedy strategy named SGG algorithm. There are three steps in their approach: First, partitions large social networks into subgraphs using the Louvain algorithm. In the second stage, the seed nodes in a group are selected using a greedy algorithm. Last, the selection of the groups is done using the principle of the inter-group already explained. By taking the advantage of the local effect, the proposal avoids the loss and the waste of resources. Thus, the good quality in the selection of the seed groups on which one can really influence the decisions of other groups. However, the algorithm has high execution time because the PageRank calculations and that the local effect decreases with the increase of the factor of influence.

The framework proposed by [126] has combined the benefit of activated groups with the diffusion cost of influence propagation to select k seed users to maximize the expected profit. They trained a information diffusion model based on Independent Cascade (IC). In addition, they formulated their description as an optimization problem, proved that it was NP-hard, and the objective function is neither submodular nor supermodular. They designed an upper bound and a lower bound that both are difference of two submodular functions. Then they proposed an Submodular–Modular Algorithm (SMA) for solving difference of submodular functions and SMA is proved to converge to local optimal. Further, they presented an randomized algorithm based on weighted group coverage maximization for GPM and apply Sandwich framework to get theoretical results. In addition, zhu et al. experiments verify the effectiveness of the proposed method, as well as the advantage against the other heuristic methods.

Zhu et al. [124] were also interested in solving the same problem by taking the advantage of groups. Indeed, they proved that it is better to maximize the number of active groups by selecting k-top seed groups in the social network which give a greater diffusion of influence than to be interested in the selection of k-top seed nodes. However, a group is only activated if $\beta\%$ of its nodes are already activated. Their solution includes the IC model in the influence propagation calculation and an objective function, neither sub-modular nor super-modular with two limits, an upper limit and a lower limit. The authors proved that it was NP-hard, and then proposed a framework to run their algorithm, called Sandwich, where GIM has shown good performance. The advantage of the algorithm is that it gives more precise results than the node-based models and the selected seed groups are of good quality, thanks to the maximization of the number of active groups. However, we had observed numerous disadvantages including the disregard of the semantic aspect.

In order to benefit the group property in social influence analysis, Ji et al. [127] conducted a study with network embedding method over the same problem. Based on the embedding vector of nodes, the diffusion proximity of nodes is easily and efficiently defined and calculated. Then the group detection is performed through network coarsening based on diffusion proximity between adjacent nodes. As a result, the nodes with large diffusion proximity are merged into the same group. The authors improved the two-stage diffusion model and also presented a fast approximate calculation formula of influence spread. Specifically, in the second stage of the diffusion model, the influence of nodes within the group is approximated to the same value and is proportional to the size of the group so that the approximate calculation of the influence spread is more accurate and reasonable. Because the objective function is proved to be monotonic and submodular, the seed nodes are selected by a greedy strategy. Their experiment results showed that the GIM is much more time efficient than classic greedy algorithm CELF with similar accuracy and achieves better accuracy than corresponding community-based influence IM algorithm with matching running time. Huang et al. [128] attempted to further improve the efficiency of group from the following two directions: (1) detected groups with the method of merging adjacent nodes which have similar influence characteristics. The nodes within groups obtained by this method have similar influence. (2) the influence spread of nodes within groups are approximated based on group structures. In this procedure, the influence spread of each node is calculated on granularities of groups. At last, the nodes with widest influence spread are selected as seed nodes with greedy algorithm. A group-based influence maximization method is proposed in [129] to solve the IM problem over the conformity-aware diffusion model which utilizes different types of conformity behaviors (where people conform to the opinions and actions of others by submitting to perceived group pressure) based on user profiles and group profiling. Recently, GIN (Group of Influential Nodes) algorithm is presented in [130]. This algorithm creates different groups of graph nodes with more connections than other groups. Then, it selects specific nodes from each group to reduce the search space to find the most influential nodes. Following the greedy method, it selects the seed nodes with the highest expected diffusion value. Experimental results show that the GIN algorithm has provided high influence spread along with low running time in comparison algorithms on all seven real-world datasets. More recently, Zhong and Guo [131] studied the problem of group influence maximization and they focused on activating groups rather than individuals. They divide an algorithm called Complementary Maximum Coverage (CMC) based on analyzing the influence of the nodes over the groups, ensuring the task of maximizing the number of activated groups. In addition, they also proposed an algorithm called Improved Reverse Influence Sampling (IRIS) via adjusting the famous Reverse Influence Sampling (RIS) algorithm for GIM. Lastly, experiments are carried out to demonstrate that our CMC and IRIS both outperform the known baselines including Maximum Coverage and Maximum Out-degree algorithms in the average number of activated groups under IC model.

In summary, the GIM problem aims to maximize the number of eventually activated groups under a model. Generally, these approaches firstly supposed that there are many definite groups and then they aim to solve how to activate these groups as more as possible, where a group is said to be activated if a certain ratio of nodes in this group are activated.

Pros: The experiment results of the proposed methods showed that on real-world datasets, indicated there are performs better than some algorithms. The major advantages of IGIM is high precision and the low complexity because the properties of the nodes and the size of the groups.

Cons: However many times, it is hard to get the group in advance in these approaches. Yet, these approaches are not observe the semantics properties of social behavior.

5 Influence maximization models facing the privacy-preserving technologies

Generally, these models based either on the individual node or based on group of nodes using the principles of information dissemination (LT, IC, Vote, Competitive, etc.) are all interested in finding only the k-top seeds in a social network while neglecting the mechanisms (automatic or manual) of privacy in real applications.

5.1 Privacy background

In today's world, the social network is growing and people profile started to explore. As a result, people can now influence friends based on the place we live, common interests, common friends, and so on. Therefore, the social influence interconnects with the people profile features. Recently, the people profile features provided with rich personal information including the date of birth, gender, profession, current area, payment method, sexual orientation, political views and religious preferences and photos. Subsequently, the public profile of a user has become an important factor of social influence. That is, without public information about a user, it is difficult to find things or people that he would be influenced. However, public information is conflicting with viruses, rumors, fake news, and this huge amount of personal data is vulnerable to attacks. Thus, users understand the risk of releasing private data and the confidence of shared data and publications is seriously diminished. Indeed, it has become not easy to let users share and propagate information despite they are influenced, because this increasing the risk of private information being leaked to unauthorized persons. For instance, malicious individuals can analyze users' lives by following their published contents and collecting other sensitive information and once users' significant information is obtained, it may lead to computer-assisted crime (harassment, kidnapping, etc.). As a result of this increased privacy anxiety, users have reservations about the transmission of their own profiles features to their neighbors. Obviously, because of privacy considerations, not all affected nodes will try to affect others by reposting the information, one possible case is that they are influenced but stop to repost the information. As a result of this increasing privacy concern, the social influence analysis has become more difficult since the above works are broadly launched in an insecure environment.

5.2 A taxonomy of privacy protection problems

User privacy protection is becoming an important part of social networks and it recently has gained wide acclaim in the research community. The data privacy not only automatic tools but also manual efforts are taken to protect privacy of users. At present, the privacy protection scheme for social networks includes privacy protection of two main graph components: nodes and edges. Privacy protection of node includes the following three types: anonymity, addition, and deletion. The next component of privacy is the link/edge privacy includes the edge perturbation. Doubtless, an extensive review of existing measures for modeling and analyzing privacy protection in the context of publishing social networks data, goes beyond the scope of the current paper and the interested reader is referred to the specialized literature, for example, [132–134]. Instead, hereafter, we will try to briefly categorize the existing works based on the privacy protection of graph components. These categories are described as follows: (1) nodes privacy protection, (2) edges privacy protection and (3) graph (nodes and edges) privacy protection.

In the first category, the graph structure is studied to prevent structure information attacks. For this, the anonymous version, based on the node privacy protection algorithms, is published instead of the original graph. Most of the existing node privacy protection algorithms are essentially similar to K-anonymity schemes. Mostly, this anonymous version should satisfy k-anonymity and it ensures that an attacker, with a certain background, cannot identify an individual with a probability greater than $\frac{1}{k}$ in the anonymous graph. According to the attacker's background, several techniques have been studied including k-degree anonymity, k-neighborhood anonymity, and k-automorphism anonymity.

The second category is the edge perturbation. This category is based on edges privacy protection, in which the original graph is modified to produce the disturbed one. The modification is generally based on methods that adds noise to the data in the form of random additions, deletions or switching of edges.

In the third category, the entire graph is considered. This category could be roughly classified into three classes, which are discussed in the following. The first class of solutions is a clustered graph and it aims to partition the graph into clusters. First, n clusters are created so that each cluster contains at least k similar nodes. Then the nodes of each cluster are modified to become indistinguishable and the edges are generalized. The second class of solutions is uncertain graph. In this last, the original deterministic graph is converted into an uncertain graph by associating probabilities to edges (the probability represents the uncertainty level of the edge presence in the original graph). Finally, the third class uses both statistics and responses to queries over the social network. To ensure that an attacker cannot deduce the presence of an individual or a link, using these statistics and responses to queries.

5.3 IM under privacy protection

Generally, privacy-preserving influence maximization is itself an important, interesting, yet often overlooked problem, in spite of a large body of work on privacy-preserving social network mining. Social influence analysis under the users' privacy is facing two major challenges for researchers. The first is related to the protection of the user's personal information and to identify the type of information to be protected through mechanisms for the protection of privacy. The second challenge is related to the models used to propagate and disseminate the information, which must be able to maximize the influence and the propagation of the influence while taking into consideration the concerns of confidentiality. Recently, this problem has started to attract the attention of researchers, but the number of models proposed to solve the problem of influence maximization while taking into account the protection of the privacy of users is seriously limited. With regard to the type of existing models, few researchers focused on IM under privacy protection, the specific are listed as follows. Tassa and Bonchi [135] were the first to study how to learn the influence strength while keeping their data secret. The authors proposed two set of protocols to jointly compute the social strength along edges of the network. A secure protocols for link influence strength perform the computations while preserving their private information. Thereafter, a secure distributed protocols for computing user influence scores. Their method is certainly appealing and useful, but it does not account for purchases in third-party stores, and hence the resultant estimations may be biased. In [136], the authors proposed novel model merging both GPS data of cyber–physical network and relationship data of online social network together in a unified framework. Then, they provided an algorithm to solve the influence maximization problem in the framework. In a different fashion, the same authors in [137] have crafted a social influence graph while protecting individual privacy. This is done by anonymization and context-related Trust, by referring to connections between users in different topics. In this graph each edge is enriched by a topic model that represents the strength of the social influence along the edge per topic. This information-rich graph is obviously much harder to anonymize than standard graphs. Besides, the identity of nodes in the network is obfuscated by randomly perturbing the network structure and the topic model. A framework that considers the GPS-enabled cyber–physical to solve the influence maximization problem with privacy reservation is presented in [138]. By employ a random walk process on cyber–physical network, a heterogeneous information framework combines both cyber–physical and online social network together is proposed. Then, to protect user's privacy of location information, a differential privacy mechanism is employed. This mechanism could protect the location together with other sensitive link privacy with a guaranteed parameter. Finally, to solve the influence maximization problem under the heterogeneous framework, an efficient and effective approach is employed. In [139], Zhu et al. proposed SHIM, a structure hole-based influence maximization algorithm which utilizes structure hole as the centrality measurement. They first identified structure hole spanner whose structure hole value is above the given threshold, followed by computing the influence capability of each structure hole spanner. After that, they selected the top-k seeds by combining the structure hole value and influence value. The work in [140] studies the IM problem towards both sensed location data and online social data with privacy concern, and proposes an algorithm. First, the new heterogeneous network is built, in which the data collected from both sensed cyber–physical world and the online social world are integrated. Through the twofold privacy protection: first integrating the sensed cyber–physical world and the online social world by random walk, and hiding the sensitive user behavior patterns; then applying differential privacy technology to protect the connection privacy within the heteromerous network. Actually, the privacy of the user in the physical world is hidden behind the heterecious network since the new construct network will not tell whether the connection between two users is their common location, common activity, or their online friendship. Finally, a martingale approach is employed to further preserve the relationship privacy for influence maximization in the heterogeneous network. Liu et al. [141] presented a privacy protection scheme for social networks based on collective influence. This collective influence is used as the noise sources on edge weight perturbation after randomization and normalization. Then, they introduced a node perturbation strategy based on edge weight perturbation strategy. Later, a context-based information diffusion model was proposed by Jing and Liu [142] to resolve the privacy issue. Based on user's decision the authors developed an monotonous and sub-modular greedy approximation algorithm with privacy consideration. The advantages of the model are the low filtering rate of the published data, and the precise and exact results in the selection of seed nodes. Contrary, one of the drawbacks is the high cost of calculation. Recently, Kukkala and Iyengar [143] presented an efficient secure multiparty computation protocols for evaluating k-Shell decomposition, PageRank and VoteRank algorithms and identify influential nodes. More recently, Qi et al. [144] proposed a network-based method with privacy-preserving for identifying influential providers in large healthcare service systems. First, the provider-interacting network is constructed by employing publicly available information on locations and types of healthcare services of providers. Second, the ranking of nodes in the generated provider-interacting network is conducted in parallel on the basis of four nodal influence metrics. Third, the impact of the top-ranked influential nodes in the provider-interacting network is evaluated using three indicators. Their experiment results showed that the provider-interacting network of healthcare service providers can be roughly created according to the locations and the publicly available types of healthcare services, without the need for personally private electronic medical claims, thus protecting the privacy of patients.

In recent times, the privacy preservation is one of the most researched problems in the IM problem. Even though a few algorithms are proposed and consider the privacy protection problem during the information diffusion process, they suffer from many drawbacks.

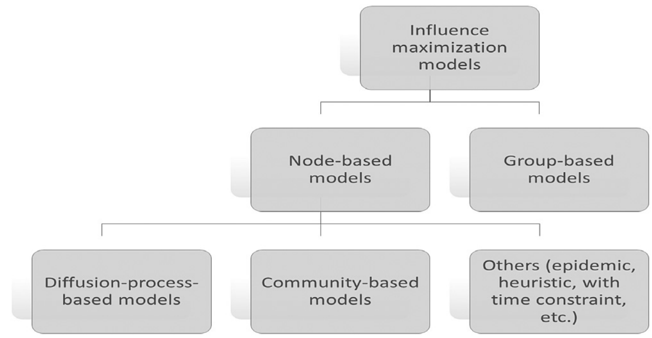

Fig. 4. General architecture of influence maximization models.

Pros: The major advantage of these approaches is studying the structural properties to solve the privacy protection in the network. Most of previous works could solve the problem efficiently and effectively.

Cons: The major problem of these approaches does not observe the propagation models but used only the structures. They usually build specific networks by combining others information. These approaches take the advantage of other structures without a specific propagation model.

In summary, a brief overview of these influence maximization models is presented in Fig. 4 and Table 5.

6 Applications of influence maximization models

For different needs and in different fields, the social influence analysis has immense application potential in several domains and it could bring many benefits. This section discusses the application of influence maximization for social networks.

In viral marketing, the advertising company gives a free sample of a product to the group of users for maximizing the product adoption using the word-of-mouth effect. So, the influential users should accept the product and increase this acceptance among the other members in a cascading mode. Today, many researches have been done in order to find influential users in social networks for the purpose of viral marketing. In all the researches, the aim is to find a set of users who are able to demonstrate enough influence among the other members of the network and reduce the expenses of marketing a product. In [1], Domingos and Richardson investigated social influence in the customer network. They proposed a model to identify customer's influence between each other in the customer network, and built a probabilistic model to mine the spread of influence for viral marketing. The paper [145] presents a survey of previous studies done on influence maximization in viral marketing.

In information recommendation, the influential users in social network are able to change the opinions and ideas of others and reversing his decisions on the network. In this context, Zhu et al. [146] presented a method to identify a set of k-top users and the complete set of experiment results are released to measure social influence in online recommender systems. Specifically, they also quantified how often people's choices are changed by others' recommendations when facing different levels of confirmation and conformity pressures. However, the recommendation can be used in viral marketing, advertising a trademark, product, person or political party, to avoid and destroy rumors, etc. For example, we find the model proposed by Goyal et al. [147] named RECMAX, which consists of identifying the set of k-top seed nodes for which a proportion is proposed and then through their recommended publications they maximize the spread of influence in social networks with the aim of maximizing the market. In [3], the authors

Table 5
Categorization of IM related works discussed in this article.

Problem	Approach	Example methods
Individual-level	Submodularity-based algorithms	[5,19–56]
	Heuristic-based algorithms	[60,63–71,76–95,99,100]
	Community-based algorithms	[59,62,102,103,105–116]
	Parallel algorithms	[97,118–123]
Group-level		[124–130]
IM under privacy protection		[135,137–144]

showed that the mined social influence and user preferences are valuable for group recommendation and viral marketing.

In rumor control, the negative information about a product propagating in OSNs can cause huge losses to a company. Similarly, the rumor that "Two bombs had exploded at the White House, injuring Barack Obama" from hacked Associated Press Twitter account resulted to 10 billion USD losses [4]. Therefore, effective methods to restrain or minimize the influence of rumors have been a popular topic. For example, there are many gossips related to coronavirus (COVID-19) spreading in the social network recently. In March 2020, someone in China posted a rumor said that the outbreak of COVID-19 might trigger a global grain crisis, we should hoard grain for at least three months since china is a big importer of grain, which caused panic among the public. Thousands of people in china went to the supermarket to scramble for grain after the rumor got out. In England, also in March 2020, some people said that the 5G signal is accelerating the spread of COVID-19. Some criminals in England went to damage the phone masts and telecoms engineers abused their power, which happened apparently as inspired by the rumor circulating online [148].

In network monitoring, the monitoring sensors can be installed to better prevent grave accident in the oil and gas pipeline network. For example, Leskovec et al. [5] proposed a novel methodology for selecting nodes to detect outbreaks of dynamic processes spreading over a graph.

In revenue maximization, the highest revenue by properly pricing the product and/or seeding customers are earned. Under the consideration of the effects of social influences, Teng et al. [7] studied the problem of revenue maximization (RM), which is one of the most important utilization of social influences in marketing.

In micro-blog marketing, the influential users in social network plays an important role in the 2016 US election where a study showed that on average 92% of people remembered pro-Trump fake news and 23% of them remembered pro-Clinton fake news. In fact, Twitter, as one of the most famous micro-blog web site, many algorithms are developed to analyze the data of it. For example, Mei et al. [149] used user interactions, and information propagation on Twitter to calculate the influence probability according to users' action history including tweet, favorite, mention/reply, and retweet. Similarly, Jendoubi et al. [150] proposed two evidential influence maximization models for Twitter social network.

In other applications, Bouguessa and Ben Romdhane [151] proposed a model to select the vertices of authority in each online community, with the aim of selecting at the end the *k*-top vertices of authority in a social network. In fact, they have shown that these selected vertices can be considered experts in the social network and they are capable of producing a continuous high quality. A probabilistic model is proposed by Tang et al. [152] for analyzed a special type of social influence which involves a change in opinion or behavior in order to fit in with a group called conformity. Recent work tries to study influence maximization by only considering the influence of users' spatio-temporal behavior on information propagation or location promotion. In [153], Chen et al. proposed a similarity-aware influence maximization model to efficiently maximize the influence spread by taking the effect of users' spatio-temporal behavior into account, which is more reasonable to describe the real information propagation. In [154], the authors investigated usage patterns and the manifestation of glass ceiling effects in different interaction types on Instagram and Facebook interaction datasets. They discovered correlations between gender and both high visibility as well as active endorsement by jointly considering the number of interaction partners and the intensity of direct and indirect interactions.

7 Challenges & future directions

During the last decade, influence maximization has been widely studied and there exist plenty of methods and techniques with promising findings that are proposed to deal with this problem from different perspectives and domains. It is not only shown the significant role of influentials in formatting and shaping opinions of others, but also in accelerating, controlling, and disseminating information flow. However, the influence measurements have become a necessity. Despite its significant evolution for several years, still has difficulties and challenges to overcome, and future objectives to achieve. This section demonstrates several challenges and open issues in IM.

7.1 Challenges and open issues

- *Dynamics*: As is known, dynamics is an unavoidable topic in social networks. In fact, the majority of existing models are models that work on static networks while ignoring the dynamism of social networks. Even the models that are interested in dynamic networks, in reality, still work on static networks because they take a snapshot of the social network each time, and they do not analyze the dynamicity of the network in real-time. Indeed, a real social network can evolve very quickly, then even using the snapshot mechanism it has become difficult with this rapid evolution because there can be big changes in the relationships between nodes and the structure of the network at each snapshot and in a different way from one moment to another. However, it is very crucial to deal with these changes. Analyzing the data every time to make a comparison and generate an effective model is not easy with real social networks. This uncertainty due to the varied evolutionary dynamics of social networks over time has attracted researchers, but the number of models offered remains limited, such as [14,155]. However, solving this problem with dynamic models operating in real-time is one of the main challenges.
- *Heterogeneity*: The diversity of social media platforms is one of the known challenges. In fact, each platform covers its own range of user activities, semantics, factors of friendship. There are numerous relevant models most of which are based on a homogeneous network. However, in the real world, we face heterogeneous networks where the nodes and edges are different types. A network is homogeneous if

and only if the edges and nodes are of the same type, and it is considered heterogeneous if the nodes and edges are different. In heterogeneous networks, there is a concept known as meta-path, which indicates the type of communication between two nodes. Moreover, the assumption of network homogeneity may lead to loss of key semantic information and failure to understand the information in its entirety. Therefore, it is necessary to define a new paradigm to study heterogeneous networks and extend the recently developed tools in the field of network science [156]. Identifying influential nodes in heterogeneous networks is recently analyzed in [157]. In this work, the authors focus on the concepts of entropy and meta-path to locate influential nodes in a homogeneous network.
- *Positive and negative influence*: Current studies of influence maximization focus almost exclusively on unsigned social networks ignoring the polarities of the relationships between users. Actually, however, the relationships own the property of the polarity, which can divided into positive relationships (e.g., friend or like) and negative relationships (e.g., enemy or dislike). The social networks including both positive relationships and negative relationships simultaneously are called as signed social networks. In the study of influence maximization, ignoring the relationship polarities between users and treating signed social networks as unsigned ones roughly lead to over-estimation of positive influence which will cause bad effect in practical applications. Influence maximization in signed social networks is a critical problem that remains pretty much open. The goal of influence maximization in signed social networks is usually to find the seed node set with maximum positive or negative influence. There is a few research work related with influence maximization in signed social networks, such as [31,40,158].

7.2 Future directions

Motivated by our detailed survey on research studies on the influence maximization in social networks, we point out possible research directions which should be considered in the future works.

- *Selfishness and privacy protection*: Protecting users' privacy is an obstacle that the vast majority of researchers have overlooked when proposing models for maximizing influence in social networks. Thus, selfishness and privacy protection problems in IM have gained extensive attention. In fact, it is remarkable that users no longer give confidence more easily to interact with the content published on the network despite the fact that they are well influenced by this content. Indeed, the main reason is that they do not want to clearly share their personal information in front of everyone, or that they do not want to give certain personal information but they can only give limited information, in other words, they only give what they want to give. However, with this lack of information provided by users and this fear of interacting with publications and published content, the dissemination of influence in networks will be limited regardless of the classic influence maximization model that ignores this obstacle. Nevertheless, how to establish effective incentive mechanisms is challenging research focus on influential user detection. As shown in Section 5, few scholars have spent a lot of effort on the selfishness and privacy protection problems.
- *Group-based influence maximization*: Group-based influence maximization is a new direction of research, although the use of communities has existed since the definition of the phenomenon of maximizing influence. Indeed, the majority of existing models are interested in the effect of the node in maximizing the propagation of information while ignoring the effect exerted by the group. On the other hand, reality and real social networks prove that the group plays a very important role in maximizing the spread of information and even more better and more effective than the role played by the effect of knots. Indeed, taking the example of buying a product, the influence of a group is stronger than a knot to convince someone to buy this product. In other words, the majority always has the strongest effect on changing and orienting a person's opinions. In fact, we can notice this clearly in the political field, when the majority support a person or political party, it can maximize the influence by encouraging other people to support this politician or this political party even if they do not know everything about this politician or this political party, more efficiently than a single node. Hence, GIM approaches are best suited for mega-scale networks and they are listed in Section 4. But their solutions remain purely structural, and the real networks are based on heterogeneity between semantics and structure. However, proposing a new measure of influence and a new model of maximizing the dissemination of information based on groups using semantics alongside the structure is another difficult and important challenge that can make orientation and evolution towards a new generation of influence maximization models.
- *Parallel methods*: Parallel algorithms are algorithms that do not follow iterative execution with a simple loop, but they perform iterations in parallel while taking advantage of the machine's graphical architecture where they typically run on multiple CPU processors and sometimes on GPU. They are not heavy and not very complex and sometimes extremely fast. Despite this advantage, most of the algorithms for identifying influential nodes presented in the literature are serial in nature. The issue of scalability in influence maximization problem can be tackled by developing distributed and parallel algorithms. To the best of the authors' knowledge, except a few parallel algorithms have been developed recently (see Section 3.4), there are no distributed algorithms existing in the literature. So, this is an open area to study the influence maximization problem and its variants under parallel and distributed settings.
- *Multiplex online social networks*: Reviewing the literature, an interesting research direction arises in which a streaming graph scenario would be considered is the influence maximization in multiplex online social networks. Nowadays, individuals are joining multiple online social networks and information simultaneously propagating amongst them, thereby creating a new dimension to the problem of influence maximization in multiplex OSNs. Recently, a few works [159–161] have investigated the IM problem in multiplex OSNs. However, as indicated by these recent studies, modern networked systems tend to have more complicated structures and multiple layers, which makes it difficult for existing seed determination techniques to deal with these networks. Thus, finding influential nodes in realistic multiplex networks remains open.

8 Conclusion

In this paper, we have reviewed current solutions from traditional individual-level influence maximization problem, towards the group-level, and then reach the influence maximization problem under privacy protection. In the taxonomy of individual-level social influence models, we distinguish four types of models: submodularity-based algorithms, heuristic-based algorithms, community-based algorithms and parallel algorithms. This survey demonstrates that by proposing different algorithms, methodologies, and frameworks, numerous research works have attempted to solve the individual-level influence maximization problem. This work is motivated by the lack of a comprehensive review on the IM group-level and IM under privacy protection. Particularly, we have explored and analyzed in detail these two issues. We also analyze the strengths and weaknesses of current models and methods. Throughout our study, we have pointed out several research challenges and outlined potential research directions. Finally, we have pointed out several research challenges and outlined potential research directions.

In social networks, mutual impact among users is common and inevitable. Identification of influential users is a critical field to accommodate the needs of social networks applications. Our study addresses and compares a broad range of many new problems and challenges to pervasive and effective social influence analysis. Furthermore, we believe many more problems and challenges will show up during the progress of the social network journey from both theoretical and practical perspectives. It is expected that this article will be useful for the readers and researchers to better understand the social influence analysis and motivate researches in the field.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] P. Domingos, M. Richardson, Mining the network value of customers, in: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2001, pp. 57–66, http://dx.doi.org/10.1145/502512.502525.
[2] D. Kempe, J. Kleinberg, É. Tardos, Maximizing the spread of influence through a social network, in: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2003, pp. 137–146, http://dx.doi.org/10.1145/956750.956769.
[3] M. Ye, X. Liu, W.-C. Lee, Exploring social influence for recommendation: A generative model approach, in: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 2012, pp. 671–680, http://dx.doi.org/10.1145/2348283.2348373.
[4] P. Wu, L. Pan, Scalable influence blocking maximization in social networks under competitive independent cascade models, Comput. Netw. 123 (2017) 38–50, http://dx.doi.org/10.1016/j.comnet.2017.05.004.
[5] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, N. Glance, Cost-effective outbreak detection in networks, in: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2007, pp. 420–429, http://dx.doi.org/10.1145/1281192.1281239.
[6] X. Tang, C.C. Yang, Ranking user influence in healthcare social media, ACM Trans. Intell. Syst. Technol. 3 (4) (2012) 1–21, http://dx.doi.org/10.1145/2337542.2337558.
[7] Y.W. Teng, C.H. Tai, P.S. Yu, M.S. Chen, Revenue maximization on the multi-grade product, in: Proceedings of the 2018 SIAM International Conference on Data Mining, 2018, pp. 576–584, http://dx.doi.org/10.1137/1.9781611975321.65.
[8] Hootsuite, The top 100 CEOs on social media, 2020, https://blog.hootsuite.com/top-100-ceos-social-media/ (Accessed January 2021).
[9] S. Peng, Y. Zhou, L. Cao, S. Yu, J. Niu, W. Jia, Influence analysis in social networks: a survey, J. Netw. Comput. Appl. 106 (2018) 17–32, http://dx.doi.org/10.1016/j.jnca.2018.01.005.
[10] Y. Li, J. Fan, Y. Wang, K. Tan, Influence maximization on social graphs: A survey, IEEE Trans. Knowl. Data Eng. 30 (10) (2018) 1852–1872, http://dx.doi.org/10.1109/TKDE.2018.2807843, 1852-1872.
[11] K. Li, L. Zhang, H. Huang, Social influence analysis: models, methods, and evaluation, Engineering 4 (1) (2018) 40–46, http://dx.doi.org/10.1016/j.eng.2018.02.004.
[12] G. Razis, I. Anagnostopoulos, S. Zeadally, Modeling influence with semantics in social networks: A survey, ACM Comput. Surv. 53 (1) (2020) 7, http://dx.doi.org/10.1145/3369780, 38 pages.
[13] S. Banerjee, M. Jenamani, D.K. Pratihar, A survey on influence maximization in a social network, Knowl. Inf. Syst. 62 (2020) 3417–3455, http://dx.doi.org/10.1007/s10115-020-01461-4.
[14] N. Hafiene, W. Karoui, L. Ben Romdhane, Influential nodes detection in dynamic social networks: A survey, Expert Syst. Appl. 159 (2020) 113642, http://dx.doi.org/10.1016/j.eswa.2020.113642.
[15] M. Granovetter, Threshold models of collective behavior, Am. J. Sociol. 83 (6) (1978) 1420–1443, https://www.jstor.org/stable/2778111.
[16] J. Goldenberg, B. Libai, E. Muller, Talk of the network: a complex systems look at the underlying process of word-of-mouth, Mark. Lett. 12 (3) (2001) 211–223, http://dx.doi.org/10.1023/A:1011122126881.
[17] G. Ritzer (Ed.), The Blackwell Encyclopedia of Sociology, Vol. 1479, Blackwell, Malden, MA, 2007, https://www.blackwellpublishing.co.uk/pdf/Sociology_Catalogue.pdf.
[18] R. Bian, Y.S. Koh, G. Dobbie, A. Divoli, Identifying top-k nodes in social networks: A survey, ACM Comput. Surv. 52 (1) (2019) 1–33, http://dx.doi.org/10.1145/3301286.
[19] D. Kempe, J. Kleinberg, É. Tardos, Influential nodes in a diffusion model for social networks, in: International Colloquium on Automata, Languages, and Programming, Springer, 2005, pp. 1127–1138, http://dx.doi.org/10.1007/11523468_91.
[20] M. Kimura, K. Saito, Tractable models for information diffusion in social networks, in: European Conference on Principles of Data Mining and Knowledge Discovery, Springer, 2006, pp. 259–271, http://dx.doi.org/10.1007/11871637_27.
[21] T. Zhu, B. Wang, B. Wu, C. Zhu, Maximizing the spread of influence ranking in social networks, Inform. Sci. 278 (2014) 535–544, http://dx.doi.org/10.1016/j.ins.2014.03.070.
[22] A. Srivastava, C. Chelmis, V.K. Prasanna, Influence in social networks: A unified model?, in: 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2014, pp. 451–454, http://dx.doi.org/10.1109/ASONAM.2014.6921624.
[23] Y. Lin, J. Lui, Algorithmic design for competitive influence maximization problems, 2014, arXiv preprint arXiv:1410.8664.
[24] C. Borgs, M. Brautbar, J. Chayes, B. Lucier, Maximizing social influence in nearly optimal time, in: Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, 2014, pp. 946–957, http://dx.doi.org/10.1137/1.9781611973402.70.
[25] Y. Tang, X. Xiao, Y. Shi, Influence maximization: near-optimal time complexity meets practical efficiency, in: Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, ACM, 2014, pp. 75–86, http://dx.doi.org/10.1145/2588555.2593670.
[26] Y. Tang, Y. Shi, X. Xiao, Influence maximization in near-linear time: A martingale approach, in: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, 2015, pp. 1539–1554, http://dx.doi.org/10.1145/2723372.2723734.
[27] H. Li, S.S. Bhowmick, A. Sun, J. Cui, Conformity-aware influence maximization in online social networks, VLDB J. 24 (1) (2015) 117–141, http://dx.doi.org/10.1007/s00778-014-0366-x.
[28] R. Mohamadi-Baghmolaei, N. Mozafari, A. Hamzeh, Trust based latency aware influence maximization in social networks, Eng. Appl. Artif. Intell. 41 (2015) 195–206, http://dx.doi.org/10.1016/j.engappai.2015.02.007.
[29] C. Wang, W. Chen, Y. Wang, Scalable influence maximization for independent cascade model in large-scale social networks, Data Min. Knowl. Discov. 25 (3) (2012) 545–576, http://dx.doi.org/10.1007/s10618-012-0262-1.
[30] G. Tong, W. Wu, S. Tang, D.Z. Du, Adaptive influence maximization in dynamic social networks, IEEE ACM Trans. Netw. 25 (1) (2017) 112–125, http://dx.doi.org/10.1109/TNET.2016.2563397.
[31] D. Li, C. Wang, S. Zhang, G. Zhou, D. Chu, C. Wu, Positive influence maximization in signed social networks based on simulated annealing, Neurocomputing 260 (2017) 69–78, http://dx.doi.org/10.1016/j.neucom.2017.03.003.
[32] W. Liu, Y. Li, X. Chen, J. He, Maximum likelihood-based influence maximization in social networks, Appl. Intell. 50 (2020) 3487–3502, http://dx.doi.org/10.1007/s10489-020-01747-8.
[33] E. Mossel, G. Schoenebeck, Reaching consensus on social networks, in: Proceedings of the the First Symposium on Innovations in Computer Science, 2009, pp. 214–229, https://repository.upenn.edu/statistics_papers/166.
[34] F. Wang, E. Camacho, K. Xu, Positive influence dominating set in online social networks, in: International Conference on Combinatorial Optimization and Applications, Springer, 2009, pp. 313–321, http://dx.doi.org/10.1007/978-3-642-02026-1_29.
[35] Y. Li, W. Chen, Y. Wang, Z.L. Zhang, Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships, in: Proceedings of the Sixth ACM International Conference on Web Search and Data Mining, 2013, pp. 657–666, http://dx.doi.org/10.1145/2433396.2433478.

[36] X. He, G. Song, W. Chen, Q. Jiang, Influence blocking maximization in social networks under the competitive linear threshold model, in: Proceedings of the 2012 Siam International Conference on Data Mining, SIAM, 2012, pp. 463–474, http://dx.doi.org/10.1137/1.9781611972825.40.
[37] H. Kim, K. Beznosov, E. Yoneki, A study on the influential neighbors to maximize information diffusion in online social networks, Comput. Soc. Netw. 2 (1) (2015) 1–15, http://dx.doi.org/10.1186/s40649-015-0013-8.
[38] W. Liu, K. Yue, H. Wu, J. Li, D. Liu, D. Tang, Containment of competitive influence spread in social networks, Knowl.-Based Syst. 109 (C) (2016) 266–275, http://dx.doi.org/10.1016/j.knosys.2016.07.008.
[39] F. Wang, G. Wang, D. Xie, Maximizing the spread of positive influence under LT-MLA model, in: Asia-Pacific Services Computing Conference, Springer, 2016, pp. 450–463, http://dx.doi.org/10.1007/978-3-319-49178-3_34.
[40] J. Sheng, L. Chen, Y. Chen, B. Li, W. Liu, Positive influence maximization in signed social networks under independent cascade model, Soft Comput. 24 (19) (2020) 14287–14303, http://dx.doi.org/10.1007/s00500-020-05195-x.
[41] A. Goyal, W. Lu, L.V. Lakshmanan, Celf++ optimizing the greedy algorithm for influence maximization in social networks, in: Proceedings of the 20th International Conference Companion on World Wide Web, 2011, pp. 47–48, http://dx.doi.org/10.1145/1963192.1963217.
[42] Y. Wang, G. Cong, G. Song, K. Xie, Community-based greedy algorithm for mining top-k influential nodes in mobile social networks, in: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2010, pp. 1039–1048, http://dx.doi.org/10.1145/1835804.1835935.
[43] A. Goyal, W. Lu, L.V. Lakshmanan, Simpath: An efficient algorithm for influence maximization under the linear threshold model, in: 2011 IEEE 11th International Conference on Data Mining, IEEE, 2011, pp. 211–220, http://dx.doi.org/10.1109/ICDM.2011.132.
[44] E. Cohen, D. Delling, T. Pajor, R.F. Werneck, Sketch-based influence maximization and computation: Scaling up with guarantees, in: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, ACM, 2014, pp. 629–638, http://dx.doi.org/10.1145/2661829.2662077.
[45] D. Kim, D. Hyeon, J. Oh, W.S. Han, H. Yu, Influence maximization based on reachability sketches in dynamic graphs, Inform. Sci. 394 (2017) 217–231, http://dx.doi.org/10.1016/j.ins.2017.02.023.
[46] T.W. Valente, Social network thresholds in the diffusion of innovations, Soc. Netw. 18 (1) (1996) 69–89, http://dx.doi.org/10.1016/0378-8733(95)00256-1.
[47] N. Chen, On the approximability of influence in social networks, SIAM J. Discrete Math. 23 (3) (2009) 1400–1415, http://dx.doi.org/10.1137/08073617X.
[48] M. Doo, L. Liu, Probabilistic diffusion of social influence with incentives, IEEE Trans. Serv. Comput. 7 (3) (2014) 387–400, http://dx.doi.org/10.1109/TSC.2014.2310216.
[49] Y. Zeng, X. Chen, G. Cong, S. Qin, J. Tang, Y. Xiang, Maximizing influence under influence loss constraint in social networks, Expert Syst. Appl. 55 (2016) 255–267, http://dx.doi.org/10.1016/j.eswa.2016.01.008.
[50] B. Zhang, L. Zhang, C. Mu, Q. Zhao, Q. Song, X. Hong, A most influential node group discovery method for influence maximization in social networks: a trust-based perspective, Data Knowl. Eng. 121 (2019) 71–87, http://dx.doi.org/10.1016/j.datak.2019.05.001.
[51] A. Zareie, A. Sheikhahmadi, M. Jalili, Identification of influential users in social networks based on users' interest, Inform. Sci. 493 (2019) 217–231, http://dx.doi.org/10.1016/j.ins.2019.04.033.
[52] M.R. Padmanabhan, N. Somisetty, S. Basu, A. Pavan, Influence maximization in social networks with non-target constraints, in: 2018 IEEE International Conference on Big Data, IEEE, 2018, pp. 771–780, http://dx.doi.org/10.1109/BigData.2018.8621973.
[53] A. Calió, R. Interdonato, C. Pulice, A. Tagarelli, Topology-driven diversity for targeted influence maximization with application to user engagement in social networks, IEEE Trans. Knowl. Data Eng. 30 (12) (2018) 2421–2434, http://dx.doi.org/10.1109/TKDE.2018.2820010.
[54] Q. He, X. Wang, Z. Lei, M. Huang, Y. Cai, L. Ma, TIFIM: A two-stage iterative framework for influence maximization in social networks, Appl. Math. Comput. 354 (2019) 338–352, http://dx.doi.org/10.1016/j.amc.2019.02.056.
[55] D. Li, W. Wang, J. Liu, Grassroots VS elites: Which ones are better candidates for influence maximization in social networks?, Neurocomputing 358 (2019) 321–331, http://dx.doi.org/10.1016/j.neucom.2019.05.053.
[56] A. Calió, A. Tagarelli, Attribute based diversification of seeds for targeted influence maximization, Inform. Sci. 546 (2021) 1273–1305, http://dx.doi.org/10.1016/j.ins.2020.08.093.
[57] K. Jung, W. Heo, W. Chen, Irie: Scalable and robust influence maximization in social networks, in: 2012 IEEE 12th International Conference on Data Mining, IEEE, 2012, pp. 918–923, http://dx.doi.org/10.1109/ICDM.2012.79.
[58] J. Kim, S.K. Kim, H. Yu, Scalable and parallelizable processing of influence maximization for large-scale social networks?, in: 2013 IEEE 29th International Conference on Data Engineering, IEEE, 2013, pp. 266–277, http://dx.doi.org/10.1109/ICDE.2013.6544831.
[59] M. Jaouadi, L. Ben Romdhane, DIN: An efficient algorithm for detecting influential nodes in social graphs using network structure and attributes, in: 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications, IEEE, 2016, pp. 1–8, http://dx.doi.org/10.1109/AICCSA.2016.7945698.
[60] X. Wang, Y. Su, C. Zhao, D. Yi, Effective identification of multiple influential spreaders by degreepunishment, Physica A (2016) 238–247, http://dx.doi.org/10.1016/j.physa.2016.05.020.
[61] H.T. Nguyen, M.T. Thai, T.N. Dinh, A billion-scale approximation algorithm for maximizing benefit in viral marketing, IEEE ACM Trans. Netw. 25 (4) (2017) 2419–2429, http://dx.doi.org/10.1109/TNET.2017.2691544.
[62] J. Shang, S. Zhou, X. Li, L. Liu, H. Wu, Cofim: A community-based framework for influence maximization on large-scale networks, Knowl.-Based Syst. 117 (2017) 88–100, http://dx.doi.org/10.1016/j.knosys.2016.09.029.
[63] M. Azaouzi, L. Ben Romdhane, An efficient two-phase model for computing influential nodes in social networks using social actions, J. Comput. Sci. Technol. 33 (2) (2018) 286–304, http://dx.doi.org/10.1007/s11390-018-1820-9.
[64] A. Şimşek, K.A.R.A. Resul, Using swarm intelligence algorithms to detect influential individuals for influence maximization in social networks, Expert Syst. Appl. 114 (2018) 224–236, http://dx.doi.org/10.1016/j.eswa.2018.07.038.
[65] A. Zareie, A. Sheikhahmadi, M. Jalili, Identification of influential users in social network using gray wolf optimization algorithm, Expert Syst. Appl. 142 (2020) 112971, http://dx.doi.org/10.1016/j.eswa.2019.112971.
[66] A. Sheikhahmadi, A. Zareie, Identifying influential spreaders using multi-objective artificial bee colony optimization, Appl. Soft Comput. 94 (2020) 106436, http://dx.doi.org/10.1016/j.asoc.2020.106436.
[67] Y. Wang, W. Dong, X. Dong, A novel ITÖ Algorithm for influence maximization in the large-scale social networks, Future Gener. Comput. Syst. 88 (2018) 755–763, http://dx.doi.org/10.1016/j.future.2018.04.026.
[68] L. Cui, H. Hu, S. Yu, Q. Yan, Z. Ming, Z. Wen, N. Lu, DDSE: A novel evolutionary algorithm based on degree-descending search strategy for influence maximization in social networks, J. Netw. Comput. Appl. 103 (2018) 119–130, http://dx.doi.org/10.1016/j.jnca.2017.12.003.
[69] Y. Yang, Y. Xu, E. Wang, K. Lou, D. Luan, Exploring influence maximization in online and offline double-layer propagation scheme, Inform. Sci. 450 (2018) 182–199, http://dx.doi.org/10.1016/j.ins.2018.03.048.
[70] Z. Aghaee, S. Kianian, Efficient influence spread estimation for influence maximization, Soc. Netw. Anal. Min. 10 (81) (2020) http://dx.doi.org/10.1007/s13278-020-00694-z.
[71] J. Tang, R. Zhang, P. Wang, Z. Zhao, L. Fan, X. Liu, A discrete shuffled frog-leaping algorithm to identify influential nodes for influence maximization in social networks, Knowl.-Based Syst. 187 (2020) 104833, http://dx.doi.org/10.1016/j.knosys.2019.07.004.
[72] M. Gong, J. Yan, B. Shen, L. Ma, Q. Cai, Influence maximization in social networks based on discrete particle swarm optimization, Inform. Sci. 367 (2016) 600–614, http://dx.doi.org/10.1016/j.ins.2016.07.012.
[73] S.S. Singh, A. Kumar, K. Singh, B. Biswas, LAPSO-IM: A learning-based influence maximization approach for social networks, Appl. Soft Comput. 82 (2019) 105554, http://dx.doi.org/10.1016/j.asoc.2019.105554.
[74] S.S. Singh, K. Singh, A. Kumar, B. Biswas, ACO-IM: maximizing influence in social networks using ant colony optimization, Soft Comput. 24 (13) (2020) 10181–10203, http://dx.doi.org/10.1007/s00500-019-04533-y.
[75] S.S. Singh, A. Kumar, K. Singh, B. Biswas, IM-SSO: Maximizing influence in social networks using social spider optimization, Concurr. Comput.: Pract. Exper. 32 (2) (2020) e5421, https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.5421.
[76] M. Alshahrani, Z. Fuxi, A. Sameh, S. Mekouar, S. Huang, Efficient algorithms based on centrality measures for identification of top-K influential users in social networks, Inform. Sci. 527 (2020) 88–107, http://dx.doi.org/10.1016/j.ins.2020.03.060.
[77] M. Kitsak, L.K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H.E. Stanley, H.A. Makse, Identification of influential spreaders in complex networks, Nat. Phys. 6 (11) (2010) 888, http://dx.doi.org/10.1038/nphys1746.
[78] N. Hafiene, W. Karoui, L. Ben Romdhane, Influential nodes detection in dynamic social networks, in: International Conference on Business Information Systems, Springer, 2019, pp. 62–73, http://dx.doi.org/10.1007/978-3-030-20482-2_6.
[79] X. Rui, X. Yang, J. Fan, Z. Wang, A neighbour scale fixed approach for influence maximization in social networks, Computing 102 (2) (2020) 427–449, http://dx.doi.org/10.1007/s00607-019-00778-5.
[80] G. Xie, Y. Chen, H. Zhang, Y. Liu, MBIC: A novel influence propagation model for membership-based influence Maximization in Social Networks, IEEE Access 7 (2019) 75696–75707, http://dx.doi.org/10.1109/ACCESS.2019.2922474.

[81] S. Chengai, N. Weinan, Q. Liqing, L. Liangyu, Scalable influence maximization based on influential seed successors, Soft Comput. 24 (8) (2020) 5921–5931, http://dx.doi.org/10.1007/s00500-019-04483-5.

[82] Z.-K. Bao, J.-G. Liu, H.-F. Zhang, Identifying multiple influential spreaders by a heuristic clustering algorithm, Phys. Lett. A 381 (11) (2017) 976–983, http://dx.doi.org/10.1016/j.physleta.2017.01.043.

[83] P. Dey, A. Chaterjee, S. Roy, Influence maximization in online social network using different centrality measures as seed node of information propagation, Sādhanā 44 (9) (2019) 205, http://dx.doi.org/10.1007/s12046-019-1189-7.

[84] Q. He, X. Wang, M. Huang, J. Lv, L. Ma, Heuristics-based influence maximization for opinion formation in social networks, Appl. Soft Comput. 66 (2018) 360–369, http://dx.doi.org/10.1016/j.asoc.2018.02.016.

[85] G. Maji, S. Mandal, S. Sen, A systematic survey on influential spreaders identification in complex networks with a focus on K-shell based techniques, Expert Syst. Appl. 161 (2020) 113681, http://dx.doi.org/10.1016/j.eswa.2020.113681.

[86] A. Zareie, A. Sheikhahmadi, EHC: Extended H-index centrality measure for identification of users' spreading influence in complex networks, Physica A 514 (2019) 141–155, http://dx.doi.org/10.1016/j.physa.2015.12.162.

[87] A. Salehi, B. Masoumi, KATZ centrality with biogeography-based optimization for influence maximization problem, J. Comb. Optim. 40 (2020) 205–226, http://dx.doi.org/10.1007/s10878-020-00580-6.

[88] J. Bae, S. Kim, Identifying and ranking influential spreaders in complex networks by neighborhood coreness, Physica A 395 (2014) 549–559, http://dx.doi.org/10.1016/j.physa.2013.10.047.

[89] L. Lü, Y.C. Zhang, C.H. Yeung, T. Zhou, Leaders in social networks, the delicious case, PLoS One 6 (6) (2011) e21202, http://dx.doi.org/10.1371/journal.pone.0021202.

[90] A. Zareie, A. Sheikhahmadi, M. Jalili, M.S.K. Fasaei, Finding influential nodes in social networks based on neighborhood correlation coefficient, Knowl.-Based Syst. 194 (2020) 105580, http://dx.doi.org/10.1016/j.knosys.2020.105580.

[91] M.E.G. Rossi, F.D. Malliaros, M. Vazirgiannis, Spread it good, spread it fast: Identification of influential nodes in social networks, in: Proceedings of the 24th International Conference on World Wide Web, ACM, 2015, pp. 101–102, http://dx.doi.org/10.1145/2740908.2742736.

[92] F. Malliaros, M.E. Rossi, M. Vazirgiannis, Locating influential nodes in complex networks, Sci. Rep. 6 (2016) 19307, http://dx.doi.org/10.1038/srep19307.

[93] L.L. Ma, C. Ma, H.F. Zhang, B.H. Wang, Identifying influential spreaders in complex networks based on gravity formula, Physica A 451 (2016) 205–212, http://dx.doi.org/10.1016/j.physa.2015.12.162.

[94] Z. Li, T. Ren, X. Ma, S. Liu, Y. Zhang, T. Zhou, Identifying influential spreaders by gravity model, Sci. Rep. 9 (1) (2019) 1–7, http://dx.doi.org/10.1038/s41598-019-44930-9.

[95] X. Huang, D. Chen, D. Wang, T. Ren, Identifying influencers in social networks, Entropy 22 (4) (2020) 450, http://dx.doi.org/10.3390/e22040450.

[96] M. Azaouzi, L. Ben Romdhane, An evidential influence-based label propagation algorithm for distributed community detection in social networks, Procedia Comput. Sci. 112 (2017) 407–416, http://dx.doi.org/10.1016/j.procs.2017.08.045.

[97] S. Kim, D. Kim, J. Oh, J.H. Hwang, W.S. Han, W. Chen, H. Yu, Scalable and parallelizable influence maximization with random walk ranking and rank merge pruning, Inform. Sci. 415 (2017) 171–189, http://dx.doi.org/10.1016/j.ins.2017.06.018.

[98] S. Kianian, M. Rostamnia, An efficient path-based approach for influence maximization in social networks, Expert Syst. Appl. (2020) 114168, http://dx.doi.org/10.1016/j.eswa.2020.114168.

[99] M. Li, Q. Zhang, Y. Deng, Evidential identification of influential nodes in network of networks, Chaos Solitons Fractals 117 (2018) 283–296, http://dx.doi.org/10.1016/j.chaos.2018.04.033.

[100] D. Liu, H. Nie, J. Zhao, Q. Wang, Identifying influential spreaders in large-scale networks based on evidence theory, Neurocomputing 359 (2019) 466–475, http://dx.doi.org/10.1016/j.neucom.2019.06.030.

[101] M. Azaouzi, D. Rhouma, L. Ben Romdhane, Community detection in large-scale social networks: state-of-the-art and future directions, Soc. Netw. Anal. Min. 9 (1) (2019) 23, http://dx.doi.org/10.1007/s13278-019-0566-x.

[102] K. Rahimkhani, A. Aleahmad, M. Rahgozar, A. Moeini, A fast algorithm for finding most influential people based on the linear threshold model, Expert Syst. Appl. 42 (3) (2015) 1353–1361, http://dx.doi.org/10.1016/j.eswa.2014.09.037.

[103] A. Bozorgi, H. Haghighi, M.S. Zahedi, M. Rezvani, INCIM: a community-based algorithm for influence maximization problem under the linear threshold model, Inf. Process. Manage. 52 (6) (2016) 1188–1199, http://dx.doi.org/10.1016/j.ipm.2016.05.006.

[104] D. Rhouma, L. Ben Romdhane, An efficient algorithm for community mining with overlap in social networks, Expert Syst. Appl. 41 (9) (2014) 4309–4321, http://dx.doi.org/10.1016/j.eswa.2014.01.002.

[105] X. Li, X. Cheng, S. Su, C. Sun, Community-based seeds selection algorithm for location aware influence maximization, Neurocomputing 275 (2018) 1601–1613, http://dx.doi.org/10.1016/j.neucom.2017.10.007.

[106] J. Shang, H. Wu, S. Zhou, J. Zhong, Y. Feng, B. Qiang, IMPC: Influence maximization based on multi-neighbor potential in community networks, Physica A 512 (2018) 1085–1103, http://dx.doi.org/10.1016/j.physa.2018.08.045.

[107] L. Qiu, W. Jia, J. Yu, X. Fan, W. Gao, PHG: A three-phase algorithm for influence maximization based on community structure, IEEE Access 7 (2019) 62511–62522, http://dx.doi.org/10.1109/ACCESS.2019.2912628.

[108] S.S. Singh, A. Kumar, K. Singh, B. Biswas, C2im: community based context-aware influence maximization in social networks, Physica A 514 (2019) 796–818, http://dx.doi.org/10.1016/j.physa.2018.09.142.

[109] A. Bozorgi, S. Samet, J. Kwisthout, T. Wareham, Community-based influence maximization in social networks under a competitive linear threshold model, Knowl.-Based Syst. 134 (2017) 149–158, http://dx.doi.org/10.1016/j.knosys.2017.07.029.

[110] H. Huang, H. Shen, Z. Meng, H. Chang, H. He, Community-based influence maximization for viral marketing, Appl. Intell. 49 (6) (2019) 2137–2150, http://dx.doi.org/10.1007/s10489-018-1387-8.

[111] H.A. Beni, A. Bouyer, TI-SC: top-k influential nodes selection based on community detection and scoring criteria in social networks, J. Ambient Intell. Humaniz. Comput. 11 (2020) 4889–4908, http://dx.doi.org/10.1007/s12652-020-01760-2.

[112] F. Cai, L. Qiu, X. Kuai, H. Zhao, CBIM-RSRW: An community-based Method for influence maximization in Social Network, IEEE Access 7 (2019) 152115-152125, http://dx.doi.org/10.1109/ACCESS.2019.2944350.

[113] Q. Wu, L. Zhou, Y. Huang, Influence maximization based on community closeness in Social Networks, in: International Conference on Web Information Systems Engineering, Springer, 2020, pp. 142–156, http://dx.doi.org/10.1007/978-981-15-3281-8_13.

[114] H. Huang, H. Shen, Z. Meng, Community-based influence maximization in attributed networks, Appl. Intell. 50 (2) (2020) 354–364, http://dx.doi.org/10.1007/s10489-019-01529-x.

[115] Q. He, X. Wang, F. Mao, J. Lv, Y. Cai, M. Huang, Q. Xu, CAOM: A community-based approach to tackle opinion maximization for social networks, Inform. Sci. 513 (2020) 252–269, http://dx.doi.org/10.1016/j.ins.2019.10.064.

[116] W. Li, K. Zhong, J. Wang, D. Chen, A Dynamic Algorithm based on cohesive entropy for influence maximization in social networks, Expert Syst. Appl. (2020) 114207, http://dx.doi.org/10.1016/j.eswa.2020.114207.

[117] W. Mnasri, M. Azaouzi, L. Ben Romdhane, Parallel social behavior-based algorithm for identification of influential users in social network, Appl. Intell. (2021) http://dx.doi.org/10.1007/s10489-021-02203-x.

[118] X. Liu, M. Li, S. Li, S. Peng, X. Liao, X. Lu, IMGPU: GPU-accelerated influence maximization in large-scale social networks, IEEE Trans. Parallel Distrib. Syst. 25 (1) (2013) 136–145, http://dx.doi.org/10.1109/TPDS.2013.41.

[119] Z. Zong, B. Li, C. Hu, DIRIEr: Distributed influence maximization in social network, in: 2014 20th IEEE International Conference on Parallel and Distributed Systems, IEEE, 2014, pp. 119–125, http://dx.doi.org/10.1109/TPDS.2013.41.

[120] G. Song, X. Zhou, Y. Wang, K. Xie, Influence maximization on large-scale mobile social network: a divide-and-conquer method, IEEE Trans. Parallel Distrib. Syst. 26 (5) (2014) 1379–1392, http://dx.doi.org/10.1109/TPDS.2014.2320515.

[121] H. Wu, K. Yue, X. Fu, Y. Wang, W. Liu, Parallel seed selection for influence maximization based on k-shell decomposition, in: International Conference on Collaborative Computing: Networking, Applications and Worksharing, Springer, 2016, pp. 27–36, http://dx.doi.org/10.1007/978-3-319-59288-6_3.

[122] M. Minutoli, M. Drocco, M. Halappanavar, A. Tumeo, A. Kalyanaraman, CuRipples: Influence maximization on multi-GPU systems, in: Proceedings of the 34th ACM International Conference on Supercomputing, 2020, 12, http://dx.doi.org/10.1145/3392717.3392750, 1-11.

[123] L. Xiao, S. Wang, G. Mei, Efficient parallel algorithm for detecting influential nodes in large biological networks on the Graphics Processing Unit, Future Gener. Comput. Syst. 106 (2020) 1–13, http://dx.doi.org/10.1016/j.future.2019.12.038.

[124] J. Zhu, S. Ghosh, W. Wu, Group influence maximization problem in social networks, IEEE Trans. Comput. Soc. Syst. 6 (6) (2019) 1156–1164, http://dx.doi.org/10.1109/TCSS.2019.2938575.

[125] Y. Dai, W. Jiang, K. Li, Group-based competitive influence maximization, in: 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), IEEE, 2018, pp. 999–1006, http://dx.doi.org/10.1109/SmartWorld.2018.00176.

[126] J. Zhu, S. Ghosh, W. Wu, C. Gao, Profit maximization under group influence model in social networks, in: International Conference on Computational Data and Social Networks, Springer, 2019, pp. 108–119, http://dx.doi.org/10.1007/978-3-030-34980-6_13.
[127] Y. Ji, L. Pan, P. Wu, Influence maximization on large-scale networks with a group-based method via network embedding, in: 2019 IEEE Fourth International Conference on Data Science in Cyberspace, IEEE, 2019, pp. 176–182, http://dx.doi.org/10.1109/DSC.2019.00034.
[128] D. Huang, L. Pan, An improved group-based influence maximization method in social networks, in: 2018 Third International Conference on Security of Smart Cities, Industrial Control System and Communications, IEEE, 2018, pp. 1–6, http://dx.doi.org/10.1109/SSIC.2018.8556647.
[129] Y. Li, X. Gan, L. Fu, X. Tian, Z. Qin, Y. Zhou, Conformity-aware influence maximization with user profiles, in: 2018 10th International Conference on Wireless Communications and Signal Processing, IEEE, 2018, pp. 1–6, http://dx.doi.org/10.1109/WCSP.2018.8555685.
[130] Z. Aghaee, S. Kianian, Influence maximization algorithm based on reducing search space in the social networks, SN Appl. Sci. 2 (12) (2020) 1–14, http://dx.doi.org/10.1007/s42452-020-03812-w.
[131] Y. Zhong, L. Guo, Group influence maximization in social networks, in: International Conference on Computational Data and Social Networks, Springer, 2020, pp. 152–163, http://dx.doi.org/10.1007/978-3-030-66046-8_13.
[132] L. Yu, S.M. Motipalli, D. Lee, P. Liu, H. Xu, Q. Liu, J. Tan, B. Luo, My friend leaks my privacy: Modeling and analyzing privacy in social networks, in: Proceedings of the 23nd ACM on Symposium on Access Control Models and Technologies, 2018, pp. 93–104, http://dx.doi.org/10.1145/3205977.3205981.
[133] S. Bourahla, M. Laurent, Y. Challal, Privacy preservation for social networks sequential publishing, Comput. Netw. 170 (2020) 107106, http://dx.doi.org/10.1016/j.comnet.2020.107106.
[134] M. Siddula, Privacy Preserving User Data Publication in Social Networks Dissertation, Georgia State University, 2020, https://scholarworks.gsu.edu/cs_diss/157.
[135] T. Tassa, F. Bonchi, Privacy preserving estimation of social influence, in: Proceedings of the 17th International Conference on Extending Database Technology, 2014, pp. 559–570, http://dx.doi.org/10.5441/002/edbt.2014.50.
[136] M. Han, J. Li, Z. Cai, Q. Han, Privacy reserved influence maximization in gps-enabled cyber-physical and online social networks, in: 2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom)(BDCloud-SocialCom-SustainCom), IEEE, 2016, pp. 284–292, http://dx.doi.org/10.1109/BDCloud-SocialCom-SustainCom.2016.51.
[137] S. Hajian, T. Tassa, F. Bonchi, Individual privacy in social influence networks, Soc. Netw. Anal. Min. 6 (1) (2016) 2:1–2:14, http://dx.doi.org/10.1007/s13278-015-0312-y.
[138] M. Han, Y. Liang, Z. Duan, Y. Wang, Mining public business knowledge: A case study in sec's edgar, in: 2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom)(BDCloud-SocialCom-SustainCom), IEEE, 2016, pp. 393–400, http://dx.doi.org/10.1109/BDCloud-SocialCom-SustainCom.2016.65.
[139] J. Zhu, Y. Liu, X. Yin, A new structure-hole-based algorithm for influence maximization in large online social networks, IEEE Access 5 (2017) 23405–23412, http://dx.doi.org/10.1109/ACCESS.2017.2758353.
[140] M. Han, Q. Han, L. Li, J. Li, Y. Li, Maximising influence in sensed heterogeneous social network with privacy preservation, Int. J. Sens. 28 (2) (2018) 69–79, http://dx.doi.org/10.1504/IJSNET.2018.096194.
[141] M. Liu, Y. Zeng, Y. Liu, Z. Liu, J. Ma, X. Zhu, Collective influence based privacy preservation for social networks, in: 2019 International Conference on Networking and Network Applications, IEEE, 2019, pp. 282–289, http://dx.doi.org/10.1109/NaNA.2019.00056.
[142] D. Jing, T. Liu, Context-based influence maximization with privacy protection in social networks, J. Wirel. Commun. Netw. 2019 (1) (2019) 142, http://dx.doi.org/10.1186/s13638-019-1405-5.
[143] V.B. Kukkala, S.R.S. Iyengar, Identifying influential spreaders in a social network (while preserving privacy), Proc. Priv. Enhanc. Technol. 2020 (2) (2020) 537–557, http://dx.doi.org/10.2478/popets-2020-0040.
[144] X. Qi, G. Mei, S. Cuomo, L. Xiao, A network-based method with privacy-preserving for identifying influential providers in large healthcare service systems, Future Gener. Comput. Syst. 109 (2020) 293–305, http://dx.doi.org/10.1016/j.future.2020.04.004.
[145] N.S.N. Abadi, M.R. Khayyambashi, Influence maximization in viral marketing with expert and influential leader discovery approach, in: 8th International Conference on E-Commerce in Developing Countries: With Focus on E-Trust, IEEE, 2014, pp. 1–8, http://dx.doi.org/10.1109/ECDC.2014.6836763.
[146] H. Zhu, B. Huberman, Y. Luon, To switch or not to switch: understanding social influence in online choices, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2012, pp. 2257–2266, http://dx.doi.org/10.1145/2207676.2208383.
[147] A. Goyal, L.V. Lakshmanan, Recmax: Exploiting recommender systems for fun and profit, in: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2012, pp. 1294–1302, http://dx.doi.org/10.1145/2339530.2339731.
[148] Q. Ni, J. Guo, C. Huang, W. Wu, Community-based rumor blocking maximization in social networks: Algorithms and analysis, Theoret. Comput. Sci. 840 (2020) 257–269, http://dx.doi.org/10.1016/j.tcs.2020.08.030.
[149] Y. Mei, W. Zhao, J. Yang, Influence maximization on twitter: A mechanism for effective marketing campaign, in: 2017 IEEE International Conference on Communications, IEEE, 2017, pp. 1–6, http://dx.doi.org/10.1109/ICC.2017.7996805.
[150] S. Jendoubi, A. Martin, L. Liétard, H.B. Hadji, B.B. Yaghlane, Two evidential data based models for influence maximization in twitter, Knowl.-Based Syst. 121 (2017) 58–70, http://dx.doi.org/10.1016/j.knosys.2017.01.014.
[151] M. Bouguessa, L. Ben Romdhane, Identifying authorities in online communities, ACM Trans. Intell. Syst. Technol. 6 (3) (2015) 1–23, http://dx.doi.org/10.1145/2700481.
[152] J. Tang, S. Wu, J. Sun, Confluence: Conformity influence in large social networks, in: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2013, pp. 347–355, http://dx.doi.org/10.1145/2487575.2487691.
[153] X. Chen, Y. Zhao, G. Liu, R. Sun, X. Zhou, K. Zheng, Efficient similarity-aware influence maximization in Geo-social Network, IEEE Trans. Knowl. Data Eng. (2020) http://dx.doi.org/10.1109/TKDE.2020.3045783.
[154] A. Teng, T.W. Li, Y.C. Liao, H.W. Chen, Y.A. Pignolet, D.N. Yang, L.Y. Chen, On Influencing the Influential: Disparity Seeding, 2020, arXiv preprint arXiv:2011.08946.
[155] N. Hafiene, W. Karoui, L. Ben Romdhane, An incremental approach to update influential nodes in dynamic social networks, Procedia Comput. Sci. 176 (2020) 781–790, http://dx.doi.org/10.1016/j.procs.2020.09.073.
[156] M. Kivelä, A. Arenas, M. Barthelemy, J.P. Gleeson, Y. Moreno, M.A. Porter, Multilayer networks, J. Complex Netw. 2 (3) (2014) 203–271, http://dx.doi.org/10.1093/comnet/cnu016.
[157] S. Molaei, R. Farahbakhsh, M. Salehi, N. Crespi, Identifying influential nodes in heterogeneous networks, Expert Syst. Appl. 160 (2020) 113580, http://dx.doi.org/10.1016/j.eswa.2020.113580.
[158] W. Ju, L. Chen, B. Li, W. Liu, J. Sheng, Y. Wang, A new algorithm for positive influence maximization in signed networks, Inform. Sci. 512 (2020) 1571–1591, http://dx.doi.org/10.1016/j.ins.2019.10.061.
[159] S. Wang, J. Liu, Y. Jin, Finding influential nodes in multiplex networks using a memetic algorithm, IEEE Trans. Cybern. (2019) http://dx.doi.org/10.1109/TCYB.2019.2917059.
[160] A. Kuhnle, M.A. Alim, X. Li, H. Zhang, M.T. Thai, Multiplex influence maximization in online social networks with heterogeneous diffusion models, IEEE Trans. Comput. Soc. Syst. 5 (2) (2018) 418–429, http://dx.doi.org/10.1109/TCSS.2018.2813262.
[161] A.I.E. Hosni, K. Li, S. Ahmad, Minimizing rumor influence in multiplex online social networks based on human individual and social behaviors, Inform. Sci. 512 (2020) 1458–1480, http://dx.doi.org/10.1016/j.ins.2019.10.063.