



# Influence maximization in social networks: a survey of behaviour-aware methods

Ahmad Zareie<sup>1</sup> · Rizos Sakellariou<sup>1</sup>

Received: 23 May 2022 / Revised: 12 March 2023 / Accepted: 27 March 2023 / Published online: 25 April 2023  
© The Author(s) 2023

## Abstract

Social networks have become an increasingly common abstraction to capture the interactions of individual users in a number of everyday activities and applications. As a result, the analysis of such networks has attracted lots of attention in the literature. Among the topics of interest, a key problem relates to identifying so-called influential users for a number of applications, which need to spread messages. Several approaches have been proposed to estimate users' influence and identify sets of influential users in social networks. A common basis of these approaches is to consider links between users, that is, structural or topological properties of the network. To a lesser extent, some approaches take into account users' behaviours or attitudes. Although a number of surveys have reviewed approaches based on structural properties of social networks, there has been no comprehensive review of approaches that take into account users' behaviour. This paper attempts to cover this gap by reviewing and proposing a taxonomy of such behaviour-aware methods to identify influential users in social networks.

**Keywords** Influence maximization · Diffusion models · Spreading process · User behaviour · Social networks

## 1 Introduction

The proliferation of network technologies, including social media, has changed people's daily activities and patterns of interaction. Various forms of social networking applications are used for different purposes: exchanging messages, broadcasting news, sharing opinions on topics of common interest, publicity, and so on. User interactions form social networks that may play an effective role in shaping an opinion, fast and widespread propagation of specific messages or news and may help establish opinions quickly. It has long been accepted that not all users play the same role or carry the same gravitas in social networks. Some users may be more active or influential or vital due to their behaviour or friends in the network. So-called influential users play an important role in social networks and can be crucial in helping spread messages quickly and widely. The *influence maximization* (IM) problem is defined as the problem of identifying a set

of users in a social network, who can influence broadly and effectively other users. This is known to be a complex problem, particularly as it requires some criterion to measure influence.

The IM problem was originally studied as an algorithmic problem by Domingos and Richardson Domingos and Richardson (2001); Richardson and Domingos (2002), while Kempe, Kleiberg and Tardos Kempe et al. (2003) were the first to formulate the problem as a discrete optimization problem. Although different studies have been dedicated to solving the IM problem, investigating the aspects of the problem that help identify influential users (Li et al. 2022; Zhang et al. 2023) and/or predict the influence of users (Gong et al. 2021; De Salve et al. 2021) is still an important research challenge due to the impact that messages propagated through social networks often have on today's society.

The most commonly used model to solve the problem is to represent a social network as a graph whose nodes represent users and edges indicate the relationships between users. Then, different criteria may be specified to measure the influence of users. In some research, user's influence is determined by the network structure, which means that influential users are identified on the basis of topological properties of the graph. In other research, apart from network structure, users' characteristics of behaviour, such as

---

✉ Ahmad Zareie  
ahmad.zareie@postgrad.manchester.ac.uk  
Rizos Sakellariou  
rizos@manchester.ac.uk

<sup>1</sup> Department of Computer Science, The University of Manchester, Manchester M13 9PL, UK

preferences or trustworthiness, are taken into account. We term the former *behaviour-agnostic* as opposed to *behaviour-aware* for the latter. In behaviour-agnostic approaches, the problem is that there is essentially a structural graph-based approach to identify influential users; differences in their individual behaviour are disregarded. Taking users' behaviour into account can match everyday realities; this is the strength (and the additional complexity) of behaviour-aware approaches.

Several surveys have reviewed various methods to address the IM problem and discussed different challenges in identifying a set of influential users. A set of surveys (Guille et al. 2013; Singh et al. 2018; Chen et al. 2013; Jaouadi and Ben Romdhane 2019) focuses on specific elements of the problem, such as diffusion models or the simulating spreading process, but without providing a comprehensive review of various methods for IM. Another set of surveys (Lü et al. 2016; Peng et al. 2018; Das et al. 2018) focuses on measures applied to rank each node in terms of its influence. These measures, often derived from the network's structure, are known as centrality measures. The ranking, obtained by centrality measures, can then inform the choice of most influential users. Other surveys focus on the methods aiming to select a set of influential users considering the network as a whole; for example, influential nodes may be chosen so that all parts of a network are covered. In Li et al. (2018), the solutions of the IM problem are reviewed from an algorithmic perspective, which relies on diffusion models that simulate the spreading process. Along the same line but using a somewhat different classification is the work in Tejaswi et al. (2016). Various classifications of methods have been proposed in other surveys (Arora et al. 2017; Al-Garadi et al. 2018; Bian et al. 2019; Yang and Pei 2019; Banerjee et al. 2020b). In Arora et al. (2017), different structural methods are systematically evaluated and compared using a benchmark platform. In Al-Garadi et al. (2018), structural users' influence detection methods are categorized and their advantages and disadvantages are discussed. Another review and classification is given in Bian et al. (2019), where some behaviour-aware methods related to reputation and trust between users are also discussed. The methods proposed to identify influential users in evolving networks are classified and reviewed in Yang and Pei (2019); Hafiene et al. (2020). In Banerjee et al. (2020b), some variants of the IM problem are reviewed and the hardness of the problem under both traditional and parameterized complexity is described.

The main characteristic of all these surveys is that they focus on behaviour-agnostic methods. Some behaviour-aware methods are briefly discussed in Lü et al. (2016); Li et al. (2018); Peng et al. (2018); Tejaswi et al. (2016); Bian et al. (2019) but without any in-depth classification or specific focus on behaviour. It is generally true that behaviour-agnostic methods have a longer history in tackling the IM

problem; this may explain why they dominate in current work. However, the starting point of behaviour-aware methods is a more realistic formulation of the IM problem. As such, behaviour-aware methods have the potential to deliver solutions of higher relevance to real-world situations. Yet, behaviour-aware methods have not been considered in any previous survey as a class on their own, that is, as a separate family of solutions deserving its own fine-grained classification. This paper addresses this gap.

In view of the above, the contributions of this paper can be described as follows:

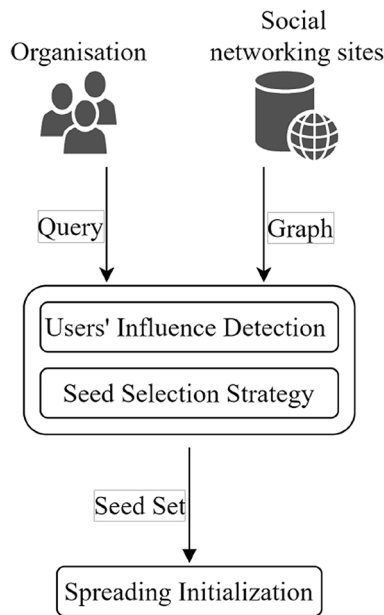
- A proposal to categorize existing methods to solve the IM problem into behaviour-agnostic and behaviour-aware, which is motivated by the realization that behavioural characteristics of users play a key role in IM.
- A taxonomy and a detailed review of behaviour-aware methods for the IM problem, which discusses the behavioural characteristics that have been taken into account to solve the IM problem, how these characteristics are modelled and how the properties of the problem are affected by taking into account these characteristics.
- A discussion of challenges for the IM problem from a behaviour-aware perspective.

The rest of the paper is organized as follows. Some basic concepts of social networks and the influence maximization problem are introduced in Sect. 2. An overview of our taxonomy is also given in this section. Section 3 covers a comprehensive overview of behaviour-aware methods for IM. Challenges and future research directions are finally presented in Sect. 4.

## 2 Preliminaries

### 2.1 Problem description

Figure 1 gives an overview of the IM problem. Somebody (say an organisation) intends to run a campaign to spread a message, advertisement, news or idea through a social network. They formulate a query to identify a set of influential users that can help spread a specific message with specific features, preferences or constraints. For example, a company may consider advertising a new model of car for sale in a special exhibition; they may target users with an age greater than 18, who are located in the vicinity of the exhibition and they may have a limited budget. This can be formulated in the query as targets, geographical preference and budget constraint. Social network sites typically provide data in relation to users and the relationships between them. Relevant user information may include shopping history, rate history, opinion, interests, geographical location and so on.



**Fig. 1** Overview of the influence maximization problem

Relationship may indicate some connection, such as friendship or common interests, between a pair of users. According to the information provided by social network sites and in line with the query formulated, influence maximization aims to identify a set of influential users as initial spreaders, known as a *seed set*, to spread the message and maximize the total number of users who are influenced. The users who are influenced are called *influenced users* or *active users*. A formal model of the IM problem is given next.

First, an organisation formulates a query to define a message, its features and related information. The query is generally modelled as  $Q = (M, AM, B)$ .  $M$  denotes the content of the message which may relate to news, an advertising slogan, an opinion, an event and so on.  $AM$  shows the different features of the message, such as topics of a product, location of an event and so on.  $B$  may be the budget for propagation that can be the number of seeds, some discount for seeds to give an incentive to propagate a message, or remuneration to a network owner (or third-party service provider) to identify influential users and initiate a spreading process.

Second, given the features of the query, information provided by social network sites is used to model the social network; this step may be done by the owners of the network or by some third-party service providers. A social network is modelled as a graph  $G = (V, E, BV, BE)$ , with users and the relationships between them represented by nodes  $V$  and edges  $E$ , respectively. If there is a relationship between two nodes  $v_i$  and  $v_j$ , it is shown by the edge  $e_{ij}$ ; two nodes that are connected by an edge are considered as neighbours.  $BV$  denotes the different behaviours of users such as location, interests, opinions and so on.  $BE$  denotes the different

features of the relationship between a pair of nodes indicated by the edges, such as trust, spread probability (also called as influence or activation probability) or common interests. Depending on the nature of relationships in the network, in some cases (for example, to represent the follow relationship in Twitter), the edges are considered as directed relationships, i.e.  $e_{ij} \neq e_{ji}$ , while in some other cases (for example, to represent friendship in Facebook) the relationship between nodes is regarded as bi-directional, and the graph is considered as undirected, i.e.  $e_{ij} = e_{ji}$ . In general, a social network graph contains no cycles, i.e.  $e_{ii} \notin E$ .

Third, according to the query  $Q$  and graph  $G$ , the influence of each node is assessed and a strategy is applied to select a set of influential nodes as seeds taking into account the budget associated with the query and with an overall goal that the selected seeds maximize the influence of the spreading process. How to determine this set of influential nodes is the core of the influence maximization problem. Finally, the spreading process is initialized using the set of seeds to propagate the message.

In what follows, an overview of the approaches applied in the literature to detect the influence of a user/a set of users is presented and the general framework of the influence maximization problem is formally defined.

## 2.2 Influence detection

Different approaches have been described in the literature to detect the influence of a user or a set of users. This paper groups these approaches into the following four families.

- Centrality measures determine the influence of each node in the social network graph based on topological properties. Different centrality measures have been proposed and extended to determine the influence of nodes (Freeman 1978, 1977; Sabidussi 1966; Zareie et al. 2017; Lü et al. 2016; Zareie and Sheikahmadi 2019; Kitsak et al. 2010).
- Simulation of the spreading process can be used to determine the influence of each node and select a set of influential nodes. Different diffusion models have been proposed to simulate the spreading process. The Independent Cascade (IC) model (Carnes et al. 2007; Goldenberg et al. 2001; Kempe et al. 2003) and the Linear Threshold (LT) model (Borodin et al. 2010; Granovetter 1978; Kempe et al. 2003) are the models that have been widely applied. In the IC model, the spreading process is *simple*, which means the propagation on the edges is mutually independent and interaction with one active node may be enough for a node to be activated (influenced). In the LT model, the spreading process is *complex*, which means a node may need to interact with

- multiple active nodes to be activated (influenced) (Chen et al. 2013).
- Reverse Influence Sampling (RIS) (Borgs et al. 2012; Tang et al. 2014; Borgs et al. 2014) relies on a random sampling technique to determine the influence of each node and identify a seed set.
- Maximum Influence Arborescence (MIA) (Chen et al. 2010) relies on the spreading probability on paths between pairs of nodes in the social network graph; the spreading probability on a path is calculated by multiplying the spreading probability on the edges of the path.

Depending on the approach applied to detect the influence, we can approximately determine the time complexity of each method. Generally speaking, in terms of time complexity, these approaches can be ranked from high to low in the order: simulation-based, MIA-based, RIS-based and centrality-based.

### 2.3 Influence maximization

The influence maximization (IM) problem aims to spread a message as widely as possible through a social network, taking into account that it is highly time-consuming and practically impossible to send the message to all users of the network. As a result some constraints are taken into account to select a small set of influential users, who have more influence than other users and can widely propagate the message. The IM problem is generally defined as in Eq. (1), where the function  $\varphi(S)$  defines the influence of a set of seeds  $S$ .

$$S^* = \underset{S \subseteq V}{\operatorname{argmax}} \varphi(S) \quad (1)$$

subject to some constraints.

Given the model discussed in Sect. 2.1, there are three different types of features related to the behavioural information which may be taken into account to select a set of influential nodes: (i) features of the users (BV) to determine the relevance of each user to the query; (ii) features of the relationships (BE) to model the spreading (influence) probability between the users based on the relevance to the query; and (iii) features of the message (AM) to target a set of relevant users in spreading process. These three types of information encapsulate the differences between behaviour-agnostic and behaviour-aware methods.

Behaviour-agnostic methods only consider the graph structure to determine the influence of nodes and identify a set of influential nodes. That is to say, behaviour-agnostic methods do not take into account anything specific for the sets BV (capturing users' behaviour) and BE (capturing edges' behaviour, that is, different features of the relationship between a pair of users). In addition, they do not consider different features of the message, as captured by the

set AM in the query message. In brief, behaviour-agnostic methods assume that: (i) all users have the same behaviour; (ii) relationships between any pairs of users are the same; and (iii) the features of the message have no impact on the propagation process. In behaviour-aware methods, some or all of these aspects may differ. It means that, in addition to graph structure, behaviour-aware methods take into account some aspects of behavioural information to propose practical approaches for real-world applications.

In this paper, we focus on behaviour-aware methods; as mentioned in Sect. 1, there are several surveys discussing behaviour-agnostic methods comprehensively.

## 3 Behaviour-aware methods

In behaviour-aware methods, besides graph structure, additional information, such as users' behaviour, relationships' features, and query content, is taken into account to improve the spreading process. For example, when a query aims to advertise a specific product, all users may not be equally interested in this product. Considering how interested each user is can improve the process of identifying a seed set and result in a successful message propagation. Based on the type of the information taken into account, behaviour-aware methods can be divided into four main categories.

User preferences and query relevance to these preferences are taken into account by *interest-aware methods*. To determine user preferences, these methods rely on preprocessing of historical records of users' activities in the networks, such as the content of their posts, the content of the posts they liked or replied to, their rating records for different products or their shopping experience and so on. In *opinion-aware methods*, the opinion of users towards a message is taken into account to find influential users that can change the attitude or opinion of other users. To determine the stance and opinion of users towards a message, these methods rely on the analysis of the sentiment and subjectivity of users in their posts or their feedback on messages posted by other users. Cost and benefit of spreading an advertising query are considered by *money-aware methods*. These methods assume that users may estimate their influence in the network and use it to negotiate a price to participate in the spreading process. Therefore, selecting different users as seeds may incur a different cost. Furthermore, the analysis of historical records of users' activities is leveraged in these methods to estimate users' valuation of a product advertised in a spreading process. This can help target users with a higher valuation and result in a higher benefit in the process. In *physical world-aware methods*, the geographical location of users is taken into account and a query aims to spread a message towards users in specific locations, for example a query may invite users to a festival taking place in a specific

geographical location. These methods assume the location of users can be modelled using historical records of users' check-in locations, GPS-enabled technologies, location-based social networks or similar techniques. These four categories are illustrated in Fig. 2 and will be further elaborated and discussed in detail in the sections that follow.

### 3.1 Interest-aware methods

Users' preferences and interests are taken into account in these methods. Spreading aims to influence the users who are potentially interested in a message. Past activities of users, such as posts, likes, ratings or shopping history, may be preprocessed to determine their interests. In these methods, the size of the seed set is fixed, that is  $|S| = k$ , and it is also considered as a constraint of the IM problem as defined in Eq. (1). We categorize these methods into three classes based on the process considered to identify the users interested in the message of the query formulated.

#### 3.1.1 Target-aware methods

In target-aware methods, whether a user is interested in the message of the query or not is indicated by a simple binary value. In other words, the behaviour of users, corresponding to BV, is modelled with binary values to indicate whether a user is interested in a message or not. Users interested in a message are considered as target nodes. Edges and the query have no further information, i.e. the sets BE and AM are not relevant in these methods. IM aims to select a seed set to spread the message to as many target nodes as possible.

In Mochalova and Nanopoulos (2014); White and Smyth (2003), several centrality measures are extended to determine the influence of each node relative to the target nodes. In these measures, the centrality is calculated just based on the target nodes; other nodes have no impact in centrality calculation. Then, the top-k central nodes are chosen as seeds. In Srinivasan et al. (2014); Wen et al. (2018d), the influence of each node is determined based on the paths between the node and other nodes. In Wen et al. (2018d), two sets, *influencers* and *followers*, are defined for each node to determine the region of influence of the node. Then, the nodes are clustered based on their region; the influential nodes are greedily selected to maximize the activation probability of the target nodes in different regions. The authors in Srinivasan et al. (2014) suggest a method based on the maximum influence arborescence method Chen et al. (2010), which guarantees a solution within a factor of the optimal solution.

Propagation of the message to target nodes is also the objective in Caliò and Tagarelli (2021); Calio et al. (2018). However, besides influence on target nodes, the diversity of the target nodes which are influenced is also taken into account. In Caliò and Tagarelli (2021), each target node is assumed to have a set of attributes; the problem is modelled as the identification of a seed set to spread the message to target nodes with diverse attributes. Then, a degree-based method is proposed to tackle the problem. The authors in Calio et al. (2018) assert that some silent nodes may be interested in a message but may not have been considered as target nodes due to lack of past activity. Spreading the message to different parts of the network can motivate such nodes. Therefore, the authors apply the reverse influence

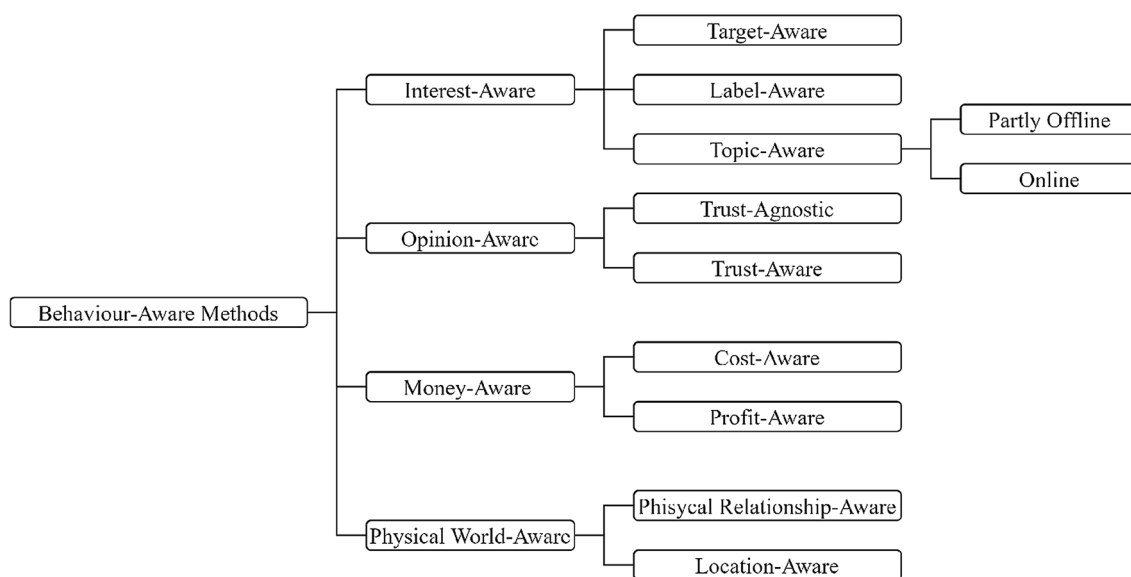


Fig. 2 A classification of behaviour-aware methods for IM

sampling method (Borgs et al. 2012) to propose a method that selects a seed set to influence target nodes that have structural diversity.

In Padmanabhan et al. (2018); Lohia et al. (2020), the authors suggest that activating non-target nodes may bring undesired costs and results. Therefore, they discuss how to select a set of influential nodes to maximize the number of activated target nodes while simultaneously minimizing the number of non-target nodes activated. The authors in Padmanabhan et al. (2018) propose a simulation-based greedy method to identify a set of influential nodes such that the number of activated target nodes is maximized, while the number of non-target activated nodes is kept below a given threshold; they propose heuristics for a time-efficient greedy method. In Lohia et al. (2020), maximizing the difference between the number of activated target nodes and activated non-target nodes is defined as the objective function of the problem. The authors propose a centrality measure to determine the influence of each node; influential nodes are selected in an iterative manner.

In summary, considering a set of nodes as target nodes prevents waste of resources (in terms of time or money) and avoids propagation of a message towards nodes who may not be interested in the topic of a message. Additional information about the diversity of nodes, as for example in Caliò and Tagarelli (2021); Calio et al. (2018), may boost the success of the spreading process. Taking into account non-target nodes, as in Padmanabhan et al. (2018); Lohia et al. (2020), helps avoid the activation of nodes that may spread an adverse message and negatively impact the IM process. However, relying on the diversity of nodes or non-target nodes needs information about users' behaviour which may not be always available. This suggests that the methods proposed in Caliò and Tagarelli (2021); Calio et al. (2018); Padmanabhan et al. (2018); Lohia et al. (2020) are less viable than the methods in White and Smyth (2003); Mochalova and Nanopoulos (2014); Wen et al. (2018d); Srinivasan et al. (2014).

### 3.1.2 Label-aware methods

Label-aware IM, also known as labelled IM, was first defined in Li et al. (2011). In label-aware IM, the features of a message, corresponding to AM in the query formulated, are modelled using a set of labels relevant to the message. The labels are taken into account to model the spread of the message in the network.

In Li et al. (2011); Tejaswi et al. (2017); Li et al. (2018), a set of labels, such as health or news, is assigned to each node. Thus, the behaviour of each user, corresponding to BV, is modelled using a set of labels in which the user is interested. Target and influential nodes are selected, according to the overlap between label sets related to the query and behaviour

of users. Two simulation-based methods are proposed in Li et al. (2011) to identify influential nodes for label-aware IM; for time efficiency, they also extend a degree-based method (Chen et al. 2009). The authors in Tejaswi et al. (2017) differentiate the state of the users between aware and adopted (in addition to inactive) and propose an extension of the linear threshold (LT) model in which every node influenced by the propagation process moves to the aware state and forwards the message, but only interested nodes can switch to adopted state and accept the message. Then, a degree-based method (Chen et al. 2009) is developed to identify a seed set which maximizes the number of users in the adopted state. The impact of friend conformity in the propagation process is taken into account in Li et al. (2018). In order to take into account friend conformity, users are divided into groups based on the similarity of labels relevant to their behaviour; a centrality measure is then developed to determine the influence of each node and group. A set of seeds is selected based on the influence of nodes and groups.

In some studies (Liu et al. 2015; Ke et al. 2018), instead of modelling the behaviour of each user, the features of the relationship between a pair of users are modelled using a set of labels; this set indicates the common interests between the pair based on their past communications. In Ke et al. (2018), the spreading (influence) probability between each pair of nodes is determined based on the similarity of the label sets relevant to the query and the relationship between the pair. Then, a reverse sampling approach is proposed to identify the most influential seeds. In Liu et al. (2015), the role of trust between users is also taken into account; the network is divided into domains based on the common labels and trust between users. Then, a degree-based method (Chen et al. 2009) is suggested to identify the influential users in each domain.

In summary, in Li et al. (2011); Tejaswi et al. (2017); Li et al. (2018), the past activity of each user is used to determine a set of labels in which the user is interested. Taking into account this set may help identify influential seeds effectively and target nodes that are potentially interested. However, to determine the set of labels relevant to the messages exchanged between the pairs, considering past interactions between pairs of users, as in Ke et al. (2018); Liu et al. (2015), may help model the propagation of a message realistically. This is because two users may share some common interests but may not influence each other in all common interests; if past interactions were not considered, the spreading probability between them for some common interests may be zero.

### 3.1.3 Topic-aware methods

In these methods, a set of network topics is first defined. The behaviour of each user, corresponding to BV, can be

modelled using a topic vector; the  $j$ -th entry of this vector is a number in  $[0, 1]$  and indicates how interested in the  $j$ -th topic the user is. Also, the behaviour of each relationship, corresponding to BE, can be modelled using a topic vector; the  $j$ -th entry of this vector is a number in  $[0, 1]$  and indicates the spreading (influence) probability on the edge for the  $j$ -th topic. The features of the message, corresponding to AM in the query, are modelled using a relevance topic vector whose  $j$ -th entry, a number in  $[0, 1]$ , denotes the relevance of the query to the  $j$ -th topic. In topic-aware methods, influential nodes are selected according to these vectors. Based on the methodology applied to identify the seeds, we divide topic-aware methods into two categories.

*Partly Offline Topic-Aware methods:* In these methods, the main idea is that there are lots of different messages related to different sets of topics and the identification of influential users for each message is accordingly time-consuming. Therefore, different seed sets for different sets of topics are selected and saved in advance using an offline approach. When a query comes up, the overlap between the topic vector of the query and the topic sets for which seeds have been saved is used to determine the most relevant saved seed sets and select the most influential seeds.

The behaviour of each user is modelled using a topic vector in Li et al. (2015); Aslay et al. (2014); Chen et al. (2016). In Li et al. (2015), reverse influence sampling (Borgs et al. 2012) is applied to determine and save offline the influence region of each node for different topics. When a query comes up, according to the influence region of nodes and the topic vector of the query, most influential nodes are identified as seeds. Some techniques are also proposed in this paper to enhance the time efficiency of the method. In Aslay et al. (2014); Chen et al. (2016), the authors try to enhance the time efficiency of seed selection by dividing the topics into a set of groups based on the similarity of the topics. In Aslay et al. (2014), the authors argue that users with similar interests are attracted by messages with similar features. Thus, topics are divided into groups based on the similarity. A set of most influential nodes for each group is identified and saved offline. When a query comes up, a set of seeds is identified based on the similarity of the topic vector of the query and predefined influential nodes for topic groups. In Chen et al. (2016), the authors observe that queries related to different topics may have a significant overlap; identifying an influential node set for each topic in advance and mixing the influential sets based on the query topics in online steps can speed up the seed selection. The authors adopt this hypothesis to propose a time-efficient method.

Instead of modelling the behaviour of users, in Chen et al. (2015) the features of the relationship between each pair of users are modelled using a topic vector, corresponding to BE, where the  $j$ -th value in the vector indicates the spreading (influence) probability of a message relevant to the  $j$ -th

topic between the pair. Then, the maximum influence arborescence method (Chen et al. 2010) is applied to develop a method to identify a set of seeds. Given the time complexity of this method, the authors apply an offline function to determine an upper bound for the influence of each node and reduce the computational complexity of seed selection.

*Online Topic-Aware Methods:* In online topic-aware methods, all steps to identify influential seeds are done online, when a query comes up; no preprocessing takes place.

In Barbieri et al. (2013); Singh et al. (2019), the features of the relationship between each pair of nodes are modelled using a topic vector, corresponding to BE. In Barbieri et al. (2013), according to the topic vector of the query, corresponding to AM, and the topic vector assigned to each edge, corresponding to BE, the spreading probability on each edge for the message is calculated. Then the IC and LT diffusion models are extended to identify a set of influential seeds. In Singh et al. (2019), the authors take into account the role of communities in the network to select a set of influential seeds covering different parts of the network. They divide the network into a set of communities and select a number of influential seeds from each community according to the size of the community.

In Zhou et al. (2014); Zareie et al. (2019); Tian et al. (2020); Li et al. (2020), the behaviour of each user is modelled using a topic vector, corresponding to BV. In Zhou et al. (2014), the authors first discuss how to determine the interest of each user in different topics based on their activities in the network. Then, they propose a simulation-based method to identify a set of influential seeds based on the topic vectors of the query and users. In Zareie et al. (2019), the similarity between the topic vector of users and the topic vector of the query is calculated using entropy divergence notation to determine a weight for each node that indicates how interested the node is in the query's message. According to the nodes' weight, a weight is then assigned to each edge, and the influential nodes are selected based on the weight of the edges connected to each node. In Tian et al. (2020), a network embedding approach is proposed to capture the interests of users; then, this approach is applied to determine the influence of each node. A set of influential nodes is identified using a reinforcement learning-based algorithm. In Li et al. (2020), the authors argue that greater similarity between the interests of a user and features of the query is an indication that the user may be influenced in the spreading process. Thus, they calculate the similarity between the topic vector of users and the topic vector of the query to determine the probability that each node is influenced in the spreading process. Then, an independent cascade model is extended to capture this property; a degree-based heuristic is proposed to determine the influence of each node and select the seeds in an iterative manner.

In summary, in some methods [partly offline methods (Li et al. 2015; Aslay et al. 2014; Chen et al. 2016) and online methods (Zhou et al. 2014; Zareie et al. 2019; Tian et al. 2020; Li et al. 2020)], the behaviour of users is taken into account to determine how interested each user is in the query's message. On the other hand, in other methods [partly offline methods (Chen et al. 2015) and online methods (Barbieri et al. 2013; Singh et al. 2019)], the log of the interactions between each pair of users is used to model the features of the relationship between the pair. As discussed before, taking into account the features of relationships may help model the propagation process in the network effectively. To compare partly offline and online methods, when a query comes up, partly offline methods may identify influential nodes to trigger the spreading process more efficiently, which is an advantage especially for breaking news spreading in a social network. However, partly offline methods need to deal with the running time for offline execution and also storage issues in relation to saving the influential seeds for queries with different topic vectors. Another challenge of partly offline methods is how to determine the potential

queries in advance to identify the influential seeds for different topic vectors.

### 3.1.4 Summary

The properties of the discussed interest-aware methods are summarized in Table 1. As mentioned, in target-aware methods, a binary value is used for each user to determine whether the user is interested in the query's message or not, something that may not properly capture the views of a user towards a message. Label-aware methods try to address this by taking into account a set of labels in which the user is interested, and these methods also use a binary state to represent whether a user is interested in a special topic or not. Conversely, topic-aware methods capture user preferences more realistically by applying a continuous value to represent the interest of each user towards a topic. However, determining the topic vector for each user demands a good record of historical data of users' past activities which may not be always available.

**Table 1** Properties of interest-aware methods, including applied behavioural features, applied method for influence detection—Centrality Measure (CM), Spreading Simulation (SS), Reverse Influence

Sampling (RIS), and Maximum Influence Arborescence (MIA)—and the type of spreading process

Category	Paper (year)	Behavioural features			Influence detection				Spreading process	
		AM	BV	BE	CM	SS	RIS	MIA	Simple	Complex
Target-aware	White and Smyth (2003)		✓		✓				✓	
	Mochalova and Nanopoulos (2014)		✓		✓				✓	
	Wen et al. (2018d)		✓					✓	✓	
	Srinivasan et al. (2014)		✓					✓	✓	
	Calio et al. (2018)		✓				✓			✓
	Caliò and Tagarelli (2021)		✓				✓		✓	
	Padmanabhan et al. (2018)		✓				✓		✓	✓
	Lohia et al. (2020)	✓		✓					✓	
Label-aware	Li et al. (2011)	✓	✓		✓	✓				✓
	Liu et al. (2015)	✓		✓	✓				✓	
	Li et al. (2018)	✓	✓	✓	✓				✓	
	Tejaswi et al. (2017)	✓		✓	✓					✓
	Ke et al. (2018)	✓		✓			✓		✓	
Partly offline topic-aware	Li et al. (2015)	✓	✓				✓		✓	✓
	Aslay et al. (2014)	✓		✓		✓			✓	
	Chen et al. (2016)	✓		✓		✓			✓	
	Chen et al. (2015)	✓		✓				✓		✓
Online topic-aware	Barbieri et al. (2013)	✓	✓			✓			✓	✓
	Singh et al. (2019)	✓		✓		✓			✓	✓
	Zareie et al. (2019)	✓	✓		✓				✓	
	Zhou et al. (2014)	✓	✓			✓			✓	
	Tian et al. (2020)	✓	✓		✓				✓	✓
	Li et al. (2020)	✓	✓		✓				✓	



### 3.2 Opinion-aware methods

In these methods, each user has an opinion or attitude about the query's message, i.e. the behaviour of each user towards the query's message is modelled using a discrete or a continuous value. Using a discrete value, opinion has a limited set of choices and can often be indicated by a binary choice between positive/negative or a choice between opinions A/B; in this case the goal of IM is to maximize the number of users subscribing to a specific opinion. Using a continuous value, the opinion of each user is denoted by a real value to express the leaning of users towards the query's message; in this case IM is about selecting a set of influential users that can potentially maximize this value for the network users.

It has been often mentioned that opinion change depends on the trust between the users. Thus, it has been proposed to model social networks as a signed graph whose edges have a positive or negative sign indicating trust or distrust between users. Yet, trust between the users is not taken into account by all methods. Thus, it is useful to divide opinion-aware methods into two categories: trust-agnostic and trust-aware. Same as with interest-aware methods, in opinion-aware methods,  $|S| = k$  is also considered as a constraint of the IM problem as defined in Eq. (1), which means the size of the seed set is fixed.

#### 3.2.1 Trust-agnostic methods

In Gionis et al. (2013); Zhang et al. (2013); Wang et al. (2018), the opinion of each user is modelled using a continuous value; a higher value corresponds to a more favourable opinion towards the query's message. In Zhang et al. (2013), the authors develop an opinion cascade model to simulate the spreading process and how users' attitude may change over this process. Then, this model is applied to suggest a greedy method to find a seed set, which, when triggered, maximizes the total opinion of influenced users. In Gionis et al. (2013), the behaviour of each user is modelled using two opinions: internal opinion, which persists over time, and expressed opinion, which is influenced by other users over time. Taking into account both internal and external opinions, the changing of expressed opinion of users is modelled; a greedy method is suggested which greedily selects the seeds in an iterative manner. The authors in Wang et al. (2018), in addition to the opinion of users, model the features of relationship between the users by taking into account both spreading probability and interaction probability on the edges. Inspired from fluid dynamics, a diffusion model is proposed to simulate the change of users' opinion over spreading process. This model is then used to describe a greedy method to find a seed set which, when triggered, maximizes the number of active users whose opinion is greater than a given threshold.

In He et al. (2021, 2023), the authors model the behaviour of each user using a continuous value; this value is considered as a threshold to determine whether the user is influenced during the spreading process. The authors assert that traditional diffusion models (independent cascade and linear threshold models) cannot properly define the dynamics of users' opinion. In He et al. (2021), the independent cascade model is used to simulate the spreading process, while the dynamics of the users' opinion is captured using a voter model; combining these two models a multi-stage diffusion model is proposed. The influence of each node is determined using a centrality measure; a greedy method along with a heuristic algorithm is suggested to identify a set of seeds. In He et al. (2023), the authors argue that, because users' opinion changes dynamically over time, seeds should be selected in an adaptive manner to guarantee the quality of the spreading process. They apply reinforcement learning theory to propose a multi-stage heuristic to identify influential seeds. In this method, the spreading process is initialized with a batch of seeds; additional seeds are selected during the spreading process.

In Chen et al. (2011); Nazemian and Taghiyareh (2012), the opinion of each user is modelled using a discrete value (negative, neutral, or positive) that models attitude towards the query's message; IM aims to maximize the number of users with positive opinion. In Chen et al. (2011), a cascade model is proposed to simulate the spreading process with negative opinions. The authors first apply this model to propose a simulation-based method; then, maximum influence arborescence (Chen et al. 2010) is applied to propose a time-efficient method. In Nazemian and Taghiyareh (2012), the authors argue that when spreading a product advertisement users' complaining behaviour should be taken into account; a user may be satisfied with a product but may still have something to complain about. Thus, in addition to users' opinion, they also model the complaining behaviour of users during the spreading process, which is considered by a simulation-based method to identify an influential seed set.

In summary, in Chen et al. (2011); Chen and He (2015), the opinion of each user is modelled using a discrete value, while in Gionis et al. (2013); Zhang et al. (2013); Wang et al. (2018); He et al. (2021, 2023), a continuous value is used to model the opinion of each user. Compared to the methods in Gionis et al. (2013); Zhang et al. (2013); Wang et al. (2018), the proposed methods in He et al. (2021, 2023) try to model more accurately how the opinion of users changes under the influence of friends. However, the dynamics of users' opinion under the influence of their friends may not be straightforward and cannot be modelled for all users in the same way.

### 3.2.2 Trust-aware methods

Users may consider other users as either friends or foes (Li et al. 2013); thus, considering both positive and negative relationships can improve opinion-aware methods. In trust-aware methods, each user has their own opinion towards the query's message; the opinion of users may change under the influence of friend or foe relationships with other users. The friend or foe relationship is typically represented by a sign; edges with a positive sign indicate friendship between users, while a negative sign indicates a foe relationship. Therefore, in trust-aware methods, the feature of the relationship between a pair of nodes is modelled using a sign describing the friend or foe relationship.

In some methods (Wang et al. 2015; Lei et al. 2016; Wang et al. 2016; Liang et al. 2019), the behaviour of each user is modelled as a continuous value, falling in  $[-1, 1]$ , to indicate the opinion of the user about the query's message; the behaviour of each relationship is modelled as a sign indicating friend or foe relationship. In Wang et al. (2015), the authors present an extension of the linear threshold (LT) model to simulate propagation in a signed network and update users' attitude over propagation. Applying this model, a simulation-based method is proposed to select a set of seeds that maximize the number of nodes with a positive opinion. In order to propose a time-efficient method for this problem, in Lei et al. (2016), a centrality measure is defined to determine the influence of each node based on the friend and foe relationships of the node in a three-hop neighbourhood; nodes with the greatest centrality are selected as seeds. In Wang et al. (2016), the authors extend the linear threshold model to take into account different states for each user during the spreading process; how the state of each user changes is determined using some given thresholds. The authors suggest a simulation-based greedy algorithm to identify an influential seed set. In Liang et al. (2019), the behaviour of each user is modelled using two continuous values: internal opinion, which does not change during the spreading process, and expressed opinion, which may be influenced and change during the propagation as a result of the influence of the other users. The authors develop a linear threshold model to simulate the propagation in this setting and propose a greedy method along with some heuristics to select a set of influential nodes.

In some other methods, the opinion of each user towards the query's message is modelled as a discrete value, such as negative/neutral/positive, A/B or red/blue. Without loss of generality, we can refer to all these as negative, neutral or positive opinion. Here, the aim of IM is to maximize the number of users with positive opinion after propagation.

In Li et al. (2014, 2017); Hosseini-Pozveh et al. (2016); Ju et al. (2020); Srivastava et al. (2015); Mohamadi-Baghmolaei et al. (2015), the independent cascade model is adopted

as the diffusion model to simulate the spreading process in the network. In Li et al. (2014) a diffusion model, named polarity-related independent cascade, is developed to simulate propagation in a signed network and change of users' opinion. Applying this model, a polarity-related method is proposed to find a set of nodes with positive opinion which, when triggered, maximize the number of nodes with positive opinion. Due to the low time-efficiency of this method, some research has tried to improve the time complexity of seed selection under the polarity-related independent cascade model; the authors in Li et al. (2017); Hosseini-Pozveh et al. (2016) apply meta-heuristic algorithms, while the authors in Ju et al. (2020) propose a path-based centrality measure to identify influential nodes. In Srivastava et al. (2015), the authors extend the problem by assuming that nodes with negative opinion can also be selected as seeds to initiate competitive propagation in the network; a centrality measure is suggested to address the problem. In Mohamadi-Baghmolaei et al. (2015), it is assumed that when a node is influenced by one of its neighbours, it takes some time until the node attempts to influence other neighbours; therefore, in addition to the trust, time latency is defined to model the behaviour of each relationship. The authors develop an extension of the independent cascade model to take into account both trust and time latency in the spreading process; then, a greedy algorithm, with some heuristics for efficiency improvement, is proposed to identify an influential seed set under the extended model.

In Ahmed and Ezeife (2013); He et al. (2016); Liang et al. (2019), the linear threshold model is adopted as the diffusion model. In Ahmed and Ezeife (2013), the authors propose an extension of the linear threshold model to simulate the propagation process and dynamics of users' opinion under the influence of neighbours with opposing opinions. A method is then proposed which greedily selects seeds using a simulation-based approach. In line with the diffusion model presented in Ahmed and Ezeife (2013) but using a different model for how the opinion of users changes during the spreading process, an extension of the linear threshold model is proposed in He et al. (2016); Liang et al. (2019); these papers apply some heuristics to efficiently identify a set of influential nodes. In He et al. (2016), a centrality measure under the extended linear threshold model is defined to detect the influence of each node; a set of influential nodes is selected using a heuristic method. The authors in Liang et al. (2019) identify a set of influential nodes using a simulation-based method incorporated with two acceleration techniques: (i) remove a set of nodes with small spreading ability and (ii) prune the number of paths to reduce the time to simulate the spreading process.

Some studies (Li et al. 2013; Chen and He 2015; Shafaei and Jalili 2014; He et al. 2021, 2022) avoid conventional diffusion models when simulating the spreading process in

order to take into account the formation of users' opinion under the influence of neighbours. A voter model is proposed in Li et al. (2013) to analyse the influence diffusion dynamics mathematically. The authors apply this model to determine the influence of nodes in long- and short-term propagation. This voter model is employed in Chen and He (2015) to develop an extension of pagerank centrality Brin and Page (2012), which determines the influence of nodes. The impact of community structure on the influence of a set in signed networks is investigated in Shafaei and Jalili (2014). For this purpose, the spreading process is simulated using a game-theoretic approach to assess the correlation between the influence of different seed sets and inter- and intra-community edges. The authors conclude that there is a significant correlation between the influence ability of a seed set and community structure. In He et al. (2021), the authors argue that the linear diffusion model may not appropriately determine the dynamics of users' opinion in the spreading process; the authors propose an extended linear threshold model that incorporates an opinion formation model. They also propose a centrality measure to determine the influence of each node in this model; a greedy method along with a heuristic algorithm is suggested to identify a set of seeds. In He et al. (2022), reinforcement learning is used to model how the users' opinion changes during the spreading process and propose an adaptive seed selection method where some seeds are selected during the spreading process.

In summary, modelling the opinion of users using continuous values (Wang et al. 2015; Lei et al. 2016; Wang et al. 2016; Liang et al. 2019) may capture their opinion more accurately compared to discrete values (Li et al. 2014, 2017; Hosseini-Pozveh et al. 2016; Ju et al. 2020; Srivastava et al. 2015; Mohamadi-Baghmolaei et al. 2015; Ahmed and Ezeife 2013; He et al. 2016; Liang et al. 2019; Li et al. 2013; Chen and He 2015; Shafaei and Jalili 2014; He et al. 2021, 2022). The methods proposed in Li et al. (2013); Chen and He (2015); Shafaei and Jalili (2014); He et al. (2021, 2022) try to model the dynamics of users' opinions differently from the spreading process. In fact, opinion dynamics is a topic that may need further attention in future research.

### 3.2.3 Summary

The properties of the discussed opinion-aware methods are summarized in Table 2. Trust-agnostic methods consider that all network relationships are positive (friend), whereas in reality networks may also include negative relationships (foe); user reaction to suggestions may differ depending on whether it comes from a friend or a foe. Overall, there is scope to model positive and negative relationships more elaborately than what has been attempted in the research literature up to now.

## 3.3 Money-aware methods

In these methods monetary aspects of the spreading process are taken into account. Some money (or equivalent, such as discount on a product) is paid to users to persuade them to take part in the spreading process as a seed. Alternatively, the activation of different users may have different monetary benefits in the spreading process. Overall, money-aware methods can be divided into two categories: cost-aware methods, where a budget constraint is considered in seed set selection, and profit-aware methods where, in addition to budget, specific decisions during the spreading process may lead to monetary benefits.

### 3.3.1 Cost-aware methods

In these methods, selection of each user as a seed is accompanied with a cost; the behaviour of each user, corresponding to  $BV$ , is modelled as a value indicating the cost that the user may ask in order to agree to take part in the spreading process as a seed. In this case, the constraint in the query, corresponding to  $B$ , is the total budget; the sum of the costs of the seeds must not exceed this budget. This implies that, in these methods, instead of specifying a fixed number of seeds, as it was the case with the methods discussed earlier, the query is driven by a budget to identify a seed set.

Influence maximization using a budget constraint, also known as budgeted-IM, was first introduced in Nguyen and Zheng (2013). The authors defined the problem under the independent cascade model; this model was applied to propose a greedy method to find seeds in several iterations. In each iteration, the node with the greatest influence to cost ratio is added to the seed set. They also applied the maximum influence arborescence method (Chen et al. 2010) to propose a time-efficient greedy algorithm to solve the problem. In Chiesse et al. (2014), the authors evaluate the impact of considering the cost of nodes and the budget constraint under the independent cascade model. They assess four well-known centrality measures to select top nodes as seeds and show that these measures cannot effectively identify a highly influential seed set in cost-aware IM.

In Wang et al. (2015); Han et al. (2014); Güneş (2019), the authors propose time-efficient methods to identify influential nodes under the independent cascade model. Two iterative functions are proposed in Wang et al. (2015) to determine the influence, susceptibility of being influenced and the persuasion cost of each node based on the properties of the node. Then, a method is proposed to select seeds iteratively; in each iteration, based on the cost and influence of nodes, a seed is selected and the susceptibility of being influenced of the other nodes is updated. In Han et al. (2014), using a centrality measure, a set of nodes is selected as candidate seeds; the set contains the nodes with greatest influence and/

**Table 2** Properties of opinion-aware methods, including applied behavioural features, applied method for influence detection—Centrality Measure (CM), Spreading Simulation (SS), Reverse Influence Sampling (RIS), and Maximum Influence Arborescence (MIA)—and the type of spreading process

Category	Paper (year)	Behavioural features			Influence detection				Spreading process	
		AM	BV	BE	CM	SS	RIS	MIA	Simple	Complex
Trust-agnostic	Gionis et al. (2013)		✓		✓					✓
	Zhang et al. (2013)		✓			✓				✓
	Wang et al. (2018)		✓			✓			✓	✓
	He et al. (2021)		✓		✓					✓
	He et al. (2023)		✓			✓			✓	
	Chen et al. (2011)		✓		✓			✓	✓	
	Nazemian and Taghiyareh (2012)		✓		✓	✓			✓	
Trust-aware	Li et al. (2013)		✓	✓	✓					✓
	Wang et al. (2015)		✓	✓		✓				✓
	Liang et al. (2019)		✓	✓		✓				✓
	Lei et al. (2016)		✓	✓	✓					✓
	Chen and He (2015)		✓	✓	✓					✓
	Li et al. (2014)		✓	✓			✓		✓	
	Li et al. (2017)		✓	✓	✓				✓	
	Ju et al. (2020)		✓	✓	✓				✓	
	Srivastava et al. (2015)		✓	✓	✓				✓	
	Hosseini-Pozveh et al. (2016)		✓	✓			✓		✓	
	Mohamadi-Baghmolaei et al. (2015)		✓	✓	✓	✓			✓	
	He et al. (2016)		✓	✓	✓					✓
	Liang et al. (2019)		✓	✓			✓			✓
	Ahmed and Ezeife (2013)		✓	✓			✓			✓
	Wang et al. (2016)		✓	✓			✓			✓
He et al. (2021)		✓	✓			✓			✓	
He et al. (2022)		✓	✓	✓					✓	

or least cost. The authors apply a meta-heuristic algorithm to develop a combination strategy to identify the best seeds as a composition of the nodes with great influence and low cost. In Güney (2019), the problem is modelled using binary integer programming. The author first applies the independent cascade model to propose a simulation-based method to identify influential seeds; then, a sampling approach is developed to propose a time-efficient method which guarantees an approximation to the optimal solution.

In de Souza et al. (2020), the problem is defined under the linear threshold model. This model is used to determine a surrounding set for each node which contains the nodes with the least cost to activate the node. This set is used to determine the influence to cost ratio of the node; then, a set of influential nodes is identified based on this ratio. In Yu et al. (2018); Du et al. (2013), the problem is solved using knapsack constraints. In Yu et al. (2018), the authors apply the history of the relationship of each user with its direct and indirect neighbours to determine the influence of the user. Seed set selection is described as an iterative process; in each iteration, the node with the greatest influence is added

to the seed set if it satisfies a particular cost-based criterion. The problem is extended in Du et al. (2013) to propagate multiple messages about multiple items. Some constraints on items are first considered and the problem is modelled using knapsack constraints. Then, based on the cost efficiency of assigning a specific item to each user, an adaptive method is proposed to identify influential seeds for each item.

The role of community structure in propagation has been taken into account in Banerjee et al. (2019, 2020). For this purpose, in Banerjee et al. (2019), the network is first divided into communities and a portion of the budget is assigned to each community, based on the number of nodes in the community. In each community, a set of nodes is iteratively selected as seeds; in each iteration the node with the maximum degree whose cost is not greater than the remaining budget of the community is selected as a seed. In Banerjee et al. (2020), the authors take into account the interests of users in each community to determine each user's influence. They assume the network contains a set of topics; users in each community are interested in a subset of these topics. Then, applying maximum influence arborescence Chen et al.

(2010), a method is proposed to identify influential users and influential topics in each community.

In Han et al. (2021), the authors address the problem from a different point of view. They consider a set of advertisers, each of whom has a limited budget and aims to spread a message to advertise a product. Given each advertiser's budget, the owner of the network selects a seed set for each advertiser to initialize the spreading process. According to the costs of seeds and also the number of influenced users for the message corresponding to each advertiser, the benefit of the owner is determined. The question is how to select a set of influential users for each advertiser so that the total benefit of the owner of the network is maximized. To tackle the problem, the authors propose sampling methods to identify the influential users that meet this goal.

### 3.3.2 Profit-aware methods

Apart from monetary incentives to persuade users to act as seeds in the spreading process, the activation of each user may bring some benefit for the organisation running the campaign. In profit-aware methods, in addition to the cost, a monetary benefit is defined to model the behaviour of each user; the benefit is a value which is gained if the user is activated during the spreading process. The goal is to maximize the overall profit based on the cost and benefit of the spreading process.

In Nguyen et al. (2016); Tang et al. (2017); Lu and Lakshmanan (2012), the behaviour of each user is modelled using two values indicating its cost if selected as a seed, and its benefit if activated in the propagation process. In Nguyen et al. (2016), the IM problem is defined as identification of a seed set whose cost is not greater than a given budget and maximizes the total benefit from all activated nodes during the spreading process. The reverse influence sampling method (Borgs et al. 2012) is then used to find influential users; nodes that have greater influence on nodes returning high benefits are considered as influential nodes. In Tang et al. (2017); Lu and Lakshmanan (2012), the problem is defined as the identification of a set of nodes to maximize profit, i.e. benefit minus cost, without any budget limitations. In Tang et al. (2017), the authors first propose a greedy method to solve the problem; they further propose a double greedy method to propose a method which guarantees an approximation of the optimal solution of the problem. The problem is further extended in Lu and Lakshmanan (2012) by defining two features (price of a product offered to each user and valuation of the user) to model the decision-making behaviour of each user. Then, in order to simulate the spreading process, the linear threshold model is extended to incorporate these features. This model is applied to propose a greedy method to identify the seeds that maximize the profit.

In Banerjee et al. (2021, 2020a), the authors assume that influencing the users who are not interested in the query's message brings no gain; therefore, a set of interested users is considered as the target of the spreading process. According to the distance (hops) between a pair of nodes, a heuristic method is proposed in Banerjee et al. (2021) to identify a set of influential nodes maximizing the total benefit. In order to improve this process, the role of communities is taken into account in Banerjee et al. (2020a). In this method, based on the number of target nodes in each community, a proportion of the budget is assigned to the community; a centrality measure is then proposed to identify influential users in each community.

Profit-aware IM is further extended in Zhu et al. (2019); Zhang et al. (2016). Instead of activating nodes, in Zhu et al. (2019), the aim is activating communities. It is assumed that the activation of each community brings a benefit; a community is activated if at least a given percentage of the nodes in the community is activated. Then, the IM problem is defined as identifying a seed set for which the profit, i.e. the total benefit of the activated community minus the total cost of the seeds, is maximized. The authors apply the reverse influence sampling method (Borgs et al. 2012; Tang et al. 2014) to propose a method which guarantees an approximation of the optimal solution. Profit-aware IM is extended in Zhang et al. (2016) by considering the query formulated to advertise a set of different products; the behaviour of each user towards the adoption of each product is modelled by an adoption threshold; the adoption of each product by the user brings different benefits. The linear threshold model is extended to incorporate these features. The authors first apply this model to propose an iterative greedy method to identify the seeds that maximize profit; in each iteration, the node with the greatest influence to cost ratio is added to the seed set as a new seed. Then, some heuristics are developed to suggest a time-efficient method. The authors also model the problem as a multi-choice knapsack problem to distribute the budget across the products optimally; then, they apply binary integer programming to solve the problem.

### 3.3.3 Summary

The properties of the discussed money-aware methods are summarized in Table 3. Cost-aware methods try to model IM in social networks by taking into account the costs of the spreading process. However, these methods assume that all influenced users bring the same benefit. On the other hand, profit-aware methods try to model the problem by taking into account monetary concerns of users in real-world scenarios. For example, in online advertisements, people have their own valuations for buying the product advertised by a campaign. Therefore, influencing different users may bring different benefits to the

**Table 3** Properties of money-aware methods, including applied behavioural features, applied method for influence detection—Centrality Measure (CM), Spreading Simulation (SS), Reverse Influence Sampling (RIS), and Maximum Influence Arborescence (MIA)—and the type of spreading process

Category	Paper (year)	Behavioural features			Influence detection				Spreading process	
		AM	BV	BE	CM	SS	RIS	MIA	Simple	Complex
Cost-Aware	Nguyen and Zheng (2013)		✓			✓			✓	
	Han et al. (2014)		✓		✓				✓	
	Banerjee et al. (2019)		✓		✓				✓	
	Banerjee et al. (2020)		✓	✓				✓	✓	
	Yu et al. (2018)		✓		✓				✓	
	Du et al. (2013)	✓	✓		✓				✓	
	Güney (2019)		✓			✓	✓		✓	
	Wang et al. (2015)		✓		✓				✓	
	de Souza et al. (2020)		✓		✓					✓
Profit-Aware	Han et al. (2021)		✓				✓		✓	
	Nguyen et al. (2016)		✓				✓		✓	✓
	Zhang et al. (2016)	✓	✓		✓	✓				✓
	Lu and Lakshmanan (2012)	✓	✓			✓				✓
	Tang et al. (2017)		✓		✓				✓	
	Banerjee et al. (2021)		✓		✓	✓			✓	
	Banerjee et al. (2020a)		✓		✓	✓			✓	
	Zhu et al. (2019)		✓				✓		✓	

campaign; considering this aspect of the problem helps profit-aware methods boost the success of the spreading process compared to cost-aware methods. It is noted that data about users' monetary concerns and valuations may not be always readily available, which means that cost-aware methods may be more relevant than profit-aware methods in practice.

### 3.4 Physical world-aware methods

All methods discussed so far considered online relationships between users. However, users may have relationships or other things in common through their activities in the physical world. For example, they may live in the same neighbourhood, and they may be regulars of the same pub, or may attend events held at the same location. In such cases, users' behaviour may also take into account aspects of their physical world presence. We term such methods as physical world-aware methods, and we divide them into two categories: (i) physical relationship-aware methods, where, alongside an online relationship, the physical world relationship between the users is considered to determine the influence of users, while messages may propagate through both online and physical world relationships; (ii) location-aware methods, where the query includes location and the probability that users visit this location is taken into account. In the latter case, the aim is to identify a set of seeds to spread a message so that the number of users visiting the location is maximized.

#### 3.4.1 Physical relationship-aware methods

As mentioned, the main assumption of these methods is that users, in addition to the online world, have some relationships in the physical world too; messages can be propagated in the physical world as well. In order to capture this, the network is modelled as a double-layered graph to show the relationship between users in both the online and physical world. Whenever two online users appear to come from the same physical location at the same time, a relationship between these two users in the physical world is implied; hence, an edge in the physical world layer (often termed as offline layer) between these two users is added. Edges in the online layer are processed in the same way as discussed previously. The physical location of users may be determined through online logging information at check-in.

In Yang et al. (2018); Li et al. (2018), if the Euclidean distance between two users over a period of time is less than a defined threshold, these users are considered as neighbours in the physical world layer; clearly the notion of neighbour (and edges) in the physical world layer is dynamic and changes over the time. In Yang et al. (2018), to select a seed set, the network is compressed into a single layer and the problem is solved using a simulation-based strategy. The authors also propose a centrality-based method to identify a set of influential seeds covering different parts of the network. In Li et al. (2018), the physical world edges are weighted; according to the ratio of attendance of two users in a common place over a period of time, a weight is assigned to each physical world edge. The two layers are

then compressed into a single layer, and the reverse influence sampling method (Borgs et al. 2012) is used to find influential seeds who can spread a message widely in both the online and physical world layers. In Chen et al. (2021), the authors take into account the effect of community structure. In their paper, the similarity between a pair of nodes is determined based on the mobility of users in physical world; then, this similarity is used to decompose the network into a set of communities. A simulation-based method and a centrality-based method are proposed to identify the influential nodes in each community.

In Li et al. (2018); Hosseinpour et al. (2019), the geographical distance between a pair of users is taken into account to model the behaviour of users. In Li et al. (2018), the authors model the network as a single-layered graph; the geographical distance between two users in the physical world along with users' interests is taken into account to model the behaviour of users; the spreading (influence) probability between the users is determined based on their behaviour. Influential nodes are then identified with the help of the reverse influence sampling method (Borgs et al. 2012); some heuristics are also proposed to improve the efficiency of the proposed method. In Hosseinpour et al. (2019), the problem is modelled as the identification of a set of influential nodes to maximize the geographical coverage of the locations around the vicinity of the location of the query formulated. The authors propose hierarchical graph modelling alongside some indices to take into account the structural information of the network and location of users. A centrality measure is proposed to identify the geographical coverage of each node and select a set of influential nodes.

In summary, the model defined in Yang et al. (2018); Li et al. (2018) makes use of a binary value to define whether there is a physical relationship between a pair of nodes. As discussed before, a binary value may not properly capture the behaviour of users, especially the location of users in the physical world which changes dynamically. Thus, in the model defined in Li et al. (2018); Hosseinpour et al. (2019), a real value is used to determine the physical distance between a pair of users and the spreading probability between them. Yet, neglecting the dynamics of the users' physical location may negatively affect the success of these methods (Li et al. 2018; Hosseinpour et al. 2019) in real-world scenarios.

### 3.4.2 Location-aware methods

The main assumption in this family of methods is that an event, such as exhibition, commemoration, sale and so on, is held in a special location. The goal is to spread a publicity message through the network aiming to maximize the number of users visiting the event.

In Li et al. (2014), according to the location of the event, a target region is included in the query formulated. The goal is to maximize the number of activated users located in the region. In order to find a set of influential nodes, the maximum influence arborescence method (Chen et al. 2010) is applied to determine the influence of each node on the nodes in the target region; a set of influential nodes is identified in an iterative manner. Users who are not in the target region but have a short distance to the region are neglected in Li et al. (2014). Thus, instead of determining a region, some studies (Song et al. 2016; Wang et al. 2019, Wang et al. 2016a, b) consider the distance of the users to the location of the event. The goal is to maximize the number of activated users that have a short distance to the location. In Wang et al. (2016a, b), for the sake of time efficiency, a partly offline strategy is adopted. In this strategy a set of sample locations is first determined; the influential set for each sample location is identified and saved offline. When a query comes up, the distance between the sample locations and the event location is considered to determine influential seeds close to the event location. The problem is further extended in Song et al. (2016); Wang et al. (2019) by taking additional behavioural information into account. In Song et al. (2016), besides event location, time constraints are also taken into account; the authors apply the reverse influence sampling method (Borgs et al. 2012) to determine a set of influential seeds. In Wang et al. (2019), event location, users' interests and spreading cost are considered; the problem is expressed as a multi-objective optimization problem, which is solved using particle swarm optimization (Eberhart and Kennedy 1995).

In Zhu et al. (2015); Zhou et al. (2015); Li et al. (2018); Su et al. (2018), the mobility of users is taken into account to determine the probability of attendance of each user in the target region. The goal is to spread the message to the users with high probability of attendance in the region. Users' check-in records are applied in these approaches; the number of times a user checked-in in the region is the measure to estimate the probability of attendance of the user in the region. In Zhu et al. (2015), the authors discuss how users' check-in behaviour may be applied to model user mobility; they define three Gaussian models to capture mobility. In Zhou et al. (2015), an extension of the independent cascade model (Kempe et al. 2003) is presented to simulate the spreading process based on the event location and the probability of attendance of users in the location. A centrality measure is then defined to determine the influence of each node in this model; the nodes with maximum centrality are chosen as seeds. The role of communities in propagation is taken into account in Li et al. (2018). For this purpose, the maximum influence arborescence method (Chen et al. 2010) is applied to determine the influence of each community based on the event location; then, the influential nodes in

each community are determined. The influence of communities and nodes is both taken into account to identify a set of influential seeds. In order to speed up the method, the authors also propose an offline tree-based model to determine and ignore the users that have no region intersections with the location defined in the query. In Su et al. (2018), in addition to the users' mobility, users' interests are also taken into account to determine influential nodes; the goal is to activate interested users with a high probability of attendance of the event. A tree-based model is defined to determine the users that have no region intersections with the location defined in the query or the users who are not interested in the event. Then, the maximum influence arborescence method (Chen et al. 2010) is extended to determine a set of influential seeds based on users' interests and probability of attendance of the event.

In Gu et al. (2021), the authors consider a situation where a query aims to spread a marketing message about a sale taking place in specific geographical locations; each user has a budget to buy some product. Given the budget and geographical distance of each user, the maximum influence arborescence method (Chen et al. 2010) is applied to define some rules to select a set of candidate seeds and a centrality measure is proposed to select the influential seeds among them.

Compared to the methods in Wang et al. (2016a, b); Song et al. (2016); Wang et al. (2019), which determine influential seeds based on the physical distance of users to the location defined in the query, the methods in Zhu et al. (2015); Zhou et al. (2015); Li et al. (2018); Su et al. (2018) may capture the behaviour of users more accurately, because they take

into account user mobility and the likelihood that users will be in a target location. Nevertheless, the physical location of users may be highly dynamic; tracking and storing all locations for the users over a period of time may bring some processing and storage challenges.

### 3.4.3 Summary

The properties of the physical world-aware methods are summarized in Table 4. The spread of a message and its influence may be affected not only by online interactions but also by physical interactions between users. Therefore, physical relationship-aware methods attempt to model the spreading process and track the influence of users not only on the basis of online interaction but by also incorporating the possible physical interactions between them. Location-aware methods aim to model the IM problem under different scenarios in which users within a specific geographical region are considered to be the target of the spreading process. Overall, although physical world-aware methods attempt to solve the IM problem in a realistic way, taking into account the physical interactions between users and their location impose an additional complexity to the problem in terms of data modelling and data availability.

## 4 Challenges and future directions

In this paper, we presented a new taxonomy of the methods proposed to solve the IM problem in social networks with a focus on methods that take into account elements of

**Table 4** Properties of physical world-aware methods, including applied behavioural features, applied method for influence detection—Centrality Measure (CM), Spreading Simulation (SS), Reverse

Influence Sampling (RIS), and Maximum Influence Arborescence (MIA)—and the type of spreading process

Category	Paper (year)	Behavioural features			Influence detection				Spreading process	
		AM	BV	BE	CM	SS	RIS	MIA	Simple	Complex
Physical relationship-aware	Yang et al. (2018)		✓		✓	✓			✓	
	Li et al. (2018)		✓				✓		✓	✓
	Chen et al. (2021)		✓		✓	✓			✓	
	Li et al. (2018)		✓				✓		✓	
	Hosseinpour et al. (2019)	✓	✓		✓				✓	
Location-aware	Li et al. (2014)	✓	✓					✓	✓	
	Wang et al. (2016)	✓	✓				✓	✓	✓	
	Wang et al. (2016)	✓	✓					✓	✓	
	Song et al. (2016)	✓	✓				✓		✓	
	Wang et al. (2019)	✓	✓				✓		✓	
	Zhou et al. (2015)	✓	✓		✓				✓	
	Li et al. (2018)	✓	✓					✓	✓	
	Su et al. (2018)	✓	✓					✓	✓	
Gu et al. (2021)	✓	✓					✓	✓		



query's features, users' behaviour and relationships' features. Research in relation to the IM problem is ongoing and attracts a significant interest as there are various challenges to address. In our view, behaviour-agnostic methods are not always accurate enough to detect influential nodes as they adopt a mechanical approach to link users through a network. Although it has been argued that users' personality may be inferred from online networks (Golbeck et al. 2011a, b), real-world behaviour of individuals is clearly far more complex than what networks may indicate or capture. Behaviour-agnostic methods essentially assume that all users behave the same and their relationships follow a common pattern, an assumption that may not be realistic in real-world scenarios. On the other hand, behaviour-aware methods try to take into account behavioural characteristics to differentiate users and their relationships based on the history of their activities in the network; these characteristics help identify influential users and initiate a successful spreading process more accurately. However, getting hold of and using users' behaviour data may pose complexity and other challenges that will be highlighted next.

*Data availability challenges:* Users' behaviour data may not exist for many real-world applications (Jalili and Perc 2017). Lots of users may be silent or may have no significant activities in a network. Also, they may not be prepared to share important details (e.g. age, gender, location) in their profile, something which may make it difficult to determine the behavioural characteristics of these users and their relationships. As a result, one may deal with a network with incomplete behavioural information in which a considerable number of users and relationships cannot be properly modelled. In addition, users' information may be governed by privacy regulations; making use of such information may not be possible without user consent and full consideration of any privacy issues that may arise.

*Data processing challenges:* Obtaining users' behaviour from historical data of users' activities has a number of big data challenges and various preprocessing steps may be necessary before such data are usable (Bello-Organ et al. 2016; Manovich 2011; Cuomo and Maiorano 2018). In general, such data may be noisy, redundant, unstructured or inconsistent. Storing, analysing and processing such data may impose costs and challenges that may be exacerbated in the context of social networks as a result of the increasing and diverse number of user interactions.

*Data modelling challenges:* In many behaviour-aware methods, the behaviour of users is modelled using numeric (quantitative) values. However, personality traits may not be simply interpreted using numeric values. For example, determining the opinion of users as a value in  $[0, 1]$ , largely determined by user posts or activities, does not fully capture user attitude. As another example, defining the relationship

between a pair of users as a friend or foe relationship is not a straightforward process.

*Data dynamicity challenges:* Users and their relationships may have dynamic behavioural characteristics in a network. The structure of the network and the features of the network components may dynamically change over the time. As a result, any data obtained are incidental and may be accompanied by uncertainty. As behavioural features of network components become important, data dynamicity may significantly affect the performance of behaviour-aware methods.

*User reaction modelling challenges:* User reaction to suggestions and influence from others is a complicated process which needs to take into account different aspects of personality traits (Aral and Walker 2012). How to model this influence in order to simulate message spreading in a network is a challenging issue. For instance, modelling user reaction to influence from friends or foes may be more complicated than simple binary states currently used by most trust-aware methods.

These challenges point to different directions, which deserve future research.

The first direction that future research will need to consider is the size of the networks. Day by day, the number of users and the volume of data exchanged between the users is increasing. The time complexity of all methods to address the IM problem, particularly when they take into account users' behaviour, needs to be considered. There may be some questions requiring further research in this case: how to propose efficient methods to identify solutions within a factor of the optimal solution, what techniques may be used to compose, effectively and efficiently, offline and online strategies for the IM problem, how to propose and implement highly parallelizable methods that can take advantage of parallel execution.

The second direction is to consider the diversity and dynamicity of the social networks themselves. Different types of networks such as temporal networks, multi-layer networks or dynamic networks are getting popular; there is a need for methods that consider such networks. Some questions to address are: how to take into account the dynamicity of the network structure and user behaviour to find solutions for IM, how to deal with incomplete data and uncertainty in the structural and behavioural information of the networks, how to deal with the diversity of multi-layer networks which are a composition of different social networks.

The third direction relates to data preparation and how somebody can make use of the data to model the networks efficiently and effectively. This aspect of the problem is not properly addressed in the literature; how to analyse and interpret behavioural data to model a network is a challenge. There are additional questions that may transcend traditional computer science boundaries: how to determine the behaviour of users based on their past activities in the network,

how to map behavioural data into numeric values to capture the behaviour of users and relationships.

Finally, information spreading in almost all methods is modelled using a conventional diffusion model. These models may not be the best way to emulate information spreading in reality. Changing the diffusion model may impact the accuracy of the methods significantly. Further research may be needed to assess the impact of users' behaviour in the spreading process and provide suitable diffusion models. A key question is how to take into account the impact that the spreading process may have in users' behaviour, as any such impact may affect subsequent stages of the spreading process. In fact, changes in user behaviour, as they happen, may need to be fed back to the diffusion model itself, something that motivates the need for adaptive and dynamic diffusion models.

**Author contribution** AZ: conceptualization, methodology, writing of the first draft. RS: conceptualization, supervision, editing. Both authors reviewed and edited the final manuscript.

## Declarations

**Conflict of interest** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Ahmed S, Ezeife C (2013) Discovering influential nodes from trust network. In: Proceedings of the 28th annual ACM symposium on applied computing, pp 121–128. ACM
- Al-Garadi MA, Varathan KD, Ravana SD, Ahmed E, Mujtaba G, Khan MUS, Khan SU (2018) Analysis of online social network connections for identification of influential users: survey and open research issues. *ACM Comput Surv (CSUR)* 51(1):16
- Aral S, Walker D (2012) Identifying influential and susceptible members of social networks. *Science* 337(6092):337–341
- Arora A, Galhotra S, Ranu S (2017) Debunking the myths of influence maximization: an in-depth benchmarking study. In: Proceedings of the 2017 ACM international conference on management of data, pp 651–666. ACM
- Aslay C, Barbieri N, Bonchi F, Baeza-Yates RA (2014) Online topic-aware influence maximization queries. In: Proceedings of the 17th international conference on extending database technology (EDBT), pp 295–306
- Banerjee S, Jenamani M, Pratihari DK (2019) Combim: a community-based solution approach for the budgeted influence maximization problem. *Expert Syst Appl* 125:1–13
- Banerjee S, Jenamani M, Pratihari DK (2020) Maximizing the earned benefit in an incentivized social networking environment: a community-based approach. *J Ambient Intell Humaniz Comput* 11(6):2539–2555
- Banerjee S, Jenamani M, Pratihari DK (2020) A survey on influence maximization in a social network. *Knowl Inf Syst* 62(9):3417–3455
- Banerjee S, Jenamani M, Pratihari DK (2021) Earned benefit maximization in social networks under budget constraint. *Expert Syst Appl* 169:114346
- Banerjee S, Pal B, Jenamani M (2020) Budgeted influence maximization with tags in social networks. In: International conference on web information systems engineering, pp 141–152. Springer
- Barbieri N, Bonchi F, Manco G (2013) Topic-aware social influence propagation models. *Knowl Inf Syst* 37(3):555–584
- Bello-Orgaz G, Jung JJ, Camacho D (2016) Social big data: recent achievements and new challenges. *Inf Fusion* 28:45–59
- Bian R, Koh YS, Dobbie G, Divoli A (2019) Identifying top-k nodes in social networks: a survey. *ACM Comput Surv (CSUR)* 52(1):22
- Borgs C, Brautbar M, Chayes J, Lucier B (2014) Maximizing social influence in nearly optimal time. In: Proceedings of the 25th annual ACM-SIAM symposium on discrete algorithms, pp 946–957. SIAM
- Borgs C, Brautbar M, Chayes JT, Lucier B (2012) Influence maximization in social networks: towards an optimal algorithmic solution. *arXiv preprint arXiv:1212.0884*
- Borodin A, Filmus Y, Oren J (2010) Threshold models for competitive influence in social networks. In: International workshop on internet and network economics, pp 539–550. Springer
- Brin S, Page L (2012) Reprint of: the anatomy of a large-scale hypertextual web search engine. *Comput Netw* 56(18):3825–3833
- Calìo A, Tagarelli A (2021) Attribute based diversification of seeds for targeted influence maximization. *Inf Sci* 546:1273–1305
- Calìo A, Interdonato R, Pulice C, Tagarelli A (2018) Topology-driven diversity for targeted influence maximization with application to user engagement in social networks. *IEEE Trans Knowl Data Eng* 30(12):2421–2434
- Carnes T, Nagarajan C, Wild SM, Van Zuylen A (2007) Maximizing influence in a competitive social network: a follower's perspective. In: Proceedings of the 9th international conference on electronic commerce, pp 351–360. ACM
- Chen S, Fan J, Li G, Feng J, Kl Tan, Tang J (2015) Online topic-aware influence maximization. *Proc VLDB Endow* 8(6):666–677
- Chen S, He K (2015) Influence maximization on signed social networks with integrated pagerank. In: 2015 IEEE international conference on smart City/SocialCom/SustainCom, pp 289–292. IEEE
- Chen W, Collins A, Cummings R, Ke T, Liu Z, Rincon D, Sun X, Wang Y, Wei W, Yuan Y (2011) Influence maximization in social networks when negative opinions may emerge and propagate. In: Proceedings of the 2011 SIAM international conference on data mining, pp 379–390. SIAM
- Chen W, Lakshmanan LV, Castillo C (2013) Information and influence propagation in social networks. *Synth Lect Data Manag* 5(4):1–177
- Chen W, Lin T, Yang C (2016) Real-time topic-aware influence maximization using preprocessing. *Comput Soc Netw* 3(1):8
- Chen W, Wang C, Wang Y (2010) Scalable influence maximization for prevalent viral marketing in large-scale social networks. In: Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining, pp 1029–1038. ACM
- Chen W, Wang Y, Yang S (2009) Efficient influence maximization in social networks. In: Proceedings of the 15th ACM SIGKDD

- international conference on knowledge discovery and data mining, pp 199–208. ACM
- Chen X, Deng L, Zhao Y, Zhou X, Zheng K (2021) Community-based influence maximization in location-based social network. *World Wide Web* 24(6):1903–1928
- Chiesse R, Figueiredo DR, Antonio AdA, Ziviani A, Niterói R, Petrópolis R (2014) Evaluation of epidemic seeding strategies under variable node costs. In: WPerformance 2014 Workshop, XXXIV Congresso Nacional da Sociedade Brasileira de Computação (CSBC), Brazil
- Cuomo S, Maiorano F (2018) Social network data analysis and mining applications for the internet of data. *Concurr Comput Pract Exp* 30(15):e4527
- Das K, Samanta S, Pal M (2018) Study on centrality measures in social networks: a survey. *Soc Netw Anal Min* 8(1):13
- De Salve A, Mori P, Guidi B, Ricci L, Pietro RD (2021) Predicting influential users in online social network groups. *ACM Trans Knowl Discov Data* 15(3):1
- de Souza R, Figueiredo D, Rocha AdA, Ziviani A (2020) Efficient network seeding under variable node cost and limited budget for social networks. *Inf Sci* 514:369–384
- Domingos P, Richardson M (2001) Mining the network value of customers. In: Proceedings of the 7th ACM SIGKDD international conference on knowledge discovery and data mining, pp 57–66. ACM
- Du N, Liang Y, Balcan MF, Song L (2013) Budgeted influence maximization for multiple products. arXiv preprint [arXiv:1312.2164](https://arxiv.org/abs/1312.2164)
- Eberhart R, Kennedy J (1995) A new optimizer using particle swarm theory. In: MHS'95. In: Proceedings of the 6th international symposium on micro machine and human science, pp 39–43. Ieee
- Freeman LC (1977) A set of measures of centrality based on betweenness. *Sociometry* 40(1):35–41
- Freeman LC (1978) Centrality in social networks conceptual clarification. *Soc Netw* 1(3):215–239
- Gionis A, Terzi E, Tsaparas P (2013) Opinion maximization in social networks. In: Proceedings of the 2013 SIAM international conference on data mining, pp 387–395. SIAM
- Golbeck J, Robles C, Edmondson M, Turner K (2011a) Predicting personality from twitter. In: 2011 IEEE 3rd international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing, pp 149–156. IEEE
- Golbeck J, Robles C, Turner K (2011b) Predicting personality with social media. In: CHI'11 extended abstracts on human factors in computing systems, pp 253–262. ACM
- Goldenberg J, Libai B, Muller E (2001) Using complex systems analysis to advance marketing theory development: modeling heterogeneity effects on new product growth through stochastic cellular automata. *Acad Mark Sci Rev* 9(3):1–18
- Gong Q, Chen Y, He X, Xiao Y, Hui P, Wang X, Fu X (2021) Cross-site prediction on social influence for cold-start users in online social networks. *ACM Trans Web* 15(2):1
- Granovetter M (1978) Threshold models of collective behavior. *Am J Sociol* 83(6):1420–1443
- Gu Y, Yao X, Liang G, Gu C, Huang H (2021) Efficient budget-distance-aware influence maximization in geo-social network. In: International conference on wireless algorithms, systems, and applications, pp 282–290. Springer
- Guille A, Hacid H, Favre C, Zighed DA (2013) Information diffusion in online social networks: a survey. *ACM SIGMOD Rec* 42(2):17–28
- Güney E (2019) On the optimal solution of budgeted influence maximization problem in social networks. *Oper Res Int J* 19(3):817–831
- Hafiene N, Karoui W, Ben Romdhane L (2020) Influential nodes detection in dynamic social networks: a survey. *Expert Syst Appl* 159:113642
- Han K, Wu B, Tang J, Cui S, Aslay C, Lakshmanan LV (2021) Efficient and effective algorithms for revenue maximization in social advertising, p 671–684. Association for computing machinery, New York, NY, USA
- Han S, Zhuang F, He Q, Shi Z (2014) Balanced seed selection for budgeted influence maximization in social networks. In: Pacific-Asia conference on knowledge discovery and data mining, pp 65–77. Springer
- He JS, Kaur H, Talluri M (2016) Positive opinion influential node set selection for social networks: considering both positive and negative relationships. In: Wireless communications, networking and applications, pp 935–948. Springer
- He Q, Fang H, Zhang J, Wang X (2023) Dynamic opinion maximization in social networks. *IEEE Trans Knowl Data Eng* 35(1):350–361
- He Q, Lv Y, Wang X, Li J, Huang M, Ma L, Cai Y (2022) Reinforcement learning based dynamic opinion maximization framework in signed social networks. *IEEE Trans Cognit Dev Syst*. <https://doi.org/10.1109/TCDS.2022.3141952>
- He Q, Sun L, Wang X, Wang Z, Huang M, Yi B, Wang Y, Ma L (2021) Positive opinion maximization in signed social networks. *Inf Sci* 558:34–49
- He Q, Wang X, Huang M, Yi B (2021) Multi-stage opinion maximization in social networks. *Neural Comput Appl* 33(19):12367–12380
- Hosseini-Pozveh M, Zamanifar K, Naghsh-Nilchi AR, Dolog P (2016) Maximizing the spread of positive influence in signed social networks. *Intell Data Anal* 20(1):199–218
- Hosseinpour M, Malek MR, Claramunt C (2019) Socio-spatial influence maximization in location-based social networks. *Futur Gener Comput Syst* 101:304–314
- Jalili M, Perc M (2017) Information cascades in complex networks. *J Complex Netw* 5(5):665–693
- Jaouadi M, Ben Romdhane L (2019) Influence maximization problem in social networks: an overview. In: 2019 IEEE/ACS 16th international conference on computer systems and applications (AICCSA), pp 1–8
- Ju W, Chen L, Li B, Liu W, Sheng J, Wang Y (2020) A new algorithm for positive influence maximization in signed networks. *Inf Sci* 512:1571–1591
- Ke X, Khan A, Cong G (2018) Finding seeds and relevant tags jointly: for targeted influence maximization in social networks. In: Proceedings of the 2018 international conference on management of data, p 1097–1111. Association for computing machinery
- Kempe D, Kleinberg J, Tardos É (2003) Maximizing the spread of influence through a social network. In: Proceedings of the 9th ACM SIGKDD international conference on knowledge discovery and data mining, pp 137–146. ACM
- Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, Stanley HE, Makse HA (2010) Identification of influential spreaders in complex networks. *Nat Phys* 6(11):888
- Lei W, Yang Q, Wang H (2016) Positive influence maximization algorithm based on three degrees of influence. In: International conference on intelligent data engineering and automated learning, pp 503–514. Springer
- Li D, Wang C, Zhang S, Zhou G, Chu D, Wu C (2017) Positive influence maximization in signed social networks based on simulated annealing. *Neurocomputing* 260:69–78
- Li D, Xu ZM, Chakraborty N, Gupta A, Sycara K, Li S (2014) Polarity related influence maximization in signed social networks. *PLoS ONE* 9(7):e102199
- Li FH, Li CT, Shan MK (2011) Labeled influence maximization in social networks for target marketing. In: 2011 IEEE international

- conference on privacy, security, risk and trust and IEEE international conference on social computing, pp 560–563. IEEE
- Li G, Chen S, Feng J, Tan KL, Li Ws (2014) Efficient location-aware influence maximization. In: Proceedings of the 2014 ACM SIGMOD international conference on management of data, pp 87–98. ACM
- Li J, Cai T, Mian A, Li RH, Sellis T, Yu JX (2018) Holistic influence maximization for targeted advertisements in spatial social networks. In: 2018 IEEE 34th international conference on data engineering (ICDE), pp 1340–1343. IEEE
- Li L, Liu Y, Zhou Q, Yang W, Yuan J (2020) Targeted influence maximization under a multifactor-based information propagation model. *Inf Sci* 519:124–140
- Li M, Wang Z, Han QL, Taylor SJE, Li K, Liao X, Liu X (2022) Influence maximization in multiagent systems by a graph embedding method: dealing with probabilistically unstable links. *IEEE Trans Cybern*. <https://doi.org/10.1109/TCYB.2022.3227805>
- Li S, Han K, Zhang J (2018) Exploring influence maximization in location-based social networks. In: International conference on collaborative computing: networking, applications and worksharing, pp 92–111. Springer
- Li X, Cheng X, Su S, Sun C (2018) Community-based seeds selection algorithm for location aware influence maximization. *Neurocomputing* 275:1601–1613
- Li Y, Chen W, Wang Y, Zhang ZL (2013) Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships. In: Proceedings of the 6th ACM international conference on web search and data mining, pp 657–666. ACM
- Li Y, Fan J, Wang Y, Tan KL (2018) Influence maximization on social graphs: a survey. *IEEE Trans Knowl Data Eng* 30(10):1852–1872
- Li Y, Gan X, Fu L, Tian X, Qin Z, Zhou Y (2018) Conformity-aware influence maximization with user profiles. In: 2018 10th international conference on wireless communications and signal processing (WCSP), pp 1–6
- Li Y, Zhang D, Tan KL (2015) Real-time targeted influence maximization for online advertisements. *Proc VLDB Endow* 8(10):1070–1081
- Liang W, Shen C, Li X, Nishide R, Piumarta I, Takada H (2019) Influence maximization in signed social networks with opinion formation. *IEEE Access* 7(2019):68837–68852
- Liang W, Shen C, Li X, Nishide R, Piumarta I, Takada H (2019) Influence maximization in signed social networks with opinion formation. *IEEE Access* 7:68837–68852
- Liu S, Jiang C, Lin Z, Ding Y, Duan R, Xu Z (2015) Identifying effective influencers based on trust for electronic word-of-mouth marketing: a domain-aware approach. *Inf Sci* 306:34–52
- Lohia P, Kannan K, Rai K, Bedathur S (2020) Ranking marginal influencers in a target-labeled network. In: Proceedings of the 7th ACM IKDD CoDS and 25th COMAD, p 254–260. Association for Computing Machinery
- Lü L, Chen D, Ren XL, Zhang QM, Zhang YC, Zhou T (2016) Vital nodes identification in complex networks. *Phys Rep* 650:1–63
- Lü L, Zhou T, Zhang QM, Stanley HE (2016) The h-index of a network node and its relation to degree and coreness. *Nat Commun* 7:10168
- Lu W, Lakshmanan LV (2012) Profit maximization over social networks. In: 2012 IEEE 12th international conference on data mining, pp. 479–488. IEEE
- Manovich L (2011) Trending: the promises and the challenges of big social data. *Debates Digit Humanit* 2:460–475
- Mochalova A, Nanopoulos A (2014) A targeted approach to viral marketing. *Electron Commer Res Appl* 13(4):283–294
- Mohamadi-Baghmolaei R, Mozafari N, Hamzeh A (2015) Trust based latency aware influence maximization in social networks. *Eng Appl Artif Intell* 41:195–206
- Nazemian A, Taghiyareh F (2012) Influence maximization in independent cascade model with positive and negative word of mouth. In: 6th international symposium on telecommunications (IST), pp 854–860. IEEE
- Nguyen H, Zheng R (2013) On budgeted influence maximization in social networks. *IEEE J Sel Areas Commun* 31(6):1084–1094
- Nguyen HT, Dinh TN, Thai MT (2016) Cost-aware targeted viral marketing in billion-scale networks. In: IEEE INFOCOM 2016-The 35th annual IEEE international conference on computer communications, pp 1–9. IEEE
- Padmanabhan MR, Somisetty N, Basu S, Pavan A (2018) Influence maximization in social networks with non-target constraints. In: 2018 IEEE international conference on big data, pp 771–780
- Peng S, Zhou Y, Cao L, Yu S, Niu J, Jia W (2018) Influence analysis in social networks: a survey. *J Netw Comput Appl* 106:17–32
- Richardson M, Domingos P (2002) Mining knowledge-sharing sites for viral marketing. In: Proceedings of the 8th ACM SIGKDD international conference on knowledge discovery and data mining, pp 61–70. ACM
- Sabidussi G (1966) The centrality index of a graph. *Psychometrika* 31(4):581–603
- Shafaei M, Jalili M (2014) Community structure and information cascade in signed networks. *N Gener Comput* 32(3–4):257–269
- Singh SS, Kumar A, Singh K, Biswas B (2019) C2im: community based context-aware influence maximization in social networks. *Physica A* 514:796–818
- Singh SS, Singh K, Kumar A, Shakya HK, Biswas B (2018) A survey on information diffusion models in social networks. In: International conference on advanced informatics for computing research, pp 426–439. Springer
- Song C, Hsu W, Lee ML (2016) Targeted influence maximization in social networks. In: Proceedings of the 25th ACM international conference on information and knowledge management, pp 1683–1692. ACM
- Srinivasan BV, Anandavelu N, Dalal A, Yenugula M, Srikanthan P, Layek A (2014) Topic-based targeted influence maximization. In: 2014 6th international conference on communication systems and networks (COMSNETS), pp 1–6. IEEE
- Srivastava A, Chelmiss C, Prasanna VK (2015) Social influence computation and maximization in signed networks with competing cascades. In: 2015 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM), pp 41–48. IEEE
- Su S, Li X, Cheng X, Sun C (2018) Location-aware targeted influence maximization in social networks. *J Am Soc Inf Sci* 69(2):229–241
- Tang J, Tang X, Yuan J (2017) Profit maximization for viral marketing in online social networks: algorithms and analysis. *IEEE Trans Knowl Data Eng* 30(6):1095–1108
- Tang Y, Xiao X, Shi Y (2014) Influence maximization: Near-optimal time complexity meets practical efficiency. In: Proceedings of the 2014 ACM SIGMOD international conference on management of data, pp 75–86. ACM
- Tejaswi V, Bindu P, Thilagam PS (2016) Diffusion models and approaches for influence maximization in social networks. In: 2016 international conference on advances in computing, communications and informatics (ICACCI), pp 1345–1351. IEEE
- Tejaswi V, Bindu P, Thilagam PS (2017) Target specific influence maximization: an approach to maximize adoption in labeled social networks. In: 2017 9th international conference on communication systems and networks (COMSNETS), pp 542–547. IEEE
- Tian S, Mo S, Wang L, Peng Z (2020) Deep reinforcement learning-based approach to tackle topic-aware influence maximization. *Data Sci Eng* 5(1):1–11

- Wang F, Jiang W, Li X, Wang G (2018) Maximizing positive influence spread in online social networks via fluid dynamics. *Futur Gener Comput Syst* 86:1491–1502
- Wang F, Wang G, Xie D (2016) Maximizing the spread of positive influence under It-mla model. In: *Asia-Pacific services computing conference*, pp 450–463. Springer
- Wang H, Yang Q, Fang L, Lei W (2015) Maximizing positive influence in signed social networks. In: *International conference on cloud computing and security*, pp 356–367. Springer
- Wang L, Yu Z, Xiong F, Yang D, Pan S, Yan Z (2019) Influence spread in geo-social networks: a multiobjective optimization perspective. *IEEE Trans Cybern*
- Wang X, Zhang Y, Zhang W, Lin X (2016a) Distance-aware influence maximization in geo-social network. In: *ICDE*, pp 1–12
- Wang X, Zhang Y, Zhang W, Lin X (2016b) Efficient distance-aware influence maximization in geo-social networks. *IEEE Trans Knowl Data Eng* 29(3):599–612
- Wang Y, Vasilakos AV, Jin Q, Ma J (2015) Pprank: economically selecting initial users for influence maximization in social networks. *IEEE Syst J* 11(4):2279–2290
- Wen YT, Peng WC, Shuai HH (2018) Maximizing social influence on target users. In: *Pacific-Asia conference on knowledge discovery and data mining*, pp 701–712. Springer
- White S, Smyth P (2003) Algorithms for estimating relative importance in networks. In: *Proceedings of the 9th ACM SIGKDD international conference on knowledge discovery and data mining*, pp 266–275. ACM
- Yang Y, Pei J (2019) Influence analysis in evolving networks: a survey. *IEEE Trans Knowl Data Eng* 33(3):1045–63
- Yang Y, Xu Y, Wang E, Lou K, Luan D (2018) Exploring influence maximization in online and offline double-layer propagation scheme. *Inf Sci* 450:182–199
- Yu Q, Li H, Liao Y, Cui S (2018) Fast budgeted influence maximization over multi-action event logs. *IEEE Access* 6:14367–14378
- Zareie A, Sheikahmadi A (2019) Ehc: extended h-index centrality measure for identification of users' spreading influence in complex networks. *Physica A* 514:141–155
- Zareie A, Sheikahmadi A, Fatemi A (2017) Influential nodes ranking in complex networks: an entropy-based approach. *Chaos Solitons Fractals* 104:485–494
- Zareie A, Sheikahmadi A, Jalili M (2019) Identification of influential users in social networks based on users' interest. *Inf Sci* 493:217–231
- Zhang H, Dinh TN, Thai MT (2013) Maximizing the spread of positive influence in online social networks. In: *2013 IEEE 33rd international conference on distributed computing systems*, pp 317–326. IEEE
- Zhang H, Zhang H, Kuhnle A, Thai MT (2016) Profit maximization for multiple products in online social networks. In: *IEEE INFOCOM 2016-The 35th annual IEEE international conference on computer communications*, pp 1–9. IEEE
- Zhang Y, Guo J, Yang W, Wu W (2023) Supplementary influence maximization problem in social networks. *IEEE Trans Comput Soc Syst*. <https://doi.org/10.1109/TCSS.2023.3234437>
- Zhou J, Zhang Y, Cheng J (2014) Preference-based mining of top-k influential nodes in social networks. *Futur Gener Comput Syst* 31:40–47
- Zhou T, Cao J, Liu B, Xu S, Zhu Z, Luo J (2015) Location-based influence maximization in social networks. In: *Proceedings of the 24th ACM international conference on information and knowledge management*, pp 1211–1220. ACM
- Zhu J, Ghosh S, Wu W, Gao C (2019) Profit maximization under group influence model in social networks. In: *International conference on computational data and social networks*, pp 108–119. Springer
- Zhu WY, Peng WC, Chen LJ, Zheng K, Zhou X (2015) Modeling user mobility for location promotion in location-based social networks. In: *Proceedings of the 21st ACM SIGKDD international conference on knowledge discovery and data mining*, pp 1573–1582. ACM

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.