Engineering 4 (2018) 40–46

Contents lists available at ScienceDirect

Engineering

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

Research
Cybersecurity—Review

Social Influence Analysis: Models, Methods, and Evaluation

Kan Li *, Lin Zhang, Heyan Huang

School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China

ARTICLE INFO

Article history:
Received 10 December 2017
Revised 5 January 2018
Accepted 8 January 2018
Available online 16 February 2018

Keywords:
Social influence analysis
Online social networks
Social influence analysis models
Influence evaluation

ABSTRACT

Social influence analysis (SIA) is a vast research field that has attracted research interest in many areas. In this paper, we present a survey of representative and state-of-the-art work in models, methods, and evaluation aspects related to SIA. We divide SIA models into two types: microscopic and macroscopic models. Microscopic models consider human interactions and the structure of the influence process, whereas macroscopic models consider the same transmission probability and identical influential power for all users. We analyze social influence methods including influence maximization, influence minimization, flow of influence, and individual influence. In social influence evaluation, influence evaluation metrics are introduced and social influence evaluation models are then analyzed. The objectives of this paper are to provide a comprehensive analysis, aid in understanding social behaviors, provide a theoretical basis for influencing public opinion, and unveil future research directions and potential applications.

1. Introduction

Online social networks such as Weibo, Twitter, and Facebook provide valuable platforms for information diffusion among their users. During this process, social influence occurs when a person's opinions, emotions, or behaviors are affected by other people [1]. Thus, changes occur in an individual's attitudes, thoughts, feelings, or behaviors as a result of interaction with other people or groups. Social influence analysis (SIA) is becoming an important research field in social networks. SIA mainly studies how to model the influence diffusion process in networks, and how to propose an efficient method to identify a group of target nodes in a network [2]. Studied questions include: Who influences whom; who is influenced; who are the most influential users, and so forth. SIA has important social significance and has been applied in many fields. Viral marketing [3–10], online recommendation [11], healthcare communities [12–14], expert finding [15–17], rumor spreading [18], and other applications all depend on the social influence effect [19–21]. Analyzing social influence can help us to understand peoples' social behaviors, provide theoretical support for making public decisions and influencing public opinion, and promote exchanges and dissemination of various activities [22].

This paper provides a comprehensive view of SIA from the aspects of models, methods, and evaluation. To this end, we identify the strengths and weaknesses of existing models and methods, as well as those of the evaluation of social influence. First, we review existing social influence models. Next, we summarize social influence methods. Finally, we analyze the evaluation of social influence.

The rest of this paper is organized as follows. In Section 2, we discuss SIA models. In Section 3, we analyze SIA methods, including influence maximization, influence minimization, flow of influence, and individual influence. We then detail social influence evaluation in Section 4. Finally, we summarize the reviewed models and methods of social influence, and discuss open questions.

2. Social influence analysis models

SIA models have been widely studied in the literature. We classify these models into two categories: microscopic and macroscopic models.

2.1. Microscopic models

Microscopic models focus on the role of human interactions, and examine the structure of the influence process. The two frequently used influence analysis models in this category are the independent cascade (IC) [23–25] and linear threshold (LT)

* Corresponding author.
E-mail address: likan@bit.edu.cn (K. Li).

https://doi.org/10.1016/j.eng.2018.02.004

[23,26] models. Since Kempe et al. [23] used these two models, they are still mainly used to assess social influence diffusion.

2.1.1. The IC and LT models

The IC model. In a social network $G = (V, E)$ with a seed set S ($S \subseteq V$), where V is the set of nodes and E is the set of edges, and S_t ($S_t \subseteq V$) is the set of nodes that are activated at step t ($t \geqslant 0$). At step $t+1$, every node $v_i \in S_t$ can activate its out-neighbors $v_j \in V \setminus \cup_{0 \leqslant i \leqslant t} S_i$ with an independent probability P_{ij}. The process ends when no node can be activated. Note that a node has only one chance to activate its out-neighbors after it has been activated, and the node remains an activated node after it is activated.

The LT model. In a social network $G = (V, E)$, the sum of the influence weights of all the neighbors of node v_i meets $\sum_{v_j \in N_{g_i^{act}}} w_{ij} \leqslant 1$, where w_{ij} is the influence weight between node v_i and its neighbor node v_j, and $N_{g_i^{act}}$ is the set of its neighbor nodes activated by node v_i. Node v_i randomly selects its own threshold θ_i, uniformly chosen from 0 to 1. Only when the sum of the influence weights of its neighbor nodes exceeds this threshold will v_i be activated.

2.1.2. Variations

For both the IC and LT models, it is usually necessary to run the Monte Carlo simulation in order to estimate a node's influence for a sufficient time period. However, this is time-consuming and unsuitable for large-scale social networks. Many researchers have proposed methods to improve the IC and LT models. Here, we divide these improvements into four categories: variations of the IC model, variations of the LT model, variations of both the IC and LT models, and models that differ from the IC or LT models and their variations. We only list representative works in this paper.

(1) Variations of the IC model. Some researchers have considered time delay and time-critical constraints for influence diffusion. Chen et al. [27] extended the IC model and proposed an IC model with meeting events (IC-M model). In the IC-M model, the activated node has the probability to meet the inactive node. Compared with the general IC model, the calculation results from this model are closer to the actual situation, although the calculation of time consumption is slightly above that of the IC model. Feng et al. [28] incorporated novelty decay into the IC model. Based on previous studies, they found that repeated exposures have diminishing influence on users. Therefore, they developed a propagation path-based algorithm to assess the influence spread of seed nodes. There are two values on each edge of a social network: influence probability and expected influence delay time. Mohamadi-Baghmolaei et al. [29] considered important time and trust factors, and proposed the trust-based latency-aware independent cascade (TLIC) model. This is the first time trust has been studied in a classic IC model. In the TLIC model, a node can change its state (i.e., as active or inactive) with different probabilities for a trusted neighbor node than for a distrusted neighbor. Budak et al. [30] introduced the multi-campaign IC model, which models the diffusion of two cascades evolving simultaneously. They studied the notion of competing campaigns, in which the good campaign counteracts the effect of a bad campaign in a social network.

(2) Variations of the LT model. Liu et al. [31] considered the containment of competitive influence diffusion in social networks, and extended the LT model to construct the diffusion-containment (D-C) model. The traditional LT model is not applicable to that a situation concerns both the diffusion and the containment of the influence. In the D-C model, the state of a node is described by the activation probability; each node is only influenced by a neighbor with a higher probability, and the sum of the probabilities of possible node states is not greater than 1. Borodin et al. [32] analyzed the competitive influence diffusion of different models based on the LT model.

(3) Variations of both the IC and LT models. Mohammadi et al. [33] considered the time delay and proposed two diffusion models: the delayed independent cascade (DIC) model and the delayed linear threshold (DLT) model. In these two models, nodes have three states: active, inactive, and latent active. To pass from an inactive state to an active state, a node must pass through the middle process of being in a latent active state. Compared with the traditional IC and LT models, the effect of influence diffusion in this model is better. However, because more than one state is possible, the calculation complexity is relatively high. In the IC and LT models, information diffusion is treated as a series of node state changes that occur in a synchronous way. However, the actual diffusion takes place in an asynchronous way, and the time stamps of the observed data are not evenly spaced out. Some models relax the synchronicity assumption of traditional IC and LT models and extend these two models to make the state change asynchronously. Saito et al. [34] proposed the asynchronous extensions: asynchronous IC (AsIC) and asynchronous LT (AsLT) models. Their learning algorithm can estimate the influence degrees of nodes in a network from relatively few information diffusion results, which avoids the overfitting problem. Guille and Hacid [35] also presented an asynchronous model—time-based AsIC model—to model the diffusion process. Other studies [30,32,36–44] have improved the classical models in specific directions. Fan et al. [38] introduced two models: the opportunistic one-activate-one (OPOAO) model, in which each person can only communicate with another person simultaneously; and the deterministic one-activate-many (DOAM) model, which has a mechanism that is similar to an information broadcast procedure. For the OPOAO model, they used the classical greedy algorithm to produce a $(1 - 1/e)$ approximation ratio; for the DOAM model, they proposed a set cover-based greedy (SCBG) algorithm to achieve an $O(\ln n)$ (n is the number of vertices) factor solution. The transmission probabilities in these two models are the same for all users. Galam [45] proposed a model that investigates opinion dynamics, followed by Lee et al. [39], who proposed a new scheme to improve Galam's model. Nevertheless, these models are confined to word-of-mouth information-exchanging process.

(4) Models that differ from the IC or LT models and their variations. Some models are different from the IC or LT models and their variations, and have solved information influence diffusion from a new point of view. Lin et al. [46] proposed a data-driven model to maximize the expected influence in the long run using meta-learning concepts. However, this model needs large amounts of data, and the accuracy of its results requires further improvement. Golnari et al. [47] proposed a heat conduction (HC) model. It considers a non-progressive propagation process, and is completely different from the previous IC or LT models, which only consider the progressive propagation process. In the HC model, the influence cascade is initiated from a set of seeds and arbitrary values for other nodes. Wang et al. [48] studied emotion influence in large-scale image social networks, and proposed an emotion influence model. They designed a factor graph model to infer emotion influence from images in social networks. Gao [49] proposed a read-write (RW) model to describe the detailed processes of opinion forming, influence, and diffusion. However, there are three main issues that this model needs to consider further: the many parameters of the model that must be inferred, the proper collection of datasets about opinion influence and diffusion, and the evaluation metrics that are suitable for this task.

Whether a classical IC or LT model or a new model, the ultimate goal for a model of social influence is to achieve faster computing speed and more accurate results, while using less memory.

However, these microscopic models are lacking in some major details. Firstly, because of the difference in characteristics (i.e., educational background and individual consciousness), different people identify with the same information to different degrees. Secondly, different individuals have different capabilities to influence other users; also, spreaders with different degrees of authority have different impacts on their neighbors and on society. Thirdly, the probability that an individual will transmit a piece of information to others is not constant, and should depend on the individual's attraction to the information. Moreover, these models use a binary variable to record whether an individual becomes infected. In addition, they assume that once an individual is infected, the individual never changes its state; however, this assumption does not reflect the realistic smoothness of the transition of individuals from one state to another.

2.2. Macroscopic models

Macroscopic models consider all users to have the same attraction to information, the same transmission probability, and identical influential power. However, since macroscopic models do not take individuals into account, the accuracy of these models' results is lower. To improve such models, therefore, the differences between individuals should be considered. Macroscopic models divide nodes into different classes (i.e., states) and focus on the state evolution of the nodes in each class. The percentage of nodes in each class is expressed by simple differential equations. Epidemic models are the most common models that are used to study social influence from a macroscopic perspective. These models were mainly developed to model epidemiological processes. However, they neglect the topological characteristics of social networks. The percentage of nodes in each class is computed by mean-field rate equations, which are too simple to depict such a complex evolution accurately. Daley and Kendall [50] analyzed the similarity between the diffusion of an infectious disease and the dissemination of a piece of information, and proposed the classic Daley–Kendall model. Since then, researchers have improved these epidemic models in general to overcome their weaknesses. Refs. [50–53] consider the topological characteristics of underlying networks in their methods. However, these scholars have ignored the impact of human behaviors in the influence diffusion process.

Researchers have recently begun to consider the role of human behaviors and different mechanisms during information influence diffusion [54–60]. Zhao et al. [59,60] proposed the susceptible–infected–hibernator–removed model as an extension from the classical susceptible–infected–removed (SIR) model in order to incorporate the forgetting and remembering mechanism; they investigated this problem in homogeneous and inhomogeneous networks. Wang et al. [54] proposed a variation of the rumor-diffusion model in online social networks that considers negative or positive social reinforcements in the acceptant probability model. They analyzed how social reinforcement affects the spreading rate. Wang et al. [55] proposed an SIR model by introducing the trust mechanism between nodes. This mechanism can reduce the ultimate rumor size and the velocity of rumor diffusion. Xia et al. [56] presented a modified susceptible–exposed–infected–removed (SEIR) model to discuss the impact of the hesitating mechanism on the rumor-spreading model. They took into account the attractiveness and fuzziness of the contents of rumors, and concluded that the more clarity a rumor has, the smaller its effects will be. Su et al. [57] proposed the microblog-susceptible–infected–removed model for information diffusion by explicitly considering users' incomplete reading behavior. In Ref. [58], motivated by the work in Refs. [56,57], Liu et al. extended the model in Ref. [56] and proposed a new SEIR model on heterogeneous networks in order to study the diffusion dynamics in a microblog.

3. Social influence analysis methods

SIA methods are used to solve the sub-problems of social network influence analysis, such as influence maximization, influence minimization, flow of influence, and individual influence. All these problems involve influence diffusion, so the same influence model can apply to all of them in some cases. However, the ultimate goal of using the model is different for each problem.

3.1. Influence maximization

Influence maximization requires finding the most influential group of members in social networks. Kempe et al. [23] formulated this problem. Given a directed graph with users as nodes, edge weights that reflect the influence between users, and a budget/threshold number k, the purpose of influence maximization is to find k nodes in a social network, so that the expected spread of the influence can be maximized by activating these nodes. This is a discrete optimization problem that is non-deterministic polynomial-time (NP)-hard for both the IC and LT models. Influence maximization is the most widely studied problem in SIA. Leskovec et al. [10] and Rogers [20] studied this problem as an algorithmic problem, and proposed some probabilistic methods. Influence maximization must achieve fast calculation, high accuracy, and low storage capacity. Most of the algorithms that consider the differences between individuals are based on greedy algorithms or heuristic algorithms.

3.1.1. Greedy algorithms

Greedy algorithms "greedily" select the active node with the maximum marginal gain toward the existing seeds in each iteration. The study of greedy algorithms is based on the hill-climbing greedy algorithm, in which each choice can provide the greatest impact value of the node using the local optimal solution to approximate the global optimal solution. The advantage of this algorithm is that the accuracy is relatively high, reaching $(1 - 1/e - \varepsilon)$ for any $\varepsilon > 0$. However, the algorithm has high complexity and high execution time, so the efficiency is relatively poor. Kempe et al. [23] were the first to establish the influence of the maximum refinement as a discrete optimization problem, and proposed a greedy climbing approximation algorithm. Leskovec et al. [61] proposed a greedy optimization method, the cost-effective lazy forward (CELF) method. Chen et al. [27] put forward new greedy algorithms, the NewGreedy and MixGreedy methods. Zhou et al. [62] proposed the upper bound-based lazy forward (UBLF) algorithm to discover top-k influential nodes. They established new upper bounds in order to significantly reduce the number of Monte Carlo simulations in greedy algorithms, especially at the initial step. Some of the algorithms that are used to study the differences between individuals are based on these greedy algorithms.

3.1.2. Heuristic algorithms

Due to the high computational complexity of greedy algorithms, many excellent heuristic algorithms have been proposed to reduce the solution-solving time and pursue higher algorithm efficiency. These heuristic algorithms iteratively select nodes based on a specific heuristic, such as degree or PageRank, rather than computing the marginal gain of the nodes in each iteration. Their drawback is that the accuracy is relatively low. The most basic heuristic algorithms are the Random, Degree, and Centrality heuristic algorithms proposed by Kempe et al. [23]. Based on the degree of the heuristic algorithm, Chen et al. [41] proposed a heuristic algorithm for the IC model: DegreeDiscount. They then proposed a new heuristic algorithm, prefix excluding maximum influence arborescence (PMIA) [63]. For the LT model, Chen et al.

[63] proposed local directed acyclic graph (LDAG) heuristic algorithm. Of course, other heuristic algorithms are also based on these heuristic algorithms, such as SIMPATH [64] and IRIE [65]. Borgs et al. [66] made a theoretical breakthrough and presented a quasi-linear time algorithm for influence maximization under the IC model. Tang et al. [67] proposed a two-phase influence maximization (TIM) algorithm for influence maximization. The expected time of the algorithm is $O[(k+\ell)(n+m)\log n/\varepsilon^2]$, and it returns a $(1-1/e-\varepsilon)$ approximate solution with at least $(1-n^{-\ell})$ probability.

The time complexity of the abovementioned algorithms is shown in Table 1 [23,41,61,63–67].

Here, we describe some representative studies that take into account individual differences or the user's own attributes. Li et al. [68] considered the behavioral relationships between humans (i.e., regular and rare behaviors) in order to simulate the effects of heterogeneous social networks to solve the influence maximization problem. They proposed two entropy-based heuristic algorithms to identify the communicators in the network, and then maximized the impact of the propagation. Although the entropy-based heuristic performance is better than those of Degree [23] and DegreeDiscount [41], this method is still based on heuristic ideas, so its accuracy is inadequate. Subbian et al. [69] proposed the individual social value: Existing influence maximization algorithms cannot model individual social value, which is usually the real motivation for nodes to connect to each other. Subbian et al. [69] used the concept of social capital to propose a framework in which the social value of the network is calculated by the number of bindings and bridging connections. The performance of the algorithm is better than that of PMIA [63], PageRank, and the weighted degree method. Li et al. [70] proposed a conformity-aware cascade (C^2) model and a conformity-aware greedy algorithm to solve the influence maximization problem. This algorithm works well when applied to the distributed platform. However, because it is based on the greedy algorithm and considers the conformity-aware calculation, the complexity of the calculation still requires improvement. Lee and Chung [71] formulated the influence maximization problem as a query processing problem for distinguishing specific users from others. Since influence maximization query processing is NP-hard, and since solving the objective function is also NP-hard, they focused on how to approximate optimal seeds efficiently. Node features are always overlooked when estimating the impact of different users. Deng et al. [72] addressed influence maximization while considering this factor. They presented three quantitative measures to respectively evaluate node features: user activity, user sensitivity, and user affinity. They combined node features into users' static effects, and then used the continuous exponential decay function to convert the strength of a user's dynamic influence between two adjacent users. They also proposed the credit distribution with node features (CD-NF) model, which redefines the credit, and designed the greedy algorithm with node features (GNF) based on the CD-NF model.

Table 1
The time complexity of different algorithms.

Algorithm	Time complexity
Hill-climbing greedy [23]	$O(knRm)$
CELF [61]	$O(knRm)$
NewGreedyIC [41]	$O(kRm)$
NewGreedyWC [41]	$O(kRTm)$
MixGreedyIC [41]	$O(kRm)$
MixGreedyWC [41]	$O(kRTm)$
DegreeDiscountIC [41]	$O(k\log n+m)$
PMIA [63]	$O[nt_{i\theta}+kn_{o\theta}n_{i\theta}(n_{i\theta}+\log n)]$
LDAG [63]	$O(\ell_v+n\log\ell_v)$
SIMPATH [64]	$O(kmn)$
IRIE [65]	$O(kmn)$
Quasilinear time algorithm [66]	$O[k\varepsilon^{-2}(m+n)\log n]$
TIM [67]	$O[(k+\ell)(m+n)\log n/\varepsilon^2]$

n: number of vertices in G; m: number of edges in G; k: number of seeds to be selected; R: number of rounds of simulations; T: number of iterations; $t_{i\theta}$, $n_{o\theta}$, $n_{i\theta}$: constants decided by θ; θ: the influence threshold; $n_{i\theta}=\max_{v\in V}\{|\mathrm{MIIA}(v,\theta)|\}$; $n_{o\theta}=\max_{v\in V}\{|\mathrm{MIOA}(v,\theta)|\}$; $\mathrm{MIIA}(v,\theta)/\mathrm{MIOA}(v,\theta)$: the maximum influence in arborescence/out arborescence of a node v; $t_{i\theta}$: the maximum running time to compute $\mathrm{MIIA}(v,\theta)$; ℓ: number of communities in the network; ε: any constant larger than 0; ℓ_v: the volume of $\mathrm{LDAG}(v,\theta)$.

When considering a certain class or influence maximization, the individual's attributes will have a great impact on the results; therefore, individual differences or the user's attributes need to be taken into account. From the descriptions above of the studies on individual differences, it is clear that most studies are based on the greedy algorithm in order to improve the accuracy. However, heuristic algorithms will be improved by a pursuit of high efficiency and low complexity, because the greedy algorithm is computationally complex. In order to reduce the running time, a considerable amount of work will be carried out on a distributed platform. Due to the individual differences, the computational complexity will be further increased compared with the original. So, for the time being, it is necessary to improve the complexity of the calculation and the accuracy of the results.

3.2. Influence minimization

Assuming that the negative information propagates in a network $G=(V,E)$ with the initially infected node set $S\subseteq V$, the goal of influence minimization is to minimize the number of final infected nodes by blocking k nodes (or vertices) of the set $D\subseteq V$, where k ($\ll|V|$) is a given constant. It can be expressed as the following optimization problem:

$$D^*=\arg\min_{D\subseteq V,|D|\leqslant k}\sigma(S|V\setminus D) \tag{1}$$

where $\sigma(S|V\setminus D)$ denotes the influence of set S when nodes in the set D are blocked.

From this notion [73], we know that influence minimization is a dual problem of influence maximization. Influence minimization mainly used to curb rumors, monitor public opinion, and so on. Yao et al. [73] proposed a method to minimize adverse effects in the network by preventing a limited number of nodes from the perspective of the topic model. When rumors and other adverse events occur in social networks and some users are already infected, the purpose of this model is to minimize the number of final infected users. Wang et al. [74] proposed a dynamic rumor influence minimization model with user experience. This model minimizes the influence of rumors (i.e., the number of users who accept and send rumors) by preventing part of the nodes from activating. Groeber et al. [75] designed a general framework of social influence inspired by the social psychological concept of cognitive dissonance, in which individuals minimize the preconditions for disharmony arising from disagreements with neighbors in a given social network. Chang et al. [76] explored the first solution to estimate the probability of successfully bounding the infected count below the out-of-control threshold; this can be logically mapped to outbreak risk, and can allow authorities to adaptively adjust the intervention cost to meet necessary risk control. They then proposed an influence minimization model to effectively prevent the proliferation of epidemic-prone diseases on the network.

3.3. Flow of influence

When information spreads, it is accompanied by influence flow [77]. In recent years, the flow of influence methods has been studied by many researchers. Subbian et al. [78] proposed a flow

pattern mining approach with the condition of specific flow validity constraints. Kutzkov et al. [79] proposed a streaming method called STRIP to compute the influence strength along each link of a social network. Teng et al. [80] examined real-world information flows in various platforms, including the American Physical Society, Facebook, Twitter, and LiveJournal, and then leveraged the behavioral patterns of users to construct virtual information influence diffusion processes. Chintakunta and Gentimis [81] discussed the relationships between the topological structures of social networks and the information flows within them. However, unlike microblogging platforms, most social networks cannot provide sufficient context to mine the flow pattern.

3.4. Individual influence

Individual influence is a relatively microscopic assessment that models the influence of a user on other users or on the whole social network. Chintakunta and Gentimis [81] proposed a method called SoCap to find influencers in a social network. They modeled influencer finding in a social network as a value-allocation problem, in which the allocated value represents the individual social capital. Subbian et al. [82] proposed an approach to identify influential agents in open multi-agent systems using the matrix factorization method to measure the influence of nodes in a network. Liu et al. [83] presented the trust-oriented social influence (TOSI) method, which considers social contexts (i.e., social relationships and social trust between participants) and preferences in order to assess individual influence. The TOSI method greatly outperforms SoCap in terms of effectiveness, efficiency, and robustness. Deng et al. [84] incorporated the time-critical aspect and the characteristics of the nodes when evaluating the influence of different users. The results showed that their approach is efficient and reasonable for identifying seed nodes, and its prediction of influence spread is more accurate than that of the original method, which disregards node features in the diffusion process.

In conclusion, considering more comprehensive user characteristics and user interaction information results in higher result accuracy.

4. Social influence evaluation

4.1. Influence evaluation metrics

Running time is a very intuitive measure of model efficiency that is easy to calculate. In general, under the same conditions, the faster a model runs, the better it is. However, the traditional greedy algorithm calculates the range of influence spreading for a given node set by a large number of repeated Monte Carlo simulations, resulting in a considerable running time. Especially in the face of current large-scale social networks, the existing algorithms cannot meet the application requirements for efficiency. Therefore, running time is an important measure of social influence evaluation.

Since the influence-spreading problem is NP-hard, it is difficult to obtain an optimal solution of the objective function. Most of the existing algorithms rely on monotonicity and submodularity of the function to achieve $(1-1/e)$ approximation [23]. However, attempts to achieve a higher approximation ratio have never stopped. Zhu et al. [85] proposed semidefinite-based algorithms in their model considering influence transitivity and limiting propagation distance.

Another metric is the number of Monte Carlo calls. Because there is no way to obtain an optimal solution, a Monte Carlo simulation is usually used to estimate the real value. Existing greedy-based algorithms demand heavy Monte Carlo simulations of the spread functions for each node at the initial step, greatly reducing the efficiency of the models. The UBLF algorithm proposed by Zhou et al. [62] can reduce the number of Monte Carlo simulations of CELF method by more than 95%, and achieved a speedup of 2–10 times when the seed set is small.

The expected spread indicates the number of nodes that the seed set can ultimately affect—and the larger the better. There are many applications in real-life scenarios where the influence spread needs to be maximized as much as possible. Typical examples of such applications are marketing and advertising. In both applications, the final expected spread represents the benefits of product promotion or the profitability of product. Therefore, exploring high expected spread algorithms is an important problem for SIA.

Robustness refers to the characteristics of a certain parameter perturbation (i.e., structure and size) that is used to maintain some other performances. Both Jung et al. [65] and Liu et al. [83] mentioned robustness in their algorithms. The IRIE algorithm proposed by Jung et al. [65] is more robust and stable in terms of running time and memory usage across various density networks and cascades of different sizes. The experimental results showed that the IRIE algorithm runs two orders of magnitude faster than existing methods such as PMIA [63] on a large-scale network, and only uses part of the memory. The TOSI evaluation method proposed by Liu et al. [83] shows superior performance in terms of robustness over the state-of-the-art SoCap [81].

Scalability refers to the ability to continuously expand or enhance the functionality of the system with minimal impact on existing systems. In social networks, scalability usually refers to the ability to expand from a small-scale network to a large-scale network. It is a common indicator used to evaluate the quality of a model. Due to the complexity of the algorithm and the long running time, the current solution algorithms only apply to small and medium-sized social networks with nodes below one million. Given today's large-scale social networks, influence analysis algorithms with good scalability must be designed to deal with the challenges posed by massive social network data.

4.2. Social influence evaluation model

The evaluation of social influence is a complex process. As a subjective attribute, a social relationship has many characteristics, including dynamics, event disparity, asymmetry, transitivity, etc. In social networks, frequent user interaction and changes in the structure of the network make the evaluation of social influence more difficult. The literature contains a few studies on the social influence evaluation model. He et al. [86] designed an influence-measuring model on the theme of online complaints; based on the entropy weight model, this model monitors and analyzes the static and dynamic properties of complaint information in real time. Enterprises can use this model to manage online group complaints. Wang et al. [12] proposed a fine-grained feature-based social influence (FBI) evaluation model, which explores the importance of a user and the possibility of a user impacting others. They then designed a PageRank algorithm-based social influence adjustment model by identifying the influence contributions of friends. The FBI evaluation model can identify the social influences of all users with much less duplication (less than 7% with the model), while having a larger influence spread with top-k influential users; it was evaluated on three datasets: HEPTH [87], DBLP [88], and ArnetMiner [89].

5. Conclusions and future work

In this paper, we survey state-of-the-art research on SIA from the aspects of influence models, methods, and evaluation. We also

analyze the strengths and weaknesses of current models and methods. Throughout our study, we unveil future research directions and potential applications.

In social influence models, we distinguish two types of models: microscopic and macroscopic models. Microscopic models consider human interactions and the structure of the influence process. Macroscopic models consider the same transmission probability and identical influential power for all users. In future, macroscopic models should focus on how to consider human behaviors and different mechanisms during information diffusion. Although many researchers have put considerable effort to improving the classical models and proposing new models from different perspectives—such as by adding constraints into models and incorporating competitive influence diffusion—there is still room for improvement.

In most existing models, a person is influenced only by the other person in a monoplex network, and the influence diffusion processes are independent. However, in real life, people often communicate with others in multiplex networks. In this situation, social influence occurs in multiplex networks, and information influence among different monoplex networks encounters cooperation and competition. The question of how to model information influence in multiplex networks is a valuable research topic. In addition, the question of how to compute information influence over time in dynamic networks should be studied. In most experiments, datasets cover up to about 100 000 nodes, so the issues inherent in applying social network analysis-related issues to massive datasets (which may include millions or tens of millions of nodes, or even more) require study. In short, there is still room for research in extending SIA models to address perceived limitations such as efficiency and scalability.

Acknowledgements

This research was supported in part by the National Basic Research Program of China (2013CB329605). The authors thank Lingling Li and Junying Shang for their helpful discussions and comments.

Compliance with ethics guidelines

Kan Li, Lin Zhang, and Heyan Huang declare that they have no conflict of interest or financial conflicts to disclose.

References

[1] Travers J, Milgram S. The small world problem. Psychol Today 1967;1:61–7.
[2] Chen W, Lakshmanan LV, Castillo C. Information and influence propagation in social networks. San Rafael: Morgan & Claypool; 2013.
[3] Freeman LC. A set of measures of centrality based on betweenness. Sociometry 1977;40(1):35–41.
[4] Baas F. A new product growth model for consumer durables. Manage Sci 1969;15(5):215–27.
[5] Brown JJ, Reingen PH. Social ties and word-of-mouth referral behavior. J Consum Res 1987;14(3):350–62.
[6] Mahajan V, Muller E, Bass FM. New product diffusion models in marketing: A review and directions for research. J Mark 1990;54(1):1–26.
[7] Domingos P, Richardson M. Mining the network value of customers. In: Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2001 Aug 26–29; San Francisco, CA, USA; 2001. p. 57–66.
[8] Goldenberg J, Libai B, Muller E. Talk of the network: a complex systems look at the underlying process of word-of-mouth. Mark Lett 2001;12(3):211–23.
[9] Richardson M, Domingos P. Mining knowledge-sharing sites for viral marketing. In: Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2002 Jul 23–26; Edmonton, AB, Canada; 2002. p. 61–70.
[10] Leskovec J, Adamic LA, Huberman BA. The dynamics of viral marketing. J ACM Trans Web 2007;1(1):5.
[11] Pálovics R, Benczúr AA, Kocsis L, Kiss T, Frigó E. Exploiting temporal influence in online recommendation. In: Proceedings of the 8th ACM Conference on Recommender Systems; 2014 Oct 6–10; Foster City, CA, USA; 2014. p. 273–80.
[12] Wang G, Jiang W, Wu J, Xiong Z. Fine-grained feature-based social influence evaluation in online social networks. IEEE Trans Parallel Distrib Syst 2014;25(9):2286–96.
[13] Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med 2007;357(4):370–9.
[14] Fowler JH, Christakis NA. Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study. Br Med J 2009;338(7685):23–7.
[15] Franks H, Griffiths N, Anand SS. Learning influence in complex social networks. In: Proceedings of the 2013 International Conference on Autonomous Agents and Multi-agent Systems; 2013 May 6–10; Saint Paul, MN, USA; 2013. p. 447–54.
[16] Dong W, Pentland A. Modeling influence between experts. In: Proceedings of the ICMI 2006 and IJCAI 2007 International Conference on Artificial Intelligence for Human Computing; 2006 Nov 3; Banff, AB, Canada; 2007. p. 170–89.
[17] Tang J, Sun J, Wang C, Yang Z. Social influence analysis in large-scale networks. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2009 Jun 28–Jul 1; Paris, France; 2009. p. 807–16.
[18] He Z, Cai Z, Wang X. Modeling propagation dynamics and developing optimized countermeasures for rumor spreading in online social networks. In: Proceedings of the 2005 IEEE 35th International Conference on Distributed Computing Systems; 2015 Jun 29–Jul 2; Columbus, OH, USA; 2015. p. 205–14.
[19] Katz E, Lazarsfeld PF. Personal influence: The part played by people in the flow of mass communications. New York: Free Press; 1965.
[20] Rogers EM. Diffusion of innovations. 5th ed. New York: Free Press; 2003.
[21] Keller E, Berry J. The influentials: one American in ten tells the other nine how to vote, where to eat, and what to buy. New York: Free Press; 2003.
[22] Peng S, Yang A, Cao L, Yu S, Xie D. Social influence modeling using information theory in mobile social networks. Inf Sci 2017;379:146–59.
[23] Kempe D, Kleinberg J, Tardos É. Maximizing the spread of influence through a social network. In: Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2003 Aug 24–27; Washington, DC, USA; 2003. p. 137–46.
[24] Leskovec J, Mcglohon M, Faloutsos C, Glance NS, Hurst M. Patterns of cascading behavior in large blog graphs. In: Proceedings of the 2007 SIAM International Conference on Data Mining; 2007 Apr 26–28; Minneapolis, MN, USA; 2007.
[25] Gruhl D, Guha R, Liben-Nowell D, Tomkins A. Information diffusion through blogspace. In: Proceedings of the 13th International Conference on World Wide Web; 2004 May 17–20; New York, NY, USA; 2004. p. 491–501.
[26] Granovetter M. Threshold models of collective behavior. Am J Sociol 1978;83(6):1420–43.
[27] Chen W, Lu W, Zhang N. Time-critical influence maximization in social networks with time-delayed diffusion process. In: Proceedings of the 26th AAAI Conference on Artificial Intelligence; 2012 Jul 22–26; Toronto, ON, Canada; 2012. p. 592–8.
[28] Feng S, Chen X, Cong G, Zeng Y, Chee YM, Xiang Y. Influence maximization with novelty decay in social networks. In: Proceedings of the 28th AAAI Conference on Artificial Intelligence; 2014 Jul 27–31; Québec City, QC, Canada; 2014. p. 37–43.
[29] Mohamadi-Baghmolaei R, Mozafari N, Hamzeh A. Trust based latency aware influence maximization in social networks. J Eng App Artif Intell 2015;41(C):195–206.
[30] Budak C, Agrawal D, Abbadi AE. Limiting the spread of misinformation in social networks. In: Proceedings of the 20th International Conference on World Wide Web; 2011 Mar 28–Apr 1; Hyderabad, India; 2011. p. 665–74.
[31] Liu W, Yue K, Wu H, Li J, Liu D, Tang D. Containment of competitive influence spread in social networks. Knowl Base Syst 2016;109(C):266–75.
[32] Borodin A, Filmus Y, Oren J. Threshold models for competitive influence in social networks. In: Proceedings of the 6th International Conference on Internet and Network Economics; 2010 Dec 13–17; Stanford, CA, USA; 2010. p. 539–50.
[33] Mohammadi A, Saraee M, Mirzaei A. Time-sensitive influence maximization in social networks. J Inf Sci 2015;41(6):765–78.
[34] Saito K, Ohara K, Yamagishi Y, Kimura M, Motoda H. Learning diffusion probability based on node attributes in social networks. In: Proceedings of the 19th International Conference on Foundations of Intelligent Systems; 2011 Jun 28–30; Warsaw, Poland; 2011. p. 153–62.
[35] Guille A, Hacid H. A predictive model for the temporal dynamics of information diffusion in online social networks. In: Proceedings of the 21st International Conference on World Wide Web; 2012 Apr 16–20; Lyon, France; 2012. p. 1145–52.
[36] Bharathi S, Kempe D, Salek M. Competitive influence maximization in social networks. In: Proceedings of the 3rd International Conference on Internet and Network Economics; 2007 Dec 12–14; San Diego, CA, USA; 2007. p. 306–11.
[37] Carnes T, Nagarajan C, Wild SM, Zuylen AV. Maximizing influence in a competitive social network: A follower's perspective. In: Proceedings of the 9th International Conference on Electronic Commerce; 2007 Aug 19–22; Minneapolis, MN, USA; 2007. p. 351–60.
[38] Fan L, Lu Z, Wu W, Thuraisingham B, Ma H, Bi Y. Least cost rumor blocking in social networks. In: Proceedings of the 2013 IEEE 33rd International

Conference on Distributed Computing Systems; 2013 Jul 8–11; Philadelphia, PA, USA; 2013. p. 540–9.
[39] Lee W, Kim J, Yu H. CT-IC: Continuously activated and time-restricted independent cascade model for viral marketing. In: Proceedings of the 2012 IEEE 12th International Conference on Data Mining; 2012 Dec 10–13; Brussels, Belgium; 2013. p. 960–5.
[40] Kostka J, Oswald YA, Wattenhofer R. Word of mouth: Rumor dissemination in social networks. In: Proceedings of the 15th International Colloquium on Structural Information and Communication Complexity; 2008 Jun 17–20; Villars-sur-Ollon, Switzerland; 2008. p. 185–96.
[41] Chen W, Wang Y, Yang S. Efficient influence maximization in social networks. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2009 Jun 28–Jul 1; Paris, France; 2009. p. 199–208.
[42] Wang Y, Wang H, Li J, Gao H. Efficient influence maximization in weighted independent cascade model. In: Proceedings of the 21st International Conference on Database Systems for Advanced Applications; 2016 Mar 27–30; Dallas, TX, USA; 2016. p. 49–64.
[43] Pathak N, Banerjee A, Srivastava J. A generalized linear threshold model for multiple cascades. In: Proceedings of the 2010 IEEE International Conference on Data Mining; 2010 Dec 13–17; Sydney, Australia; 2010. p. 965–70.
[44] Bharathi S, Kempe D, Salek M. Competitive influence maximization in social networks. In: Proceedings of the 3rd International Workshop on Web and Internet Economics; 2007 Dec 12–14; San Diego, CA, USA; 2007. p. 306–11.
[45] Galam S. Modelling rumors: the no plane Pentagon French hoax case. Phys A 2003;320:571–80.
[46] Lin SC, Lin SD, Chen MS. A learning-based framework to handle multi-round multi-party influence maximization on social networks. In: Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2015 Aug 10–13; Sydney, Australia; 2015. p. 695–704.
[47] Golnari G, Asiaee A, Banerjee A, Zhang ZL. Revisiting non-progressive influence models: Scalable influence maximization. In: Proceedings of the 31st Conference on Uncertainty in Artificial Intelligence; 2015 Jul 12–16; Amsterdam, The Netherlands; 2015.
[48] Wang X, Jia J, Tang J, Wu B, Cai L, Xie L. Modeling emotion influence in image social networks. IEEE Trans Affect Comp 2015;6(3):286–97.
[49] Gao D. Opinion influence and diffusion in social network. In: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information; 2012 Aug 12–16; Portland, OR, USA; 2012. p. 997.
[50] Daley DJ, Kendall DG. Epidemics and rumors. Nature 1964;204(4963):1118.
[51] Moreno Y, Nekovee M, Pacheco AF. Dynamics of rumor spreading in complex networks. Phys Rev E Stat Nonlin Soft Matter Phys 2004;69(6 Pt 2):066130.
[52] Nekovee M, Moreno Y, Bianconi G, Marsili M. Theory of rumor spreading in complex social networks. Phys A 2007;374(1):457–70.
[53] Zhou J, Liu Z, Li B. Influence of network structure on rumor propagation. Phys Lett A 2007;368(6):458–63.
[54] Wang H, Deng L, Xie F, Xu H, Han J. A new rumor propagation model on SNS structure. In: Proceedings of the 2012 IEEE International Conference on Granular Computing; 2012 Aug 11–13; Hangzhou, China; 2012. p. 499–503.
[55] Wang Y, Yang X, Han Y, Wang X. Rumor spreading model with trust mechanism in complex social networks. Commum Theor Phys 2013;59(4):510–6.
[56] Xia L, Jiang G, Song B, Song Y. Rumor spreading model considering hesitating mechanism in complex social networks. Phys A 2015;437:295–303.
[57] Su Q, Huang J, Zhao X. An information propagation model considering incomplete reading behavior in microblog. Phys A 2015;419:55–63.
[58] Liu Q, Li T, Sun M. The analysis of an SEIR rumor propagation model on heterogeneous network. Phys A 2017;469:372–80.
[59] Zhao L, Wang J, Chen Y, Wang Q, Cheng J, Cui H. SIHR rumor spreading model in social networks. Phys A 2012;391(7):2444–53.
[60] Zhao L, Qiu X, Wang X, Wang J. Rumor spreading model considering forgetting and remembering mechanisms in inhomogeneous networks. Phys A 2013;392(4):987–94.
[61] Leskovec J, Krause A, Guestrin C, Faloutsos C, VanBriesen J, Glance N. Cost-effective outbreak detection in networks. In: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2007 Aug 12–15; San Jose, CA, USA; 2007. p. 420–9.
[62] Zhou C, Zhang P, Zang W, Guo L. On the upper bounds of spread for greedy algorithms in social network influence maximization. IEEE Trans Knowl Data Eng 2015;27(10):2770–83.
[63] Chen W, Wang C, Wang Y. 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; 2010 Jul 25–28, Washington, DC, USA; 2010. p. 1029–38.
[64] Goyal A, Lu W, Lakshmanan LVS. SIMPATH: An efficient algorithm for influence maximization under the linear threshold model. In: Proceedings of the 2011 IEEE 11th International Conference on Data Mining; 2011 Dec 11–14; Vancouver, BC, Canada; 2012. p. 211–20.
[65] Jung K, Heo W, Chen W. IRIE: a scalable influence maximization algorithm for independent cascade model and its extensions. Rev Crim 2011;56(10):1451–5.
[66] Borgs C, Brautbar M, Chayes J, Lucier B. Maximizing social influence in nearly optimal time. In: Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms; 2014 Jan 5–7; Portland, OR, USA; 2014. p. 946–57.
[67] Tang Y, Xiao X, Shi Y. Influence maximization: Near-optimal time complexity meets practical efficiency. In: Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data; 2014 June 22–27; Snowbird, UT, USA; 2014. p. 75–86.
[68] Li CT, Lin SD, Shan MK. Influence propagation and maximization for heterogeneous social networks. In: Proceedings of the 21st International Conference on World Wide Web; 2012 Apr 16–20; Lyon, France; 2012. p. 559–60.
[69] Subbian K, Sharma D, Wen Z, Srivastava J. Social capital: The power of influencers in networks. In: Proceedings of the 2013 International Conference on Autonomous Agents and Multi-agent Systems; 2013 May 6–10; Saint Paul, MN, USA; 2013. p. 1243–4.
[70] Li H, Bhowmick SS, Sun A. CINEMA: Conformity-aware greedy algorithm for influence maximization in online social networks. In: Proceedings of the 16th International Conference on Extending Database Technology; 2013 Mar 18–22; Genoa, Italy; 2013. p. 323–34.
[71] Lee JR, Chung CW. A query approach for influence maximization on specific users in social networks. IEEE Trans Knowl Data Eng 2015;27(2):340–53.
[72] Deng X, Pan Y, Shen H, Gui J. Credit distribution for influence maximization in online social networks with node features. J Intell Fuzzy Syst 2016;31(2):979–90.
[73] Yao Q, Zhou C, Shi R, Wang P, Guo L. Topic-aware social influence minimization. In: Proceedings of the 24th International Conference on World Wide Web; 2015 May 18–22; Florence, Italy; 2015. p. 139–40.
[74] Wang B, Chen G, Fu L, Song L, Wang X. DRIMUX: dynamic rumor influence minimization with user experience in social networks. IEEE Trans Knowl Data Eng 2017;29(10):2168–81.
[75] Groeber P, Lorenz J, Schweitzer F. Dissonance minimization as a microfoundation of social influence in models of opinion formation. J Math Sociol 2014;38(3):147–74.
[76] Chang CW, Yeh MY, Chuang KT. On the guarantee of containment probability in influence minimization. In: Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining; 2016 Aug 18–21; San Francisco, CA, USA; 2016. p. 231–8.
[77] Faisan MM, Bhavani SD. Maximizing information or influence spread using flow authority model in social networks. In: Proceedings of the 10th International Conference on Distributed Computing and Internet Technology; 2014 Feb 6–9; Bhubaneswar, India; 2014. p. 233–8.
[78] Subbian K, Aggarwal C, Srivastava J. Content-centric flow mining for influence analysis in social streams. In: Proceedings of the 22nd ACM International Conference on Information & Knowledge Management; 2013 Oct 27–Nov 1; San Francisco, CA, USA; 2013. p. 841–6.
[79] Kutzkov K, Bifet A, Bonchi F, Gionis A. STRIP: Stream learning of influence probabilities. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2013 Aug 11–14; Chicago, IL, USA; 2013. p. 275–83.
[80] Teng X, Pei S, Morone F, Makse HA. Collective influence of multiple spreaders evaluated by tracing real information flow in large-scale social networks. Sci Rep 2016;6(1):36043.
[81] Chintakunta H, Gentimis A. Influence of topology in information flow in social networks. In: Proceedings of the 2016 Asilomar Conference on Signals, Systems and Computers; 2016 Nov 6–9; Pacific Grove, CA, USA; 2017. p. 67–71.
[82] Subbian K, Sharma D, Wen Z, Srivastava J. Finding influencers in networks using social capital. In: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining; 2013 Aug 25–28; Niagara Falls, ON, Canada; 2013. p. 592–9.
[83] Liu G, Zhu F, Zheng K, Liu A, Li Z, Zhao L, et al. TOSI: a trust-oriented social influence evaluation method in contextual social networks. Neurocomputing 2016;210:130–40.
[84] Deng X, Pan Y, Wu Y, Gui J. Credit distribution and influence maximization in online social networks using node features. In: Proceedings of the 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery; 2015 Aug 15–17; Zhangjiajie, China; 2016. p. 2093–100.
[85] Zhu Y, Wu W, Bi Y, Wu L, Jiang Y, Xu W. Better approximation algorithms for influence maximization in online social networks. J Comb Optim 2015;30(1):97–108.
[86] He J, Hu M, Shi M, Liu Y. Research on the measure method of complaint theme influence on online social network. Expert Syst Appl 2014;41(13):6039–46.
[87] Gehrke J, Ginsparg P, Kleinberg J. Overview of the 2003 KDD cup. ACM SIGKDD Explor Newslett 2003;5(2):149–51.
[88] Yang J, Leskovec J. Defining and evaluating network communities based on ground-truth. Knowl Inf Syst 2015;42(1):181–213.
[89] Tang J, Zhang J, Yao L, Li J, Zhang L, Su Z. ArnetMiner: Extraction and mining of academic social networks. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2008 Aug 24–27; Las Vegas, NV, USA; 2008. p. 990–8.