International Journal of Modern Physics C
Vol. 32, No. 6 (2021) 2150079 (18 pages)

DOI: 10.1142/S0129183121500790

A sequential seed scheduling heuristic based on determinate and latent margin for influence maximization problem with limited budget

Jianxin Tang*,‡, Fuqing Zhao*, Ruisheng Zhang†, Baoqiang Chai†
and Shilu Di*

*School of Computer and Communication
Lanzhou University of Technology
Lanzhou, Gansu 730050, P. R. China

†School of Information Science and Engineering
Lanzhou University
Lanzhou, Gansu 730000, P. R. China
‡tangjx16@lzu.edu.cn

Received 10 September 2020
Accepted 26 January 2021
Published 3 March 2021

The influence maximization problem in social networks aims to select a subset of most influential nodes, denoted as seed set, to maximize the influence diffusion of the seed nodes. The majority of existing works on this problem would ignite all the seed nodes simultaneously at the beginning of the diffusion process and let the influence diffuses passively in the network. However, it cannot depict the practical dynamics exactly of viral marketing campaigns in reality and fails to provide driving policies to control over the diffusion. In this paper, we focus on the dynamic influence maximization problem with limited budget to study the scheduling strategies including which influential node is to be seeded during the diffusion process and when to seed it at the right time. A time-dependent seed activating feedback scheme is modeled firstly by considering the time factor and its impact on the influence obligation in diffusion process. Then a scheduling heuristic based on determinate and latent margin is proposed to evaluate the marginal return of candidate nodes and activate the right seed node to promote the viral marketing. Extensive experiments on four social networks show that the proposed algorithm achieves significantly better results than a typical static influence maximization algorithm based on swarm intelligence and can improve the influence propagation under the time-dependent diffusion model comparing with the centrality-based scheduling heuristics.

Keywords: Social networks; scheduling influence maximization; seed activating strategy; latent effect; viral marketing.

PACS Nos.: 87.85.Xd, 89.20Ff.

‡Corresponding author.

1. Introduction

With the rapid development of Web2.0, social networks including Facebook, Twitter, Google+, WeChat, etc., have found their irreplaceable applications in information diffusion, viral marketing, epidemic inhibition and crowd-sourcing.[1–3] The underlying actuator of these fundamental applications is the social influence, which can be utilized to reshape one's emotions, opinions and even the behaviors. As one of the interesting research topics of social network analysis, the influence maximization problem, which was first modeled by Domingos and Richardson[4] and then elegantly formulated as a discrete optimization problem by Kempe and Kleinberg,[5] aims to select a subset of k influential nodes, denoted as seed set, such that the number of activated nodes triggered by the seed nodes through an influence diffusion model is maximal.

The majority of existing works on influence maximization generally select the targeted number of influential seed nodes and ignite them "simultaneously" at the beginning of the influence diffusion process, during which the influence is diffused passively in the network.[6,7] A toy example is the viral marketing, the marketing sponsors tend to expose all the discount or free products to a set of most influential consumers to profit maximal marginal return by leveraging the effect of "word of mouth". However, the idealized influence diffusion models cannot well depict the sophisticated influence diffusion scenarios in reality for the significant role of marketing sponsors, of which the goal is to control over the diffusion process optimally of the marketing campaigns, is always overlooked. Therefore, the results achieved by conventional influence maximization algorithms usually turn out to be unsatisfactory, especially on large-scale heterogeneous social networks, where consumers are always exposed to the social phenomenon of information overload. To make full use of the limited budget resource and take control of the vital marketing process, marketers have to proactively exert external operations to prolong the product promoting campaign to profit from the consumers in practical marketing scenarios.

An alternative way is to estimate and activate the seed nodes well-timed according to the dynamic marginal return of the candidate nodes during the diffusion process. This advanced influence diffusion was firstly formulated as scheduling influence maximization problem by Chierichetti *et al.*[8] and Lin *et al.*[9] pointed out that social network providers play an important role in deciding when the information should be brought to the attention of individual users especially when the user attention is limited. Samadi *et al.*[10] addressed that the duration of activation and reactivation of seeds are important in many real-world viral marketing situations, and therefore the decision-makers planning for long campaigns need not only initiate the seed spreaders but also control the information process over time to achieve maximal impact at special time point. To better understand the dynamics of the influence diffusion in practical scenarios and study the strategies for evaluating and scheduling seed nodes with constraint of limited budget, this paper firstly analyzes the traits of two typical time-independent influence diffusion models, independent cascade model and linear threshold model, under the limited user attention paradigm and herd behavior effect, and a time-dependent influence diffusion model is then presented to sketch the dynamics in real-world scenarios. Finally, a seed scheduling influence maximization algorithm comes with a seed activating strategy is proposed, and its performance is validated on four social networks.

The contribution of this work can be summarized along the following two axes:

- A time-dependent seed activating feedback (TSAF) model is presented to depict the dynamics of practical information spreading process with limited budget resource constraint. In the model, the latent effect of influence in the IC model and the accumulation characteristic of influence in the LT model are both inherited.
- A seed scheduling heuristic based on determinate and latent margin (SDLM) is proposed to solve the scheduling influence maximization problem under the time-dependent model. To improve the influence diffusion, a determinate and latent margin (DLM) scheme is conceived to estimate the expected influence margin of a node and seed the influential node with the most DLM score in the right time stamp. Extensive experiments are conducted on four social networks to validate the performance of the proposed algorithm.

The organization of the reminder of this paper is structured as follows. An overview of related work on static influence maximization problem and the efforts on scheduling influence maximization problem are given in Sec. 2. Section 3 formulates a time-dependent influence diffusion model and proposes a seed scheduling influence maximization heuristic. The performance of the proposed scheduling algorithm is validated by conducting experiments on four social networks in Sec. 4. Section 5 concludes this paper with future work.

2. Related Work

The existing studies on influence maximization problem can be categorized into two typical classes including static influence maximization and dynamic scheduling influence maximization according to the strategies for activating the seed nodes.

2.1. *Works on static influence maximization problem*

According to the conventional framework for influence maximization formulated by Domingos and Richardson,[4] once the k most influential nodes are activated, their influence would diffuses passively during the whole process and ends when there are no more new activated nodes. To obtain the seed nodes, Kempe *et al.*[5] transformed the viral marketing into a mathematical optimization problem by idealizing the information diffusion dynamics in social network, and proposed a greedy algorithm which features a theoretical bound of $(1 - \frac{1}{e} - \epsilon)$ on the optimal solution. However, the simple heuristic turns out to be computationally challenging on large-scale networks. Subsequent researches following the seminal work mainly focus on developing effective and efficient algorithms for the estimation and identification of influential nodes.[11–13] Goyal *et al.*[14] presented an effective way to promote the simple greedy algorithm to be scalable to large-scale networks by reordering the candidate nodes according to the marginal return of each node, which can dramatically reduce the number of Monte–Carlo (MC) simulation times. To avoid the MC simulation mechanism for estimating the marginal return of candidate nodes, Tang *et al.*[15] treated each node in the network topology as a virtual bat individual and proposed a discrete bat algorithm by redefining the evolutionary rules for the bat population to optimize the candidate seed set. Experiments show that it is a promising way to solve influence maximization problem in large-scale networks based on swarm intelligence. To identify the most influential nodes efficiently in large-scale networks, Shang *et al.*[16] suggested to partition the whole network into independent communities by assuming that the information diffusion between communities is sparse.

In addition, variant influence maximization problem has been explored in many works. Mohamadi *et al.*[17] emphasized that trust aspect of the diffusion process plays a major role in reshaping the dynamics of the networks, and a trust-based latency aware influence maximization model was proposed by considering time and trust aspects simultaneously. Morone and Makse[18] indicated that specific structural nodes, which hinge on the frame of social network and whose interconnections are usually sparse, play a significant role in the spreading of information and proposed to identify a minimal set of influential nodes to maximize the influence diffusion. Li *et al.*[19] took the location information of the viral marketing promotion into consideration and assumed that each user in the social network has different preferences on different locations, a location-aware influence maximum arborescence model was conceived to estimate the marginal return of candidate nodes. Lü *et al.*[20] gave an informative overview of the related works on the identification of vital nodes.

However, empirical results of influence maximization in practical scenarios show that the performance of existing methods tends to be unsatisfactory due to the role played by the seed activation scheduling in enhancing the marginal return of influence diffusion is overlooked in the studies on the static influence maximization problem.[21] Therefore, the interesting problems that how to depict the diffusion cascading behaviors of influential nodes' influence in the practical marketing scenarios in a more accurately way and what external operations can be exerted to unfold the influence propagation actively in the social networks and control over the influence diffusion processes dynamically need to be further explored and addressed.

2.2. *Dynamic scheduling influence maximization problem*

Activating the seed nodes in multiple stages is an alternative solution to control over the diffusion of node influence in the social network. Jankowski *et al.*[22] analyzed how the ratio of seeds used in the initialization of the diffusion process affects the performance in terms of number of activated nodes and its duration. The experimental results showed that minimizing the number of seeds in the first beginning stage would increases the duration of the diffusion process. Ni[23] pointed out that marketers have to make sequential decisions in reality to optimize the influence spread as the diffusion situation evolves and varies. Under the constraint condition of minimizing the complete influence diffusion time, the author formulated the sequential seeding problem as a Markov decision process.

Sela *et al.*[24] are among the first to investigate the problem of finding not only which nodes to be seeded but also when to seed them when taking into account the timing aspect under the constraint of limited budget. Furthermore, for the situation that people from social networks can only accept limited amount of information due to the limited strength of mind, which has been discovered as an intrinsic physiological property of human by social science, the authors addressed that the influence diffusion characteristics of the activated nodes in the independent cascade model or the linear threshold model cannot well depict the diffusion dynamics in practical viral marketing scenarios. The targeted k influential nodes were identified and scheduled sequentially according to the evolutionary status of the influence diffusion, and the experimental results proved that the proposed scheduling approach by taking into account the timing aspect can improve the rates of the influence spread by over 23% compared to conventional seeding methods. Furthermore, Sela *et al.*[21] believed that the adoption of promotional products in real-world commercial activities relies on continuous active promotion efforts by the marketer, and presented a novel active viral marketing model to specify the dynamics of the influence diffusion over the social network. By analyzing the stochastic dynamics, diminishing social effect and the state-dependent seeding properties of the commonly used contagion models experimentally, Goldenberg *et al.*[25] validated the effectiveness of scheduled seeding approach, which selects influential nodes in right time stamp, over the conventional static influence maximization approaches. Samadi *et al.*[26] studied the strategies to identify the seed nodes based on a blogger-centric problem on a two-level network over a given time horizon, and proposed a partial parallel cascade model to estimate the timing of seed activation by paying the prominent bloggers under a given budget constraint. Tang and Yuan[27] studied the adaptive influence maximization problem to balance the delay and performance tradeoff, and proposed a novel adaptive policy with bounded approximation ratio by adjusting a controlling parameter.

Extensive experimental results of the previous literatures have proved the effectiveness and necessity of scheduling seeds activation by exerting external operations on the influence diffusion process. However, the presented dynamic activating models and approaches to seeding influential nodes need to be further explored to better depict the dynamics of the influence diffusion process in reality. In this paper, we focus on the advanced scheduling influence maximization problem by integrating the influence decaying factor with the timing aspect and limited budget constraint into a time-dependent influence diffusion model, and a scheduling influence maximization algorithm by utilizing the activation knowledge of the diffusion process is developed to provide the marketer an ability to gain a higher activation rate as the process unfolds.

3. Problem Formulation and the Proposed Method

3.1. *Time-dependent seed activating feedback model*

Limited user attention paradigm. The majority of existing influence maximization algorithms simulate the information diffusion behaviors mainly based on the independent cascade (IC) model and linear threshold (LT) model, in which there are two states for each node, *active* and *inactive* and *inactive* nodes can be activated by *active* nodes and switched into *active* nodes, but not vice versa. In the IC model, an *active* node activated in time t has only one opportunity to influence its *inactive* neighbors with a successful probability p in the next time step $t+1$. In the LT model, an *inactive* node u turns to be *active* when the total influence accumulated from its *active* neighbors exceeds its critical activation threshold θ_u, and node u remains *active* till the diffusion process ends. The above two models are both time-independent and assume that node's influence never diminishes or decays with time during the diffusion process. Whereas, nodes from social networks in real-world are usually exposed to overloaded information, such as various promotional products and other marketing campaigns, which is far beyond one's processing capabilities. So, it is very important and necessary to determine the right time to seed the influential nodes in the overloaded information circumstances, especially when the node's attention is limited, a well-documented psychological and cognitive concept from social science, of which the role in affecting people's behaviors and their interactions in social media has been confirmed by recent researches.[28,29] Under the limited user attention paradigm, nodes can only pay transitory attention and processing capabilities to the received information.[28] The timing aspect in the limited attention paradigm has emerged as an essential and necessary factor in modeling the time-dependent behaviors of information diffusion.

Herd effect. As addressed above, each node's attention in the social network is limited, so an effective way to capture one's attention is through the creation of social hypes, which was termed as herd behavior tendency by Sela *et al.*[24] In the herd effect, the herd messages received from adjacent time periods would potentially ignite a cascade diffusion process by leveraging the effect of "word-of-mouth" in viral marketing.

Latent marginal return. As illustrated in Refs. 30 and 31, social proximity is believed to be an effective model to predict the tendencies of nodes in social networks. According to the mechanism formulated in the LT model, the influence on one inactive node from its surround active neighbors will be accumulated for a certain period until the current node is activated, and during the period, the node is more likely to purchase the recommended product with the number of its active neighbors increases, and the other is the opposite. Therefore, seeding the nodes with more latent marginal return in the right time is a more realistic setting.

Limited budget constraint. In general, the limited adverting budget that the marketing sponsors could afford is a crucial factor that should be considered in identifying influential nodes, as well as in selecting an influential individual as a seed node to promote the diffusion of information in the viral marketing campaign. And the cooperation costs are always in proportion to the number of friendships of the candidate node. Therefore, a time-dependent influence diffusion model with limited budget constraint can better simulate the diffusion process of information.

Based on the above statements, a novel TSAF model is conceived in this paper. In the TSAF model, each node in the topology has one of the following three states, including *Noninfected*, *Infected* & *Infectious* and *Infected* & *Noninfectious* at time t. *Noninfected* node can be selected as a seed node and its state will be turned into *Infected* & *Infectious* state, in which the node has chances to infect its noninfected neighbors in the following oblivion O time period. With time elapsing, more and more neighbor nodes around a *Noninfected* node u are infected or seeded, and the *Noninfected* node will be activated once the number of its neighbor nodes which are infectious is up to the predefined infection threshold C, such that $\sum_{t=\Delta t}^{\Delta t+O} I_{v,u} \geq \theta_u$, where node v is an *Infected* & *Infectious* node from the neighbor set Γ_u of u and $I_{v,u}$ is defined as the influence on node u from node v. The influence diffusion process ends when there are no more nodes could be infected after the limited budget B is worn out.

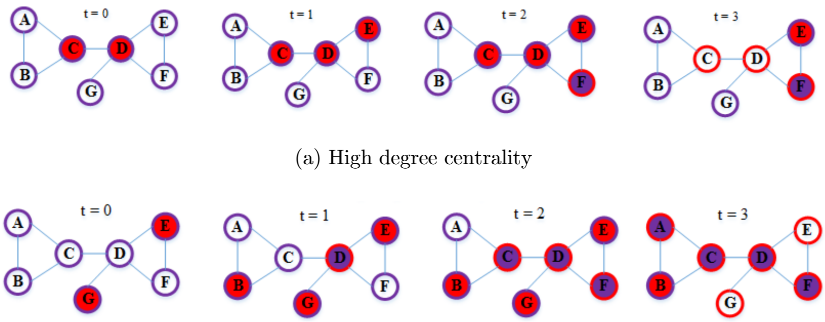
Based on the TSAF model, we define the scheduling influence maximization problem as follows.

Definition 1. Given a social network $G = (V,\ E)$, where V is the set of nodes and E is the set of edges, the time-dependent seed activating feedback model and limited budget B, scheduling influence maximization problem is targeted to seed a subset of influential nodes into the seed set S at the right timing during the diffusion process, where seeding an influential node needs one idealized unit budget, such that the influence diffusion is maximum when the diffusion process ends.

3.2. *A scheduling influence maximization algorithm*

As addressed above, the presented TSAF model inherits the traits of the IC model and the LT model. However, unlike the IC model, in which an infected node has only one chance to activate its uninfected neighbors, in the TSAF model, a node in the state of *Infected* & *Infectious* is permitted to diffuse its influence within a predefined oblivion O time period. Meanwhile, similarly to the "accumulation effect" of LT model, *Noninfected* nodes in the TSAF model will be infected once the number of its neighbors in the state of *Infected* & *Infectious* is up to the predefined activation threshold C, which is enlightened by the herd effect. Consequently, the required conditions for seeding a node is that a proportional of its neighbors have been in the state of *Infected* & *Infectious*.

A toy example illustrating the diffusion processes of node influence based on different seeding strategies is shown in Fig. 1. There are seven nodes tied by eight edges in the artificial social network, and the related parameters B, O and C are set to 3, 2 and 2, respectively. In Fig. 1(a), the high degree centrality-based heuristic is



(a) High degree centrality

(b) A laten-based scheduling heuristic

Fig. 1. (Color online) Comparisons on diffusion processes under different seed nodes activating strategies.

adopted to select influential nodes with the highest degree value into the seed set S to promote the influence propagation. Along the description of the TSAF model, two nodes C and D are seeded initially to ignite the diffusion process, however, additional node(s) would be infected only when the third alternative node E is seeded at time $t = 2$, then the state of activated nodes C and D will become *Infected* & *Noninfectious* at time $t = 3$ and the diffusion process stops because the limited budget is exhaust. However, the performance of the influence diffusion is brilliant when an external effect is exerted on the dynamic process. As illustrated in Fig. 1(b), if we initially select nodes E and G as the seed at time $t = 0$, node D will be activated for there are two infectious nodes within its direct neighbors at time $t = 1$. Meanwhile, if we seed B as an influencer at time $t = 1$, then nodes C and D will be activated separately at time $t = 2$. At time $t = 3$, node A will be infected by its two infectious neighbor nodes B and C, then the diffusion process stops for there are no uninfected nodes in the network. Therefore, it is of great significance to make further exploration on the scheduling influence maximization problem so as to better understand the diffusion dynamics behaviors in the evolution and control of social networks.

Comparing with the conventional influence maximization algorithms that activate the targeted seed nodes simultaneously at the beginning of the diffusion process and let the influence diffuses passively in the networks, scheduling influence maximization algorithms need external effort to find not only the best nodes to be seeded but also the right timing to perform these seedings effectively to maximize the diffusion scale of the influence.

For simplicity, we denote that

- The state of node v at time t as $\zeta^t(v) = \begin{cases} 0 & \text{Noninfected,} \\ 1 & \text{Infected \& Infectious,} \\ 2 & \text{Infected \& Noninfectious.} \end{cases}$
- $U_X^t = \{u | u \in V \wedge \zeta^t(v) = X\}$ as the set of nodes which are in state X at time t.

- $\Gamma(v) = \{u|(u,v) \in E\}$ as the neighbors of node v.
- $\Gamma^t_X(v) = \Gamma(v) \cap V^t_X$ as the neighbors of node v which are in state X at time t.

To seed the nodes at the right time under the TSAF model, an estimator naming Determinate-Latent Margin (DLM for short) is conceived to evaluate and identify seed nodes. More specifically, the marginal return can be evaluated by the DLM estimator according to the following two parts: ① Determinate marginal return.

According to the diffusion dynamics of the TSAF model, for a *Noninfected* node u, when the number of neighbors, of which the state is being $X = 1$ at time t, is up to the critical activation threshold C, node u will be activated spontaneously. If the number of infectious neighbors equals $C - 1$, then selecting an infected node v that is one of the neighbors of node u can bring a determinate margin including itself, denoted as in Eq. (1).

$$\mathrm{DM}(u) = 1 + \sum_{v \in N^{(1)}(U^t_{X=1})} (|(u,v)||(u,v) \in E). \tag{1}$$

② Latent marginal return. Besides the determinate margin, seeding node u can bring a latent influence diffusion margin, as addressed in Sec. 3.1. A direct way to estimate the latent margin is the number of *Noninfected* nodes that are adjacent to the infectious nodes at time t. Meanwhile, taking account of the influence decay of the infectious nodes, a coefficient is adopted to prune the latent margin. So the latent margin of seeding node u can be calculated by Eq. (2), where parameter λ is a tunable exponent factor with a constant value $\lambda = O$ in this paper

$$\mathrm{LM}(u) = \frac{1}{C^\lambda} \sum_{w \in N^{(1)}(u)/\{v|(u,v)\in E, v \in N^{(1)}(U^t_{X=1})\}} (|(u,w)||(u,w) \in E). \tag{2}$$

Algorithm 1. A DLM-based seed scheduling influence maximization algorithm

Input: Social network $G = (V, E)$, budget B, threshold C, oblivion time O.
Initialize: $t \leftarrow 0$, $S \leftarrow \Phi$

select and ignite an initial seed node s_0 based on DC centrality
$b \leftarrow 1$
while $b <= B$ **do**
  $t \leftarrow t + 1$
  update the state of u, where $|\Gamma^t_{X=1}(u)| \geq C$
  $s_{b+1} \leftarrow \arg\max(DLM(v))$, where $\Gamma^t_{X=1}(u) = C - 1$
  $S \leftarrow S \cup s_{b+1}$
  $b \leftarrow b + 1$
  update the state of s_{b+1} according to the O^t_i
end while
return The seed set S to be scheduled.

Therefore, the mathematical returned margin by the DLM estimator can be evaluated according to Eq. (3).

$$\mathrm{DLM}(u) = \mathrm{DM}(u) + \mathrm{LM}(u). \tag{3}$$

As described above, the framework of the scheduling influence maximization heuristic is described in Algorithm 1.

4. Experiments and Statistical Analysis

4.1. *Preliminary*

4.1.1. *Datasets and statistical characteristics*

To simulate the diffusion dynamics of the presented TSFA model and validate the performance of the proposed scheduling influence maximization heuristic, simulative experiments on influence diffusion are conducted on four undirected social networks selected from SNAP.[a] Table 1 shows the statistical characteristics of the four social networks. The node degree distribution of the four social networks is illustrated in Fig. 2, respectively.

4.1.2. *Baseline algorithms*

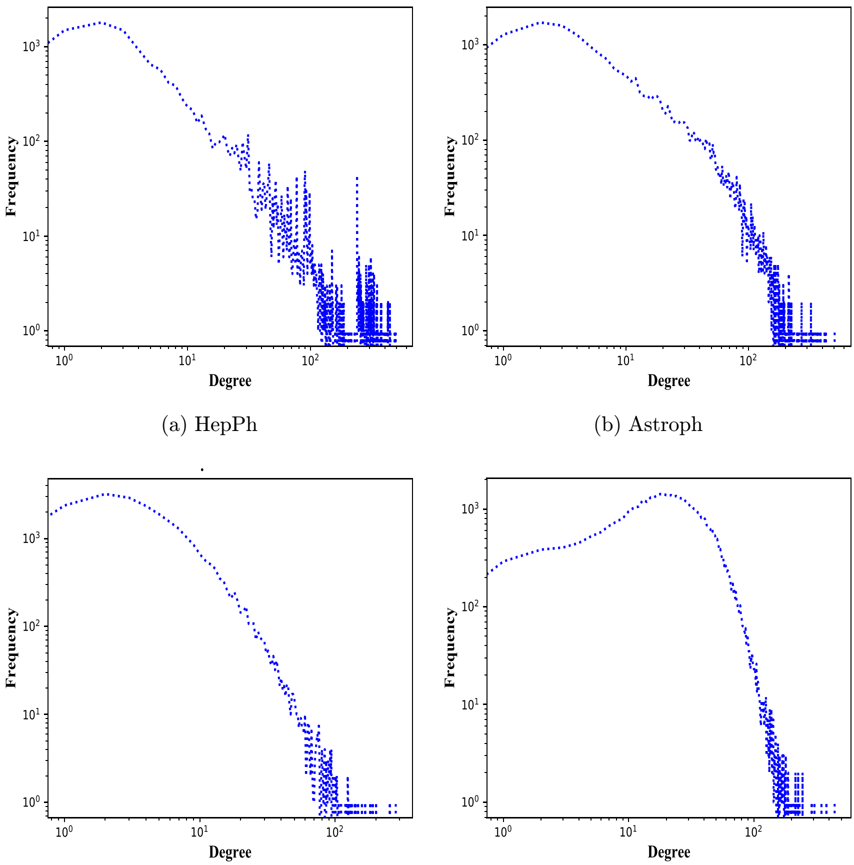
To simulate and show the performance of SDLM on influence diffusion, three seed scheduling methods, including the novel SSH algorithm,[21] classical degree centrality and the random heuristic method, and a novel influence maximization algorithm[32] are employed to conduct the compared experiments on the four social networks.

- Scheduled Seeding Heuristic (SSH) recommends seeding the influential nodes with the most utility score based on a novel Active Viral Marketing model at each time step.

Table 1. Statistical characteristics of the four social networks. $|V|$ and $|E|$ represent the number of nodes and edges in the network, respectively. $\langle k \rangle$ is the average node degree, $\bar{d}$ is the average shortest path distance, C represents the average clustering coefficient, and AC represents the assortativity coefficient.

Networks	$\lvert V\rvert$	$\lvert E\rvert$	$\langle k \rangle$	$\bar{d}$	C	AC
HepPh	10680	24316	4.554	7.486	0.440	0.238
Astroph	18772	198110	21.107	4.194	0.677	0.205
CondMat	23133	186936	16.162	5.352	0.055	0.135
Deezer	54573	846915	31.038	3.609	0.071	0.172

[a] http://snap.stanford.edu/snap/index.htm.



(a) HepPh (b) Astroph

(c) CondMat (d) Deezer

Fig. 2. (Color online) Node degree distribution of the four different social networks.

- Degree Centrality (DC), a topology-based method, seeds the unactivated node with the highest out-degree value from the network as the current most influential node to propagate the influence.
- Random method (RD) seeds a node that is uniformly drawn from the network topology and has not yet been selected into the seed set or not activated in previous time steps.
- Discrete Shuffled Frog-Leaping Algorithm (DSFLA) is a meta-heuristic population-based evolution algorithm that selects the targeted seed nodes effectively and efficiently by optimizing an expected influence diffusion function without using Monte–Carlo simulations. DSFLA is a typical static algorithm that activates all the seed nodes simultaneously at the beginning of the influence propagation process.

4.2. *Experimental simulations and performance analysis*

4.2.1. *Comparison on influence diffusion of the five algorithms*

In this phase, we mainly focus on the influence diffusion dynamics and the total number of activated nodes in the network under different budget constraints between advanced scheduling influence maximization algorithms and the static seed selecting algorithm. For the parameter settings of SDLM, the bounded budget is set to 100, the critical activation threshold C and oblivion time O are set to 2, respectively, and the number of initialized seed is set to 1. The simulation in the experiments of SSH runs 400 times for each combination, the recursion depth is set to 2, the activation threshold θ_v and the infection time are set to 3 and 10, respectively. As to the DSFLA algorithm, all the parameter setting strategies are set according to the original literature, and the propagation probability p under the IC model is set to 0.05.

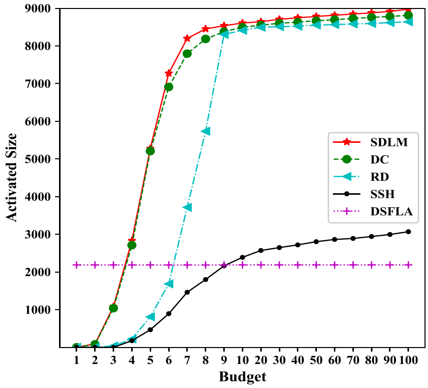
Figure 3 shows the comparison on the influence diffusion between scheduling influence maximization algorithms and the static metaheuristic DSFLA on the four different social networks. As illustrated in Fig. 3, we can see that there is distinct difference in the size of activated nodes between the two kind of seed selecting algorithms. The activated size shows that all the four scheduling algorithms are superior to the static influence maximization algorithm DSFLA, though it achieves satisfying influence diffusion at a large propagation probability $p = 0.05$.

More specifically, when the targeted seed set size is small, DSFLA performs better than the scheduling algorithms and achieves satisfied influence diffusion needing only limited seed nodes under the large propagation probability. On the contrary, the scheduling algorithms can only activate few nodes restrained by the activating threshold and the decaying oblivion time of node influence. However, the size of activated nodes by the seeded nodes of the scheduling algorithms improves with the increase of budget and is significantly larger than DSFLA. Obviously, the super-performance of the scheduling algorithms benefits from the time-dependent influence diffusion model and the scheduling influence maximization heuristics.

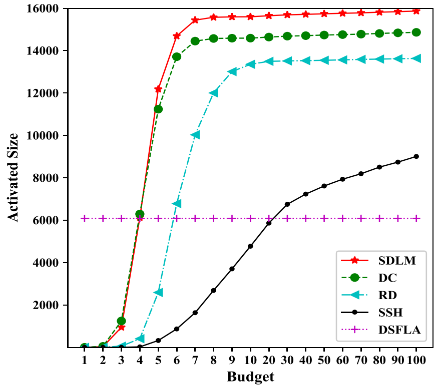
In addition, we can see that the proposed SDLM outperforms the SSH algorithm under different budget constraint scenarios. The highest proportion of activated nodes in the network achieved by the SSH algorithm is as high as 48%, as shown in Fig. 3(b), on the four social networks, while the SDLM algorithm activates almost 82% of the nodes in the network.

4.2.2. *Performance on variable number of initialized seeds*

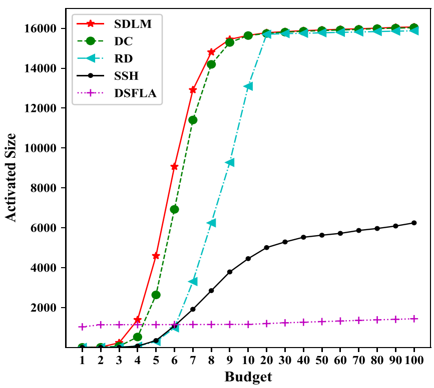
The simulations depicted in Fig. 3 also show that there are distinct difference among the four scheduling influence maximization algorithms, and the proposed SDLM achieves the best performance comparing to the other three scheduling heuristics. So, the effect of variable number of initialized seeds on the performance of the three scheduling heuristics, including SDLM, DC and RD, is further investigated under the same TSAF model in this phase. For simplification, we simulate the influence



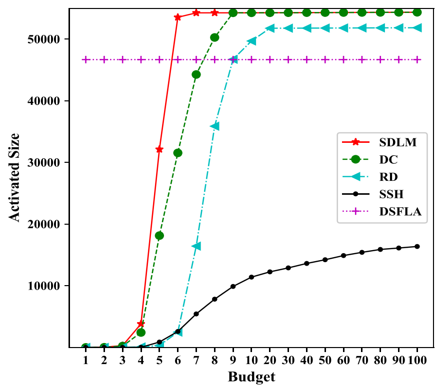
(a) HepPh



(b) Astroph



(c) CondMat

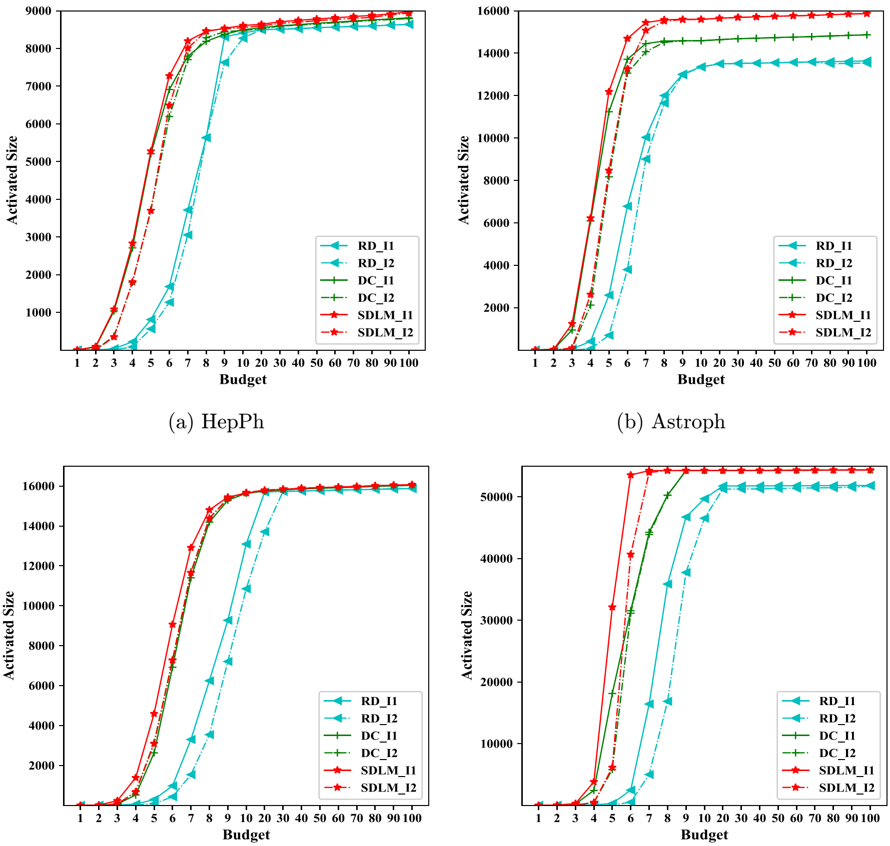


(d) Deezer

Fig. 3. (Color online) Performance comparison on the influence diffusion between scheduling influence maximization algorithms and the static metaheuristic DSFLA on the four different social networks.

diffusion of the three scheduling heuristics under the initialized number of seeds at scenarios where $I = 1$ and $I = 2$, respectively.

Figure 4 shows the influence diffusion simulation of the three scheduling heuristics at different number of initialized seeds. We can see from Fig. 4 that SDLM performs as the best scheduling algorithm at both of the two parameter setting scenarios comparing to the other two heuristics. The random method, which seeds a node that is uniformly drawn from the network topology and has not been selected into the seed set or not activated in previous time steps, achieves the worst influence diffusion according to the activated node size on the four social networks. More specially, when the number of initialized seed nodes is set to $I = 1$, the three scheduling heuristics can only diffuse the influence to merely few adjacent infectious nodes of the seed nodes. Then the size of activated nodes rises exponentially with the increase of the budget, and increases slowly when most of the nodes in the networks are activated.



(a) HepPh (b) Astroph

(c) CondMat (d) Deezer

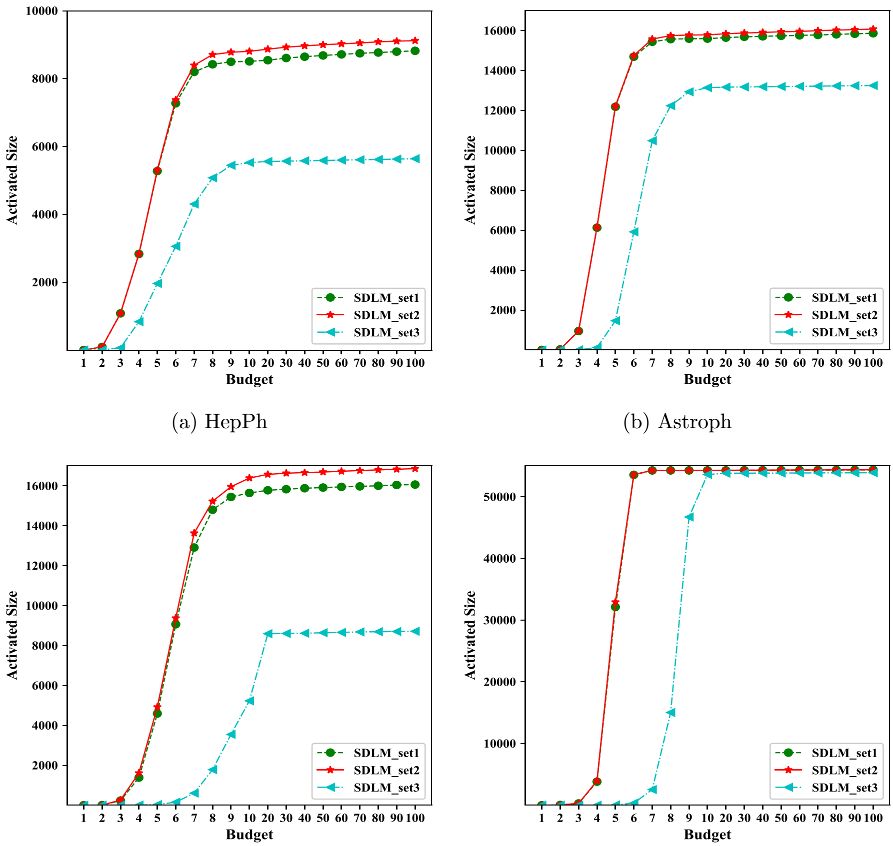
Fig. 4. (Color online) Performance of the scheduling heuristic SDLM comparing to the DC and RD methods under variable number of initialized seeds on the social networks.

At the scenario $I = 2$, all the three algorithms infect fewer nodes at initial stages than the former scenario. The reason for the inferiority is that the initialized seed nodes selected according to the heuristic metrics are scattered into the network, so it needs consume more budget to meet the activation threshold. However, all the three heuristics can still achieve comparable activated size to the first scenario where the number of initialized seed nodes is set to 1.

4.2.3. *Performance on variable combinatorial scheduling parameter settings*

The simulations depicted in Figs. 3 and 4 show that there are distinct difference among the three scheduling influence maximization algorithms, and the proposed SDLM achieves the best performance comparing to the other scheduling heuristics under the TSAF model. Therefore, the effectiveness of variable combinatorial parameter setting schemes on the performance of SDLM is further explored in this phase. According to the correlated relation among the three dominating parameters including node's influence oblivion time O, activation threshold C and the number of initialized seed I, three typical parameter setting strategies ((i) $I = 1$, $O = 2$, $C = 2$. (ii) $I = 1$, $O = 3$, $C = 2$. (iii) $I = 1$, $O = 2$, $C = 3$) are conceived to simulate the influence diffusion of SDLM.

The illustrations shown in Fig. 5 prove that variable parameter setting strategies have remarkable effect on the performance of SDLM. It is observed that larger oblivion time would contribute more activated nodes during the influence diffusion process. Consequently, the SDLM under the second parameter setting strategy performs the most satisfied influence diffusion. Instead, when the infectious threshold is larger, it needs more inflected and infectious nodes from adjacent neighborhood to infect the current uninfected node, so it would result in the phenomenon that the



(a) HepPh (b) Astroph

(c) CondMat (d) Deezer

Fig. 5. (Color online) Performance of the SDLM under variable combinatorial parameter setting strategies on the four social networks.

infected and infectious nodes become noninfectious ones before their influence decays to 0, as particularly shown in Figs. 5(a) and 5(c).

Though the experimental results show that parameter setting strategies in the scheduling influence maximization algorithms based on the seed activating feedback model play an important role in promoting the diffusion of node influence in the network, the extensive experiments conducted on the four social networks prove that the scheduling influence maximization algorithms show their superior performance to conventional ones that activate all the seed nodes simultaneously at the beginning of the influence diffusion process. It is a promising way to make further exploration on the scheduling influence maximization problem to better understand the influence diffusion dynamics, customer behavior modeling and control over the cascading process such as in active viral marketing and information diffusion prediction of practical scenarios, etc.

5. Conclusions

In the static influence maximization problem that would activate all the selected seed nodes simultaneously and let the influence diffuses passively in the network without any external intervention, it is hard to depict precisely the diffusion dynamics in reality where the marketers usually impose interventions to control the diffusion process. To better model the user behaviors and control over the influence diffusion process, this paper presents a time-dependent seed activating feedback model by considering the influence oblivion factor and the budget constraint to depict the influence diffusion dynamics performed in the network. And a scheduling influence maximization heuristic is proposed based on the determinate and latent margin of the candidate seed nodes to optimize the scheduling seed set. Extensive experiments on four social networks show that the proposed SDLM outperforms typical centrality-based scheduling heuristics and the conventional influence maximization algorithms. It provides positive chance to control the influence diffusion over the network such as in active viral marketing campaigns and information diffusion prediction of practical scenarios by implementing scheduling seed activation strategies.

Acknowledgments

This work was financially supported by the National Key Research and Development Plan of China under Grant Number 2020YFB1713600 and the Zhejiang Provincial Natural Science Foundation under Grant Number LQ20F020011.

References

1. E. Stephen, G. Hasan, V. S. A. Kumar, M. V. Marathe, S. Aravind, T. Zoltn and W. Nan, *Nature* **429**, 180 (2004).
2. J. Yan, H. He and Y. Sun, *IEEE Trans. Inform. Foren. Security* **9**, 451 (2014).

3. J. Gao, B. Barzel and A. L. Barabasi, *Nature* **530**, 307 (2016).
4. P. Domingos and M. Richardson, Mining the network value of customers, in *ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, ACM, New York, NY, USA, 2001, pp. 57–66.
5. D. Kempe and J. Kleinberg, Maximizing the spread of influence through a social network, in *ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, Washington, DC, USA, 2003, pp. 137–146.
6. S. Peng, Y. Zhou, L. Cao, S. Yu, J. Niu and W. Jia, *J. Netw. Comput. Appl.* **106**, 17 (2018).
7. Y. Li, J. Fan, Y. Wang and K. Tan, *IEEE Trans. Knowl. Data Eng.* **30**, 1852 (2018).
8. F. Chierichetti, J. Kleinberg and A. Panconesi, *SIAM Journal on Computing(SIAM J. COMPUT.)* **43**, 1906 (2014).
9. S. Lin, Q. Hu, F. Wang and P. S. Yu, Steering information diffusion dynamically against user attention limitation, in *2014 IEEE Int. Conf. Data Mining*, Shenzhen, China, 2014, pp. 330–339.
10. M. Samadi, R. Nagi, A. Semenov and A. Nikolaev, *Omega* **77**, 96 (2018).
11. S. K. Pal, S. Kundu and C. A. Murthy, *Fundam. Inf.* **130**, 317 (2014).
12. S. Kundu, C. A. Murthy and S. K. Pal, A new centrality measure for influence maximization in social networks, in *Proc. 4th Int. Conf. Pattern Recognition and Machine Intelligence, PReMI'11* (Springer-Verlag, Berlin, Heidelberg, 2011), pp. 242–247.
13. T. Cao, X. Wu, S. Wang and X. Hu, Oasnet: An optimal allocation approach to influence maximization in modular social networks, in *Proc. 2010 ACM Symp. Applied Computing, SAC '10* (ACM, New York, NY, USA, 2010), pp. 1088–1094.
14. A. Goyal, W. Lu and L. V. S. Lakshmanan, Celf++: optimizing the greedy algorithm for influence maximization in social networks, in *Int. Conf. Companion on World Wide Web*, Hyderabad, India, 2011, pp. 47–48.
15. J. Tang, R. Zhang, Y. Yao, Z. Zhao, P. Wang, H. Li and J. Yuan, *Knowl.-Based Syst.* **160**, 88 (2018).
16. J. Shang, H. Wu, S. Zhou, J. Zhong, Y. Feng and B. Qiang, *Phys. A: Stat. Mech. Appl.* **512**, 1085 (2018).
17. R. Mohamadi-Baghmolaei, N. Mozafari and A. Hamzeh, *Eng. Appl. Artif. Intell.* **41**, 195 (2015).
18. F. Morone and H. A. Makse, *Nature* **527**, 544 (2015).
19. X. Li, X. Cheng, S. Su and C. Sun, *Neurocomputing* **275**, 1601 (2018).
20. L. Lü, D. Chen, X. L. Ren, Q. M. Zhang, Y. C. Zhang and T. Zhou, *Phys. Rep.* **650**, 1 (2016).
21. A. Sela, D. Goldenberg, I. Ben-Gal and E. Shmueli, *Exp. Syst. Appl.* **107**, 45 (2018).
22. J. Jankowski, R. Michalski, P. Bródka and A. Karczmarczyk, Increasing coverage of information diffusion processes by reducing the number of initial seeds, in *Proc. 2017 IEEE/ACM Int. Conf. Advances in Social Networks Analysis and Mining 2017, ASONAM '17* (ACM, New York, NY, USA, 2017), pp. 713–720.
23. Y. Ni, *Appl. Soft Comput.* **56**, 730 (2017).
24. A. Sela, I. Ben-Gal, A. S. Pentland and E. Shmueli, Improving information spread through a scheduled seeding approach, in *Proc. 2015 IEEE/ACM Int. Conf. Advances in Social Networks Analysis and Mining 2015, ASONAM '15* (ACM, New York, NY, USA, 2015), pp. 629–632.
25. D. Goldenberg, A. Sela and E. Shmueli, *IEEE Trans. Comput. Soc. Syst.* **5**, 621 (2018).
26. M. Samadi, R. Nagi, A. Semenov and A. Nikolaev, *Omega* **77**, 96 (2018).
27. S. Tang and J. Yuan, *Oper. Res. Lett.* **48**, 24 (2020).
28. L. Weng, A. Flammini, A. Vespignani and F. Menczer, *Sci. Rep.* **2**, 335 (2012).

29. N. O. Hodas and K. Lerman, How visibility and divided attention constrain social contagion, in *2012 Int. Conf. Privacy, Security, Risk and Trust and 2012 Int. Conf. Social Computing*, 2012.
30. S. Aral, L. Muchnik and A. Sundararajan, *Netw. Sci.* **1**, 125 (2013).
31. A. Sela, D. Goldenberg, E. Shmueli and I. Ben-Gal, Scheduled seeding for latent viral marketing, in *Proc. 2016 IEEE/ACM Int. Conf. Advances in Social Networks Analysis and Mining, ASONAM '16* (IEEE Press, San Francisco, CA, USA, 2016), pp. 642–643.
32. J. Tang, R. Zhang, P. Wang, Z. Zhao, L. Fan and X. Liu, *Knowl.-Based Syst.* **187**, 104833 (2020).