

## Research Article

# On the Shoulders of Giants: Incremental Influence Maximization in Evolving Social Networks

**Xiaodong Liu, Xiangke Liao, Shanshan Li, Si Zheng, Bin Lin, Jingying Zhang, Lisong Shao, Chenlin Huang, and Liquan Xiao**

*School of Computer, National University of Defense Technology, Changsha 410073, China*

Correspondence should be addressed to Xiaodong Liu; liuxiaodong@nudt.edu.cn

Received 13 March 2017; Revised 4 July 2017; Accepted 1 August 2017; Published 28 September 2017

Academic Editor: Piotr Brodka

Copyright © 2017 Xiaodong Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Influence maximization problem aims to identify the most influential individuals so as to help in developing effective viral marketing strategies over social networks. Previous studies mainly focus on designing efficient algorithms or heuristics on a static social network. As a matter of fact, real-world social networks keep evolving over time and a recalculation upon the changed network inevitably leads to a long running time. In this paper, we propose an incremental approach, IncInf, which can efficiently locate the top- $K$  influential individuals in evolving social networks based on previous information instead of calculation from scratch. In particular, IncInf quantitatively analyzes the influence spread changes of nodes by localizing the impact of topology evolution to only local regions, and a pruning strategy is further proposed to narrow the search space into nodes experiencing major increases or with high degrees. To evaluate the efficiency and effectiveness, we carried out extensive experiments on real-world dynamic social networks: Facebook, NetHEPT, and Flickr. Experimental results demonstrate that, compared with the state-of-the-art static algorithm, IncInf achieves remarkable speedup in execution time while maintaining matching performance in terms of influence spread.

## 1. Introduction

The increasing popularity of online social network has promoted the diffusion of information, opinions, adoption of new products, and so forth and provided great opportunities for intelligent viral marketing. To benefit best from the word-of-mouth effect, influence maximization (IM) is one fundamental and important problem that aims to identify a small set of influential individuals so as to develop effective viral marketing strategies to maximize the influence over a given social network [1]. As a matter of fact, real-world social networks keep evolving over time. For example, in Facebook, new people might join, while old ones might withdraw, and people might make new friends with each other. Moreover, real-world social networks are evolving in a rather surprising speed; it is reported that as much as 1 million new accounts are created in Twitter every day [2]. Such massive evolution of network topology, on the contrary, may

lead to a significant transformation of the network structure, thus raising a natural need of efficient reidentification.

Existing researches and solutions on influence maximization focus mainly on developing effective and efficient algorithms on a given static social network. Although one could possibly run any of the static influence maximization methods, such as [3–6], to find the new top- $K$  influential individuals when the network is updated, this approach has some inherent drawbacks that cannot be neglected: (1) the running time of a specific static method can be extremely long and unacceptable, especially on large-scale social networks, and (2) whenever the network topology is changed, we need to recalculate the influence spreads for all the nodes, which leads to very high costs. Can we quickly and efficiently identify the influential nodes in evolving social networks? Can we incrementally update the influential nodes based on previously known information instead of frequently recalculating from scratch?

Unfortunately, the rapidly and unpredictably changing topology of a dynamic social network poses several challenges in the reidentification of influential users, which we list as follows. On one hand, the interconnections between edges in real-world social graphs are rather complicated; as a result, even one small change in topology may affect the influence spreads of a large number of nodes, not to mention the massive changes in large-scale social networks. It is very difficult to efficiently compute the changes of influence spreads for all the nodes after the evolution. On the other hand, since there are a great number of nodes in large-scale social networks, how to effectively limit the range of potential influential nodes and reduce the amount of calculation to the maximum is a very challenging problem.

To well address these challenges, we investigate the dynamic characteristics exhibited during the evolution of real-world social networks. Through tests on three real-world dataset traces, Facebook, NetHEPT, and Flickr, we observe that, first, the growth of social network is mainly based on the preferential attachment principle [7]; that is, the new-coming edges prefer to attach to nodes with higher degree, which naturally leads to the “rich-get-richer” phenomena; and, second, the top- $K$  influential nodes are mainly selected from those high-degree nodes. Inspired by such observations, we know that the influence changes of some nodes will have no impact on the top- $K$  selection and thus can be pruned to reduce the amount of calculation. Motivated by this, we propose IncInf, an incremental method to identify the top- $K$  influential nodes in evolving social networks instead of recalculating from scratch, thus significantly improving the efficiency and scalability to handle extraordinarily large-scale networks. To sum up, the main contributions of IncInf are as follows.

First, we design an efficient approach to quantitatively analyze the influence spread changes from network topology evolution by adopting the idea of localization. A tunable parameter is provided for tradeoff between efficiency and effectiveness.

Second, we propose a pruning strategy that could effectively narrow the search space into nodes only experiencing major increases or with high degrees based on the changes of influence spread and the previous top- $K$  information.

Third, we conduct extensive experiments on three dynamic real-world social networks. Compared with the state-of-the-art static algorithm, IncInf achieves remarkable speedup in execution time while providing matching influence spread. Moreover, IncInf provides better scalability to scale up to extraordinarily large-scale networks.

A preliminary version of this paper appears in [8], where we presented the basic idea of IncInf algorithm. In this paper, we make the following additional contributions. First, we add corresponding experiments to compare IncInf with IMM [9] in terms of influence spread and running time. Second, we test the effect of our pruning strategy to demonstrate its effectiveness. Third, we add a new experiment to evaluate the sensitivity of the localization parameter  $\theta$  and pruning threshold  $\eta$  in terms of influence spread and running time.

The remainder of this paper is organized as follows. In Section 2, we show the related work. Section 3 presents

related preliminaries and problem definition. Section 4 shows the structural evolution characteristics of dynamic social networks that we observe from three datasets: Facebook, NetHEPT, and Flickr. Section 5 details the design of our incremental algorithm IncInf. The performance of IncInf is evaluated by comprehensive experiments in Section 6. We conclude the paper in Section 7.

## 2. Related Work

Influence maximization on static networks has attracted a great deal of attention. The hill-climbing greedy algorithm proposed by Chen et al. suffers from low efficiency, and many efficient algorithms have been proposed recently to address this problem. Leskovec et al. [5] exploit the submodularity of influence spread function and develop an optimized greedy algorithm, CELF, which is much faster than basic greedy algorithm. Chen et al. [3] propose MixGreedy, which computes the influence spread for each seed set in one single simulation and incorporates the CELF optimization. MIA [4] uses local arborescence structures of each node to approximate the influence spread, thereby gaining efficiency by restricting computations and updates only to the local regions. However, MIA only considers static networks, while in this paper we specifically design an incremental algorithm for evolving social networks. Recently, Wang et al. [10] propose a Community Greedy Algorithm (CGA) that took community property into account. Goyal et al. propose CELF++ [11], which further exploits the property of submodularity of the spread function to avoid unnecessary recomputations of marginal gains and considerably improves the efficiency of CELF algorithm. IRIE [12] is also a heuristic proposed by Jung et al., which incorporates influence ranking algorithm with influence estimation method to achieve scalability. Liu et al. [13] design a new framework to accelerate the influence maximization by leveraging the parallel processing capability of GPU. Chen et al. [14] develop a community-based framework to tackle the influence maximization problem with an emphasis on the efficiency issue. Tang et al. [9] design a martingale approach that tries to find the top- $K$  nodes in near-linear time. And, in [15], Wang proposes a method to obtain each node’s marginal contribution by Owen value and deploys it in online terrorist network analysis. Lu et al. study the complexity of the influence maximization problem in deterministic linear threshold model in [16]. In [17], Lu et al. show how to efficiently estimate the influence spread for influence maximization under the linear threshold model. In [18], Nguyen and Zheng focus on the budgeted influence maximization (BIM) problem that aims to select seed nodes at a total cost no more than the fixed budget. Han et al. [19] study the influence maximization in timeliness networks and design a novel algorithm that incorporates time delay for timeliness and opportunistic selection for acceptance ratio. Liu et al. [20] propose the time-constrained influence maximization problem and develop a set of parallel algorithms for achieving more time savings. Pei et al. [21] take

advantage of the concept of subcritical path and propose CI-TM, a collective influence algorithm of optimal percolation for second-order transitions.

The influence maximization problem on dynamic social networks still remains largely unexplored to date. Habiba et al. [22] and Michalski et al. [23] propose a dynamic social network model that is different from ours. In their proposal, the network keeps evolving during the process of influence propagation, and their goal is to find the top- $K$  influential nodes over such a dynamic network. When compared to [22, 23], our work is based on snapshot graph model and our goal is to incrementally identify top- $K$  influential nodes based on the topology changes of two adjacent snapshots. Chen et al. [24] extend the IC model to incorporate the time delay aspect of influence diffusion among individuals in social networks and consider time-critical influence maximization, in which one wants to maximize influence spread within a given deadline. Meanwhile, in [25], the authors consider a continuous time formulation of the influence maximization problem in which information or influence can spread at different rates across different edges. Aggarwal et al. [26] try to discover influential nodes in dynamic social networks and they design a stochastic approach to determine the information flow authorities with the use of a globally forward approach and a locally backward approach. Their influence model and target are different from ours. Zhuang et al. [27] argue that the evolution of online social network could not be fully observed and design a probing strategy so that the actual influence diffusion process can be best uncovered with the probing nodes. Tong et al. [28] mainly focus on the fact that the diffusion processes in real-world dynamic social networks have many aspects of uncertainty and propose a method that selects seed users in an adaptive manner.

### 3. Preliminaries and Problem Statement

In this section, we illustrate the definition of social network and the influence diffusion model that we will use throughout the paper and then give the problem definition of influence maximization in evolving networks.

#### 3.1. Preliminaries on Influence Maximization

*Social Network.* A social network is formally defined as a directed graph  $G = (V, E, P)$ , where node set  $V = \{v_1, v_2, \dots, v_n\}$  denotes entities in the social network. Each node can be either active or inactive and will switch from being inactive to being active if it is influenced by other nodes. Edge set  $E \subset V \times V$  is a set of directed edges representing the relationship between different users. Take Twitter as an example. A directed edge  $(v_i, v_j)$  will be established from node  $v_i$  to  $v_j$  if  $v_i$  is followed by  $v_j$ , which indicates that  $v_j$  may be influenced by  $v_i$ .  $P$  denotes the influence probability of edges; each edge  $(v_i, v_j) \in E$  is associated with an influence probability  $p(v_i, v_j)$  defined by function  $p : E \rightarrow [0, 1]$ . If  $(v_i, v_j) \notin E$ , then  $p(v_i, v_j) = 0$ .

```

(1) Initialize  $S = \emptyset$ 
(2) for  $i = 1$  to  $K$  do
(3)   Select  $v = \arg \max_{v_i \in (V \setminus S)} (\sigma(S \cup v_i) - \sigma(S))$ 
(4)    $S = S \cup \{v\}$ 
(5) end for

```

ALGORITHM 1: Basic greedy algorithm.

*Independent Cascade (IC) Model.* IC model is a popular diffusion model that has been well studied in [3, 6, 10, 29]. Given an initial set  $S$ , the diffusion process of IC model works as follows. At step 0, only nodes in  $S$  are active, while other nodes stay in the inactive state. At step  $t$ , for each node  $v_i$  that has just switched from being inactive to being active, it has a single chance to activate each currently inactive neighbor  $v_j$  and succeeds with a probability  $p(v_i, v_j)$ . If  $v_i$  succeeds,  $v_j$  will become active at step  $t + 1$ . If  $v_j$  has multiple newly activated neighbors, their attempts in activating  $v_j$  are sequenced in an arbitrary order. Such a process runs until no more activations are possible [29]. We use  $\sigma(S)$  to denote the influence spread of the initial set  $S$ , which is defined as the expected number of active nodes at the end of influence propagation.

*Basic Greedy Algorithm.* Domingos and Richardson [1, 30] first introduced the influence maximization problem on static networks in 2001. In [29], Kempe et al. propose a basic hill-climbing greedy algorithm as shown in Algorithm 1. The proposed greedy algorithm works in  $K$  iterations, starting with an empty set  $S$  (line (1)). In each iteration, a node  $v_i$  that brings the maximum marginal influence spread  $\sigma_S(v_i) = \sigma(S \cup v_i) - \sigma(S)$  is selected to be included in  $S$  (lines (3) and (4)). The process ends when the size of  $S$  reaches  $K$  (line (2)). However, this algorithm has a serious efficiency drawback due to the compute-intensive influence spread calculation. Several recent studies [3–6, 10, 12, 31–35] aimed at addressing this efficiency issue.

#### 3.2. Formal Definition of IM Problem in Evolving Networks.

This paper distinguishes itself from previous works by considering the dynamic nature of online social networks. As a matter of fact, the real-world social networks are not wholly static but keep evolving gradually over time. The evolution of large social networks has raised new sets of questions; among them one interesting yet challenging problem is how to quickly identify the top- $K$  influential users when the topology of the network is changed.

To solve such a problem, we define an evolving network  $\zeta = (G^0, G^1, \dots, G^t)$  as a sequence of network snapshots evolving over time, where  $G^t = (V^t, E^t, P^t)$  is the network snapshot at time  $t$ .  $\Delta G^t = (\Delta V^t, \Delta E^t, \Delta P^t)$  denotes the structural change of network graph  $G^t$ . Obviously, we have  $G^{t+1} = G^t \cup \Delta G^t$ . And the influence maximization problem is defined as follows:

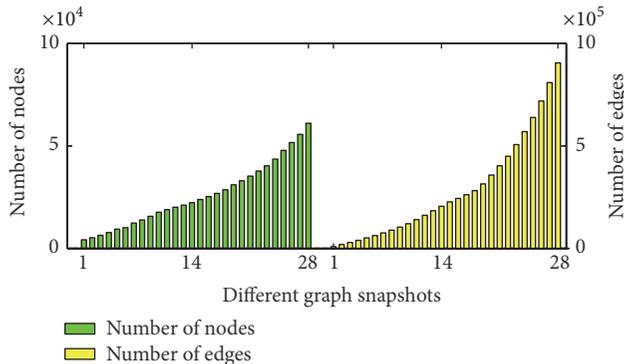


FIGURE 1: Number of nodes and edges per month of the Facebook dataset.

*Given.* One has the social network  $G^t$  at time  $t$ , the top- $K$  influential nodes  $S^t$  in  $G^t$ , and the structural evolution  $\Delta G^t$  of graph  $G^t$ .

*Objective.* The aim is to identify the influential nodes  $S^{t+1} \subset V^{t+1}$  of size  $K$  in  $G^{t+1}$  at time  $t + 1$ , such that the influence spread  $\sigma(S^{t+1})$  is maximized at the end of influence diffusion.

#### 4. Observations of Social Network Evolution

In this section, we study some patterns of social network evolution. The numbers of nodes and edges are firstly investigated in Section 4.1 to examine the growth of users and interconnections over time. Then, we look into the degree distribution of nodes and the preferential attachment rule for new edges in Section 4.2. We further examine the relation between the influence and the degree of node in Section 4.3. We study three network traces, namely, Facebook, NetHEPT, and Flickr, whose detailed description can be found in Section 6. Here we only show the results on Facebook, since the evolution trends on the other datasets are qualitatively similar and thus were omitted.

*4.1. How Fast Does the Network Evolve?* Nodes and edges are the basic elements of the social network topology. In this subsection, we use the numbers of nodes and edges to examine the growth of users and interconnections over time. Figure 1 illustrates the numbers of nodes and edges over the entire trace period on the Facebook dataset; we take a snapshot per month. From Figure 1, we observe a linear increase in the number of nodes, which indicates a steady number of new users who joined the network per month. Meanwhile, in terms of edges, the number goes up almost exponentially. The number of edges after 14 months is 25.6x of that in the initial graph while the number rises to 112.9x after 28 months. Such rapid growth of nodes and edges raises a natural need to efficiently find the most influential nodes after the topology evolution.

*4.2. What Is the Pattern of Network Topology Evolution?* Understanding the pattern of the network topology evolution is of primary importance to design efficient influence

maximization algorithms for evolving social networks. In this subsection, we further investigate the degree distribution of nodes and the preferential attachment rule [7, 36, 37] for new coming edges. Figure 2(a) shows the degree distribution of the Facebook final graph in log-log scale. As expected, it mainly follows the well-known power-law distribution. A large percent of the users have only a small number of links with other users, while there exist some “hub” nodes with extremely large number of connections. This is consistent with the real-world networks.

We also study the preferential attachment rule or, in other words, the “rich-get-richer” rule [38], which postulates that when a new node joins the network, it creates a number of edges, where the destination node of each edge is chosen proportional to the destination’s degree. This means that new edges are more likely to connect to nodes with high degree than ones with low degree. This is reasonable in reality; Lady Gaga gains 30,000 new followers on average every day [39], which can never imagine for any common individual. The results on the Facebook dataset are demonstrated in Figure 2(b), where  $x$ -axis is the degree of different nodes and  $y$ -axis is the average number of new edges attached to nodes of different degree. Note that both  $x$ -axis and  $y$ -axis are in log scale. From Figure 2(b), we can see that the degree of users in Facebook is linearly correlated with the number of new links created. This suggests that high-degree nodes get super preferential treatment. Consequently, the influence spread change should be considerably great for the influential nodes, while there may be only small or even no change for ordinary people.

*4.3. What Is the Relation between Influence and Degree?* Examining the relation between the influence and the degree of node can help us understand the effect of degree changing on the influence spread of nodes. For this reason, we run the static MixGreedy algorithm [3] on the final graph and identify the top-50 influential nodes. The results on the Facebook dataset are illustrated in Figure 3, where  $x$ -axis is the rank of degrees of different nodes (we only show the top 150). Obviously, all the selected influential nodes have a large degree. In particular, among the 50 nodes, 48 nodes rank in top 100 of the whole 61,096 nodes in terms of degree, and the

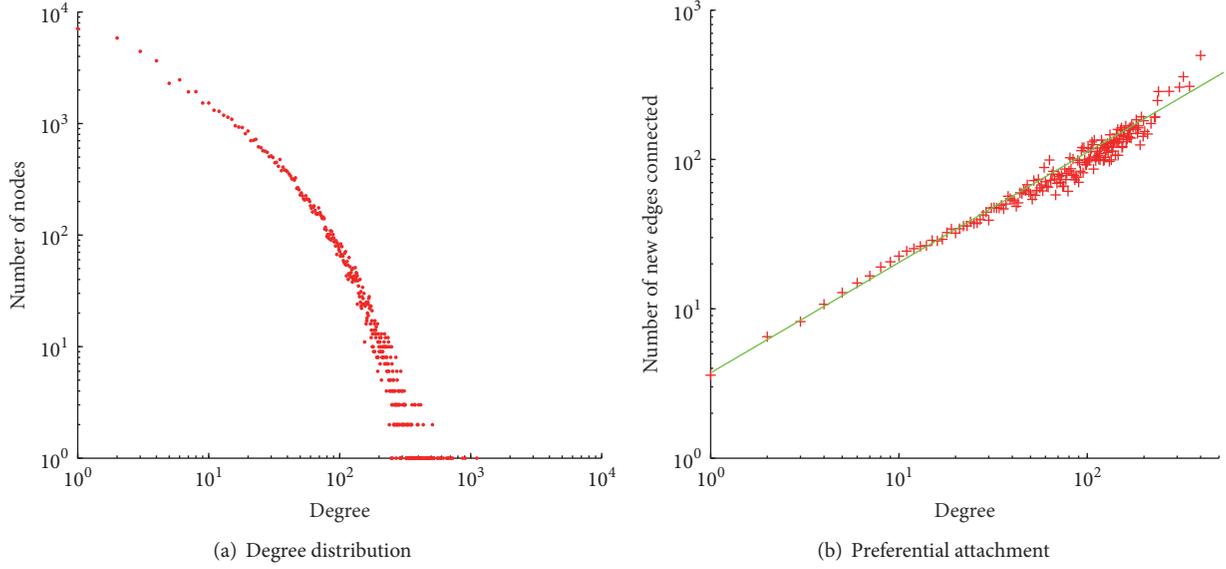


FIGURE 2: Degree distribution and preferential attachment on Facebook.

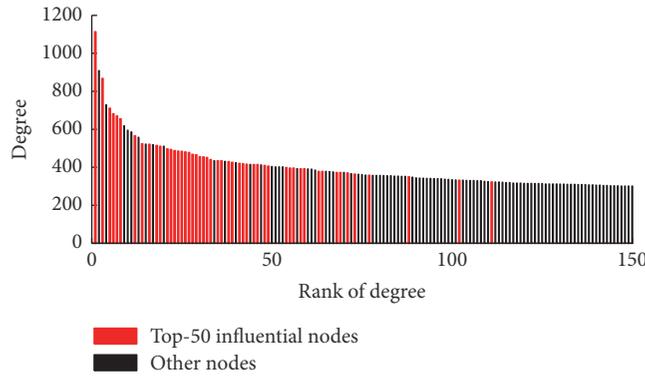


FIGURE 3: The relation between the influence spread and the degree in Facebook.

other two nodes rank 102 and 111, respectively. Meanwhile, on the NetHEPT and Flickr datasets, the top-50 influential nodes are selected from the top 1.79% and 0.84% nodes in degree, respectively. This demonstrates that the top- $K$  influential nodes are mainly selected from those with large degrees. However, it is worth noting that the top- $K$  influential nodes in influence maximization are usually not the top- $K$  nodes ranking in degree, since the influence spreads of different nodes may overlap with each other.

## 5. IncInf Design

In this section, we present the detailed design of IncInf, an incremental approach to solve the influence maximization problem on dynamic social networks. The main idea of IncInf is to take full use of the valuable information that is inherent in the network structural evolution and previous influential nodes so as to substantially narrow the search space of influential nodes. In this way, IncInf can significantly reduce the computation complexity and improve the efficiency. Figure 4

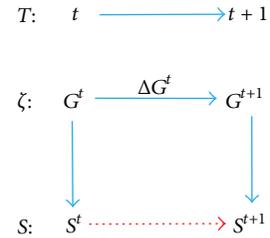


FIGURE 4: IncInf design.

briefly illustrates the general idea of IncInf in dynamic social networks. The top- $K$  influential nodes  $S^{t+1}$  of  $G^{t+1}$  at time  $t+1$  are incrementally identified based on the previous influential nodes  $S^t$  at time  $t$  and the structural change  $\Delta G^t$  from  $G^t$  to  $G^{t+1}$ . In particular, we design an efficient method to quantitatively analyze the impact of different structural changes on the influence spread of nodes by adopting the idea of localization (Section 5.2) and propose a pruning strategy to reduce the number of potential influential nodes

(Section 5.3). We first describe six types of basic operation of topology evolution in dynamic networks in Section 5.1.

**5.1. Basic Operations of Topology Evolution.** The evolution of social network, when reflected into its underlying graph, can be summarized into six categories, which are inserting or removing a node, introducing or deleting an edge, and increasing or decreasing the influence probability of an edge. We denote the six types of topology change as *addNode*, *removeNode*, *addEdge*, *removeEdge*, *addWeight*, and *decWeight*. The detailed descriptions and their effects on influence spread are shown in Table 1.

It should be noted that only after the *addNode* operation can node  $u$  establish links (*addEdge*) or sever links (*removeEdge*) with other nodes, and node  $u$  can only be removed when all its associated edges are deleted. Moreover, the weight operation can be equivalently decomposed into two edge operations. For example, *addWeight* ( $u, v, \Delta w$ ) can be divided into *removeEdge* ( $u, v$ ) and *addEdge* ( $u, v, w + \Delta w$ ), supposing that the previous weight of edge ( $u, v$ ) is  $w$ .

**5.2. Influence Spread Changes.** As discussed above, whenever an edge ( $u, v$ ) is introduced into or removed from the social network, the influence spread of all the nodes that can reach node  $u$  may be changed. However, as a matter of fact, the real-world social networks exhibit small-world network characteristics and the connections between nodes are highly complicated. As a result, even one small change in topology, such as an edge addition or removal, may affect the influence spread of a large number of nodes, thus introducing massive recalculations. In order to reduce the amount of computation, we design an approach to efficiently calculate the changes on the influence spread of nodes, which adopts the localization idea [4] and tries to restrict the influence spread to the local regions of nodes.

The main idea of localization is to use the local region of each node to approximate its overall influence spread. In particular, we use the maximum influence path to approximate the influence spread from node  $u$  to  $v$ . Here the maximum influence path  $\text{MIP}(u, v, G)$  from node  $u$  to  $v$  in graph  $G$  is defined as the path with the maximum influence probability among all the paths from node  $u$  to  $v$  and can be formally described as follows:

$$\text{MIP}(u, v, G) = \arg \max_{p \in P(u, v, G)} \{\text{prob}(p)\}, \quad (1)$$

where  $\text{prob}(p)$  denotes the propagation probability of path  $p$  and  $P(u, v, G)$  denotes all the paths from node  $u$  to  $v$  in graph  $G$ . For a given path  $p = \{u_1, u_2, \dots, u_m\}$ , the propagation probability of path  $p$  is defined as follows:

$$\text{prob}(p) = \prod_{i=1}^{m-1} p(u_i, u_{i+1}). \quad (2)$$

Moreover, an influence threshold  $\theta$  is set to tradeoff between accuracy and efficiency. During the propagation process, we only consider paths whose influence probability is larger than  $\theta$  while ignoring those with probability smaller than  $\theta$ . By

doing this, the influence is effectively restricted to the local region of each node.

Similarly, in our proposal, we localize the impact of topology changes on influence spread into local regions and thus reduce the amount of computation. Among six types of topology change, *addNode* (or *removeNode*) is the most straightforward, since it simply sets the influence spread of the node to 1 (or 0); *addWeight* and *decWeight* as well as *removeEdge* are methodologically similar to *addEdge*. Consequently, in the following, we take *addEdge* as an example to show which nodes' influence spreads need to be updated and how to determine those changes when a new edge is added to the graph.

Consider the case when a new edge  $e = (u, v, w)$  is introduced between two existing nodes  $u$  and  $v$ . We denote the graphs before and after such a topology change as  $G^t$  and  $G^{t'}$ , and the current seed set is  $S$ . The detailed algorithm is described in Algorithm 2. According to the principle of localization [4], if the propagation probability  $w$  is smaller than the specified threshold  $\theta$  or not bigger than the probability of  $\text{MIP}(u, v, G^t)$ , edge  $e$  can be simply neglected and there is no need to update any node's influence spread (lines (1)–(3)). Otherwise, the newly added edge  $e$  would become  $\text{MIP}(u, v, G^{t'})$ . As a result, each node  $i$  whose maximum influence path to  $u$  has an influence probability larger than  $\theta$  is likely to experience a rise in terms of influence spread (line (4)) because node  $i$  may influence more nodes through the new edge  $e$ . So, we then check the probability of the maximum influence path from  $i$  to  $v$  and its successors in  $G^t$  and  $G^{t'}$ . Based on the two probabilities, we divide the problem into two small cases.

The first case is when the probability of maximum influence path from  $i$  to  $j$  in  $G^t$  is smaller than  $\theta$ , while that in  $G^{t'}$  is larger than  $\theta$  (lines (5)–(6)). Here  $j$  denotes the node whose probability of  $\text{MIP}(v, j, G^t)$  is larger than  $\theta$ . In such a case, node  $i$  builds a new path to  $j$  through the new edge  $e$ , which increases the influence spread of  $i$  by  $\text{prob}(\text{MIP}(i, j, G^{t'})) \times (1 - \text{prob}(j, S))$  (line (7)). Here  $\text{prob}(j, S)$  is the probability of the fact that node  $j$  is influenced by the current seed set  $S$ , which is defined as follows:

$$\begin{aligned} \text{prob}(j, S) &= \begin{cases} 1, & \text{if } j \in S \\ 1 - \prod_{w \in n(j)} 1 - \text{prob}(w, S) \cdot p(w, j), & \text{if } j \notin S. \end{cases} \quad (3) \end{aligned}$$

Here  $n(j)$  denotes the in-neighbour set of  $j$ .

The second case is when the probability of maximum influence path from  $i$  to  $j$  is larger than  $\theta$  in both  $G^t$  and  $G^{t'}$  (lines (9)–(11)). In this case, the influence increase of node  $i$  is  $(\text{prob}(\text{MIP}(i, j, G^{t'})) - \text{prob}(\text{MIP}(i, j, G^t))) \times (1 - \text{prob}(j, S))$ .

We treat the network dynamics from  $G^t$  to  $G^{t+1}$  as a finite change stream  $c_1, c_2, \dots, c_i, \dots$ , where each change  $c_i$  is one of the six topology changes we described previously. When all the changes in the change stream are processed, we can obtain the influence spread change for all the nodes. Next, we will

TABLE I: Details of six types of basic operation.

Operation	Description	Impact on influence spread
$addNode(u)$	Add a new node $u$ to the current network	The influence spread of $u$ is set to 1
$removeNode(u)$	Delete an existing node $u$ from the network	The influence spread of $u$ is set to 0
$addEdge(u, v, w)$	Introduce a new edge $(u, v)$ with $p(u, v) = w$	The influence spread of all the nodes that can reach $u$ may be increased
$removeEdge(u, v)$	Remove an existing edge $(u, v)$ from the network	The influence spread of all the nodes that can reach $u$ may be decreased
$addWeight(u, v, \Delta w)$	Increase $p(u, v)$ by $\Delta w$	The influence spread of all the nodes that can reach $u$ may be increased
$decWeight(u, v, \Delta w)$	Reduce $p(u, v)$ by $\Delta w$	The influence spread of all the nodes that can reach $u$ may be decreased

**Input:** a new edge  $e = (u, v, w)$ , graph  $G^t$ .

**Output:** The influence spread changes of nodes in  $G^{t+1}$ .

```

(1) if  $w < \theta$  or  $w \leq \text{prob}(\text{MIP}(u, v, G^t))$  then
(2)   return;
(3) end if
(4) for each node  $i$  with  $\text{prob}(\text{MIP}(i, u, G^t)) > \theta$  do
(5)   for each node  $j$  with  $\text{prob}(\text{MIP}(v, j, G^t)) > \theta$  do
(6)     if  $\text{prob}(\text{MIP}(i, j, G^t)) < \theta$  and  $\text{prob}(\text{MIP}(i, j, G^{t+1})) > \theta$  then
(7)        $\text{deltaInf}[i] += \text{prob}(\text{MIP}(i, j, G^{t+1})) \times (1 - \text{prob}(j, S))$ 
(8)     end if
(9)     if  $\text{prob}(\text{MIP}(i, j, G^t)) > \theta$  and  $\text{prob}(\text{MIP}(i, j, G^{t+1})) > \theta$  then
(10)       $\text{deltaInf}[i] += (\text{prob}(\text{MIP}(i, j, G^{t+1})) - \text{prob}(\text{MIP}(i, j, G^t))) \times (1 - \text{prob}(j, S))$ 
(11)    end if
(12)  end for
(13) end for

```

ALGORITHM 2: Edge addition.

show how to effectively find the top- $K$  influential nodes in the new graph  $G^{t+1}$  based on these influence spread changes and the previous influential nodes information.

**5.3. Potential Top- $K$  Influential Users Identification.** Inspired by the observations of Section 4, we design a pruning strategy to reduce the search space of potential influential nodes in this subsection. It is assumed that we only know which are the top- $K$  influential nodes in graph  $G^t$ , but their detailed influence spreads are beyond our knowledge. The reasons are mainly twofold. First, several influence maximization algorithms, such as DegreeDiscount [3], do not calculate the influence spread information to identify influential users so that such information is unavailable. Second, even though this information is ready, storing it will cost as much as  $O(nK)$  memory space, where  $n$  is the number of nodes in  $G^t$ . Since real-world social networks are typically of large scale, this will introduce serious storage overhead and directly affect the scalability.

From the preferential attachment rule, we know that the influence spread changes of those high-degree nodes

should be much greater than the ordinary nodes. Moreover, according to the power-law distribution, such high-degree nodes only account for a small part of the whole nodes. Consequently, we can pick out nodes only experiencing major increases or with high degrees because these nodes are of great potential to become the top- $K$  influential nodes in  $G^{t+1}$ . Then we only calculate the actual influence spreads for these selected nodes while ignoring the others. In this way, a large percent of nodes are pruned and the search space is largely narrow. It should be noted that a smart pruning strategy is of key importance, since a poor selection might either affect the efficiency or reduce the accuracy in terms of influence spread. We describe the details of our pruning strategy as follows:

- (1) In the  $i$ th iteration, if the influence spread of the previous influential node  $S_i^t$  increases in  $G^{t+1}$ , the chosen nodes are those with a larger influence spread change than  $\text{deltaInf}[S_i^t]$ .

In most cases, the influential nodes will attract a great number of new nodes and establish new links. Thus, their influence spreads will increase drastically. In such a case, it is impossible for the nodes

```

Input:  $G^t, S^t$ , and  $G^{t+1}$ .
Output: the top- $K$  influential nodes  $S^{t+1}$  in  $G^{t+1}$ .
(1) Initialize  $S^{t+1} = \emptyset$ ;
(2) for  $i = 1$  to  $K$  do
(3)   for each topology change  $c_j$  from  $G^t$  to  $G^{t+1}$  do
(4)     calculate the influence spread change  $\text{deltaInf}[\cdot]$ ;
(5)   end for
(6)   select a set of potential nodes as  $pn$  according to pruning strategy;
(7)   for each node  $v_j \in pn$  do
(8)     calculate the marginal influence spread  $\sigma_{S^{t+1}}(v_j)$ ;
(9)   end for
(10)  select  $v_{\max} = \arg \max_{v_j \in pn} (\sigma_{S^{t+1}}(v_j))$ ;
(11)   $S = S \cup v_{\max}$ ;
(12) end for

```

ALGORITHM 3: IncInf.

whose influence spread changes are smaller than the influential nodes to become the most influential nodes in  $G^{t+1}$ . Therefore, when the influence spread of the previous influential nodes increases, we only select those whose influence spread changes are larger than the influential nodes in  $G^t$ . According to the preferential attachment rule, such a pruning method can greatly narrow the search space and reduce the amount of computation.

- (2) In the  $i$ th iteration, if the influence spread of the previous influential node  $S_i^t$  decreases in  $G^{t+1}$ , in addition to qualification 1, the nodes are further selected to hold a sufficiently large degree or experience a sufficiently great increase. In order to formally define “large degree” and “great increase,” here we set a threshold  $\eta$  for tradeoff between running time and influence spread. Here the nodes with sufficiently large degrees (or great increase) are defined as the set of nodes  $v_j$  whose degree (or degree increase ratio) is among the top  $\eta$  percent of all nodes in  $G^{t+1}$ . The degree increase ratio of  $v_j$  is defined as  $\text{degree}_j^{t+1} / \text{degree}_j^t$ , where  $\text{degree}_j^t$  denotes the degree of node  $v_j$  in graph  $G^t$ . Experimental results in Section 6 will demonstrate that 5% may stand as a good tradeoff between running time and influence spread.

It should be noted that although the case where the influence spread of a previous influential node decreases during the evolution rarely happens, we consider it here for completeness. In this case, except for qualification 1, we further select nodes because the number of nodes satisfying qualification 1 is relatively large, which leads to mass computation. Meanwhile, in reality, a node with small degree has only very low probability to become an influential node. In order to select only the most potential nodes, we refine the requirement and additionally select the nodes with large degree and large increase. Consequently,

the search space is strictly circumscribed and the computational complexity is greatly reduced.

After the potential nodes are selected, we calculate the actual influence spread of these nodes in  $G^{t+1}$  and select the one with the maximum influence spread in each iteration. Algorithm 3 outlines the design of our proposed algorithm IncInf. IncInf iterates for  $K$  round (line (2)) and in each round selects one node, providing the maximum marginal influence spread. Lines (3)–(5) calculate the influence spread change of each node caused by the topology evolution. Nodes with great potential to become top- $K$  influential are selected (line (6)) and their influence spreads are computed in  $G^{t+1}$  (lines (7)–(9)). And then the node providing the maximal marginal gain will be selected and added to the set  $S^{t+1}$  (lines (10)–(11)).

## 6. Experiments

In this section, we present the experimental results of our algorithm on identifying top- $K$  influential nodes in dynamic social networks. We examine two metrics, running time and influence spread, for evaluating the effectiveness as well as the execution efficiency of different algorithms. The experimental results are detailed in Sections 6.2, 6.3, and 6.4.

*6.1. Experimental Setup.* We choose three real-world social networks: Facebook social network, NetHEPT citation network, and Flickr social network (Table 2 summarizes the statistical information of the datasets):

- (i) Facebook: this dataset is the friendship relationship network among New Orleans regional network on Facebook, spanning from September 2006 to January 2009 [40]. There are more than 60 K users connected together by as much as 1.5 M links in the social network. 41.4% of these edges contain no time information and are thus discarded. In our experiments, the nodes and links from September 2006 to April 2007 are used as the first snapshot and then network snapshots are recorded every 3 months

TABLE 2: Summary information of the real-world social networks.

Datasets	Nodes			Edges		
	Initial number	Final number	Growth	Initial number	Final number	Growth
Facebook	12,364	61,096	394%	73,912	905,665	1125%
NetHEPT	5,802	29,555	409%	57,765	352,807	511%
Flickr	1,620,392	2,570,535	58.6%	17,034,807	33,140,018	94.5%

- (ii) NetHEPT: this is an academic citation network [41] extracted from “High Energy Physics-Theory” section of the arXiv over the period from 1992 to 2003 and covers the citations within a dataset of 28 K papers with 352 K edges. In our experiments, the citation links of the first three years (i.e., from 1992 to 1994) are considered as the basic graph and the network snapshots are recorded once a year
- (iii) Flickr: this dataset [42] contains the user-to-user links crawled from the Flickr social network daily over the period from 2 November 2006 to 3 December 2006 and again from 3 February 2007 to 18 May 2007, representing a total of 104 days of growth. There are totally 2.5 M Flickr users and 33 M links. During this period of observation, over 9.7 million new links are formed and over 950,000 new users joined the network. In our experiments, we use the network before 2 November 2006 as the basic graph and another five snapshots are recorded on 3 December, 3 February, 3 March, 3 April, and 18 May

We compare our algorithm with five static algorithms: *MixGreedy*, *ESMCE*, *MIA*, *IMM*, and *Random*. *MixGreedy* is an improved greedy algorithm on the IC model proposed by Chen et al. in [3]. *ESMCE* is a power-law exponent supervised estimation approach designed by Liu et al. in [6]. *MIA* is a heuristic that uses local arborescence structures of each node to approximate the influence propagation [4]. *IMM* is an algorithm designed by Tang et al. based on the martingales estimation techniques and is able to run in near-linear time [9]. *Random* is a basic heuristic that randomly selects  $K$  nodes from the whole datasets.

The propagation probability of the IC model is selected randomly from 0.1, 0.01, and 0.001 for each network snapshot. The parameters of the evaluated algorithms are set as suggested by their authors. For *IMM*, the parameters  $\epsilon$  and  $\iota$  are set to 0.5 and 1, respectively, in the experiments. For the setting of  $\theta$  in *MIA*, as stated in Section 6.4, the knee point of the influence spread curve can serve as a good tuning point of  $\theta$  for tradeoff between efficiency and effectiveness. For example, the knee points of influence spread curves of Facebook and NetHEPT are 1/180 and 1/200, respectively. The value of  $\theta$  for a particular dataset in the experiment is selected according to the knee point to get the best tradeoff. For *ESMCE*, the confidence level is set to 95% and the maximal Monte Carlo error threshold is set to 5%. For *IncInf*, the value of  $\theta$  is set similar to that of *MIA*, that is, the knee point of the influence spread curve, while the value of  $\eta$  is set to 5% as suggested in Section 6.4 for tradeoff between running time and influence spread.

**6.2. Efficiency Study.** In this subsection, the efficiency of our proposed algorithm is studied and compared with corresponding static algorithms, *MixGreedy* and *MIA*, through experiments on the Facebook, NetHEPT, and Flickr datasets. The experiments are conducted on a PC with Intel Core i7 920 CPU @2.67 GHz and 6 GB RAM. The running times of four algorithms are measured by selecting 50 seeds from the whole dataset.

The time costs of different algorithms are illustrated in Figure 5, where we record the total time cost for each snapshot of the three datasets. Since incremental and static algorithms have the same time cost in the initial snapshot, they are omitted in the figure. The experimental results show that the time costs of our algorithm on each snapshot are obviously less than those of static algorithms. Obviously, *MixGreedy* takes the longest time among four kinds of influence maximization algorithms. It takes *MixGreedy* more than as much as 6 hours to identify the top-50 influential nodes on the final NetHEPT dataset, while the time is even longer on the larger dataset Facebook. Moreover, *MixGreedy* is not feasible to run on the largest dataset Flickr due to the unbearably long running time. *ESMCE*, benefiting from its sampling estimation method, runs much faster than *MixGreedy*, but it still takes as much as 3511 seconds on average to run on the five snapshots of Flickr. Compared with two greedy algorithms, the heuristic *MIA* and the martingales approach *IMM* perform much better. It only takes *MIA* 23.8 seconds (*IMM* 8.2 seconds, resp.) to run on the final Facebook graph. When running on the Flickr dataset with as much as 2.5 M nodes and 33 M edges, however, the speedup of *IMA* is still far from being satisfactory, since it still needs more than 45 minutes to finish. Comparatively, *IMM* performs better than *MIA* and it takes *IMM* nearly 10 minutes to locate the influential nodes. Meanwhile our proposed algorithm, *IncInf*, outperforms all the static algorithms in terms of efficiency. In particular, *IncInf* is almost four orders of magnitude faster than the *MixGreedy* algorithm on the Facebook dataset. Compared with the *MIA* heuristic, the speedup of *IncInf* is 8.41x and 6.94x on the Facebook and NetHEPT datasets, respectively. What is more, when applied on the largest dataset Flickr, *IncInf* can achieve as much as 20.65x speedup on average. The efficiency of *IncInf* is only slightly better than *IMM* on our small dataset NetHEPT, but it performs considerably better than *IMM* on Flickr dataset: the running time of *IncInf* is only 40% on average of *IMM* on each snapshot. This is because *IncInf* only computes the incremental influence spread changes and adaptively identifies the influential nodes based on the previous influential nodes and the current influence spread changes. The experimental results clearly validate the efficiency advantage of our incremental algorithm *IncInf*.

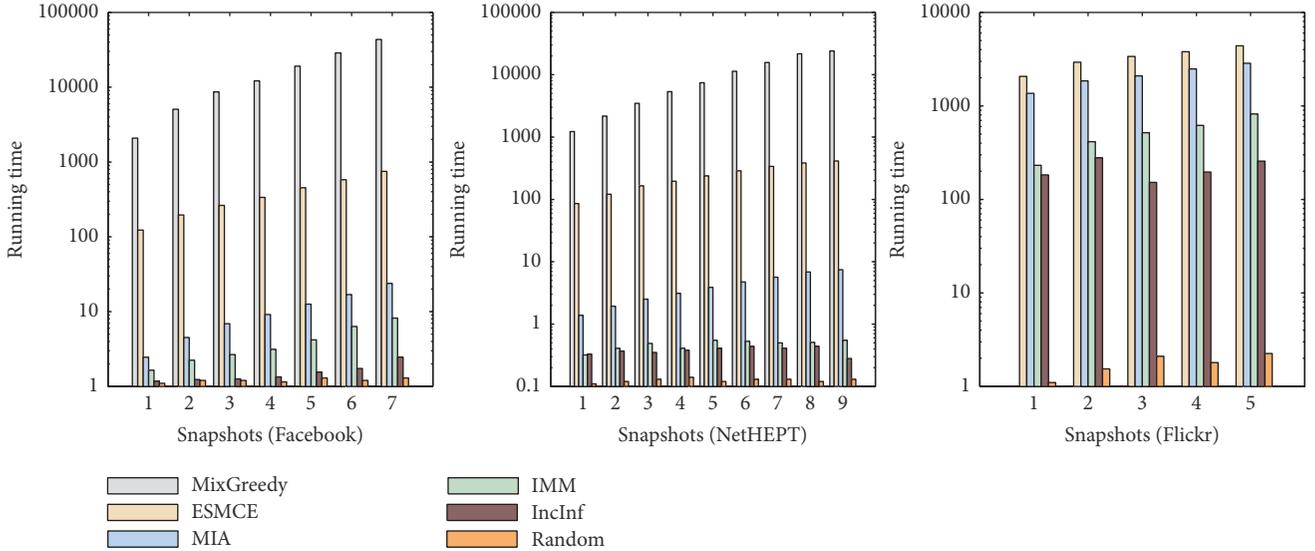


FIGURE 5: The time costs of different algorithms on three real-world datasets.

We can also observe that the running time of IncInf is not monotone like other algorithms as the time evolves. This is because the running time of IncInf is closely related to the topology change between two graph snapshots. An evident change in topology will usually lead to a relatively long running time and vice versa. Without doubt, Random runs the fastest among all the algorithms. However, as we will show in Section 6.3, its accuracy is much worse and unacceptable when developing real-world viral marketing strategies.

We also test the effect of our pruning strategy. Here we take the Facebook dataset as an example; the results on other datasets are similar and thus are omitted. Different from other experiments, we recorded the Facebook graph from September 2006 to October 2007 (14 months) as snapshot  $A$  in this experiment. After that, we take snapshots every month as snapshot  $B$ . We use IncInf to find the top- $K$  influential nodes in snapshot  $B$  based on ones in snapshot  $A$ . The result is shown in Figure 7.  $x$ -axis is the time interval between snapshots  $A$  and  $B$ , and  $y$ -axis is the ratio of the number of nodes after pruning to the total number of nodes in snapshot  $B$ . The minimum and maximum pruning ratios are 3.90% and 5.86%, respectively, with a mean ratio of 4.72% on all the 14 time intervals between snapshots  $A$  and  $B$ . This demonstrates that our pruning strategy can effectively limit the search space into a small percent of nodes. We can also see in Figure 7 that, with the increase of time interval, the ratio, although not monotone, generally becomes larger. This is mainly because a longer time interval means a larger amount of topology changes, and basically it will be possible for more nodes to become influential nodes.

**6.3. Effectiveness Study.** In this subsection, we study the influence spread of the top- $K$  influential nodes selected by our algorithm as well as other static algorithms. The influence spreads of different algorithms are measured as the number of nodes that are influenced by the top-50 influential nodes

selected. Obviously, the higher the influence spread, the better the effectiveness. We have not tested the performance of MixGreedy on the Flickr dataset as the running time is excessively long.

Figure 6 shows the experimental results. MixGreedy outperforms all the other algorithms in terms of influence spread. However, the efficiency issue limits its application to large-scale dataset such as Flickr. The performances of ESMCE, MIA, IMM, and IncInf almost match MixGreedy on the Facebook dataset, while on NetHEPT the gaps become larger but remain acceptable (only 3.4%, 4.7%, 4.5%, and 5.1% lower than MixGreedy on average). When applied to the Flickr dataset, ESMCE performs the best, since ESMCE strictly controls the error threshold by iterative sampling. The influence spread of MIA almost matches IMM on the three datasets and is slightly lower than ESMCE. Compared with MIA, IncInf shows very close performance and is only 2.87% lower on average of all five snapshots, which demonstrates the effectiveness of our proposal. Random, as the baseline heuristic, clearly performs the worst on all the graphs. Actually, the influence spread of Random is only 15.6%, 12.1%, and 10.9% of those of IncInf on Facebook, NetHEPT, and Flickr, respectively.

We shall note that the reason IncInf has slightly lower influence spread is mainly twofold. First, IncInf restricts the influence into local regions to speed up the computation of influence spread changes, which will affect the effectiveness. Second, a pruning strategy is designed to narrow down the search space based on the influence spread changes and previous top- $K$  information. Despite slight loss in effectiveness, IncInf gains remarkable improvement in efficiency, as mentioned before.

**6.4. Tuning of Parameters  $\theta$  and  $\eta$ .** First, we study how effectively the localization parameter  $\theta$  of IncInf represents a

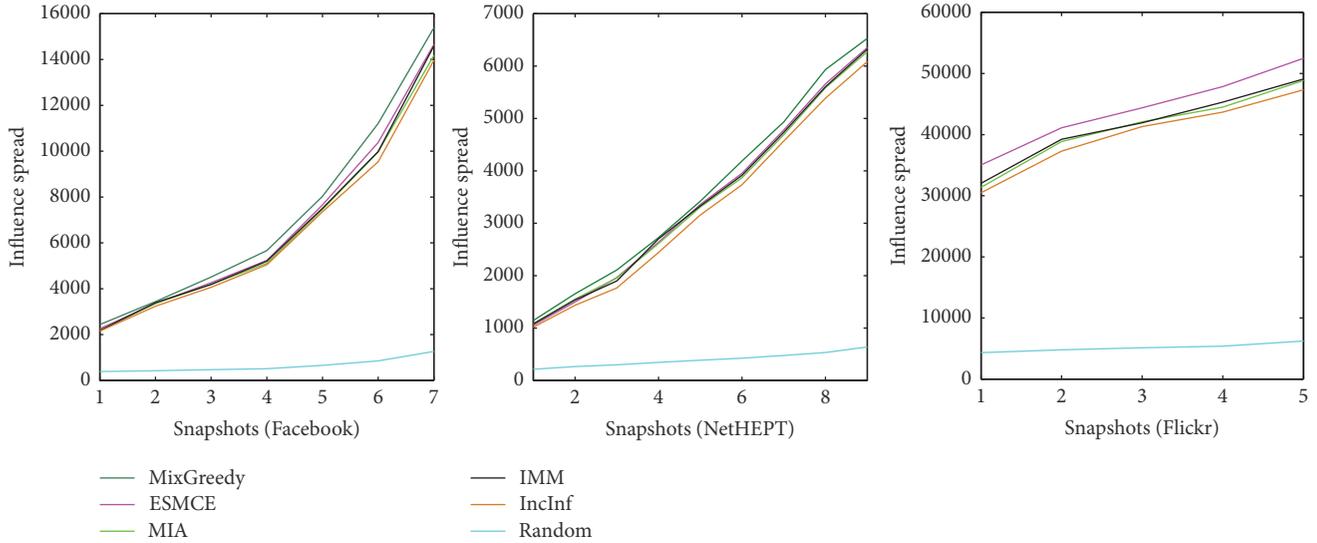


FIGURE 6: The influence spread of different algorithms on three datasets.

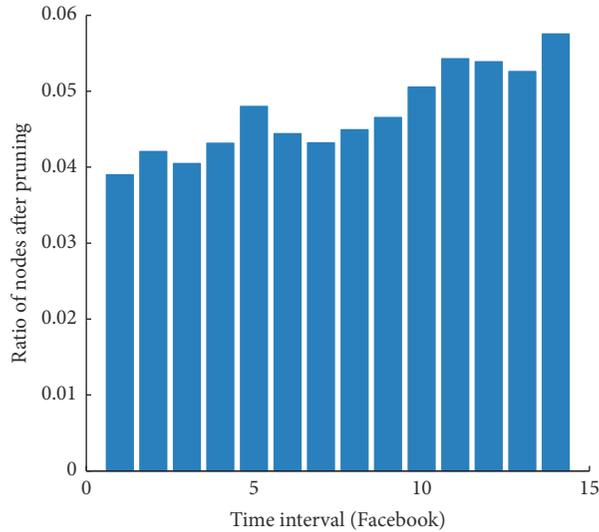


FIGURE 7: The effect of pruning strategy on the Facebook dataset.

tradeoff between efficiency and effectiveness. We run IncInf with different values of  $\theta$  on the final Facebook and NetHEPT graphs. The running time and influence spread are measured based on seed size  $K = 50$ .

The experimental results are shown in Figure 8. Note that  $x$ -axis represents the reciprocal of  $\theta$ . We observe that  $\theta$  acts as a tradeoff between efficiency and effectiveness: with the decrease of  $\theta$ , IncInf and MIA achieve better influence spread. However, this is gained at the cost of longer running time, that is, poor efficiency. For example, when we reduce  $\theta$  from  $1/200$  to  $1/500$  on the Facebook dataset, the influence spread of IncInf increases by 15.4%, while the running time is 1.12x longer. Moreover, we can observe that the influence spread of IncInf almost matches that of MIA in all values of  $\theta$ . For example, IncInf is only 1.87% lower than MIA in influence spread when  $\theta$  is set to  $1/200$  in the NetHEPT dataset. But

IncInf shows overwhelming advantages in terms of running time. When  $\theta$  is set to  $1/500$  in Facebook, IncInf needs only 5 seconds to identify the top-50 influential nodes, while it takes MIA more than 150 seconds to finish the same work. More importantly, with the decrease of  $\theta$ , the influence spread increases sharply at the beginning but the increase is no longer that significant after  $\theta$  is lowered to a certain level. On the contrary, the running time is almost linear to  $1/\theta$ . This suggests that the knee point of the influence spread curve can serve as a good tuning point of  $\theta$ , where we could obtain the best gain from both influence spread and running time.

Then, we will evaluate the sensitivity of pruning threshold  $\eta$  in terms of influence spread and running time. The results are illustrated in Figure 9. From Figure 9, we can see that, with the increase of  $\eta$ , the running time increases gently at the beginning and then turns into a sharp boost. For example,

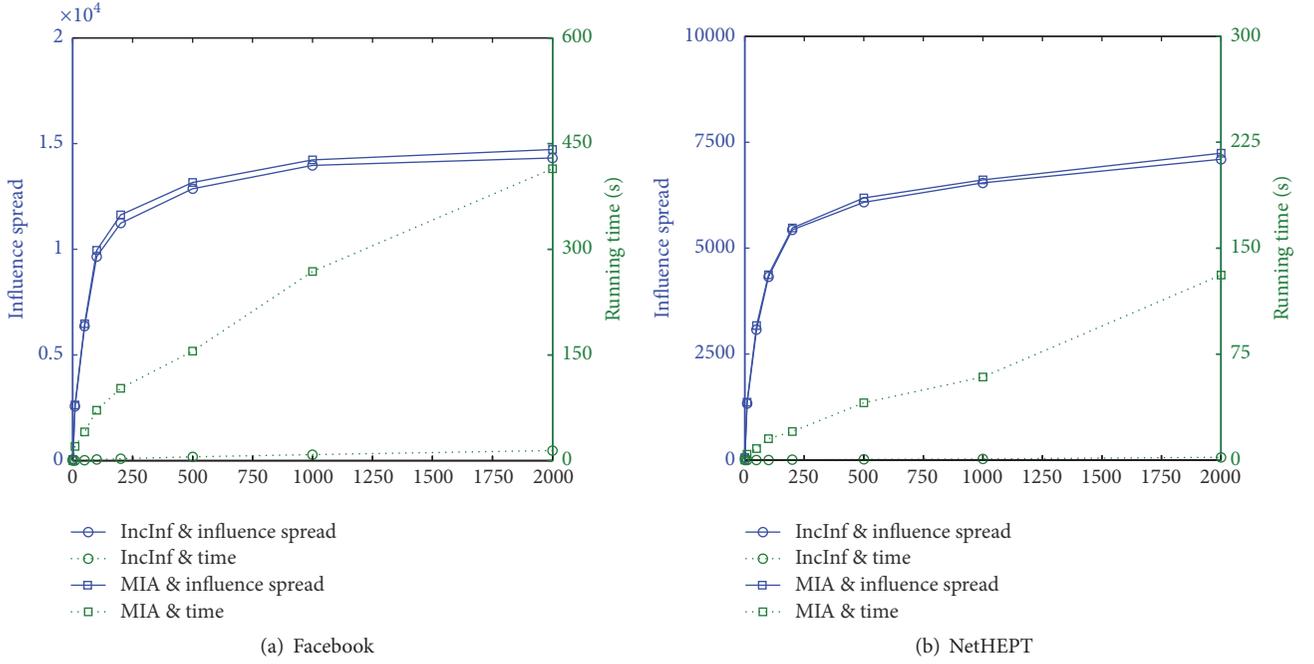


FIGURE 8: The effect of tuning of  $\theta$  on running time and influence spread.

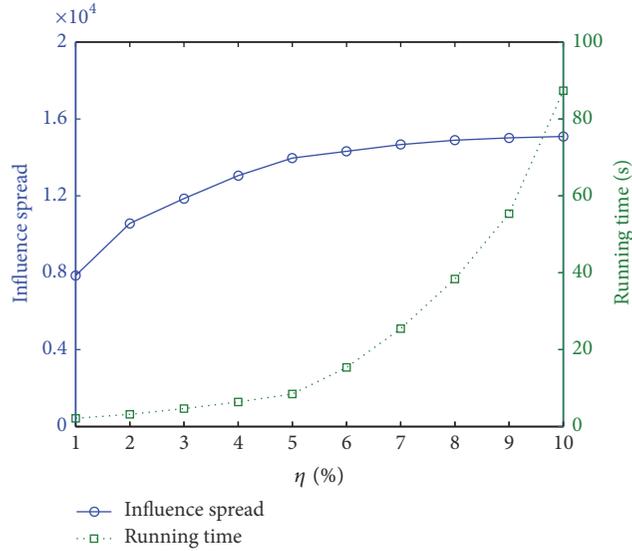


FIGURE 9: The effect of tuning of  $\eta$  on running time and influence spread.

when we increase  $\eta$  from 1% to 5%, the running time of IncInf on the Facebook dataset only increases from 2.13 s to 8.47 s, while it dramatically increases from 8.47 s to 87.35 when  $\eta$  is tuned from 5% to 10%. This phenomenon is closely related to the power-law distribution of degree in social network; when  $\eta$  is set to a large number, a relatively large number of potential nodes would be selected.

In terms of influence spread, with the increase of  $\eta$ , more nodes are selected as potential nodes, which will guarantee better influence spread. Different from the running time,

the influence spread grows rather rapidly at the beginning and then gradually slows down. The influence spread on the Facebook dataset is 7854 when  $\eta$  is set to 1% and rapidly grows to 13967 when the maximum error threshold is 5%. After that, the growth trend slows down and the influence spread is about 15091 as  $\eta$  increases to 10%. The reason to explain such phenomenon is that the top- $K$  influential nodes are mainly selected from high-degree nodes. Therefore, when  $\eta$  becomes larger, although more nodes would be selected, their contribution to influence spread is relatively small; thus the

growth trend slows down. Based on the above observation, here we suggest that 5% may stand as a good tradeoff between running time and influence spread.

**6.5. Discussions.** Experimental results demonstrate that our proposed IncInf algorithm significantly reduces the execution time of state-of-the-art static influence maximization algorithm while maintaining satisfying accuracy in terms of influence spread. Although IncInf performs better, it has a few limitations for further improvement.

First, IncInf directly depends on previous information of top- $K$  influential nodes for effective pruning, while sometimes such information is incomplete or even unavailable. We plan to study this problem later. Second, IncInf is designed for the IC model, which may somehow limit its application. But we believe that our idea of incremental computation for influence maximization could be properly extended to other influence diffusion models.

## 7. Conclusion and Future Work

In this paper, we consider the influence maximization problem in evolving social networks and propose an incremental algorithm, IncInf, to efficiently identify top- $K$  influential nodes in dynamic social networks. Taking advantage of the structural evolution of networks and previous information on individual nodes, IncInf substantially reduces the search space and adaptively selects influential nodes in an incremental way. Extensive experiments demonstrate that IncInf significantly reduces the execution time of state-of-the-art static influence maximization algorithm while maintaining satisfying accuracy in terms of influence spread.

There are several future directions for this research. First, IncInf has large potential to fit into modern parallel computing framework. This is because IncInf restricts the computation of influence spread changes into local regions, which could ease the partition of social graph for parallel computation. Moreover, the proposed pruning strategy could be effectively performed in parallel. Second, our current IncInf algorithm is derived from the basic IC model. We believe that the conception of incremental computation for influence maximization could be properly extended to other influence diffusion models, such as another classic LT model. Third, although there have been a few researches [43] about how to measure the propagation probability, this problem is not well addressed yet, especially for large-scale dynamic social networks.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

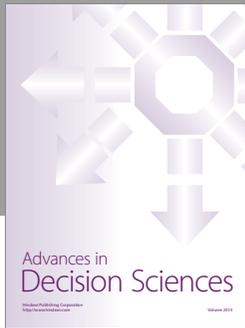
## Acknowledgments

This research was supported by NSFC under Grant no. 61402511.

## References

- [1] P. Domingos and M. Richardson, "Mining the network value of customers," in *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '01)*, pp. 57–66, ACM Press, San Francisco, Calif, USA, August 2001.
- [2] Twitter 2012 Facts and Figures, 2012, <http://www.website-monitoring.com/blog/2012/11/07/>.
- [3] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09)*, pp. 199–208, ACM, July 2009.
- [4] C. Wang, W. Chen, and Y. Wang, "Scalable influence maximization for independent cascade model in large-scale social networks," *Data Mining and Knowledge Discovery*, vol. 25, no. 3, pp. 545–576, 2012.
- [5] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. Vanbriesen, and N. Glance, "Cost-effective outbreak detection in networks," in *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '07)*, pp. 420–429, New York, NY, USA, August 2007.
- [6] X. Liu, S. Li, X. Liao, L. Wang, and Q. Wu, "In-Time Estimation for Influence Maximization in Large-Scale Social Networks," in *Proceedings of the ACM EuroSys Workshop on Social Network Systems*, pp. 1–6, Bern, Switzerland, 2012.
- [7] A.-L. Barabasi and R. Albert, "Emergence of scaling in random networks," *American Association for the Advancement of Science. Science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [8] X. Liu, X. Liao, S. Li, and B. Lin, "Towards efficient influence maximization for evolving social networks," in *Proceedings of the 18th Asia Pacific Web Conference*, pp. 232–244, 2016.
- [9] Y. Tang, Y. Shi, and X. Xiao, "Influence maximization in near-linear time: a martingale approach," in *Proceedings of the ACM SIGMOD International Conference on Management of Data, (SIGMOD '15)*, pp. 1539–1554, Melbourne, Australia, June 2015.
- [10] Y. Wang, G. Cong, G. Song, and K. Xie, "Community-based Greedy algorithm for mining top- $K$  influential nodes in mobile social networks," in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '10)*, pp. 1039–1048, ACM, July 2010.
- [11] A. Goyal, W. Lu, and L. V. S. Lakshmanan, "CELF++: optimizing the greedy algorithm for influence maximization in social networks," in *Proceedings of the 20th International Conference Companion on World Wide Web, (WWW '11)*, pp. 47–48, Hyderabad, India, April 2011.
- [12] K. Jung, W. Heo, and W. Chen, "IRIE: A Scalable Influence Maximization Algorithm for Independent Cascade Model and Its Extensions," <https://arxiv.org/abs/1111.4795>.
- [13] X. Liu, M. Li, S. Li, S. Peng, X. Liao, and X. Lu, "IMGPU: GPU-accelerated influence maximization in large-scale social networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 1, pp. 136–145, 2014.
- [14] Y.-C. Chen, W.-Y. Zhu, W.-C. Peng, W.-C. Lee, and S.-Y. Lee, "CIM: community-based influence maximization in social networks," *ACM Transactions on Intelligent Systems and Technology*, vol. 5, no. 2, article no. 25, 2014.
- [15] X.-G. Wang, "An algorithm for critical nodes problem in social networks based on owen value," *The Scientific World Journal*, vol. 2014, Article ID 414717, 2014.
- [16] Z. Lu, W. Zhang, W. Wu, J. Kim, and B. Fu, "The complexity of influence maximization problem in the deterministic linear

- threshold model,” *Journal of Combinatorial Optimization*, vol. 24, no. 3, pp. 374–378, 2012.
- [17] Z. Lu, L. Fan, W. Wu, B. Thuraisingham, and K. Yang, “Efficient influence spread estimation for influence maximization under the linear threshold model,” *Computational Social Networks*, vol. 1, no. 1, 2014.
- [18] H. Nguyen and R. Zheng, “On budgeted influence maximization in social networks,” *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 6, pp. 1084–1094, 2013.
- [19] M. Han, M. Yan, Z. Cai, and Y. Li, “An exploration of broader influence maximization in timeliness networks with opportunistic selection,” *Journal of Network and Computer Applications*, vol. 63, pp. 39–49, 2016.
- [20] B. Liu, G. Cong, Y. Zeng, D. Xu, and Y. M. Chee, “Influence spreading path and its application to the time constrained social influence maximization problem and beyond,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 8, pp. 1904–1917, 2014.
- [21] S. Pei, X. Teng, J. Shaman, F. Morone, and H. A. Makse, “Efficient collective influence maximization in cascading processes with first-order transitions,” *Scientific Reports*, vol. 7, Article ID 45240, 2017.
- [22] Y. Y. Habiba, T. Y. Berger-Wolf, and J. Saia, “Finding Spread Blockers in Dynamic Networks,” in *Advances in Social Network Mining and Analysis*, vol. 5498 of *Lecture Notes in Computer Science*, pp. 55–76, Springer, Berlin, Germany, 2010.
- [23] R. Michalski, T. Kajdanowicz, P. Bródka, and P. Kazienko, “Seed selection for spread of influence in social networks: Temporal vs. static approach,” *New Generation Computing*, vol. 32, no. 3–4, pp. 213–235, 2014.
- [24] W. Chen, W. Lu, and N. Zhang, “Time-critical influence maximization in social networks with time-delayed diffusion process,” in *Proceedings of the Proceedings of the 26th Conference on Artificial Intelligence*, Toronto, Canada, 2012.
- [25] M. Gomez-Rodriguez and B. Scholkopf, “Influence Maximization in Continuous Time Diffusion Networks,” in *Proceedings of the 29th International Conference on Machine Learning*, Scotland, UK, 2012.
- [26] C. Aggarwal, S. Lin, and P. S. Yu, “On influential node discovery in dynamic social networks,” in *Proceedings of the 12th SIAM International Conference on Data Mining (SDM '12)*, pp. 636–647, SIAM, Calif, USA, April 2012.
- [27] H. Zhuang, Y. Sun, J. Tang, J. Zhang, and X. Sun, “Influence maximization in dynamic social networks,” in *Proceedings of the 13th IEEE International Conference on Data Mining, ICDM 2013*, pp. 1313–1318, Dallas, Tex, USA, December 2013.
- [28] G. Tong, W. Wu, S. Tang, and D.-Z. Du, “Adaptive Influence Maximization in Dynamic Social Networks,” *IEEE/ACM Transactions on Networking*, vol. 25, no. 1, pp. 112–125, 2017.
- [29] D. Kempe, J. Kleinberg, and E. Tardos, “Maximizing the spread of influence through a social network,” in *Proceedings of the Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 137–146, Washington, DC, USA, 2003.
- [30] M. Richardson and P. Domingos, “Mining knowledge-sharing sites for viral marketing,” in *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '02)*, pp. 61–70, ACM, Edmonton, Canada, July 2002.
- [31] Y.-C. Chen, W.-C. Peng, and S.-Y. Lee, “Efficient algorithms for influence maximization in social networks,” *Knowledge and Information Systems*, vol. 33, no. 3, pp. 577–601, 2012.
- [32] M. Heidari, M. Asadpour, and H. Faili, “SMG: fast scalable greedy algorithm for influence maximization in social networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 420, pp. 124–133, 2015.
- [33] F. Morone and H. A. Makse, “Influence maximization in complex networks through optimal percolation,” *Nature*, vol. 524, no. 7563, pp. 65–68, 2015.
- [34] G. Song, X. Zhou, Y. Wang, and K. Xie, “Influence maximization on large-scale mobile social network: a divide-and-conquer method,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 5, pp. 1379–1392, 2015.
- [35] C. Zhou, P. Zhang, W. Zang, and L. Guo, “On the upper bounds of spread for greedy algorithms in social network influence maximization,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 10, pp. 2770–2783, 2015.
- [36] A. Capocci, V. D. P. Servedio, F. Colaiori et al., “Preferential attachment in the growth of social networks: The internet encyclopedia Wikipedia,” *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 74, no. 3, Article ID 036116, 2006.
- [37] J. Leskovec, L. Backstrom, R. Kumar, and A. Tomkins, “Microscopic evolution of social networks,” in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2008*, pp. 462–470, Las Vegas, Nev, USA, August 2008.
- [38] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*, Cambridge University Press, Cambridge, Mass, USA, 2010.
- [39] V. P. CORONA, “Memory, Monsters, and Lady Gaga,” *The Journal of Popular Culture*, vol. 46, no. 4, pp. 725–744, 2013.
- [40] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi, “On the evolution of user interaction in Facebook,” in *Proceedings of the 2nd ACM SIGCOMM Workshop on Online Social Networks*, pp. 37–42, Barcelona, Spain, August 2009.
- [41] “ArXiv NetHEPT dataset,” <http://www.cs.cornell.edu/projects/kddcup/datasets.html>.
- [42] A. Mislove, H. S. Koppula, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, “Growth of the flickr social network,” in *Proceedings of the 1st ACM SIGCOMM Workshop on Social Networks (WOSN '08)*, pp. 25–30, ACM, August 2008.
- [43] J. Tang, J. Sun, C. Wang, and Z. Yang, “Social influence analysis in large-scale networks,” in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '09*, pp. 807–815, Paris, France, July 2009.



# Hindawi

Submit your manuscripts at  
<https://www.hindawi.com>

