An Exploration of Broader Influence Maximization in Timeliness Networks with Opportunistic Selection

Meng Han, Mingyuan Yan, Zhipeng Cai, Yingshu Li*

The Department of Computer Science
Georgia State University
Atlanta, Georgia 30303

Abstract

The goal of classic influence maximization in Online Social Networks (OSNs) is to maximize the spread of influence with a fixed budget constraint, *e.g.* the size of seed nodes is pre-determined. However, most existing works on influence maximization overlooked the information timeliness. That is, these works assume the influence will not decay with time and the influence could be accepted immediately, which are not practical. Secondly, even the influence could be passed to a specific node in time, whether the influence could be delivered (influence take effect) or not is still an unknown question. Furthermore, if let the number of users who are influenced as the depth of influence and the area covered by influenced users as the breadth, most of research results are only focus on the influence depth instead of the influence breadth. Timeliness, acceptance ratio and breadth are three important factors neglected but strong affect the real result of influence maximization. In order to fill the gap, a novel algorithm that incorporates time delay for timeliness, opportunistic selection for acceptance ratio and broad diffusion for influence breadth has been investigated in this paper. In our model, the breadth of influence is measured by the number of communities, and the tradeoff between depth and breadth of influence could be balanced by a parameter φ. Empirical studies on different large real-world

*Corresponding author

Email addresses: mhan@cs.gsu.edu (Meng Han), myan2@cs.gsu.edu (Mingyuan Yan), zcai@gsu.edu (Zhipeng Cai), yli@cs.gsu.edu (Yingshu Li)

Preprint submitted to Journal of Network and Computer Applications *September 23, 2015*

social networks show that our model demonstrates that high depth influence does not necessarily imply broad information diffusion. Our model, together with its solutions, not only provides better practicality but also gives a regulatory mechanism for influence maximization as well as outperforms most of the existing classical algorithms.

Keywords: Social Networks, Influence Maximization, Algorithm, Data Mining
2015 MSC: 00-01, 99-00

1. Introduction

Each month, more than 1.3 billion users are active on Facebook, and 190 million unique visitors are active on Twitter. Furthermore, 48% of 18-34 year old Facebook users check their online personal web pages when they wake up, and 98% of 18-24 year old people are involved in at least one kind of social media[1]. Since customers are the most important foundation of business, Online Social Networks (OSNs) have become one of the most effective and efficient solutions for marketing and advertising. But there is still no specific answer for how to handle and utilize data from OSNs. The development of OSNs and the resultant of a huge volume of data bring both opportunities and computation challenges.

Influence maximization, as one of the most popular topics in OSNs, attracts a lot of interest recently. Several models have been proposed in the literatures [4, 5] to model influence diffusion. However, because of the complexity and diversity of social phenomenons, many important features have been ignored, resulting the practical influence diffusion is still not well modeled. We are facing a lot of challenges such as timeliness, acceptance ratio and breadth while analyzing and maximizing influence in OSNs. *Timeliness* refers to the phenomena that the effect of influence would decay with time; *acceptance ratio* measures the percentage of influence which gets response; and influence *breadth* is aims at

[1] http://www.statisticbrain.com/facebook-statistics/

maximizing influence not only by having more users, but also by achieving a broader user distribution in reality.

In the viral marketing and media domain, it is very common that many limited-time promotions and immediacy news exist where the influence and spreading of them decay with time. During the process of advertisement promotion or marketing strategies, the fact that a message could be passed on to someone never means the message could be accepted by the receivers (acceptance means the receivers take actions or response to the message). Therefore, receiving and accepting would be two procedures of influence. From this point of view, takeing the acceptance ratio into account for influence would make the model more practical than the traditional naive way. The expectation of the influence model traditionally formulated is considered as the depth of influence. Another important issue is how broad area the influence could be from the s-elected source seeds: the breadth of influence. Breadth relies not only on the number of influenced nodes, but also on the size of the area that could be covered by the influenced nodes. Surprisingly, although most researchers consider the path or routing of influence spreading based on network structure, as far as we know, there is not any existing work considering the range (breadth) of the influence yet. Therefore, the question appears: which one is more important for influence maximization? influence more users in depth [2] or in breadth?

Let us take a conventional social network activity as an example to discuss influence diffusion in daily life. Assume there is one user on Facebook sharing a new song or movie. This action results in an influence diffusion process. That is, friends or followers of the action initiator will have similar behaviors - be influenced. Considering one instance, user *Mike* posts a new status "*I got a new iPhone 6 plus from Apple Store with student promotion. It is awesome!*" with pictures on Facebook. All of *Mike*'s friends and followers will get this information from their Facebook's news feed or related search results. For timeliness,

[2]depth might result in "rendezvous problem", which is a term from mathematics to state the overcrowded of seeds selection

the effect of this influence will be weakened as time goes on. For acceptance ratio, obviously not all the neighbors who see the post will forward it, although some of *Mike*'s friends might have already be influenced and begun to take next step to purchase an iPhone, but some of his friends might have simply ignored this post. We consider the receiving of that post as the first step of influence, and all the users having a friend relationship with *Mike* have a probability to receive this influence. But only the neighbors who comment, forward this status, or take response action regarding this post could be considered as accepting the influence, which is the second step of the influence. For the breadth of influence, one possibility is a lot of *Mike*'s friends are studying in the same department of the same university. If we evaluate the influence ability of *Mike* in the whole social network, he might not be as good as another user *Michael* who has fewer friends studying in many different universities. Compared with *Mike*, *Michael* has a good chance to pass the influence much more broader than *Mike*. Thus, all the three aforementioned factors we mentioned above should be taken into account.

Additionally, how to evaluate influence in OSNs is still an open problem. Although several models have been proposed to evaluate, influence by analyzing the history logs [9] or learning users' behaviors [20], there is still few literatures considering the impact between users in a timeliness model with respect to the influence decaying process and the optimistic selection for a better acceptance ratio. Therefore, different from the most traditional influence models which only focus on the simple traditional influence expectation result or the efficiency of the algorithm [6, 7, 8], we deal with influence maximization from a much more practical and comprehensive perspective.

In this paper, we address the problem of identifying the node set which maximizes influence in practical social networks. Our model incorporates influence decay function, opportunistic selection and broader maximization accommodating to three factors: timeliness, acceptance ratio and breadth. More specifically, our contributions are summarized as follows:

1. We formulate the problem of influence maximization with opportunistic selection in a timeliness model *ICOT*. The model incorporates the timeliness feature and considers the decaying of influence diffusion.
2. We propose opportunistic selection to deal with the acceptance ratio which represents the real reception of influence transmission in practice.
3. We show the NP-hardness of the problem together with the monotone and submodular properties of the object function. Our model is generalizable to other influence maximization problem by using a different influence diffusion model. The analysis result shows that the classical models (e.g. *IC*) are special cases of our model.
4. Considering the coverage of influence diffusion, we take the first step to explore the relationship between the breadth and depth of influence and propose the model *BICOT*. Specifically, in the extended version of our model, we use the number of communities to measure the breadth of the influence, which is novel.
5. The experiment results on several real data sets show that our solution can significantly improve the practicability and accuracy against several baseline methods. Especially on the aspect of influence spreading range.

The rest of the paper is organized as follows. Section 2 reviews the related works. Section 3 presents the preliminaries and problem definition, then we introduce our model with analysis and the algorithm in Section 4. The evaluation results based on real and synthetic data sets are shown in Section 5. Section 6 concludes the paper.

2. Related Work

To maximize influence in OSNs, the *IC* model [4] and another threshold model *LT* together with their extensions set the foundation for most of the existing cascading algorithms. Since Kempe et al. [4] formulated the influence maximization problem as an optimization problem, a series of empirical studies have been performed on influence learning [9, 10], algorithm optimizing

[8, 11, 12], scalability promoting [13, 15], and influence of group conformity [16]. Leskovec et al. [5] modeled the outbreak detection problem and proved that the influence maximization problem is a special case of their new problem. A Cost-Effective Lazy Forward (*CELF*) scheme is proposed which uses the submodular property achieving 700 times speedup in selecting seed vertices compared with the basic greedy algorithm [4]. As indicated in [15], *CELF* still faces the serious scalability problem. Therefore, Chen et al. proposed some new heuristics algorithms based on the arborescence structure which could handle million-sized graphs. The proposed algorithm spreads influence as the greedy algorithm while is more than six orders of magnitude faster than the greedy one. In [25], the authors proposed algorithm *IRIE* where *IR* is for influence ranking and *IE* is for influence maximization in both the classical *IC* model and the extension *IC-N* model considering negative opinions [14]. They claimed that their algorithms scale better than *PMIA* [15] with up to two orders of magnitude speedup and significant savings on memory usage, while maintaining the same or even better influence.

Besides the fundamental influence maximization problem and several variants mentioned above, there are two kinds of previous works related to ours: dynamic network models and structural analysis for influence diffusion. The phenomena of time delay in influence diffusion has been explored in statistics. Timeliness concerned by us, different from time decay, emphasize more on delivery time of influence. The observation in [22] shows that the heterogeneity of human activities has important effect on influence diffusion. Thang et al. [23] modeled influence maximization by limiting the influence of nodes that are within d hops from the seeding for some constant $d \geq 1$. The authors proposed algorithm VirAds which guarantees a relative error bound of $O(1)$ when the network follows power-law. They also provided theoretical analysis to show the hardness of the model. They further extended the previous algorithm to obtain a near optimal solution within a ratio better than $O(\log n)$. Chen et al. [24] proposed the Independent Cascade model with meeting events (*IC-M*) to capture time-delay. Differently, our model not only considers the time decay and acceptance ratio of influence in dynamic networks, but also take structural breadth of a network into account. Zhuang et al. [26] consider the structure changing over a network, aiming at probing a subset of nodes in social network to estimate the actual influence diffusion process.

Wang et al. [17] tried to reduce the computation cost by dividing a network into many communities. They first run the greedy algorithm in each community and calculate the expected influence increase of each community. A dynamic programming algorithm is proposed to select the optimal community first, then the most influential nodes from each community are chosen. This process runs iteratively until the top-k influential nodes are obtained. Different from our work, they do not consider timeliness in their model. Besides, they partition a network into disjoint communities only for the purpose of reducing computation cost.

To the best of our knowledge, none of the existing approaches considers the time sensitivity of influence, acceptance ratio and both the influence spreading breadth and depth together.

3. Preliminaries and Problem Definition

Kempe et al. [4] formulated the influence maximization problem as a discrete optimization problem: given a network with a node influence probability (weight) on each edge, a node set with a fixed size is initially activated as seeds and these seeds begin to influence other nodes under a certain model. The objective is to find the optimal node set which could maximize the expected number of final active nodes. Formally, we can model a network as a directed graph $\mathcal{N} = (V, E, W)$ where V, E, W represents the vertices, edges, and weights, respectively. Let function $\delta(\cdot)$ be the expected number of active nodes at the end of the influence process. Out purpose is to identify a seed set S of size up to k which devote such S which can maximize $\delta(S)$. Denote such S as:

$$S^{\star} = \arg \max_{S \subseteq V, |S| \leq k} \delta_{IC}(S) \tag{1}$$

Table 1: Notations adopted in sections

Notation	Description
$\mathcal{G}$	A weighted directed graph
V	The vertices set
E	The edge set
W	The weights set on edges
O	The opportunistic acceptance ratio set
k	The number of influential nodes to be mined
S	The set of influential nodes
τ	The influence decaying ratio
$d_\tau(t)$	The decrease ratio of influence at time t
$f_o(\cdot)$	The information diffusion ratio for current step
$\tilde{T}_o$	Threshold of opportunistic selection ratio
$\delta_{ICOT}(\cdot)$	The objective function for $ICOT$ model
$\delta_{BICOT}(\cdot)$	The objective function for $BICOT$ model
$P_C(v)$	The percentage of communities node v influenced
$i(v)$	The initialize PageRank score for node v
φ	Tradeoff parameter for depth and breadth

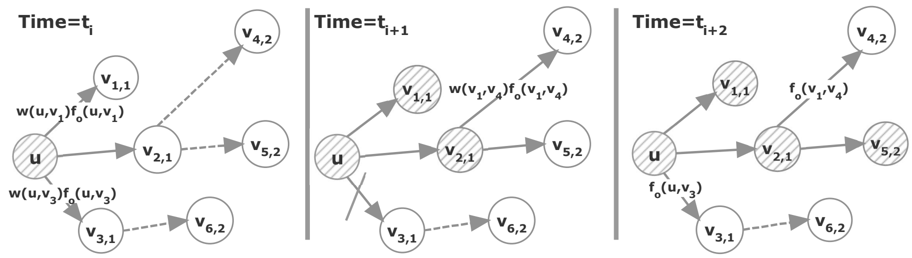

Figure 1: Models of Social Influence. (shaded circle represents an activated node, a blank circle represents an inactivated node, solid line represents an influence attempt with probability $w(u, v)f_o(u, v)$, and a dash line changes to a solid line only when the start node becomes active.)

The diffusion process under the *Independent Cascade* (*IC*) model works in discrete time $t_0, t_1, t_2, \ldots$. Initially, all the seeds in set S are activated at t_0, while all the other nodes are inactive. As the process continues to time t_i $(i > 0)$, any active u in the prior time t_{i-1} is given a single chance to active any of its currently inactivate neighbors with independent probability $w(u, v) \in W$. Once a node is activated, it stays and will not change status any more. The stochastic process iteratively continues until no new activated node appears.

The general idea behind *IC* is to measure influence ability by the number of activated nodes. It targets at finding the optimal seed set which can maximize the global influence in the network. As mentioned in Section 1, in practice, the influence diffusion process has to face opportunistic selection and time decay. Thus, function $\delta(\cdot)$ should also be improved to adapt to the reality.

We first extend the *IC* model to a dynamic network with time decay and opportunistic selection, then we propose a utility function to measure influence breadth.

Formally, we introduce our *ICOT* ($\underline{IC}$ model with $\underline{O}$pportunistic selection and $\underline{T}$ime decay) model. We define $\delta_{ICOT} : 2^V \to \mathcal{R}$ as the objective function such that $\delta_{ICOT}(S)$ with $S \subseteq V$ is the final expected number of activated nodes under *ICOT* model.

$$S^{\dagger} = \arg \max_{S \subseteq V, |S| \leq k} \delta_{ICOT}(S, o, \tau) \tag{2}$$

where o is the opportunistic acceptance ratio set controlling the acceptance of influence, and τ is the influence decaying ratio controlling the decaying process as time goes on.

The influence maximization problem with opportunistic selection under the *ICOT* model is the problem of finding the optimal seed set S with at most k seeds such that the expected number of activated nodes is maximized.

The extended version of *ICOT* is *BICOT* (***B***roadly influence maximization problem under the *ICOT* model). Different from *IC* which only maximizes the influence expectation in depth, *BICOT* considers both depth and breadth of influence. We will provide more properties and details of this model in the next section.

$$S^{\ddagger} = \arg \max_{S \subseteq V, |S| \leq k} \delta_{BICOT}(S, o, \tau, \varphi) \tag{3}$$

where φ is the parameter leveraging depth and breadth of influence.
As a summary, the two proposed models could be formolized as follows. Let M be the influence model. Our purpose is to find the optimal node set such that:

$$S^{\S} = \arg \max_{S \subseteq V, |S| \leq k} \delta_{M}(\cdot) \tag{4}$$

Problem Statement:

Input: Directed graph G, parameters (τ and $\tilde{T_o}$ for *ICOT* or α, β, ϵ, τ, $\tilde{T_o}$, and φ for *BICOT*), influence model type M (*ICOT* or *BICOT*).

Output: Optimal seed set $S^{\S}$ which maximizes influence in G under M.

4. Model Analysis and Algorithm

This section introduces the details and properties of the *ICOT* model and the *BICOT* model.

4.1. Model Analysis

We model a social network as a directed graph $\mathcal{G} = (V, E, W, O)$. We may learn the influence probability weight $w(u,v) \in W$ on each edge from practice initially. O denotes the set of opportunistic acceptance ratio functions where $f_o(u,v) \in O$ represents an independent probability indicating whether the target could accept the influence or not (in this paper we use the same weight $w(u,v)$ as an example, $f_o(u,v)$ could also be learned according to further information related to real data). $d_\tau(t)$ is a decaying function representing the decrease of influence, where t is the beginning time when only the selected seeds turn active, $t_{current}$ is the current time, and τ is the decaying coefficient.

$$d_\tau(t) = \frac{t_{current} - t}{\tau} \tag{5}$$

In *ICOT*, due to time decay and influence decrease, for each step of influence diffusion, an opportunistic acceptance function $f_o(\cdot)$ is designed to model the latest step of the information diffusion with continues time decaying.

$$f_o(u,v) = w(u,v)^{d_\tau(t)} \tag{6}$$

The acceptance ratio between nodes u and v denoted by $f_o(u,v)$ is an independent probability different from $w(u,v)$. In *ICOT*, the probability that u's influence reaches v is measured by $w(u,v)$, the opportunity whether v accepts this influence or not is decided by both $w(u,v)$ and $f_o(u,v)$. Furthermore, the final objective function is also improved to $\delta_{ICOT}(\cdot)$ which includes the weight all the active nodes try to influence their neighbors at the end (all the neighbors of the active nodes in the last step) with acceptance ratio greater or equal to threshold $\tilde{T_o}$. Those nodes will also be marked as activated according to our case study in Section 1.

Fig. 1 shows an example of influence diffusion under the *ICOT* model. Node v_{a,t_d} denotes the status of v_a in the diffusion time slot t_d. As shown in the example, at the beginning time t_i, only node u is *active* and all the links from u to its neighbors indicate the chance (attempt) of influence (solid line) from u to other nodes (e.g. v_1, v_2, and v_3). If v_1, v_2, and v_3 could be influenced (received

$(w(u, v))$ and accepted $(f_o(u, v))$ the influence) successfully, their status will change to *activate* and they continue influence others in the next step as shown by the dashed link from them. At time t_{i+1}, nodes v_1 and v_2 are influenced successfully by u, but node v_3 is not. Because link (u, v_3) is the only link between u and v_3, and v_3 does not receive the influence from u by $w(u, v_3)$ successfully. u will not try to influence v_3 by $w(u, v_3)$ anymore but will attempt to influence v_3 by $f_o(u, v_3)$ again at the end of the diffusion.

Several possibilities could be considered in mapping the decay and opportunistic selection into *ICOT* in practice. As mentioned above, user *Mike*'s promotion on Facebook for his new iPhone 6 will diffuse to all his followers, but whether and when they can be influenced and when and whether they would continue to pass this information to others are uncertain events. The decay and the opportunistic receiving selection phenomenon are very common in our daily life. Therefore, the model considers influence from both the receiving and accepting aspects is very important to capture the natural characteristics of influence diffusion in practice.

Theorem 1:. The Influence Maximization Problem under the *ICOT* model is NP-hard.

Proof:. The original influence maximization problem for the *IC* model is NP-hard. The *IC* model is a special case of the *ICOT* model with opportunistic acceptance ratio being constant 1 (without the effect of decaying function), and the threshold of opportunistic selection for the final step being constant 0. This leads to the hardness result of Theorem 1.

There are two choices: designing a heuristic algorithm which has no theoretical performance guarantee or an approximation algorithm with nice approximation ratio which can guarantee the solution results. Since influence maximization has been widely employed in OSNs, a solution results in real cost. Thus, a better accuracy leads to a better profit for a company entity. In this paper, we try to find a solution with theoretical guarantee and incorporate various optimization strategies to improve efficiency.

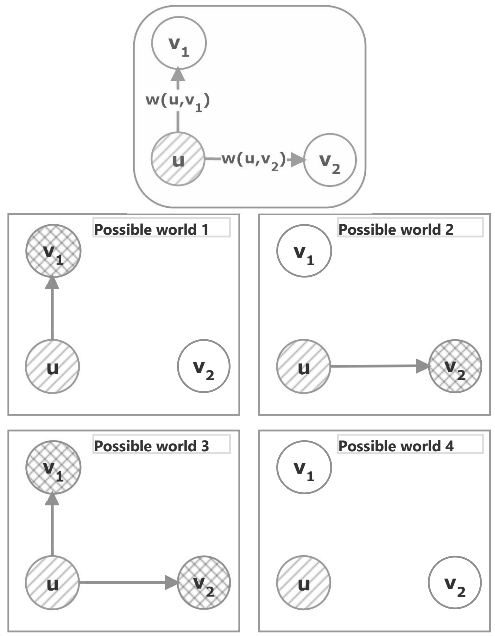

Figure 2: An instance of possible world sematic.

Given function $\delta(\cdot) : 2^V \to \mathcal{R}$, the function is *monotone* iff $\delta(S_1) \leq \delta(S_2)$ whenever $S_1 \subseteq S_2$. Also, function $\delta(\cdot)$ is *submodular* iff $\delta(S_1 + x) - \delta(S_1) \geq \delta(S_2 + x) - \delta(S_2)$ whenever $S_1 \subseteq S_2 \subset V$ and $x \in V \setminus S_2$ where V is the set of the vertices.

As shown in [4], IC model is *monotone* and *submodular* which allows us to develop a hill-climbing-style greedy algorithm to achieve $(1-1/e-\epsilon)$ approximation ratio. Since the IC model is a special case of our $ICOT$ model, the objective function of $ICOT$ can also satisfy both monotonicity and submodularity. ■

Theorem 2:. Influence function $\delta_{ICOT}(\cdot)$ is *monotone* and *submodular* under the $ICOT$ model.

Proof:. We use the "*possible worlds*" semantic to prove the theorem. As shown in Fig. 2, the top graph $< v_1, u, v_2 >$ is a small fragment of the whole network (we use $\mathcal{G}$ to denote this uncertain graph fragment) and the four graph instances are possible world semantics generated from $\mathcal{G}$. For each possible world instance, based on the weight on each edge, each instance with different generation probability could be presented as a corresponding determined graph. All the possible world instances are generated by a cascade process. We could directly assume that before the cascade process starts, the outcomes for all the opportunistic selection and time decaying process have already been determined. For each possible world W_x, the existing probability is

$$P(\mathcal{G} \Rightarrow W_x) = \prod_{e \in E(W_x)} p(e) \prod_{e \in E(\mathcal{G})) \setminus E(W_x)} (1 - p(e)) \tag{7}$$

Specifically, each cascade step could be viewed as an individual coin-flip event with probability $f_o(u, v)$ which determines if u will influence v at the corresponding time t successfully or not. Since all coin-flip events are independent, a determined set of the coin-flip events could be mapped to a *possible world* W_x. Assume there is an edge (u, v) in W_x, under the traditional IC model, without opportunistic selection and time decaying, u could directly reach v via one hop with probability 1. In the $ICOT$ model, to be more practical and accurate, u has to pass through opportunistic selection and decaying process when it tries to influence v. Since the time decaying process will not stop unless the distance between two nodes approaches to 0, it would be a limited process for opportunistic selection. On the other hand, node v is reachable from a seed set S if and only if there exists at least one path from S to v consisting of all active links (each node on the link is active). Let S_1 and S_2 be two arbitrary sets such that $S_1 \subseteq S_2 \subseteq V$. Since $\delta_{ICOT}(S)$ is the number of the nodes reachable from S in possible world W_x, if there is any node reachable from S_1, the active path will also be included in S_1's super set S_2. We can get the monotonicity of $\delta_{ICOT}(S)$.

For submodularity, based on Eq. 7, let all the probabilities related to our opportunistic selection and decaying process equal to 1. Different from IC, to take the decaying and delaying phenomenon into account, $ICOT$ tries to influence all the neighbors of activated nodes by $f_o(\cdot)$ for the last time (as accepting step) even no new activated node appears. Consider one instance of the accepting step of influence diffusion, the relationship between the number of neighbors in the last step and the number of nodes could be activated is just linear. If let the acceptance function $f_o(\cdot)$ equal to 0 at this point, IC and $ICOT$ could be unified. Considering node u reachable from $S_2 \cup \{w\}$ (w is another active node not in S_2) but not reachable from S_2, which means u is not reachable from S_1 either. Thus, w has to be the source of the active path to u, and u should be reachable from $S_1 \cup \{w\}$. For the margin increase for both S_1 and S_2, we have

$$\delta_{ICOT}(S_1 \cup \{w\}) - \delta_{ICOT}(S_1) \geq \delta_{ICOT}(S_2 \cup \{w\}) - \delta_{ICOT}(S_2) \tag{8}$$

Then consider the opportunistic selection and time decaying process, we have

$$\delta_{ICOT}(S) = \sum_{\mathcal{G} \Rightarrow W_x} Pr(W_x)\delta^{W_x}_{ICOT}(S) \tag{9}$$

Since $\delta_{ICOT}(S)$ is a nonnegative linear combination of $\delta^{W_x}_{ICOT}(S)$ which are monotone and submodular functions, $\delta_{ICOT}(S)$ keeps the same property, that is, submodular. ■

Based on the result of Nemhauser et al. [29], function $\delta(\cdot)$ suggests an approximate greedy algorithm with factor $1 - 1/e$. However, the hardness of computing $\delta(\cdot)$ for the IC model is $\#P$-hard[15]. If we apply the proof result to the $ICOT$ model, for a large scale network, even if a greedy approximate algorithm is applied by using Monte-Carlo simulations, the computation cost is still unacceptable. Considering the influence breadth, we apply a community detection algorithm [31] in the network to find different communities with overlap, then calculate the best influential k nodes taking both individual influence and global influence into account by applying a dynamic programming algorithm.

Our goal of influence maximization is to influence more nodes and larger area. In this case, besides the objective function $\delta_{ICOT}(\cdot)$, we take a further step to make influence diffusion as broad as possible.

Fig. 3 shows an example of the breadth of influence. The two circles represent two communities, and the influence is diffused according to the directed links. Assume we measure the influence by the number of outgoing links. Node v_{10} has the most outgoing links, and it should be selected in the next step based

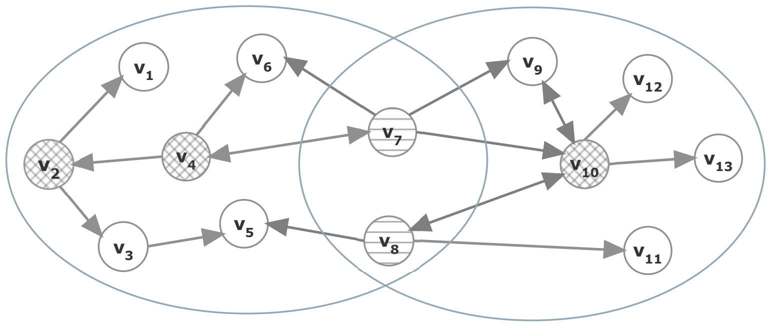

Figure 3: An example of social influence.

on the current measurement. Suppose that the algorithm has selected the best $k-1$ influential nodes including v_{10}. If v_2, v_4, and v_8 provide the same influence increase, and v_2, v_4, and v_8 all have 3 outgoing links, since v_8 connects two different communities, v_8 has significant advantages than the other two, considering the breadth of influence.

Next, we discuss the *BICOT* model. Suppose network $\mathcal{G}$ has m communities $\mathcal{C} = \{C_1, C_2, \ldots, C_m\}$. The more communities the influence could cover, the broader influence this model could achieve. We borrow the similar idea in [28] mining structural hole spanners in a network. Different from structural hole spanners which only consider the minimal value of user's importance scores in different communities, we try to find the nodes that maximize the influence globally and affect as more communities as possible. Formally, let N_c be the number of communities the algorithm could cover under *ICOT*.

Intuitively, we expect the node's individual influence in its community to be similar to its influence in the whole network. Although the gap between local community and global influential node sets exists, as the monotone we proved, the influence diffusion is built on unit node activities from local to global. The social network is strong community-based organization, and the influential node set in local from a very large extent represents the global result. We try to find the best k influential seeds in each community first, then by comparing the difference between local and global, we iteratively fill the gap by further optimization algorithms. Let $P_C(v)$ be the number of communities node v influenced divided by the number of all communities, and $S \subseteq \mathcal{C}$ denotes the subset containing more than one community, then a utility function $Q(\cdot)$ is defined for each node to measure its contribution in maximizing the influence breadth. Let $A(v, S)$ be the structural score of v in S.

$$Q(v, C_i) = \max_{e_{u,v} \in E, S \subseteq \mathcal{C} \wedge C_i \in S} \{P_C(v) Q(v, C_i),\ \alpha_i Q(v) + \beta_S A(v, S)\} \tag{10}$$

$$A(v, S) = \min_{C_i \in S} \{Q(v, C_i)\} \tag{11}$$

In Eq. 10, α_i and β_S are two tunable parameters. The contribution function $Q(\cdot)$ is computed as the combination of the importance score of v's friends and the structural score of v itself. Since $Q(\cdot)$ is the influence measurement of individual node, we use the famous PageRank [30] to initialize score $i(v)$ for each node v in each community, then continue the iteration until the converge based on the two reinforce Eq. 10 and Eq.11 stable. Same as [28], for all the node v not belongs to community C_i, we set their influential score to 0, that is:

$$\begin{aligned} Q(v, C_i) =& i(v), v \in C_i \\ Q(v, C_i) =& 0, v \notin C_i \end{aligned} \tag{12}$$

Theorem 3:. For α_i and β_S, the function scores of $Q(v, C_i)$ and $A(v, S)$ exist for any graph if and only if,

$$\max_{C_i \in S} \{\alpha_i + \beta_S\} \leq P_C(v) \tag{13}$$

Proof:. Suppose community $C_i \in \mathcal{C}$ and $C_i \in S$ such that $\alpha_i + \beta_S > P_C(v)$. Considering nodes v_1 and v_2 which connected to each other with the PageRank score $i(v_1) = i(v_2) = 1$, where $v_1 \in \cap_{C_j \in S} C_j$ and $v_2 \in C_i$. We have $Q(v_1, C_i) = P_C(v_1)$. Then by Eq. 11, $A(v_1, S) = \min_{C_i \in S}\{Q(v_1, C_i)\} = P_C(v_1)$. According to Eq. 10, $P_C(v_1) Q(v_2, C_i) \geq \alpha_i Q(v_1, C_i) + \beta_S A(v_1, S) = P_C(v_1)(\alpha_i + \beta_S) >$

$P_C(v_1)$, which means product of two positive fraction is larger than one of the fractions, which is impossible.

For the if direction, $\{\alpha_i + \beta_S\} \leq P_C(v)$. Suppose in the first iteration $Q^0(v, C_i)P_C(v) \leq P_C(v)$ and k-th iteration later $Q^k(v, C_i)P_C(v) \leq i(v)P_C(v) \leq P_C(v)$. In the $(k+1)$-th iteration, for each $C_i \in S$, we have $Q^{k+1}(v, C_i)P_C(v) \leq \alpha_i Q^k(u, C_i) + \beta_S A^k(u, S) \leq P_C(v_1)$. ■

We narrow the bound of the result in [28] α and β from $\{\alpha_i + \beta_S\} \leq 1$ to $\{\alpha_i + \beta_S\} \leq P_C(v)$. We also improve the performance of the *ICOT* model by incorporating the number of communities which can be globally covered by one node.

As shown in Algorithm 1, through finite iterations we can get a rank of all the nodes based on their own ability to influence others within their communities. By the configuration of parameters α and β, we can control the balance of influence depth and influence breadth. Let $r(v, C_i)$ be the rank of node v in community C_i, and $Rank(v, C_i)$ be the rank of node v in the network.

$$Rank(v) = \frac{\sum \frac{r(v,C_i)}{|C_i|}}{\text{Number of communities involving } v} \times 100\% \tag{14}$$

By Eq. 14, we assign a percentage value $Rank(v)$ with a control parameter φ to each node v, and calculate the influence spreading process on each edge by $\varphi Rank(v)w(\cdot)$. Thus, we can conclude our *BICOT* shown in Eq. 3.

4.2. Algorithm

The difference between *ICOT* and *BICOT* is whether taking breadth as a measurement for influence. Besides breadth, we adopt heuristic strategies in [6] in terms of a dynamic programming algorithm for both models. First, we detect communities in a network allowing overlap between different communities. Second, Algorithm 1 is applied to get the rank of each node. Through parameter φ, we control the balance of breadth and depth. Then, consider the updated weight of each node. We incorporate the strategies in [6] to model to find the seed set.

Algorithm 1: Iteration algorithm

Input: Graph G, α_i, β_S, and convergence threshold ϵ

Output: Function convergence result $Q(v, C_i), A(v, S)$

Initialize $Q(v, C_i)$ according to Eq. 12

while $\max |Q'(v, C_i) - Q(v, C_i)| \geq \epsilon$ **do**

  for $v \in V$ **do**

    for $C_i \in \mathcal{G}$ **do**

      $t(v, C_i) = \max_{C_i \in S}\{\beta_S A(v, S) + \alpha_i Q(v)\}$

    end

    if $u \in N(v)$ *&* $t(u, \cdot) \neq t'(u, \cdot)$ **then**

      `/*` t' `is the previous value of t which monitors the change of` v`'s neighbors */`

      for $v \in V$ **do**

        for $C_i \in \mathcal{G}$ **do**

          $Q'(v, C_i) = \max_{C_i \in S}\{P_C(v)Q(v, C_i), \max\{t(v, C_i)\}\}$

        end

        for $C_i \in \mathcal{G}$ **do**

          $A'(v, S) = \min_{C_i \in S}\{Q(v, C_i)\}$

        end

      end

    end

  end

  Update $Q = Q'$ and $A = A'$

end

In [6], Chen et al. designed a heuristic strategy which builds a tree-like structure for influence. Then influence spreading path is maximized through a greedy algorithm. We use the same idea, but our model considers the opportunistic selection and influence ability decrease over time. When calculating and finding the seeds which have the largest incremental result in $ICOT$ and $BICOT$, if the margin increases less than or equal to $\tilde{T_o}$, we regard this path as disconnected. The algorithm for $BICOT$ is shown as follows:

Algorithm 2: Algorithm for model $BICOT$

Input: Graph G, α_i, β_S, ϵ, φ, τ and $\tilde{T_o}$

Output: Seed set for maximizing influence $S^\dagger$

1 Do community detection by Algorithm 1 from;

2 Algorithm by $1(\alpha_i,\, \beta_S,\, \epsilon)$ to get the value of $Rank(\cdot)$ by Eq. 14 for each node;

3 By parameter φ with Eq. 14 to control the tradeoff between influence breadth and depth;

4 Calculate the influence maximization seed set based on the $BICOT$ model with parameters τ and $\tilde{T_o}$;

For model $ICOT$, we only consider the opportunistic selection and time delay, reducing the step for calculating the influence breadth for each node (Lines 2, and 3 in Algorithm 2). Then the seed finding process does not need to be incorporated with Eq. 14. The detailed algorithm is ignored due to space limitation.

5. Empirical Evaluations

We perform the experiments forwards the following data sets.

Table 2: Amazon Dataset

Data	Nodes	Edges	Diameter
Amazon0302 (A1)	262111	1234877	29
Amazon0312 (A2)	400727	3200440	18
Amazon0505 (A3)	410236	3356824	21
Amazon0601 (A4)	403394	3387388	21

5.1. Data and Observations

Epinions [3] is a Who-trust-whom network, where nodes are members of the web site and a directed edge from user u to v means u has influence to v (v trusts u). The network includes 75,879 nodes and 508,837 edges.

Twitter [4] is one of most notable micro-blogging services. Twitters can publish tweets. We use the dataset obtained from [18]. The subnetwork includes 112,044 nodes (users of Twitter), and 468,238 edges (following relationships) and 2,409,768 tweets posted by them.

Inventor is a network of inventors, obtained from [19] extracted from USPTO[5]. The network consists of 2,445,351 nodes and 5,841,940 edges (co-inventing relationships).

Amazon Dynamic Networks

Table 2 is derived from the *Customers Who Bought This Item Also Bought* feature of the Amazon website. The four networks are from March to May in 2003. Connection is established in a network from i to j if product i is frequently co-purchased with product j [5].

Fig. 4 shows the average degree of all the seven data sets. The probability on each edge is learned from the networks in later time, which means the probabilities of the first network come from the second one, and the probabilities of the last network come from the first three networks based on the linear predic-

[3] http://www.epinions.com/
[4] http://www.twitter.com
[5] http://www.uspto.gov/

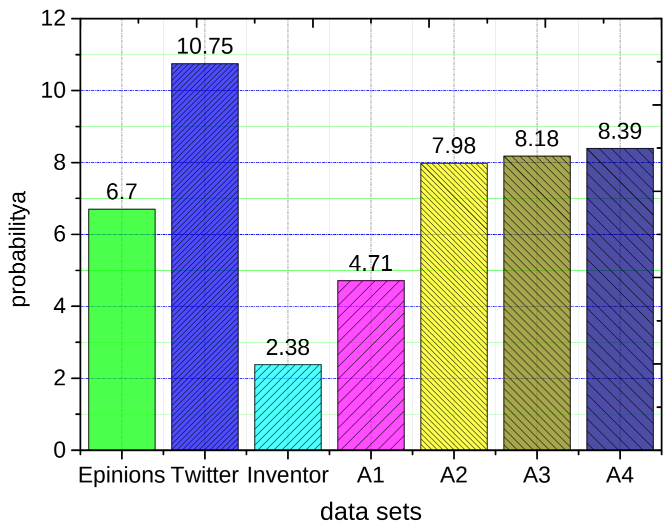

Figure 4: Average degree of data sets.

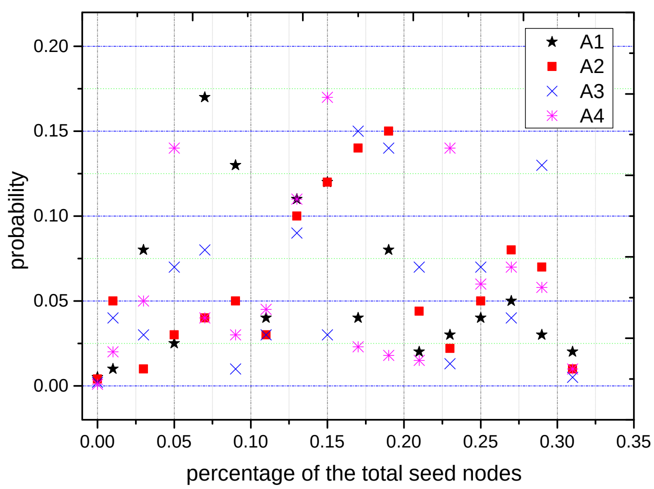

Figure 5: Probability distribution of 4 Amazon networks.

tion. The probability distribution of the four networks from Amazon is shown in Fig. 5. As shown, the probability distribution of 4 Amazon networks are mainly in range of 0.02-0.05. The reason of this range is the social characters of the relationship based on co-purchased network. And this probability dis-

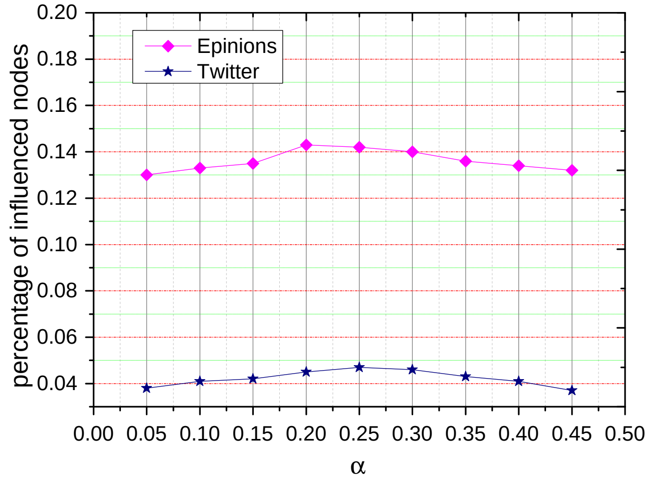

Figure 6: Effect of α for influence diffusion

tribution also shows that the Amazon co-purchased are overall loose networks. Most research literatures assume that the probabilities or the weights on links and the thresholds are given. However, as pointed out by Goyal et al.[9], learning those probabilities and thresholds is a non trivial problem. Therefore, we use a learning algorithm on the raw input data [27] to get the balance between complexity and practicability. For the Amazon data set, since there are a series of snapshots of the networks, we generate the real influence spreading trend by comparing our model to the real learning algorithm [21] which initially treats the data as a user log then solves the influence maximization problem.

All the codes are implemented in C++, and all the experiments are performed on a PC running Ubuntu 14.04 LTS with Intel(R)2 Quad CPU 2.83GHz and 6GB memory.

We examine how the parameters affect influence spread in Algorithm 1. As shown in Fig. 6 and Fig. 7, the performance of Algorithm 1 is insensitive to the variation of α and β. Consider the difference between two networks Epinions and Twitters, the average degree is 6.7 for Epinions, and 4.17 for Twitter. Thus, the main factor affect parameter α and β is the sparsity of the network.

We first evaluate the number of the influenced nodes under different models.

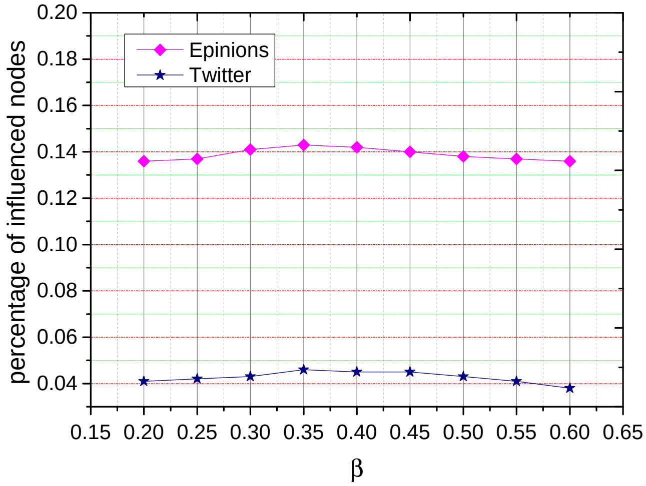

Figure 7: Effect of β for influence diffusion

As shown in Fig. 8, Fig. 9, and Fig. 10, we compare the traditional IC model [15] with our two models on the three static networks. From the three plots, we can see that the proposed model on the static network show very similar trend like the traditional *IC* model. Our models consider the optimistic selection and time decaying. We also proposed a method to calculate the final influence expectation which include more nodes when the influence spread process ends. We set the default value of $\tau = 0.5$ giving the influence breadth and depth the same weights.

To show our contributions in a convincing way, we compare our model with the up-to-date experiment based algorithm in [21] on the aspect of the real influence spread. We run our algorithm on the first Amazon co-purchase network, and run Goyal's algorithm called *CD* based on the four networks since their algorithm requires users' log. Meanwhile, we compare with traditional *IC* model towards on Amazon network 1 and network 4. As shown in Fig. 11, although all the curves follow similar trends, for a larger k, *CD* which is based on learning has slower increase which is more practical since it learns the knowledge from four data sets. Apparently, our models are more approximate to model *CD*

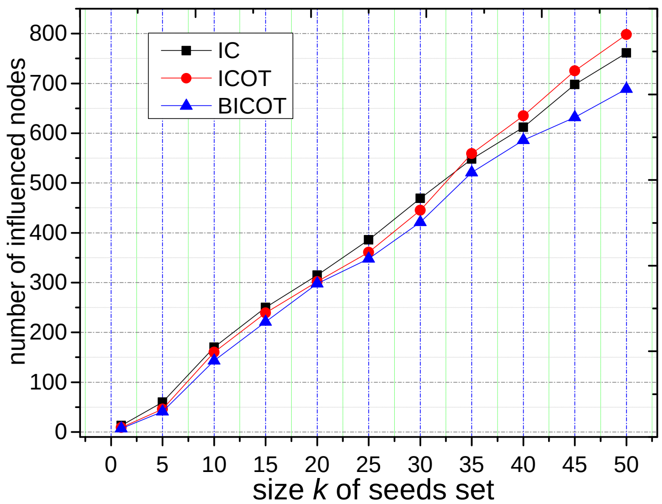

Figure 8: IC VS ICOT VS BICOT in Epinions.

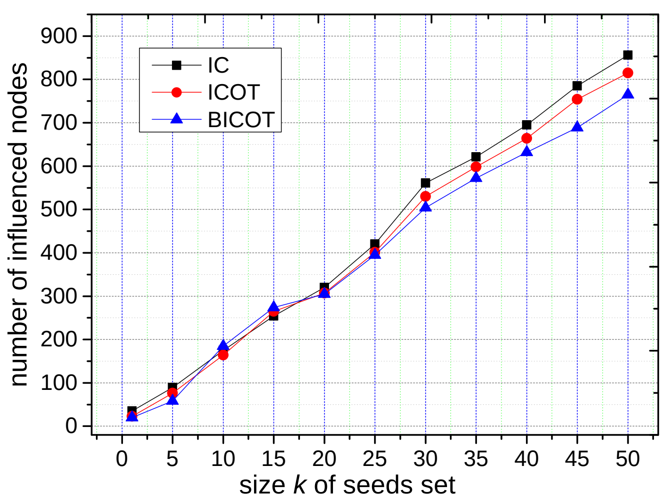

Figure 9: IC VS ICOT VS BICOT in Twitter.

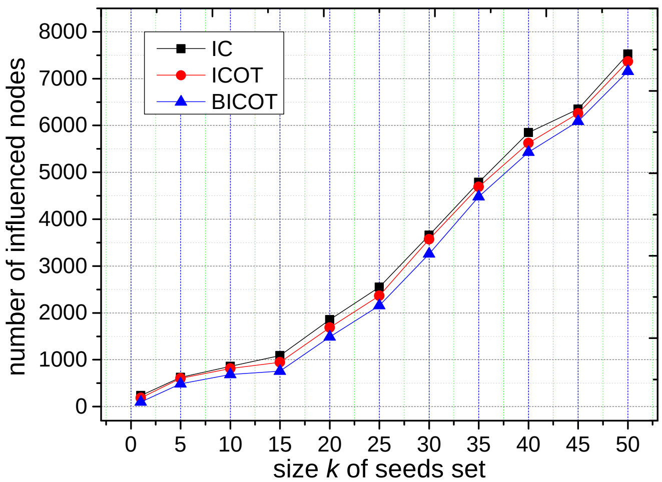

Figure 10: IC VS ICOT VS BICOT in Inventor.

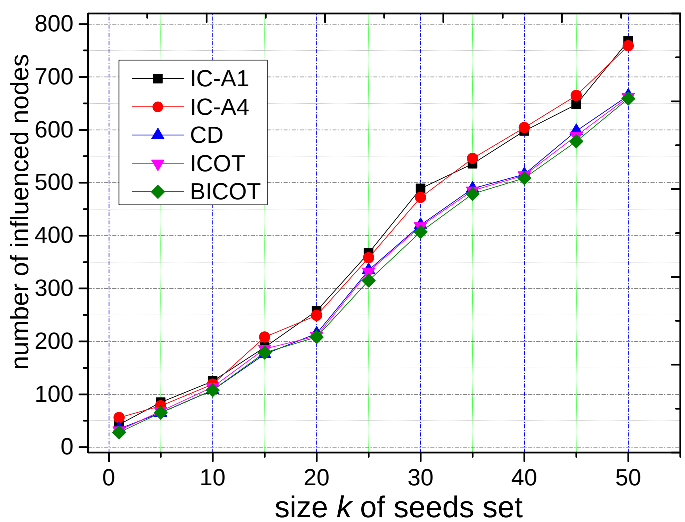

Figure 11: Influence spread by different algorithms

which means that our models are closer to the influence spreading in practice. Contrast to Fig. 11, Fig. 12 shows the number of the communities covered by each algorithms. Obviously, our *BICOT* covers much more communities than the *IC* and *CD*. The advantage of our model is as well as we have a similar

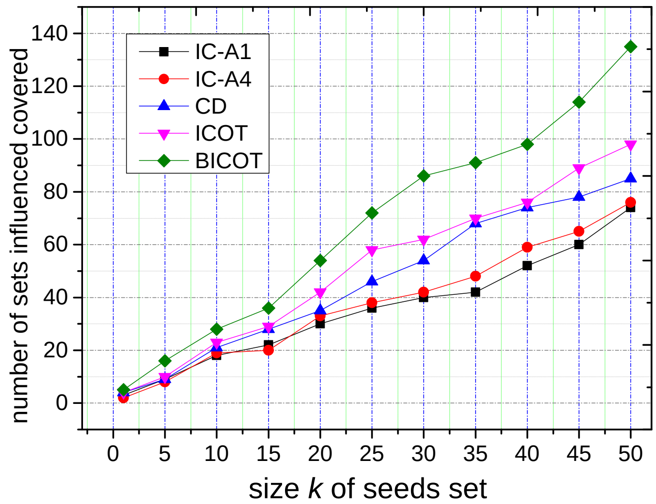

Figure 12: Communities covered by different algorithms

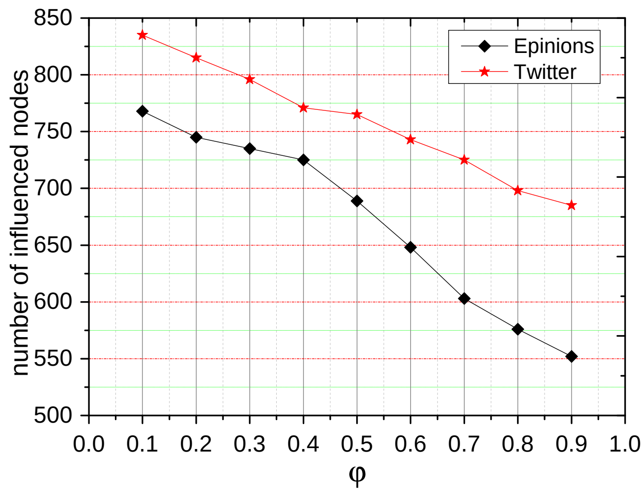

Figure 13: Influence performances for different φ

result of influence maximization follow the real diffusion, community-based algorithm give a much better efficiency to the influence maximization problem. Further more, our model cover more communities indicating a broader influence diffusion.

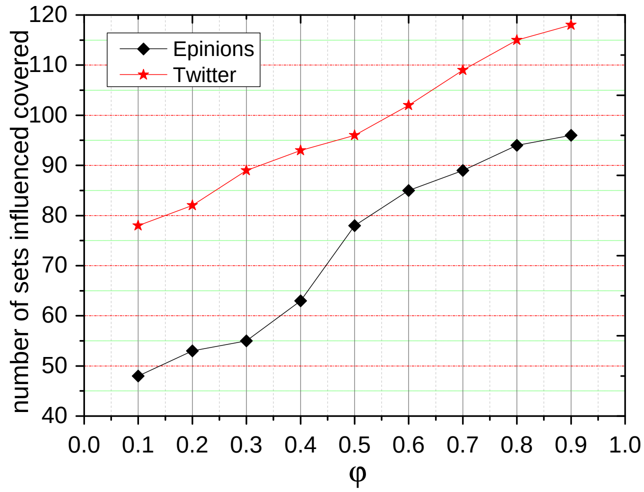

Figure 14: Communities covered for different φ

To evaluate the relationship between influence depth and breadth, we change parameter φ from 0.1 which cares more about influence depth to 0.9 which emphasizes more on the breadth.

Fig. 13 shows the influence spread for different φ. We can see that as φ increases, the influence is decreased. This is because by the definition of our objective function, we care more about breadth than depth. With the same parameter setting, we can derive from Fig. 14 that although the influence spread has been reduced, the number of the communities covered by our algorithm is increased.

6. Conclusion

In this work, based on the observations from real data and application, we propose model *ICOT* which incorporates both diffusion decay and opportunistic acceptance selection for dynamic networks. In addition, we develop model *BICOT* to control the balance between influence depth and breadth. We take the first step to explore the potential of broad influence maximization. Through comprehensive experiments results, we show that our model can achieve a comparable influence diffusion result like the learning-based algorithm which has a more strict input requirement, and our models have a broader influence coverage.

References

[1] Alzahrani, Bander A., et al. Scalability of information centric networking using mediated topology management. Journal of Network and Computer Applications (2015) 126–133.

[2] Xu, Mingwei, et al. SIONA: a service and information oriented network architecture. Journal of Network and Computer Applications (2014) 80–91.

[3] Liu, Jieyao, et al. "Application partitioning algorithms in mobile cloud computing: Taxonomy, review and future directions." Journal of Network and Computer Applications 48 (2015) 99-117.

[4] D. Kempe, J. Kleinberg, V. Tardos, Maximizing the spread of influence through a social network, in: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, Washington, D.C., 2003, pp. 137–146.

[5] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, N. Glance, Cost-effective outbreak detection in networks, in: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, San Jose, California, USA, 2007, pp. 420–429.

[6] W. Chen, Y. Wang, S. Yang, Efficient influence maximization in social networks, in: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining,ACM, Paris, France, 2009, pp. 199–208.

[7] A. Goyal, F. Bonchi, L. V. Lakshmanan, S. Venkatasubramanian, On minimizing budget and time in influence propagation over social networks, Social Network Analysis and Mining (2013) 1–14.

[8] J. Tang, J. Sun, C. Wang, Z. Yang, Social influence analysis in large-scale networks, in: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, Paris, France, 2009, pp. 807–816.

[9] A. Goyal, F. Bonchi, L. V. S. Lakshmanan, Learning influence probabilities in social networks, in: Proceedings of the third ACM international conference on Web search and data mining, ACM, New York, New York, USA, 2010, pp. 241–250.

[10] K. Saito, M. Kimura, K. Ohara, H. Motoda, Learning asynchronous-time information diffusion models and its application to behavioral data analysis over social networks.

[11] A. Goyal, W. Lu, 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, ACM, Hyderabad, India, 2011, pp. 47–48.

[12] A. Goyal, F. Bonchi, L. V. Lakshmanan, S. Venkatasubramanian, Approximation analysis of influence spread in social networks, 2005.

[13] C. Wang, W. Chen, Y. Wang, Scalable influence maximization for independent cascade model in large-scale social networks, Data Mining and Knowledge Discovery 25 (3) (2012) 545–576.

[14] W. Chen, A. Collins, R. Cummings, T. Ke, Z. Liu, D. Rincon, X. Sun, Y. Wang, W. Wei, Y. Yuan, Influence maximization in social networks when negative opinions may emerge and propagate., 2011, pp. 379–390.

[15] W. Chen, C. Wang, Y. Wang, Scalable influence maximization for prevalent viral marketing in large-scale social networks, in: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, Washington, DC, USA, 2010, pp. 1029–1038.

[16] J. Tang, S. Wu, J. Sun, Confluence: conformity influence in large social networks, in: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, Chicago, Illinois, USA, 2013, pp. 347–355.

[17] Y. Wang, G. Cong, G. Song, K. Xie, Community-based greedy algorithm for mining top-k influential nodes in mobile social networks, in: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, Washington, DC, USA, 2010, pp. 1039–1048.

[18] J. Hopcroft, T. Lou, J. Tang, Who will follow you back?: reciprocal relationship prediction, 2011, pp. 1137–1146.

[19] J. Tang, B. Wang, Y. Yang, P. Hu, Y. Zhao, X. Yan, B. Gao, M. Huang, P. Xu, W. Li, Others, Patentminer: topic-driven patent analysis and mining, 2012, pp. 1366–1374.

[20] J. Zhang, B. Liu, J. Tang, T. Chen, J. Li, Social influence locality for modeling retweeting behaviors, in: Proceedings of the Twenty-Third international joint conference on Artificial Intelligence, AAAI Press, Beijing, China, 2013, pp. 2761–2767.

[21] A. Goyal, F. Bonchi, L. V. S. Lakshmanan, A data-based approach to social influence maximization, Proc. VLDB Endow. 5 (1) (2011) 73–84.

[22] J. E. L. Iribarren, E. Moro, Impact of human activity patterns on the dynamics of information diffusion, Physical review letters 103 (3) (2009).

[23] T. N. Dinh, D. T. Nguyen, M. T. Thai, Cheap, easy, and massively effective viral marketing in social networks: truth or fiction?, in: Proceedings of the 23rd ACM conference on Hypertext and social media, ACM, Milwaukee, Wisconsin, USA, 2012, pp. 165–174.

[24] W. Chen, W. Lu, N. Zhang, Time-critical influence maximization in social networks with time-delayed diffusion process, 2012.

[25] K. Jung, W. Heo, W. Chen, Irie: A scalable influence maximization algorithm for independent cascade model and its extensions.

[26] H. Zhuang, Y. Sun, J. Tang, J. Zhang, X. Sun, Influence maximization in dynamic social networks, 2013, pp. 1313–1318.

[27] K. Saito, R. Nakano, M. Kimura, Prediction of information diffusion probabilities for independent cascade model, 2008, pp. 67–75.

[28] T. Lou, J. Tang, Mining structural hole spanners through information diffusion in social networks, 2013, pp. 825–836.

[29] G. L. Nemhauser, L. A. Wolsey, M. L. Fisher, An analysis of approximations for maximizing submodular set functionsi, Mathematical Programming 14 (1) (1978) 265–294.

[30] L. Page, S. Brin, R. Motwani, T. Winograd, The pagerank citation ranking: Bringing order to the web.

[31] K. Macropol, A. Singh, Scalable discovery of best clusters on large graphs, Proceedings of the VLDB Endowment 3 (1-2) (2010) 693–702.