Community-Based Acceptance Probability Maximization for Target Users on Social Networks

Ruidong Yan[1], Yuqing Zhu[2], Deying Li[1(✉)], and Yongcai Wang[1]

[1] School of Information, Renmin University of China, Beijing 100872, China
{yanruidong,deyingli,ycw}@ruc.edu.cn

[2] Department of Computer Science, California State University at Los Angeles, Los Angeles, CA 90032, USA
yuqing.zhu@calstatela.edu

Abstract. Social influence problems, such as *Influence Maximization* (IM), have been widely studied. But a key challenge remains: How does a company select a small size seed set such that the acceptance probability of target users is maximized? In this paper, we first propose the *Acceptance Probability Maximization* (APM) problem, i.e., selecting a small size seed set S such that the acceptance probability of target users T is maximized. Then we use classical *Independent Cascade* (IC) model as basic information diffusion model. Based on this model, we prove that APM is NP-hard and the objective function is monotone non-decreasing as well as submodular. Considering community structure of social networks, we transform APM to *Maximum Weight Hitting Set* (MWHS) problem. Next, we develop a pipage rounding algorithm whose approximation ratio is $(1 - 1/e)$. Finally, we evaluate our algorithms by simulations on real-life social networks. Experimental results validate the performance of the proposed algorithm.

Keywords: Social influence · Community structure
Seed selection · Submodularity · Approximate algorithm

1 Introduction

In recent years, with the rapid development of the internet and computer technology, some significant social networks have been widely integrated into our daily life, such as Facebook, Twitter and Google+. These online social networks have become significant platforms for disseminating useful content such as news, ideas, opinions, innovations, interests, etc. In viral market, *Influence Maximization* (IM) has been extensively studied. This research has been found useful in market recommendations through the powerful word-of-mouth effect in social networks. Specifically, a company launches a kind novel product and wants to market it by social network. Due to limited budget, it can only choose a small number of initial clients (seeds) to use it (by giving them free samples). The

S. Tang et al. (Eds.): AAIM 2018, LNCS 11343, pp. 293–305, 2018.
https://doi.org/10.1007/978-3-030-04618-7_24

company hopes that these initial clients like this product and recommend it to their friends on the social network. Similarly, their friends influence their friends of friends and so on. Finally, the company wants to maximize the number of clients who adopt the products.

However, in some scenarios, one may consider the maximizing acceptance probability of target users. Specifically, assume each user on social network has a potential value for a company. The company pay more attention to the users with higher potential value. We call these higher value users as *target users* (Selecting them as seeds requires very high cost since they are influential and authoritative). Intuitively, the company will benefit a great deal if it can maximize acceptance probability of target users. In this situation, the company aims at finding an optimal seed set within a budget such that the sum of acceptance probability of target users is maximized. We call this problem as *Acceptance Probability Maximization* (APM). It's obvious that IM problem is different from APM problem. The former selects a seed set from all nodes in network within a budget such that the expected number of nodes influenced by seed set through information diffusion is maximized. However, the latter selects a seed set from all nodes except target nodes within a budget such that the sum of acceptance probability of target users is maximized.

In fact, this problem is challenging. Intuitively, it should select global influential users or target users' neighbors as seeds. However, this intuitive choice may not be effective: (1) Target users are far away from global influential nodes. The influence from global influential nodes is less than it from local influential nodes that are close to target users. (2) Although target users' neighbors have highly influence on target users, it is impossible to choose all the neighbors of target users as seeds with restriction of small size seed set. Considering these two points, we should focus on local (community) influential nodes. Further, APM can be applied to most applications, such as personalized services, targeted advertising, targeted information dissemination, recommendation system, etc.

To the best of our knowledge, only a few studies focus on APM problem even though it plays an essential role in viral marketing. The similar studies have been done, such as [6,13]. Guo et al. [6] propose a problem to find the top-k most influential nodes to a given user. They develop a simple greedy algorithm. We expand their work and solve APM from different perspective. In [13], Yang et al. advocate recommendation support for active friending, where a user actively specifies a friending target. In other worlds, to maximize the probability that the friending target would accept an invitation from the source user. The difference between APM and previous works are: (1) Instead of [6,13], APM has multiple target users; (2) APM requires the acceptance probability instead of expected number of influenced nodes. We summarize main contributions as follows:

- We propose the *Acceptance Probability Maximization* (APM) problem and prove it's NP-hard. And we show that computing APM is $\#p - hard$.
- We prove objective function is monotone non-decreasing and submodular.

– Considering community structure of social networks, we transform APM to *Maximum Weight Hitting Set* (MWHS) problem. Then we propose a pipage rounding algorithm for APM and prove approximation ratio is $(1 - 1/e)$.
– We run the proposed algorithm and compare with other existing methods.

The rest of this paper is organized as follows. In Sect. 2, we introduce related work. In Sect. 3, influence diffusion model is presented. In Sect. 4, we state problem description. In Sect. 5, we show the properties of objective function. Algorithm is designed in Sect. 6. The experiment results are shown in Sect. 7. We draw our conclusions in Sect. 8.

2 Related Work

Kempe et al. [8] model viral marketing as a discrete optimization problem, which is named *Influence Maximization* (IM). They propose a greedy algorithm with $(1-1/e)$-performance ratio since the function is submodular under *Independent Cascade* (IC) or *Linear Threshold* (LT) model. Previous researches without target users, which cannot be directly transplanted to APM, such as [9]. In [9], Kuhnle et al. consider *Threshold Activation Problem* (TAP) which finds a minimum size set triggering expected activation of at a certain threshold. They exploit the bicriteria nature of solutions to TAP and control the running time by a parameter.

The related work involves the target users such as [2,10,11,14]. In [14], Zhou et al. study a new problem: Give an activatable set A and a targeted set T, finding the k nodes in A with the maximal influence in T. They give a greedy algorithm with guarantee of $(1 - 1/e)$. In [10], Song et al. formalize the problem targeted influence maximization in social networks and adopt a login model where each user is associated with a login probability and he can be influenced by his neighbors only when he is online. Moreover, they develop a sampling based algorithm that returns a $(1-1/e-\varepsilon)$-approximate solution. In [11], Temitope et al. extend the fundamental *Influence Maximisation* (IM) problem with respect to a set of target users on a social network. In doing so, they formulate the *Minimal Influencer for Target Users* (MITU) problem and compare with state of the art algorithms. Unfortunately, they don't have any theoretical analysis. In [2], Chang et al. study a novel problem: Given a period of promotion time and a set of target users, each of which can be activated by its neighbors multiple times, they aim at maximizing the total acceptance frequency of these target users by initially selecting k most influential seeds. They propose a generalized diffusion model called the *Multiple Independence Cascade* (MIC) and a greedy algorithm for solving this problem.

3 Influence Diffusion Model

We briefly introduce influence diffusion model: *Independent Cascade* (IC) model. Given a social network $G = (V, E, w)$, where V represents node set, $E \subseteq V \times V$

represents edge set, and w_{uv} of edge (u, v) denotes the probability that node u can activate v successfully. We call a node as *active* if it adopts the products or information from other nodes, *inactive* otherwise. Influence propagation process unfolds discrete time steps t_i, $(i = 0, 1, \ldots)$. Initial seed set $S_{t_0} = S$. Let S_{t_i} denote *active* nodes in time step t_i, and each node u in S_{t_i} has single chance to activate each *inactive* neighbor v through its out-edge with probability w_{uv} at time step t_{i+1}. Repeat this process until no more new nodes can be activated. A node can only switch from *inactive* to *active*, but not in the reverse direction.

4 Preliminaries and Problem Description

4.1 Preliminaries

A set function f is monotone increasing if $f(A) \leq f(B)$ whenever $A \subseteq B$. Submodular functions have a natural diminishing returns property. If V is a finite set, a submodular function is a set function $f : 2^V \to \Re$, where 2^V denotes the power set of V, which satisfies the following condition: for every $A \subseteq B \subseteq V$ and $x \in V \backslash B$, $f(A \cup \{x\}) - f(A) \geq f(B \cup \{x\}) - f(B)$.

Further, we introduce some basic definitions for later discussion. *Set cover* [7]: Given a ground set $U = \{u_1, u_2, \ldots, u_n\}$ and a collection of subsets of $C = \{C_1, \ldots, C_m\}$. The set cover problem is to identify the smallest sub-collection C' from C such that C' covers all elements in U, i.e., $\bigcup_{C_i \in C'} C_i = U$.

s-t connectedness [12]: Given a directed graph G and arbitrary two nodes s and t. The *s-t connectedness* finds number of subgraphs of G in which there is a directed path from s to t.

4.2 Problem Description

Given a directed social network $G = (V, E, w)$, an information diffusion model $\mathcal{M}$, a target users set $T = \{T_1, T_2, \ldots, T_q\}$, and a positive integer budget b, where V denotes all users, $E \subseteq V \times V$ denotes the relationships between users, and w_{uv} of edge (u, v) means the probability that u activates v successfully. We assume that the acceptance probability of a node v is equal to the v's activation probability. We define the acceptance probability of a node $v \in V$ when given a seed set S under IC model as follow

$$Pr_{\mathcal{M}}(v, S) = \begin{cases} 1, & \text{if } v \in S \\ 0, & \text{if } N^{in}(v) = \emptyset \\ 1 - \prod\limits_{u \in N^{in}(v)} (1 - Pr_{\mathcal{M}}(u, S) w_{uv}), & \text{otherwise.} \end{cases} \tag{1}$$

where $N^{in}(v)$ is the set of in-neighbors of v and $Pr_{\mathcal{M}}(u, S)w_{uv}$ represents the probability u successfully activates v under the diffusion model $\mathcal{M}$ (IC model). As we can clearly see the acceptance probability of a node v depends on the acceptance probability of its in-neighbors u. Then we define the acceptance probability for target users $v \in T$ from seed set S under the diffusion model as

$$Pr_{\mathcal{M}}(T,S) = \sum_{v \in T} Pr_{\mathcal{M}}(v,S). \tag{2}$$

Now, we can formally define the *Acceptance Probability Maximization* (APM). Given a social network $G = (V, E, w)$, a target users set T, an information diffusion model $\mathcal{M}$, and a positive integer budget b, APM aims to find a seed set S^* such that

$$S^* = \arg \max_{S \subseteq V \setminus T, |S|=b} Pr_{\mathcal{M}}(T,S). \tag{3}$$

In particular, we omit the subscript $\mathcal{M}$ if the context is clear.

5 Properties of APM

We show the properties of APM problem as following theorems.

Theorem 1. *APM problem is NP-hard under the IC model even if* $|T| = 1$.

Proof. We prove it with reduction from the set cover problem [7]. We construct a new network $G' = (V', E', w')$. V' includes three parts: (1) Create a node C_i for each C_i; (2) Create a node u_j for each u_j; (3) Create a target node T. E' is defined as follows. If C_i contains u_j, then add a directed edge (C_i, u_j) from node C_i to node u_j with $w'_{C_i u_j} = 1$. Moreover, for each node u_j, add a directed edge (u_j, T) from node u_j to node T with $w'_{u_j T} = p$. Obviously, the above transformation can be done in polynomial time.

We prove that there is a subset $C' \subseteq C$ covering all nodes in U in the set cover problem if and only if there is a solution with acceptance probability $1-(1-p)^{|U|}$[1] when selecting $b = |C'|$ nodes as seeds. We first prove the sufficient condition. If there exists a subset C' covering all node in U, which obtains acceptance probability $1-(1-p)^{|U|}$ for target nodes. We then prove the necessary condition. If there exists a seed set C' with $|C'| = b$ obtaining acceptance probability $1-(1-p)^{|U|}$, then C' must covering all nodes in U. If the set cover problem is solvable, then APM problem is also solvable. As we all know, the former is NP-hard, therefore the latter is also NP-hard.

Theorem 2. *Given a seed set S and a target set T, computing acceptance probability from seed set S to target set T is $\#p-hard$ under the IC model.*

Proof. We prove this theorem with reduction from a classical $\#p - complete$ problem named *s-t connectedness* [12]. For simplicity, we let $T = \{z\}$. We assign the probability of each edge as 0.5 to guarantee each subgraph with equal probability. Therefore it's straightforward to see that *s-t connectedness* is equivalent to compute the path probability from s to t.

[1] Notice that $1-(1-p)^{|U|}$ is maximum probability of T under the IC model.

Let $Pr(T, S, G)$ denote the acceptance probability of T from a given seed set S on G under the IC model. First, let $S = \{s\}$ and $w_{uv} = 0.5$ for all $(u, v) \in E$. Therefore $P_1 = Pr(T, S, G)$. Next, we add a new node z' and two directed edges (z', z) with $w_{z'z} = 0.5$ and (t, z') with $w_{tz'} = 1$, obtaining a new graph G'. Then, we compute $P_2 = Pr(T, S, G')$. Therefore, $P_2 - P_1 = (1 - P_1) \cdot Pr(t, s) \cdot w_{tz'} \cdot w_{z'z}$, which is related to the probability $Pr(t, s)$ that s is connected to t. As we all know, s-t connectedness is $\#p-complete$ and thus theorem follows immediately.

Theorem 3. *The objective function (3) is monotone non-decreasing and submodular under the IC model.*

Proof. Obviously, increasing the seed nodes does not reduce the objective function value, so we omit its proof. We show that $Pr(T, S)$ is submodular. Consider following two cases.

Case 1: If $A = B$, for an arbitrary node v, then $Pr(T, A \cup \{v\}) - Pr(T, A) = Pr(T, B \cup \{v\}) - Pr(T, B)$ always stands.

Case 2: If $A \subset B$, let $\triangle A_v = Pr(T, A \cup v) - Pr(T, A)$ denote the marginal influence on the target nodes that are not already in the union $\bigcup\limits_{u \in A} Pr(T, u)$. Let $\triangle B_v = Pr(T, B \cup v) - Pr(T, B)$ denote the marginal influence on the target nodes that are not in the union $\bigcup\limits_{u \in B} Pr(T, u)$. Obviously, $\triangle A_v$ is no less than $\triangle B_v$, that is, $Pr(T, A \cup v) - Pr(T, A) \geq Pr(T, B \cup v) - Pr(T, B)$ which $Pr(T, S)$ is submodular with respect to S.

6 Algorithm

From Theorem 2, computing the APM is $\#p - hard$. Therefore we need to find an approximate method to calculate it. Intuitively, computing APM in local structures, such as communities, allows efficient computation. Further, in each community, constructs a local tree structure [3] and approximates local influence diffusion to the target nodes.

Give a social network $G = (V, E, w)$, a community set $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_m\}$ where $\bigcup_{1 \leq j \leq m} \mathcal{C}_j = V$, a target user set $T = \{T_1, T_2, \ldots, T_q\}$ and a positive integer budget b. For a path $Path(u, v) =< u = p_1, p_2, \ldots, p_l = v >$ from u to v in $\mathcal{C}_j$, we define the probability of this path as $\mathcal{P}(u, v) = \prod_{i=1}^{i=l-1} w_{p_i p_{i+1}}$. If u successfully activates v through path $Path(u, v)$, u must activate all the nodes along this path. Let $Path_{\mathcal{C}_j}(u, v)$ denote the set of all paths from u to v in $\mathcal{C}_j$.

Definition 1. *(Maximum Influence Path (MIP)). For a community $\mathcal{C}_j$, we define $MIP_{\mathcal{C}_j}(u, v)$ from u to v in $\mathcal{C}_j$ as*

$$MIP_{\mathcal{C}_j}(u, v) = \arg \max_{Path(u,v)} \{\mathcal{P}(u, v) | Path(u, v) \in Path_{\mathcal{C}_j}(u, v)\}. \qquad (4)$$

Note that if we transform w_{uv} to $1/w_{uv}$ for each edge (u, v), $MIP_{\mathcal{C}_j}(u, v)$ is equivalent to the shortest path from u to v in $\mathcal{C}_j$. The shortest path problem has polynomial time algorithms, e.g., Floyd-Warshall and Dijkstra algorithms. For a target node $v \in T$, we create a tree structure which is the union of MIPs to v, to estimate the acceptance probability to v from other nodes. Moreover, we use a threshold θ to delete MIPs which have small probabilities.

Definition 2. *(Maximum Acceptance Probability Tree (MAPT)). For a threshold θ, the maximum acceptance probability tree of a target node $v \in T$ in $\mathcal{C}_j$, $MAPT(v, \theta)$, is*

$$MAPT(v,\theta) = \bigcup_{u \in \mathcal{C}_j, MIP\mathcal{C}_j(u,v) \geq \theta} MIP_{\mathcal{C}_j}(u,v). \tag{5}$$

In fact, we assume that the influences only propagate within communities and their propagation in these communities are independent of each other. With this assumption, we can calculate the acceptance probability that $v \in T$ is activated when given a seed set S exactly. Considering community structure that plays a vital role in propagation [1,4], we transform APM problem into *Maximum Weight Hitting Set* (MWHS) problem.

Definition 3. *(Maximum Weight Hitting Set (MWHS)). Given an element set V, a family of subsets $\mathcal{C} \subseteq 2^V$, a weight function $w : \mathcal{C} \mapsto \Re^+$, and a positive integer b. MWHS finds a subset $S \subseteq V$ and $|S| = b$ such that maximizes the total weight of subsets in $\mathcal{C}$ hit by S. (S hits $\mathcal{C}$, which means $S \cap \mathcal{C} \neq \emptyset$.)*

Let S denote seed set. We should do following two steps so that APM can be transformed into MWHS. (1) We say that seed set S hits community $\mathcal{C}_j$ if seed set $S \bigcap \mathcal{C}_j \neq \emptyset$. (2) Let $w_j = w(\mathcal{C}_j) = \sum_{v \in \mathcal{C}_j, S_j \in \mathcal{C}_j} Pr(v, S_j)$, where S_j is the set of seed nodes in community $\mathcal{C}_j$ and $Pr(v, S_j)$ is probability that S_j activates v successfully. Now, we formalize APM as following integer programming.

$$\begin{aligned} \max H(v) = \sum_{j=1}^{m} w_j \cdot \min\{1, \sum_{i \in \mathcal{C}_j} v_i\} \\ s.t. \sum_{v_i \in V \setminus T} v_i = b \\ v_i \in \{0,1\}, v_i \in V \setminus T, i = 1, \ldots, |V \setminus T|. \end{aligned} \tag{6}$$

where $v_i = 1$ if $v_i \in S$ or $v_i = 0$ otherwise. We label all nodes except the target nodes from 1 to $|V \setminus T|$ and $i \in \mathcal{C}_j$ denotes the node label belonging to the community $\mathcal{C}_j$. Note that $v_i = 0$ or 1, $H(v) = \sum_{j=1}^{m} w_j \cdot \min\{1, \sum_{i \in \mathcal{C}_j} v_i\}$ can rewrite as $F(v) = \sum_{j=1}^{m} w_j \cdot (1 - \prod_{i \in \mathcal{C}_j}(1 - v_i))$. Therefore we have

$$\begin{aligned} \max F(v) = \sum_{j=1}^{m} w_j \cdot (1 - \prod_{i \in \mathcal{C}_j} (1 - v_i)) \\ s.t. \sum_{v_i \in V \setminus T} v_i = b \\ v_i \in \{0,1\}, v_i \in V \setminus T, i = 1, \ldots, |V \setminus T|. \end{aligned} \tag{7}$$

We consider the relaxed problem of (6), i.e., $0 \le v_i \le 1$. This relaxed problem can be found an optimal solution in polynomial time [5]. Based on this, we propose the pipage rounding algorithm for APM to obtain an integer solution.

Algorithm 1. Pipage Rounding Algorithm (PRA)

Input: $G = (V, E, w)$, a community set $\mathcal{C} = \{\mathcal{C}_1, \ldots, \mathcal{C}_m\}$, a target user set $T = \{T_1, \ldots, T_q\}$, a parameter θ, a inter budget b and an influence diffusion model $\mathcal{M}$.
Output: seed set S.
1: Find an optimal solution $\mathcal{S} = \{v_1, \ldots, v_b\}$ to relaxed problem;
2: $S \leftarrow \mathcal{S}$;
3: **for** each community $\mathcal{C}_l$ **do**
4:   **if** there exists a target node $v \in \mathcal{C}_l$ **then**
5:     Create $MAPT(v, \theta)$;
6:     **while** S has an non-integral component **do**
7:       Choose $0 < v_k, v_j < 1$ in $MAPT(v, \theta)$;
8:       Define $S(\varepsilon)$ by

$$v_i(\varepsilon) = \begin{cases} v_i, & \text{if } i \neq k, j, \\ v_j + \varepsilon, & \text{if } i = j, \\ v_k - \varepsilon, & \text{if } i = k; \end{cases}$$

9:       Let $\varepsilon_1 \leftarrow \min\{v_j, 1 - v_k\}$;
10:       Let $\varepsilon_2 \leftarrow \min\{1 - v_j, v_k\}$;
11:     **end while**
12:     **if** $F(S(-\varepsilon_1)) \ge F(S(\varepsilon_2))$ **then**
13:       $S \leftarrow S(-\varepsilon_1)$;
14:     **else**
15:       $S \leftarrow S(\varepsilon_2)$;
16:     **end if**
17:   **end if**
18: **end for**
19: **return** S.

Round one or two non-integer components of optimal solution to relaxed problem in each iteration, which does not cause the objective function value decreasing. We have following theorem.

Theorem 4. *The approximation ratio of Algorithm 1 is* $(1 - 1/e)$.

Proof (Proof of Theorem 4). Let $\mathcal{S}$ denote the optimal solution to relaxed problem and round the $\mathcal{S}$ to get an integer solution S_I for (6). Since $F(S(\varepsilon))$ is convex with respect to ε, $\max\{F(S(-\varepsilon_1)), F(S(\varepsilon_2))\} \ge F(S)$ if ε_1, $\varepsilon_2 > 0$. Thus, the value of $F(S)$ is non-decreasing in the loop of Algorithm 1. Therefore, $F(S_I) \ge F(\mathcal{S})$. We note that S_I has only integer components, and $F(S_I) = H(S_I)$. According to [5], it follows that $H(S_I) = F(S_I) \ge F(\mathcal{S}) \ge (1 - \frac{1}{e})H(\mathcal{S})$.

Let us analyze the complexity of the Algorithm 1. Finding an optimal solution to relaxed problem can be done in $O(|V|m)$ on G. The loop from line 3 to 18 at most runs $O(m)$ times. In each iteration, there are at most q target users. The inner loop runs at most b times. Therefore, the time complexity is $O(|V|m+mqb)$.

7 Experiments

In this section, we evaluate our algorithm on real-life networks. We first describe the datasets and experiment setup, and then show the results. Furthermore, we compare with other popular approaches.

7.1 Experiment Setup

Datasets: We use three real-life networks with various scale from (SNAP)[2]. Table 1 provides the details of these datasets. Further, 'CC' represents clustering coefficient and '#Community' represents the number of communities. Note that Amazon and Youtube are undirected networks. Therefore we transform these two undirected networks into directed networks. Specifically, for an undirected edge (u, v) on Amazon and Youtube networks, we randomly generate a directed edge (u, v) or (v, u) with probability of 0.5 respectively. According to [3], we let $\theta = 0.03$ in all experiments.

Table 1. The statistics of data sets

Dataset	#Node	#Edge	CC	#Community
E-mail	1K	25.6K	0.399	42
Amazon	334.8K	925.8K	0.396	75K
Youtube	1134.8K	2987.6K	0.080	8K

E-mail. This network is generated using email data from a large European research institution. Each node represents a researcher and each directed edge (u, v) means that u sent at least one email to v.

Amazon. It is based on *Customers Who Bought This Item Also Bought* feature of the Amazon website. Each node is a product. If a product u is frequently co-purchased with product v, thus there is an edge (u, v) between u and v.

Youtube. This network is a video-sharing social network. Each node is a user on network. Users form friendship if they share same videos.

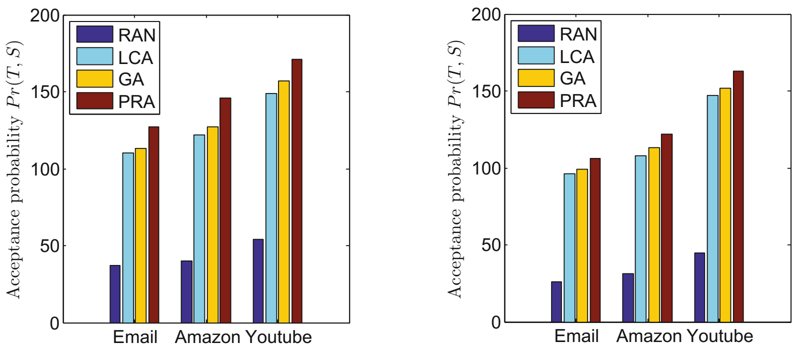
[2] http://snap.stanford.edu/data.

Comparison Methods: To compare with existing methods, other methods are as comparison methods: Local Cascade Algorithm (LCA) [6] and Greedy Algorithm (GA) [14]. Our pipage rounding algorithm is abbreviated as PRA.

Random (RAN) means that it randomly selects seed nodes.

Local Cascade Algorithm (LCA) [6]**.** LCA constructs a local cascade community consists of only the shortest paths between each node and the target node, then restricts computations within the shortest path community.

Greedy Algorithm (GA) [14]**.** The influence spread of seed set S in the targeted set $T \subseteq V$ is the expected number of activated nodes in T by S. And the greedy algorithm iteratively selects a new seed v that maximizes the incremental change of function, to be included into the seed set S, until b seeds are selected.



(a) $w = 0.5$ (b) $w = TRI$

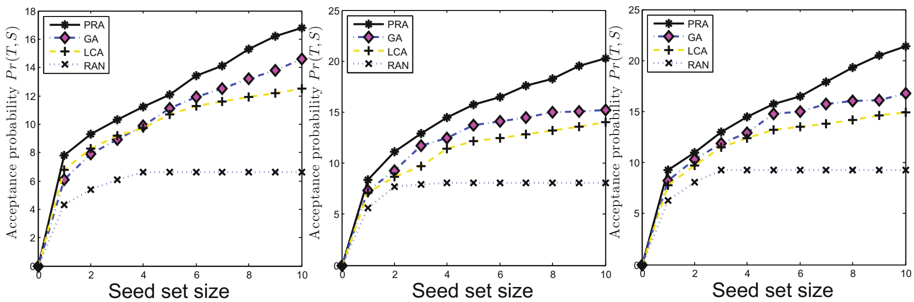
Fig. 1. The total acceptance probability of target users under IC model: target users set size $|T| = 1000$ on Amazon and Youtube networks or $|T| = 500$ on E-mail network, $w = 0.5$ or $w = TRI, \theta = 0.03$ and $b = 30$.

7.2 Results

The Acceptance Probability of Target Users: We calculate the total acceptance probability of target users when $|T| = 1000$ or $|T| = 500$ and $b = 30$ on each network with different methods. Figure 1 illustrates the results. The horizontal axis represents the names of social networks. The vertical axis represents the total acceptance probability of target users. We compare RAN, PRA, LCA and GA where RAN means randomly selecting seed nodes. In both subfigures, total acceptance probabilities of target users show similar trends on each network. Specifically, on each network, total acceptance probabilities satisfy following relationship: RAN<LCA<GA<PRA. On the other hand, instead of utilizing Monte-Carlo simulation, it indicates MAPT is a more effective approximation to calculate the acceptance probability than other methods. In Fig. 1(a), we

let $w = 0.5$. PRA is 8.19%–13.01% more than GA, 12.87%–16.44% more than LCA, and 68.42%–72.60% more than RAN. In Fig. 1(b), we let $w = TRI$[3]. PRA is 6.60%–7.38% more than GA, 9.43%–11.48% more than LCA, and 72.39%–75.47% more than RAN. Although GA has higher acceptance probability than LCA, it's too time consuming in experiments.

Seeds Size vs. Acceptance Probability of Target Users: In this part, for PRA, we analyze how the number of seeds affects the acceptance probability of target users when given a target user set in a fixed community. In fact, we randomly choose $|T| = 1000$ nodes on Amazon and Youtube networks as target users and choose a fixed community whose target users size is greater than 50. In particular, we randomly choose $|T| = 500$ on Email network and select a community whose target users size is greater than 40. Figure 2 shows the results. The horizontal and vertical axis indicate seeds size and acceptance probability of target users in the fixed community, respectively. Note that all methods (PRA, LCA, GA and RAN) show the property of diminishing marginal return. More precisely, acceptance probability sharply increases when seed size increases from $|S| = 1$ to $|S| = 5$. While it increases slowly from $|S| = 5$ to $|S| = 10$. Our PRA method is the best and RAN is the worst one because it has performance guarantee as we analyzed before. RAN randomly selecting seeds with high probability can not activate target users that leads to its worst.



(a) Email, $w = 0.5$ (b) Amazon, $w = TRI$ (c) Youtube, $w = TRI$

Fig. 2. The relationship between total acceptance probability of target users and seeds size on each network under the IC model: $w = 0.5$ for Email network and $w = TRI$ for Amazon and Youtube networks, $|T| > 50$ in a fixed community for Amazon and Youtube networks, $|T| > 40$ in a fixed community for Email network, $\theta = 0.03$.

8 Conclusion

In this paper, we study a novel problem called *Acceptance Probability Maximization* (APM) problem that finds a small size seed set such that the acceptance

[3] We uniformly at random select a probability from $\{0.1, 0.3, 0.5\}$.

probability of target users is maximized. Based on IC model, we show APM is NP-hard and compute it is $\#p-hard$. And we prove objective function satisfies monotonicity and submodularity. Considering the community structure of social networks, we transform our APM problem into MWHS problem. We develop a pipage rounding algorithm which has a $(1-1/e)$ approximation ratio. In order to evaluate our proposed methods, extensive experiments have been conducted. The experiment results show that our method outperforms comparison approaches.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China Grant No. 11671400, 61672524. The Fundamental Research Funds for the Central University, and the Research Funds of Renmin University of China, 2015030273, and the Research Funds of Renmin University of China 16XNH116.

References

1. Algesheimer, R., Dholakia, U.M., Herrmann, A.: The social influence of brand community: evidence from European car clubs. J. Mark. **69**(3), 19–34 (2005)
2. Chang, C.W., Yeh, M.Y., Chuang, K.T.: On influence maximization to target users in the presence of multiple acceptances. In: 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 1592–1593. IEEE (2015)
3. Chen, W., Wang, C., Wang, Y.: Scalable influence maximization for prevalent viral marketing in large-scale social networks. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1029–1038. ACM (2010)
4. Crandall, D., Cosley, D., Huttenlocher, D., Kleinberg, J., Suri, S.: Feedback effects between similarity and social influence in online communities. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 160–168. ACM (2008)
5. Du, D.Z., Ko, K.I., Hu, X.: Design and Analysis of Approximation Algorithms, vol. 62. Springer, New York (2011). https://doi.org/10.1007/978-1-4614-1701-9
6. Guo, J., Zhang, P., Zhou, C., Cao, Y., Guo, L.: Personalized influence maximization on social networks. In: Proceedings of the 22nd ACM International Conference on Information and Knowledge Management, pp. 199–208. ACM (2013)
7. Karp, R.: Reducibility among combinatorial problems. In: Miller, R.E., Thatcher, J.W., Bohlinger, J.D. (eds.) Complexity of Computer Computations, pp. 85–103. Springer, Boston (1972). https://doi.org/10.1007/978-1-4684-2001-2_9
8. Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 137–146. ACM (2003)
9. Kuhnle, A., Pan, T., Alim, M.A., Thai, M.T.: Scalable bicriteria algorithms for the threshold activation problem in online social networks. In: INFOCOM 2017-IEEE Conference on Computer Communications, pp. 1–9. IEEE (2017)
10. Song, C., Hsu, W., Lee, M.L.: Targeted influence maximization in social networks. In: Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, pp. 1683–1692. ACM (2016)
11. Temitope, O.A.S., Ahmad, R., Mahmudin, M.: Influence maximization towards target users on social networks for information diffusion. J. Telecommun. Electron. Comput. Eng. (JTEC) **10**(1–10), 17–24 (2018)

12. Valiant, L.G.: The complexity of enumeration and reliability problems. SIAM J. Comput. **8**(3), 410–421 (1979)
13. Yang, D.N., Hung, H.J., Lee, W.C., Chen, W.: Maximizing acceptance probability for active friending in online social networks. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 713–721. ACM (2013)
14. Zhou, C., Guo, L.: A note on influence maximization in social networks from local to global and beyond. Procedia Comput. Sci. **30**, 81–87 (2014)