# Behavioral Intentions Maximization for Multiple Products and Rumors in Online Social Networks

Chung-Wei Lee, Shih-Hsuan Huang, and Ming-Jer Tsai
Department of Computer Science
National Tsing Hua University
Hsinchu, Taiwan
E-mail: mjtsai@cs.nthu.edu.tw

***Abstract*—Marketing through online social networks is convenient, low-cost, and beneficial for companies seeking to expand their customer numbers. In the literature, many studies address the influence maximization problem with one or multiple products, which selects initial consumers (seeds) to spread one or multiple product information such that the number of consumers receiving these product information (the influenced consumers) is maximized. However, to date, none of these schemes take the rumors and the beliefs of other persons that could significantly change the consumer's behavioral intention into account at once. In this paper, we fill this gap by proposing a new variant of the influence maximization problem with multiple products, the Budgeted Behavioral Intentions Maximization problem, which asks for a set of seeds with the total cost not greater than a given budget in online social networks such that the total expected behavioral intentions of the consumers influenced by the selected seeds and the rumors are maximized. In addition, we propose an approximation algorithm for the Budgeted Behavioral Intentions Maximization problem. We also conduct simulations to evaluate the performance of our algorithm using real traces and synthesis data. Experimental results show that our algorithm outperforms several greedy algorithms.**

## I. Introduction

Online social networks, including Facebook, Twiter, and YouTube are among the top-ten most visited websites on the Internet [1], provide a convenient platform for fast information propagation and people interactions. Such interactions, including product commercials and recommendations could directly and efficiently influence people's consuming behaviors. Recent studies showed that network-based marketing has a direct effect on increasing product adoption. For example, State bicycle, a bicycle company through Facebook commercials obtains towards 500,000 USD in incremental sales every year [1]. Therefore, for companies, through online social networks to expand their consumer numbers is convenient, low-cost, and beneficial. For example, due to a limited budget, a company may select a small number of initial consumers (by making them payments) in online social networks to spread the product information and then expect that a large number of consumers will receive the product information. The problem here is to select a given number of initial consumers (seeds) to spread the product information such that the largest number of consumers receive the product information; this is known as the influence maximization problem.

Domingos and Richardson were the first to define the influence maximization problem as an algorithmic problem and propose probabilistic methods for the problem defined [2], [3]. In addition, Kempe et al. were the first to formulate the problem as a discrete optimization problem [4]. They showed the optimization problem is NP-hard and proposed an approximation algorithm for it.

In the literature, several studies address the variants of the influence maximization problem [5], [6]. Instead of selecting a given number of seeds, Tang et al. [5] introduced a problem of selecting seeds subject to a budget constraint and proposed an approximation algorithm for the problem defined, where each consumer has a cost to pay when selected as a seed. Furthermore, nowadays, a consumer can be on multiple social networks, such as Facebook, Twitter, and Google+; therefore Nguyen et al. [6] studied the influence maximization problem on multiple networks instead of a single one. They proposed a method to couple multiple networks into a single network such that any solution to the influence maximization problem on a single network could be a solution without compromising the quality on multiple networks. However, these studies [5], [6] only took a single propagation into account, which indicates that there is only one considered product. Unlike [5], [6], Zhang et al. [1] thought that for a company, there is not only one product. There are all kinds of products to satisfy the various demands of consumers. For example, HUAWEI produces both cheap ordinary phones and expensive smart phones. Therefore, Zhang et al. [1] took multiple propagation into account, which indicates that there are multiple considered products. They studied a **P**rofit **M**aximization with **M**ultiple **A**doptions (PM$^2$A) problem, which asks for a seeding set within a limited budget to massively influence users and achieve the goal of profit maximization. However, none of the methods in [1], [4]–[6] take the beliefs of other persons into account.

For a company, in reality, there are many adversaries for business profit. Thus, when a company would market their products through online social networks, its adversaries could propagate the negative information (rumor) to attack it. These rumors could affect consumers behaviors and make this company image be damaged. Thus, for a company, it should respect this issue. According to the past studies [7], they pointed that such rumor enables the network unreliable and may cause panic in population. For example, the rumor of swine flu propagated in Twitter caused widespread terror in 2009. In 2011, the Twitter account of Fox news was hacked, and this account posted the news that the president of the United states has been shot dead endlessly. Therefore, for a company, effective policies for rumor containment are required in online social networks. The problem here is to ask for $k$ seed users to trigger the spread of the positive diffusion under a budget $k$ such that the number of the users who are not influenced by the rumors could be maximized; this is known as the influence blocking maximization problem.

Budak et al. were the first to define the influence blocking maximization problem which is a variant of the influence maximization problem as an algorithmic problem and proposed a greedy algorithm for the probem defined [8]. He et al. [9] were the first to formulate the influence blocking maximization problem as a discrete optimization problem. They showed the optimization problem is NP-hard and proposed an approximation algorithm for it. Moreover, since the algorithm proposed by He et al. [9] demanded considerable time to execute, Ping et al. [10] proposed a fast heuristic without a provable approximation ratio, and Tong et al. [7] proposed a fast approximation algorithm for the optimization problem individually. Like [1], [4]–[6], none of the methods in [7]–[10] take the beliefs of other persons into account.

Fishbein investigated the behavior of a person from the viewpoint of social psychology [11]. Through a long-term study, he found that the behavior of a person is strongly related to his/her behavioral intention (i.e, the relative strength of intention to perform a behavior). In addition, Fishbein found that the behavioral intention of a person for an object depended on his/her belief toward the object. If a person's belief toward an object was positive, then the person would be willing to accept or adopt the object, that is, the behavioral intention of the person for the object would be high. Moreover, Fishbein also found that the behavioral intention of a person was changed by the beliefs of other persons. Thus, when we predict the behavioral intention of a person, we must take this factor into account. However, a person might care little for what others think; that is, the behavioral intention of a person may be changed very little by others' beliefs. Thus, when we predict the behavioral intention of a person, we must not only combine his/her beliefs with the beliefs of other persons, but also take the weights of the beliefs into account. Based on these arguments, Fishbein proposed a model, Extended Fishbein Model (EFM), to quantify them. One of the practical applications of the EFM is to estimate the degree of consumers' shopping desires toward a product (e.g., Music CD) in the marketing. A company could make various marketing decisions through these estimated data. Besides, Seligman and Csidszentmihalyi proposed positive psychology that is the branch of psychology [12]. The famous argument of the positive psychology is that we should respect the positive emotion or belief of a person rather than the negative one. Then, we could find and understand the factors of the positive behavior of a person. Therefore, one of the goal of the positive psychology is to investigate the positive emotion or belief of a person in their daily life. A practical application of positive psychology is to assist people in identifying what they really want at heats so as to help them get their desires.

To the best of our knowledge, no existing methods of the influence maximization problem with multiple products take the rumors, the beliefs of other persons, and the argument of the positive psychology into account for predicting the consumer's behavioral intention. In this paper, we fill this gap by proposing a new variant of the influence maximization problem, the **B**udgeted **B**ehavioral **I**ntentions **M**aximization ($B^2$IM) problem, which asks for a set of seeds with the total cost not greater than a given budget in online social networks such that the total expected behavioral intentions of the consumers influenced by the selected seeds and the rumors are maximized.

In the remainder of this paper, we will study a scenario concerning the spread of a company's multiple product information in online social networks and introduce the $B^2IM$ problem based on the scenario studied in Section II. Subsequently, we propose an approximation algorithm for the $B^2IM$ problem in Section III. Using simulations, we evaluate the performance of our algorithm in Section IV. Related works are presented in Section V. Finally, we conclude this paper in Section VI.

## II. Budgeted Behavioral Intentions Maximization

### A. Scenario

A company develops multiple new products and wants to market them through online social networks. As a result of the business profit, its adversaries would propagate the negative information (rumor) to attack it. Therefore, on a condition of the existence of the rumors and a limited budget, this company can only select a small set of consumers as seeds by some social network platform (e.g., Facebook) to spread these product information and desires that the consumers who have highest behavioral intentions toward these products will be able to obtain these product information. In online social networks, two consumers $u$ and $v$ contact each other with probability $p(u,v) \in [0,1]$. Each consumer $u$ is associated with $w(u) > 0$ denoting the cost of selecting $u$ as a seed to spread the product information. $R_s$ is the set of rumors. Based on [7], we assume $R_s$ is fixed. Based on the argument of the positive psychology [12], each consumer has his/her own positive belief with scale $b(u) \geq 0$ toward the product (see Table I for the scale). In addition, each consumer has his/her behavioral intention $B(u) \geq 0$ toward the product, which is evaluated dependent on the set of seeds and $R_s$. Our goal is to

TABLE I
THE SCALE OF THE DEGREE OF A CONSUMER'S OWN POSITIVE BELIEF TOWARD A PRODUCT ($b(u)$)

| Degree | heaviest | heavier | heavy | light | lighter | lightest | Not at all |
|---|---|---|---|---|---|---|---|
| Scale | +6 | +5 | +4 | +3 | +2 | +1 | 0 |

TABLE II
THE SCALE OF THE DEGREE OF THE CHANGE OF OTHER CONSUMERS' POSITIVE BELIEFS ON THE CONSUMER'S BEHAVIORAL INTENTION TOWARD A PRODUCT ($d(u, v)$)

| Degree | Strongly | Moderately | Slightly | Not at all |
|---|---|---|---|---|
| Scale | +3 | +2 | +1 | 0 |

find a small set of consumers as seeds to spread the multiple product information in online social networks under a budget and the existence of the rumors constraints, such that the total expected behavioral intentions of the consumers who obtain these product information are maximized.

### B. Rumor Propagation

When a consumer receives the rumor information toward the product, the rumor always polishes itself to be convincing [7]. Thus, the behavioral intention of a consumer toward the product would be declined. Based on the above argument and [7], we assume a consumer receives the rumor information, his/her behavioral intention $B(u)$ would become $\lfloor \frac{B(u)}{2} \rfloor$. Moreover, when a consumer receives more rumor information, his/her behavioral intention $B(u)$ would be more declined, and so on. Once a consumer behavioral intention $B(u)$ become 0, it indicates that he/she is not interested in that product not at all.

### C. User/Seed-Set Behavioral Intentions

We have $q$ different kinds of product inforamtion need to be propagated. Based on [1], we could view this propagation process as unfolding on $q$ seperate networks. Therefore, for each seperate network, given a set of seeds $V'$, the expected behavioral intention of a consumer in an online social network is evaluated in three steps. First, the online social network induced by $r$-th random process, $OSN_r$, is established based on the random process method proposed by Kempe et al. [4]. Let $P_r$ denote a sequence of the real numbers for all pairs of consumers $u$ and $v$, $t(u, v)$, that are randomly chosen from the interval $[0, 1]$ at $r$-th random process. Then, $t(u, v) \in P_r$ is used as the threshold value for the existence of edge $(u, v)$ in $OSN_r$. That is, edge $(u, v)$ is in $OSN_r$ if $p(u, v) \geq t(u, v)$; otherwise, edge $(u, v)$ is not in $OSN_r$. Subsequently, the influence scope (i.e., the scope of the spread of the product information) of $V'$ by $r$-th random process, $IS_r(V')$, is obtained from $OSN_r$ by removing all vertices (and their incident edges) that are unreachable from the vertices in $V'$. That is, in $IS_r(V')$, there is a path from each vertex to a vertex in $V'$.

Second, for the influence scope of $V'$ induced by $r$-th random process, $IS_r(V')$, the behavioral intention of a consumer $u$, $B_r(u)$, is evaluated based on the Extended Fishbein Model (EFM) [11], a famous model of investigating consumer behavior in the marketing domain. More specifically, the consumer $u$'s behavioral intention is evaluated by the sum of the change of consumer $u$'s own positive belief on his/her behavioral intention ($W_1(u) \cdot b(u)$) and the total change of each his/her neighboring consumer $v$'s positive belief on his/her behavioral intention ($W_2(u) \cdot \sum_{v \in NC_r(u)} (b(v) \cdot d(u, v))$), where the change of consumer $v$'s positive belief on consumer $u$'s behavioral intention is evaluated by the product of the consumer $v$'s positive belief ($b(v)$) and the degree of the change of consumer $v$'s positive belief on consumer $u$'s behavioral intention ($d(u, v)$, see Table II for the scale), as described in the following equation.

$$B_r(u) = W_1(u) \cdot b(u) + W_2(u) \cdot \sum_{v \in NC_r(u)} (b(v) \cdot d(u, v)), \quad (1)$$

where $NC_r(u)$ denotes the set of the consumer $u$'s neighbors (i.e. the set of nodes able to contact with $u$) in $IS_r(V')$, $W_1(u) \in [0, 1]$ and $W_2(u) \in [0, 1]$ denote the weights of the changes of the consumer $u$'s own positive belief and the other consumers' positive beliefs on the consumer $u$'s behavioral intention toward the product, respectively, and $W_1(u) + W_2(u) = 1$. The total behavioral intentions of the consumers influenced by the selected seeds in $V'$ at $r$-th random process, $F_r(V')$, are evaluated as follows.

$$F_r(V') = \sum_{u \in IS_r(V')} B_r(u). \quad (2)$$

Moreover, there are $q$ different kinds of products, thus, the expected behavioral intention of a consumer $u$, $B_\mu(u)$, over all online social networks induced by random processes is evaluated as $\sum_{i=1}^{q} \lim_{N \to \infty} \frac{1}{N} \cdot \sum_{r=1}^{N} B_r^i(u)$. The total expected behavioral intentions of the consumers influenced by the selected seeds in $V'$, $F_\mu(V')$, are evaluated in the following equation.

$$F_\mu(V') = \sum_{i=1}^{q} \lim_{N \to \infty} \frac{1}{N} \cdot \sum_{r=1}^{N} F_r^i(V'). \quad (3)$$

### D. The Problem Definition and Hardness

Based on [1], we model an online social network as a directed graph $G = (V, E)$, where vertex $u$ in $V$ denotes consumer $u$, and edge $(u, v)$ in $E$ denotes the relationship between consumers $u$ and $v$. Based on the studied scenario, the *$B^2$IM* problem is described in Definition 1 as follows.

**Definition 1.** Given two positive number $B$ and $q$, the set of rumors $R_s$, and a directed graph $G = (V, E)$ with weights $W_1(u) \in [0, 1]$, $b(u)$, and $w(u)$ associated with all vertices $u \in V$ and weights $p(u, v) \in [0, 1]$ and $d(u, v)$ associated with all edges $(u, v) \in E$, the **B**udgeted **B**ehavioral **I**ntentions **M**aximization (B$^2$IM) problem asks for a set of seeds $V' \subseteq V$ to propagate $q$ product information with the total cost not greater than $B$ such that the total expected behavioral intentions of the vertices influenced by $R_s$ and the selected seeds in $V'$, $F_\mu(V')$, are maximized.

Remark that the weights, including $b(u)$ (i.e., the product of the consumer $u$'s positive belief) and $d(u, v)$ (i.e., the degree of the change of consumer $v$'s positive belief on consumer $u$'s behavioral intention), could be learned and evaluated from the real data that are collected from real-world networks and websites, such as Facebook [1] and Last.fm [13], as studied in Section IV-A.

By a polynomial-time reduction from the **I**nfluence **M**aximization (IM) problem [4], we can prove the *$B^2$IM* problem is NP-hard in Theorem 1. We omit the proof of Theorem 1 due to the page limit. The proof of Theorem 1 could be referred to Theorem 1 of our technical report [14].

**Theorem 1.** *The $B^2$IM problem is NP-hard.*

## III. Approximation Algorithm

### A. Algorithm

The primary idea of Algorithm 1 is based on the partial enumeration technique [15], [16]. Algorithm 1 enumerates all subsets $V'$ of $V$ that contain less than $y$ vertices with the total cost $w(V')$ at most $B$ (line 1), and selects the one with the greatest average total estimated behavioral intentions of the influenced vertices (over $N$ random processes) as a candidate solution $H_1$ (line 3). In addition, Algorithm 1 enumerates all subsets $V'$ of $V$ that contain exactly $y$ vertices with the total cost at most $B$ (lines 6 and 17), expands all sets $V'$ by iteratively including the vertex with the maximum ratio of the incremental average total estimated behavioral intentions of the influenced vertices to the cost until the budget is exceeded (lines $7-13$), and then selects the one with the greatest average total estimated behavioral intentions of the influenced vertices as a candidate solution $H_2$ (lines $14-16$). Algorithm 1 outputs $H_1$ if the average total estimated behavioral intentions of the vertices influenced by the vertices in $H_1$ is not smaller than that of the vertices influenced by the vertices in $H_2$; otherwise, Algorithm 1 outputs $H_2$ (lines 18 − 22).

### B. Analysis

By (1) and (2), we could show the function $F_r$ of the *$B^2$IM* problem is non-negative, monotone, and submodular in Lemma 1. We omit the proof of Lemma 1 due to the page limit. The proof could be referred to Lemmas 1, 2, 3 of our technical report [14].

**Algorithm 1** Approximation Algorithm for the *$B^2$IM* Problem

**Input:** A graph $G = (V, E)$ with $W_1(u)$, $b(u)$, $w(u)$ associated with all vertices $u \in V$ and $p(u, v)$ and $d(u, v)$ associated with all edges $(u, v) \in E$, and numbers $q$, $y$, $B$, $N$, and a set $R_s$

1: Evaluate $\hat{F}(V')$ for all $V' \subseteq V \setminus R_s$ with $|V'| < y$ and $w(V') \leq B$ by Procedure 1 with input parameters $G$, $q$, $y$, $N$, $R_s$ and $V'$;
2: **for** $i = 1$ to $q$ **do**
3: $\quad H_1 \leftarrow \underset{V'_i : V' \subseteq V \setminus R_s, |V'| < y, w(V') \leq B}{\arg\max} \hat{F}(V'_i)$;
4: **end for**
5: $H_2 \leftarrow \varnothing$;
6: **for** each $V' \subseteq V \setminus R_s$ with $|V'| = y$ and $w(V') \leq B$ **do**
7: $\quad$ **while** there exists $u \in V \setminus V' \cup R_s$ such that $w(V' \cup \{u\}) \leq B$ **do**
8: $\quad\quad$ Evaluate $\hat{F}(V' \cup \{u\})$ for all vertices $u \in V \setminus V' \cup R_s$ by Procedure 1 with input parameters $G$, $q$, $y$, $N$, $R_s$ and $V' \cup \{u\}$;
9: $\quad\quad$ **for** $i = 1$ to $q$ **do**
10: $\quad\quad\quad u \leftarrow \underset{u : u \in V \setminus V' \cup R_s, w(V' \cup \{u\}) \leq B}{\arg\max} \dfrac{\hat{F}(V'_i \cup \{u\}) - \hat{F}(V'_i)}{w(u)}$
11: $\quad\quad$ **end for**
12: $\quad\quad V' \leftarrow V' \cup \{u\}$;
13: $\quad$ **end while**
14: $\quad$ **if** $\hat{F}(V') > \hat{F}(H_2)$ **then**
15: $\quad\quad H_2 \leftarrow V'$;
16: $\quad$ **end if**
17: **end for**
18: **if** $\hat{F}(H_1) \geq \hat{F}(H_2)$ **then**
19: $\quad$ return $H_1$;
20: **else**
21: $\quad$ return $H_2$;
22: **end if**

**Definition 2.** [1] Non-negativity: Let $S$ be a non-empty finite set and $Z$ a function from $2^S$ (the power set of $S$) to $\mathbb{R}$. If $Z(A) \geq 0$ for all $A \subseteq S$, then $Z$ is non-negative.

**Definition 3.** [1] Monotonicity: Let $S$ be a non-empty finite set and $Z$ a function from $2^S$ to $\mathbb{R}^+ \cup \{0\}$. If $Z(A) \leq Z(B)$ for all $A \subseteq B \subseteq S$, then $Z$ is monotone.

**Definition 4.** [1] Submodularity: Let $S$ be a non-empty finite set and $Z$ a function from $2^S$ to $\mathbb{R}^+ \cup \{0\}$. If $Z(A \cup \{e\}) - Z(A) \geq Z(B \cup \{e\}) - Z(B)$ for all $e \in S \setminus B$ and all $A \subseteq B \subseteq S$, then $Z$ is submodular.

**Lemma 1.** *The function $F_r : 2^V \to \mathbb{R}$ in the $B^2$IM problem is non-negative, monotone, and submodular.*

**Theorem 2.** *The objective function $F_\mu$ in the $B^2$IM problem is non-negative, monotone, and submodular.*

*Proof.* By Lemma 1, the function $F_r : 2^V \to \mathbb{R}$ is non-negative, monotone, and submodular. Since $F_\mu$ is a positive linear combination of $F_r$ by (3), $F_\mu$ is non-negative, monotone, and submodular. $\square$

By [16], for a problem like the **B**udgeted **M**aximum

**Procedure 1** Evaluation of Estimated Behavioral Intentions

**Input:** A graph $G = (V, E)$ with $W_1(u)$, $b(u)$, and $w(u)$ associated with all vertices $u \in V$ and $p(u, v)$ and $d(u, v)$ associated with all edges $(u, v) \in E$, and numbers $q$, $y$, $N$, and a set $R_s$, $V'$

**for** $i = 1$ to $q$ **do**
  $V'_i = V'$;
  $\hat{F}(V'_i) = 0$;
  **for** $r \leftarrow 1$ **to** $N$ **do**
    Obtain $P_r$ by randomly choosing $t(u, v)$ from the interval $[0, 1]$ for all pairs of vertices $u$ and $v$;
    Obtain $OSN_r$ from $G$ by removing all edges $(u, v)$ with $p(u, v) < t(u, v)$;
    Obtain $IS_r(V'_i)$ from $OSN_r$ by removing all vertices (and their incident edges) unreachable from the vertices in $V'_i$;
    **for** all vertices $u$ in $IS_r(V'_i)$ **do**
      $\hat{F}(V'_i) = \hat{F}(V'_i) + B_r(u)$, where $B_r(u)$ is evaluated by (1);
    **end for**
  **end for**
  return $\hat{F}(V'_i) = \frac{\hat{F}(V'_i)}{N}$;
**end for**

**C**overage (BMC) problem [15], if the objective function is non-negative, monotone, and submodular, and the objective function value can be estimated with a bounded error, then the partial enumeration method with $y \geq 3$ can be an approximation algorithm. Lemma 2 demonstrates that the objective function value of the *B*$^2$*IM* problem can be estimated with a bounded error if a large number of random processes is performed. As a result of the page limit, we omit the proof of Lemma 2. The proof could be referred to Lemma 4 of our technical report [14].

**Lemma 2.** *In Algorithm 1, if* $N \geq \frac{b_{max}^4 m^4 q^2}{\delta^3}$, *then* $P(|\hat{F}(V') - F_\mu(V')| < \delta) \geq 1 - \delta$ *for any* $V' \subseteq V$, *where* $b_{max}$ *is the maximum scale of the consumer's own positive belief toward a product,* $m = |V|$, $q$ *is the number of products,* $\delta$ *is a small positive number,* $N$ *is the number of random processes,* $\hat{F}(V') = \frac{\sum_{i=1}^{q}\sum_{r=1}^{N} F_r^i(V')}{N}$ *is an estimated value of* $F_\mu$, *and* $F_r$ *is the outcome of the* $r$*-th random process.*

**Theorem 3.** *Algorithm 1 with* $y \geq 3$ *and* $N \geq \frac{b_{max}^4 m^4 q^2}{\delta^3}$ *computes a seed set* $V'$ *such that* $F_\mu(V') \geq (1 - \frac{1}{e})F_\mu(V^*) - \epsilon$ *with probability at least* $1 - \delta$, *where* $e$ *is the base of natural logarithm,* $V^*$ *is the optimal solution of the B*$^2$*IM problem,* $\delta \in [0, 1]$, *and* $\epsilon = \frac{2B}{\min_{u \in V}\{w(u)\}} \cdot \delta$.

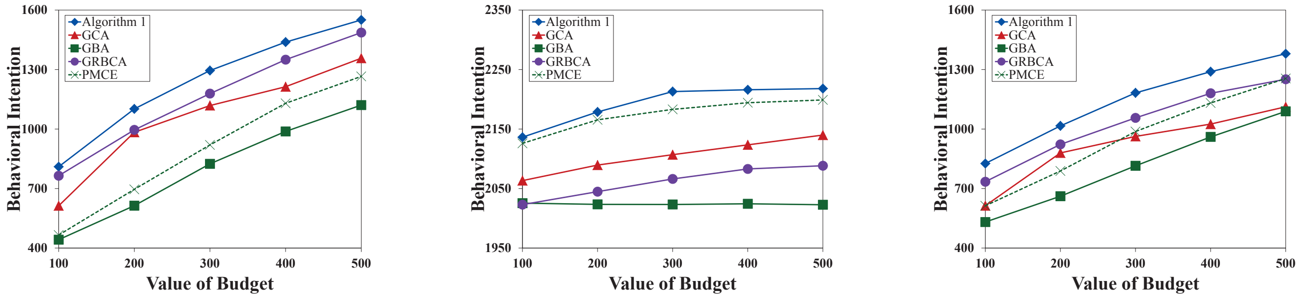
*Proof.* The objective function $F_\mu$ in the *B*$^2$*IM* problem is non-negative, monotone, and submodular by Theorem 2. In addition, if $N \geq \frac{b_{max}^4 m^4 q^2}{\delta^3}$, $|\hat{F}(V') - F_\mu(V')| < \delta$ for any seed set $V' \subseteq V$ under a fixed budget $B$ with probability at least $1 - \delta$ by Lemma 2. Thus, by Corollary 5 of [16], Algorithm 1, a partial enumeration method, with $y \geq 3$ and $N \geq \frac{b_{max}^4 m^4 q^2}{\delta^3}$ computes a seed set $V'$ such that $F_\mu(V') \geq (1 - \frac{1}{e})F_\mu(V^*) - \frac{2B}{\min_{u \in V}\{w(u)\}} \cdot \delta$ with probability at least $1 - \delta$, as desired. □

## IV. Numerical Result

### A. Simulation Settings

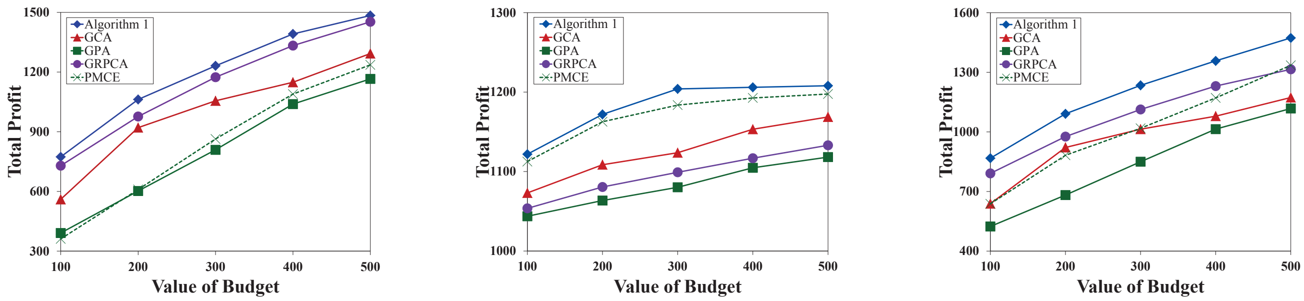
We are the first to investigate the *B*$^2$*IM* problem in an online social network. Therefore, we compare our algorithm (Algorithm 1) with several greedy algorithms, including the greedy on cost algorithm (GCA), the greedy on behavioral intention algorithm (GBA), and the greedy on the ratio of behavioral intention to cost algorithm (GRBCA). The GCA, GBA, and GRBCA iteratively expand the seed set by including the seed with the minimum cost, the maximum behavioral intention, and the maximum ratio of behavioral intention to cost, respectively, until the budget is exceeded. In addition, we also compare Algorithm 1 with the **P**rofit **M**aximization with **C**ost **E**ffectiveness (PMCE) method of the **P**rofit **M**aximization with **M**ultiple **A**doptions (PM$^2$A) problem [1], which is a variant of the influence maximization problem with multiple products by modifying its objective function. The PMCE iteratively expands two isolated sets through including the seed with the maximum behavioral intention, and the maximum ratio of behavioral intention to square of cost, respectively, until the budget is exceeded. Then, it selects the better one from these two sets as the final solution. Like [1], [7], simulations are conducted using the real traces, including NetS, BlogCatalog, and Facebook, collected from real-world networks [1], and the log-based consumer profiles collected from Last.fm [13].

The real traces collected from real-world networks record the contact information of people in different social plaforms over a certain number of months. NetS is a Co-authorship Network in Network Science, with nodes representing authors and edges representing co-authorship. BlogCatalog is a social blog directory website, with nodes representing bloggers and edges representing the relationship among these bloggers. The dataset of Facebook records friendship information among New Orleans regional network, with nodes representing users and edges representing the friendship among them. In our simulations, 1000 consumers were randomly chosen from NetS, BlogCatalog, and Facebook in default. Like [7], we randomly selected 100 consumers as rumors to propagate the negative information toward the product among 1000 consumers in default. Let $F_{uv}$ denote the number that consumer $u$ contacts $v$, and $\lambda = F_{uv}/(T_e - T_s)$ denote the rate that consumer $u$ contacts $v$, where $T_s$ and $T_e$ denote the start and end time, respectively, of the period during which the traces are recorded. Like [17], we assume the contact process between consumers is a homogeneous Poisson process, and set the probability that consumer $u$ contacts $v$ ($p(u, v)$) to $1 - e^{-\lambda \cdot T_p}$, where $T_p$ denotes the period of spreading the romor and product information and was set to six months in our simulations. Let $H_{uv}$ denote the number of contacts between consumers $u$ and $v$, and $H_u$ the total number of contacts of consumer $u$. Then, the scale of the degree of the change of consumer $v$'s positive belief on consumer $u$'s behavioral intention toward the product ($d(u, v)$) was set to $\lfloor \frac{H_{uv}}{H_u + 1} \cdot d_{max} \rfloor$, where $d_{max}$ denotes the maximum scale of the degree of the change of the




(a) (b) (c)

Fig. 1. Impact of the value of budget on the total behavioral intentions of the influenced consumers using (a) NetS, (b) BlogCatalog, and (c) Facebook.




(a) (b) (c)

Fig. 2. Impact of the value of budget on the total profit of the influenced consumers using (a) NetS, (b) BlogCatalog, and (c) Facebook.

other consumer's positive belief on the consumer's behavioral intention toward a product and was set to 3 in the simulations based on TABLE II.

The log-based consumer profiles collected from Last.fm record the music listening habits of 8000 consumers. For each consumer, the 100 songs that were listened to the most were recorded. In the simulations, two product information are propagated, one is the CD of female singer, and the other is the CD of male singer. Each consumer $u$ was associated with a randomly-selected listening profile and sets the scale of the consumer's own positive belief toward the song by the female and male singer ($b(u)$) to $\lfloor \frac{n}{101} \cdot b_{max} \rfloor$ and $\lfloor \frac{m}{101} \cdot b_{max} \rfloor$ respectively, where $n$ and $m$ denote the number of songs by female and male singers among the 100 songs recorded respectively, and $b_{max}$ denotes the maximum scale of the consumer's own positive belief toward the song by the female and male singers, which was set to 6 in default based on TABLE I. In addition, the cost of selecting consumer $u$ as a seed to spread the product information ($w(u)$) was randomly chosen from the interval $[1, 10]$ in default based on [5], and the weight of the change of consumer $u$'s own positive belief on consumer $u$'s behavioral intention toward the product ($W_1(u)$) was randomly chosen from the interval $[0.8, 1]$ for each consumer $u$ in default based on [11]. Since Algorithm 1 has an approximation ratio of around $(1 - \frac{1}{e})$ with high probability for the *B*²*IM* problem by Theorem 3, $y$ was set to 3 in default. Recall that in Algorithm 1, given a set $V' \subseteq V$, we need to perform the random process enough times ($N$) to bound the difference between $\hat{F}(V')$ and $F_{\mu}(V')$. Like [5], through the simulation, $N$ was set to 1000 for NetS, BlogCatalog, and Facebook in the remaining simulations.

## B. Simulation Results

We first study the effect of the different values of budget on the total behavioral intentions of the influenced consumers in Fig. 1. Subsequently, for showing that Algorithm 1 also outperforms the PMCE method for the *PM*²*A* problem [1], we study the effect of the different values of budget on the total profit of the influenced consumers in Fig. 2 by modifying our objective function and adjusting several settings in our simulations. The objective function is to ask for a set of seeds under a fixed budget, such that the total expected profit of the consumers influenced by the selected seeds and the rumors with all kinds of products is maximized. We omit the settings of the simulations due to the page limit. The settings could be referred to IV-B of our technical report [14]. In addition, we also add several greedy algorithms, including the greedy on cost algorithm (GCA), the greedy on profit algorithm (GPA), and the greedy on the ratio of profit to cost algorithm (GRPCA) to be compared. The GCA, GPA, and GRPCA iteratively expand the seed set by including the seed with the minimum cost, the maximum profit, and the maximum ratio of profit to cost, respectively, until the budget is exceeded.

In the different values of budget on the total behavioral intentions, Algorithm 1 outperforms GCA, GBA, GRBCA, and PMCE. This is reasonable because Algorithm 1 avoids selecting the rumors as seeds and selects the consumers with the greatest incremental average total estimated behavioral intentions of the influenced consumers (over $N$ random processes) as seeds. In addition, among all traces, the average contact probability of two consumers is the greatest; thus, the average behavioral intention of a consumer is the greatest using BlogCatalog. Therefore, among all traces, each algorithm has the greatest total behavioral intentions of the influenced consumers using BlogCatalog. Similarly, in the different values of budget on the total profit, Algorithm 1 also outperforms GCA, GPA, GRPCA, and PMCE. This is because Algorithm 1 selects a greater number of seeds under a certain budget and avoids selecting the rumors as seeds.

Furthermore, as the value of budget increases, the number of seeds that can be selected increases; thus, the performance of each algorithm increases, as shown in Figs. 1 and 2, respectively. It was noted that as the value of budget increases from 300 to 500, the performance of Algorithm 1 insignificantly increases, as shown in Figs. 1(b) and 2(b), respectively. This is because when the value of budget is around 300, most of the consumers with great incremental average total estimated behavioral intentions and profit of the influenced consumers have been already selected as seeds by Algorithm 1 in the simulations using BlogCatalog, respectively.

## V. Related Works

The **B**udgeted **M**aximum **C**overage (BMC) problem [15] asks for a collection of sets $S' \subseteq S$ such that the total costs of the elements in $S'$ does not exceed the given budget constraint, and the total weights of the elements covered by $S'$ is maximized. Khuller et al. [15] showed that the *BMC* problem is NP-hard and proposed an approximation algorithm for the *BMC* problem. In the *BMC* problem, the value of the objective function (i.e., the total weights of the elements covered by the selected collection of sets) can be evaluated exactly. By contrast, the value of the objective function (i.e., the total expected behavioral intentions of the consumers influenced by the selected seeds and the rumors) in our problem cannot be determined in polynomial time. In [16], Krause et al. showed that, if the value of the objective function in a problem like the *BMC* problem could be estimated with a bounded error, then the algorithm for the *BMC* problem could be applicable to the problem. However, no method of evaluating the value of the objective function with a bounded error is presented in [16]. In this paper, the algorithm for our problem is obtained from the algorithm for the *BMC* problem by proposing a method of evaluating the value of the objective function for our problem with a bounded error. In [5], Tang et al. proposed the **B**udgeted **I**nformation **P**ropagation **M**aximization (BIPM) problem, which is a variant of **I**nfluence **M**aximization (IM) problem. They showed that the *BIPM* problem is NP-hard and adopted the Hill-Climbing method with an approximation ratio of around $1/2 \cdot (1 - \frac{1}{e})$. However, since the objective function of our problem is different from that of the *BIPM* problem, the Hill-Climbing method for the *BIPM* problem cannot be employed for our problem. In [7], Tong et al. proposed the **R**umor **B**locking (RB) problem, which is also a variant of *IM* problem. They showed that the *RB* problem is NP-hard and proposed an approximation algorithm. Like the *BIPM* problem, the algorithm for the *RB* problem cannot be employed for our problem due to their different objective functions.

## VI. Conclusion

In this paper, we investigated the budgeted behavioral intention maximization problem, termed B$^2$IM, in an online social network. To the best of our knowledge, the *B$^2$IM* problem is the first of the influence maximization problems with multiple products to take the beliefs of other persons and the rumors into account for predicting the consumer's behavioral intention in the literature so far. We showed that the *B$^2$IM* problem is NP-hard and proposed an algorithm (Algorithm 1) with an approximation ratio of around $(1 - \frac{1}{e})$ with high probability for the *B$^2$IM* problem. We conducted simulations to evaluate the performance of Algorithm 1 using the real traces and synthesis data. The simulation results showed that Algorithm 1 outperforms several greedy algorithms for the *B$^2$IM* problem.

## References


[1] H. Zhang, H. Zhang, A. Kuhnle, and M. T. Thai, “Profit maximization for multiple products in online social networks,” in *IEEE INFOCOM*, 2016.

[2] P. Domingos and M. Richardson, “Mining the network value of customers,” in *ACM SIGKDD*, 2001.

[3] M. Richardson and P. Domingos, “Mining knowledge-sharing sites for viral marketing,” in *ACM SIGKDD*, 2002.

[4] D. Kempe, J. Kleinberg, and E. Tardos, “Maximizing the spread of influence through a social network,” in *ACM SIGKDD*, 2003.

[5] S. Tang, J. Yuan, X. Mao, X. Y. Li, W. Chen, and G. Dai, “Relationship classification in large scale online social networks and its impact on information propagation,” in *IEEE INFOCOM*, 2011.

[6] D. T. Nguyen, S. Das, and M. T. Thai, “Influence maximization in multiple online social networks,” in *IEEE GLOBECOM*, 2013.

[7] G. Tong, W. Wu, L. Guo, D. Li, C. Liu, B. Liu, and D.-Z. Du, “An efficient randomized algorithm for rumor blocking in online social networks,” in *IEEE INFOCOM*, 2017.

[8] C. Budak, D. Agrawal, and A. E. Abbadi, “Limiting the spread of misinformation in social networks,” in *ACM WWW*, 2011.

[9] X. He, G. Song, W. Chen, and Q. Jiang, “Influence blocking maximization in social networks under the competitive linear threshold model,” in *SIAM SDM*, 2012.

[10] Y. Ping, Z. Cao, and H. Zhu, “Sybil-aware least cost rumor blocking in social networks,” in *IEEE GLOBECOM*, 2014.

[11] L. G. Schiffman and L. L. Kanuk, *Consumer Behavior*. Prentice HalL, 2003.

[12] S. Martin E. P. and C. Mihaly, *Positive Psychology: An Introduction*. American Psychologist, 2000.

[13] “Last.fm,” [Online]. http://www.last.fm.

[14] [Online]. https://www.dropbox.com/s/4nakwxpq5k1grbn/BBIM.pdf?dl=0.

[15] S. Khuller, A. Moss, and J. Naor, “The budgeted maximum coverage problem,” in *ACM IPL*, 1999.

[16] A. Krause and C. Guestrin, “A note on the budgeted maximization of submodular functions,” in *Technical Report. CMU-CALD-05-103*, 2005.

[17] W. Gao, G. Cao, T. L. Porta, and J. Han, “On exploiting transient social contact patterns for data forwarding in delay-tolerant networks,” in *IEEE TMC*, 2013.