Coupon Advertising in Online Social Systems: Algorithms and Sampling Techniques

Guangmo (Amo) Tong, Student Member, IEEE, Weili Wu, Member, IEEE, and Ding-Zhu Du

Abstract—Online social systems have become important platforms for viral marketing where the advertising of products is carried out with the communication of users. After adopting the product, the seed buyers may spread the information to their friends via online messages e.g. posts and tweets. In another issue, electronic coupon system is one of the relevant promotion vehicles that help manufacturers and retailers attract more potential customers. By offering coupons to seed buyers, there is a chance to convince the influential users who are, however, at first not very interested in the product. In this paper, we propose a coupon based online influence model and consider the problem that how to maximize the profit by selecting appropriate seed buyers. The considered problem herein is markedly different from other influence related problems as its objective function is not monotone. We provide an algorithmic analysis and give several algorithms designed with different sampling techniques. In particular, we propose the RA-T and RA-S algorithms which are not only provably effective but also scalable on large datasets. The proposed theoretical results are evaluated by extensive experiments done on large-scale real-world social networks. The analysis of this paper also provides an algorithmic framework for non-monotone submodular maximization problems in social networks.

Index Terms—Social networks, social advertising, approximation algorithm.

1 INTRODUCTION

Online social networks have become important platforms for viral marketing as they allow fast spread of brand-awareness information. As a typical diffusion process, the influence starts to spread from seed users and then goes round by round from active users to inactive users. The classic influence maximization (IM) problem aims to find the seed users that can maximize the number of influenced nodes. In the seminal work [1], Kempe et al. propose the triggering model and formulate the IM problem as a combinatorial optimization problem. A set function \( h() \) over a ground set \( \mathcal{N} \) is submodular if

\[
h(\mathcal{N}_1 \cup \{v\}) - h(\mathcal{N}_1) \geq h(\mathcal{N}_2 \cup \{v\}) - h(\mathcal{N}_2)
\]

for any \( \mathcal{N}_1 \subseteq \mathcal{N}_2 \subseteq \mathcal{N} \) and \( v \in \mathcal{N} \setminus \mathcal{N}_2 \). It turns out that IM is a size-constraint monotone submodular maximization problem and therefore the natural greedy algorithm yields a good approximation [2]. In the past decade, a huge body of works has been devoted to influence related problems such as rumor blocking, active friending and effector detection, just to name a few. The process of influence diffusion generalizes the spread of many kinds of information such as news, ideas and advertisements. In this paper, we crystallize the influence diffusion process and in particular consider the advertising of product in online social systems.

Influenced node Vs. Adopter. Whereas IM is motivated by viral marketing, maximizing the influence may not result in a high profit because an influenced node is not necessarily to be an adopter of the product. As mentioned in Kalish’s work [3], the adoption of a new product is characterized by two steps: awareness and adoption. The traditional influence diffusion models characterize the spread of brand-awareness information but they do not take account of the issue of product adoption. Among the factors that affect the customers’ behavior, the price is undoubtedly the most relevant one. In this paper, we consider a simplified adoption process by assuming that, once being aware of the product, a customer becomes an adopter if they believe the product is fairly priced, and only an adopter can bring more adopters by passing on the product information further to their friends [4].

Coupon Systems. A real advertising campaign is often conducted with promotional tools such as coupon, reward, free sample and etc. Coupon has been demonstrated to be an effective promotional method as it creates short-term excitement [4] and therefore immediately impacts the customers’ purchase behavior. Hence, in online social marketing, coupon enables the seller to convince the influential users who were not able to be adopters without a coupon. In this paper, we consider the marketing process where the seed buyers are offered with coupons. From the perspective of sellers, coupons are not free because they essentially lower the price of the product. Therefore, increasing the number of seed buyers does not necessarily bring more profit, which leads the non-monotonicity of the objective function of the considered problem.

This paper. By integrating coupon system into the general triggering model [1], we in this paper propose the triggering-coupon (T-C) model for product advertising process with coupon in social networks. Based on the T-C model, we consider the problem of selecting seed buyers such that the profit can be maximized, which is termed as

1. The analysis in paper can be applied to other models where the adoption process is independent from the influence process.
profit maximization (PM) problem. As discussed later, the difficulty in solving the PM problem lies in two aspects. First, as a function of the seed users, the profit under the T-C model is submodular but not monotone increasing anymore. Consequently, the well-known greedy algorithm does not work and we have to look for other approximation algorithms for unconstrained submodular maximization (USM). With the very recent research progress, USM admits the realization-based profit maximization (PM) problem. As discussed later, the existing USM algorithms assume that there is an oracle of the objective function while, unfortunately, computing the profit under the T-C model is a #P-hard problem. Since the exact value of the objective function is hard to compute, one naturally seeks help from its estimates obtained by sampling. 

2. The PM problem is NP-hard and therefore approximation algorithms are the best possible solutions with performance guarantees.

3. As introduced later in Sec. 2, the so-called realization is a derandomization of a triggering model.

In this subsection, we give a formal description of the considered model.

### Influence Model

The structure of a social network is given by a directed graph $G = (V, E)$ where $V = \{v_1, ..., v_n\}$ denotes the set of users and $E$ is edge set. Let $n$ and $m$ be the number of users and edges, respectively. If there is an edge $(u, v)$ in $E$, we say $u$ (resp. $v$) is an in-neighbor (resp. an out-neighbor) of $v$ (resp. $u$). Let $N^+_v$ and $N^-_v$ be the set of the out-neighbors and in-neighbors of $v$, respectively. From the view of influence diffusion, we say a user is active if they successfully receive the target information from their neighbors. Depending on the ways of influence diffusion, many influence models have been developed. In this paper, we consider the general triggering model of which the famous IC and LT models are special cases.
Definition 1 (Triggering Model). An influence model is a triggering model if its diffusion process can be equivalently described as follows. Each node \( v \) independently chooses a random "triggering set" \( T_v \) according to some distribution over subsets of its in-neighbors. To start the process, we target a set of users for initial activation. After the initial iteration, an inactive node \( v \) becomes active at time step \( t \) if it has an in-neighbor in its chosen triggering set \( T_v \) that is active at time step \( t - 1 \).

By Def. 1, a triggering model is given by a pair \((G, D)\) where \( G = (V = \{v_1, …, v_n\}, E) \) is the network structure and \( D = \{D_{v_1}, …, D_{v_n}\} \) is a set of distributions. In particular, \( D_{v_i} \) specifies the distribution over subsets of the in-neighbors of user \( v_i \). As two special triggering models, IC model and LT model are defined as follows.

Definition 2 (IC model). The IC model assumes that an active user has one chance to activate each of their neighbors with a certain probability and the activations of different pairs of users are independent. Under this model, each edge \((u, v)\) is assigned a real number \( p(u, v) \in (0, 1] \) which is the propagation probability from \( u \) to \( v \). Under the IC model, the distribution \( D_{v_i} \) is given by the probability of the edges. For a node \( v \), an in-neighbor \( u \) of \( v \) has the probability \( p(u, v) \) to appear in the triggering set \( T_v \).

Definition 3 (LT model). Under the general threshold model, a user becomes active if they have received sufficient influence from their neighbors. Specifically, each edge \((u, v)\) has a weight \( w(u, v) > 0 \) and each user \( v \) holds a threshold \( \theta_v > 0 \). The user \( v \) becomes active if \( \sum_{u \in A(v)} w(u, v) \geq \theta_v \) where \( A(v) \) is the set of active in-neighbors of user \( v \). For the linear threshold model, we further require that, for each user \( v_i \), \( \sum_{u \in N^- v_i} w(u, v_i) \leq 1 \) and the threshold \( \theta_{v_i} \) is selected uniformly at random from \([0, 1]\). As shown in 1, taking LT model as a triggering model, the triggering set \( T_v \) of a node \( v \) is decided as follows: \( v \) selects at most one of its in-neighbors at random where an in-neighbor \( u \) has the probability of \( w(u, v) \) to be selected and with probability \( 1 - \sum_{u \in N^- v_i} w(u, v_i) \) that \( v \) does not pick any edges.

Note that the diffusion process under triggering model is stochastic as the triggering sets are generated at random. An important property of triggering model is shown as follows.

Theorem 1. Under a triggering model, the expected number of active nodes is a submodular function with respect to the seed set.

For most of the triggering models, the process defined in Def. 1 is not the real diffusion process. For example, instead of generating the triggering sets in advance, the real diffusion process of IC model goes round by round from inactive users to active users until no user can be further activated.

Coupon System. In this paper, we consider the very basic coupon system where a product is associated with two features, \( P \) and \( C \), which, respectively, denote the price of the product and the value of the coupon. Let \( I_v \) be the intrinsic value held by \( v \) for the considered product. After receiving the information of the product, a user becomes an adopter if they believe the product is fairly priced, i.e., \( I_v \geq P \). Without coupons, only the users with \( I_v \geq P \) can be considered as seed users, while once given a coupon the user \( v \) with \( I_v \geq P - C \) can also be the seed adopters. In equivalent, one may say the intrinsic value of user \( v \) increases to \( I_v + C \) if a coupon is offered. In a social network, there can be some influential users whose intrinsic value is less than the price of the product, but coupon allows the seller to activate such powerful users and consequently raises the total profit. We denote by \( r = \frac{P-C}{P} \) the ratio of the price \( P \) to \( P - C \). One can take \( r \) as a normalized discount ratio of the product.

**T-C Model.** Combining the influence model with coupon system, we have the triggering-coupon (T-C) model of which the marketing process unfolds as follows.

- Initially all the nodes are inactive and a seed set \( S \subseteq V \) is decided.
- At each time step \( t \), an inactive node \( v \) becomes active if (a) \( v \) is activated by its neighbors and (b) \( I_v \geq P \).
- The process terminates when there is no user can be further activated.

Under the T-C model, the active users correspond to the adopters of the product. To distinguish it from the pure influence model, we will use adopter instead of active user in the rest of this paper. With the coupon system, the diffusion process is slightly different from the traditional spread of influence. However, the T-C model itself forms a new triggering model.

Lemma 1. A triggering model combined with the coupon system yields a new triggering model.

Proof. Let \( D = \{D_{v_1}, …, D_{v_n}\} \) be the set of the distributions given by the triggering model. With the coupon system, the T-C model can be taken as another triggering model where the new triggering distribution \( D_v \) of each node \( v \) is defined as

\[
D_v = \begin{cases} 
D_v & \text{if } I_v \geq P \\
\text{empty distribution} & \text{otherwise.}
\end{cases}
\]

where the empty distribution always returns the empty set. That is, for a node \( v \), if \( I_v \geq P \) then the distribution remains unchanged and \( T_v \) is decided according to \( D_{v} \), otherwise, \( T_v \) is always the empty set as \( v \) can never be an adopter.

An instance of the T-C model is called a T-C network. A T-C network \( \Upsilon \) consists of a graph topology, the price of the product, the value of the coupon, and a distribution of triggering set for each node defined in Eq. (2). In this paper, we use the general T-C model for analysis, and adopt the IC-coupon (IC-C) and LT-coupon (LT-C) models for experiments.

### 3.2 Problem Definition

Given a set \( S \) of seed users and a T-C network \( \Upsilon \), let \( \pi(S) \) be the expected number of adopters under \( S \). For a user \( v \), if \( P > I_v + C \), then \( v \) cannot be an adopter even if offered with a coupon. Without loss of generality, we can remove such nodes from the network in advance and therefore assume
Lemma 2. \( f \) gives the key property of our objective function. The following lemma 
\[ \text{Lemma 3} \] Given an objective function \( h() \) over a ground set \( N \), the set produced by Alg. 1 is a \( \frac{1}{2} \)-approximation provided that \( h() \) is submodular and \( h(\emptyset) + h(\bar{N}) \geq 0 \).

The existing algorithms assume the availability of an oracle of the objective function. However, it turns out that computing the exact value of \( \pi(S) \) is \#P-hard [13]. As seen in the recent works on influence related problems, the main difficulty in solving such problems has shifted from approximation algorithm design to the estimating of the expected influence \( \pi(S) \). Since the value of \( f() \) is hard to compute, Alg. 1 cannot be directly applied. Given that the diffusion process is stochastic, for a certain seed set \( S \), \( f(S) \) can be naturally estimated by sampling. In this paper, we discuss several sampling methods and show how they can be integrated into the existing approximation algorithms. In particular, we use the Buchbinder’s algorithm (Alg. 1) for illustration. For a small value \( \epsilon \in (0, \frac{1}{2}) \) and a large value \( N > 0 \), an algorithm is a \( (\frac{1}{2} - \epsilon, 1 - \frac{2\epsilon}{N}) \)-approximation if it is able to produce a solution with an approximation ratio of \( \frac{1}{2} - \epsilon \) with at least 1 - \( 1/N \) probability. We aim to design the algorithms with such a performance guarantee and assume that \( \epsilon \) and \( N \) are fixed in the rest of this paper.

Sampling methods can be generally classified into two categories, forward sampling and reverse sampling. As the name suggests, the forward sampling methods obtain an estimate of \( \pi(S) \) by directly running the diffusion process while the reverse sampling methods simulate the diffusion process in the reverse direction. The algorithms designed based on such sampling techniques are shown in next two sections.

4 Forward Sampling

Before discussing the PM problem, let us first consider the general submodular maximization problem in the case that the exact function value is hard to compute. For Alg. 1, suppose that an estimate \( \tilde{h}() \) of \( h() \) is employed whenever \( h() \) is called. One can expect that the approximation ratio can be arbitrarily close to \( \frac{1}{2} \) as long as the estimates are sufficiently accurate. The modified algorithm is shown in Alg. 2 which slightly differs from Alg. 1 and takes a lower bound \( L^* \) of the optimal value as an extra input. The following key theorem indicates the relationship between the precision of the estimates and the approximation ratio.

Theorem 2. For any instance \((h(), N = \{v_1, \ldots, v_n\})\) and any \( \epsilon \in (0, \frac{1}{2}) \), Alg. 2 is a \( (\frac{1}{2} - \epsilon) \)-approximation, if
\[ |\tilde{h}(S) - h(S)| \leq \frac{\epsilon}{n} \cdot L^* \]
holds for each \( S \) inspected by Alg. 2, where \( \tilde{h}(S) \) is the estimate of \( h(S) \) and \( L^* \) is a lower bound of the optimal value of \( h() \).

Proof. See Appendix B.1

5. Strictly speaking, Alg. 1 provides an expected \( \frac{1}{2} \)-approximation as it is a randomized algorithm. For simplicity, we will take such a solution as an exact \( \frac{1}{2} \)-approximation. This does not affect the fundamental issues of our analysis.

6. In [5], it is assumed that \( h() \) is non-negative while in fact we only need that \( h(\emptyset) + h(\bar{N}) \geq 0 \). The proof is similar to that of Theorem 2 shown later. A detailed discussion is provided by J. Tang et al. [21].

7. \( \frac{1}{2} \) is the best possible ratio for general unconstrained submodular maximization [22].
Algorithm 2 Forward Sampling Framework

1: Input: $(\tilde{h}(\cdot),N = \{v_1,...,v_n\},\epsilon,L^*)$;
2: $X_0 \leftarrow 0$, $Y_0 \leftarrow N$;
3: for $i = 1:n$ do
4: \quad $\tilde{a}_i \leftarrow \tilde{h}(X_{i-1} \cup \{v_i\}) - \tilde{h}(X_{i-1}) + \frac{2n}{\epsilon} \cdot L^*$;
5: \quad $\tilde{b}_i \leftarrow \tilde{h}(Y_{i-1} \setminus \{v_i\}) - \tilde{h}(Y_{i-1}) + \frac{2n}{\epsilon} \cdot L^*$;
6: \quad $\tilde{a}_i \leftarrow \max\{\tilde{a}_i,0\}$; $\tilde{b}_i \leftarrow \max\{\tilde{b}_i,0\}$;
7: \quad $\text{rand} \leftarrow$ a random number from 0 to 1 generated in uniform;
8: \quad if $\text{rand} \leq \tilde{a}_i/\tilde{a}_i + \tilde{b}_i$ or $\tilde{a}_i + \tilde{b}_i = 0$ then
9: \quad \quad $X_i \leftarrow X_{i-1} \cup \{v_i\}$, $Y_i \leftarrow Y_{i-1}$;
10: \quad else
11: \quad \quad $X_i \leftarrow X_{i-1}$, $Y_i \leftarrow Y_{i-1} \setminus \{v_i\}$;
12: \quad Return $X_n$.

Alg. 2 provides a framework for maximizing a submodular function $\tilde{h}(\cdot)$ when the exact value of $\tilde{h}(\cdot)$ is hard to compute and an accurate estimator $\tilde{h}(\cdot)$ is obtainable. In the rest of this section, we introduce two estimators of $\pi(\cdot)$ and apply the above framework to solve the PM problem.

4.1 SPM Algorithm

A straightforward way to estimate $\pi(S)$ is to use Monte Carlo simulation. That is, for a given seed set $S$, we simulate the diffusion process under the considered T-C model to obtain the samples of $\pi(S)$ and take the sample mean as the estimator. For a seed set $S \subseteq V$, let $\bar{\pi}_l(S)$ be the sample mean obtained by running $l$ times of simulations and define that $\bar{f}_l(S) = \frac{P \cdot \bar{\pi}_l(S) - C \cdot |S|}{n}$.

Algorithm 3 Realization-based Profit Maximization

1: Input: $\Pi, \epsilon$ and $N$;
2: $l \leftarrow \frac{\ln 8 + \ln n + \ln N}{(n^2 + \epsilon n)/2}$;
3: Generate $l$ realizations $G_l = \{g_1,...,g_l\}$ by sampling triggering sets for each node;
4: $V^* \leftarrow$ Forward Sampling Framework($\bar{f}_{G_l}(\cdot), V, \epsilon, f(V)$);
5: Return $V^*$.

Theorem 3. For any $\epsilon \in (0,\frac{1}{2})$ and $N > 0$, to achieve a $(\frac{1}{2} - \epsilon)$-approximation with probability at least $1 - 1/N$, the SPM algorithm demands the running time of $O(mn^3 \ln n)$.

Proof. Since there are $4n$ sets inspected by Alg. 2 by Lemma 4 and the union bound, $|\bar{f}_l(S) - f(S)| \leq \frac{\epsilon}{n} f(V)$ holds for every inspected set with probability at least $1 - 1/N$ provided that $l \geq \delta_0$. Therefore, by Theorem 3 the approximation ratio is $\frac{1}{2} - \epsilon$ if $l = \delta_0$. Because $\delta_0$ simulations are used for each estimating and there are totally $4n$ estimated function values, the number of executed simulations is $4\delta_0 n$. The total running time is therefore $O(4\delta_0 nm)$ as one simulation costs $O(m)$ in the worst case.

Note that executing one simulation for $f(S)$ costs $O(m)$ in the worst case but in fact it is proportional to $f(S)$ which is usually much smaller than $m$.

4.2 RPM Algorithm

Now let us consider another estimator of $f()$. As mentioned in Sec. 3 the real diffusion process of a triggering model, namely IC and LT model, is equivalent to the one described in Def. 4. Therefore, instead of simulating the real diffusion process, one can obtain an estimate of $\pi(S)$ by sampling the triggering set of each node. When the triggering set of each node is determined, we obtain a realization of the given network.

Definition 4 (Realization). For a T-C network $\Pi$, a realization $g = \{T_1,...,T_{\gamma}\}$ is a collections of triggering sets of the nodes where $T_{\gamma}$ is sampled according to distribution given by $\Pi$. Equivalently, a realization can be taken as a deterministic T-C network.

The following lemma shows the number of simulations needed to meet the accuracy required by Eq. (5).

Lemma 4. For a certain set $S \subseteq V$ and $\frac{1}{2} > \epsilon > 0$, we have

$$P_r[|\bar{f}_l(S) - f(S)| > \frac{\epsilon f(V)}{n}] \leq \frac{1}{4nN},$$

if $l \geq \delta_0$ where

$$\delta_0 = \frac{4\ln 8 + \ln n + \ln N}{(2n^2 + \epsilon n)/2}.$$  

Proof. It can be proved by directly applying the Chernoff bound. By rearrangement, $P_r[|\bar{f}_l(S) - f(S)| > \frac{\epsilon f(V)}{n}]$ is equal to $P_r[l \cdot \bar{\pi}_l(S) - l \cdot \frac{\pi(S)}{n} > l \cdot \frac{\pi(S)}{n} \cdot \frac{\epsilon f(C)}{P \cdot \pi(S)}]$. Applying Chernoff bound, this probability is no larger than

$$2 \exp\left(-\frac{l \cdot \frac{\pi(S)}{n} \cdot \frac{\epsilon f(C)}{P \cdot \pi(S)}}{2} \left(1 + \frac{\epsilon f(C)}{P \cdot \pi(S)}\right)^2\right).$$

By Eq. (5) and $\pi(S) \leq n$, this probability is at most $\frac{1}{4nN}$.

Given a seed set $S$, the following two processes are equivalent to each other with respect to $\pi(S)$.

- Process a. Execute the real diffusion process on $G$.
- Process b. Randomly generate a realization $g$ and execute the diffusion process on $g$ according to Def. 4.

For a seed set $S$, let $\bar{\pi}_g(S)$ be the number of adopters under $S$ on realization $g$ and correspondingly we define that $\bar{f}(g,S) = P \cdot \bar{\pi}_g(S) - C \cdot |S|$. According to Lemma 5 one can estimate $\pi(S)$ by repeatedly executing the Process b in Lemma 5 and calculating the mean of $\bar{\pi}_g(S)$. Given a set of $l$ realizations $G_l = \{g_1,...,g_l\}$ generated independently, for a seed set $S$, let

$$\bar{\pi}_G(S) = \frac{\sum_{i=1}^l \bar{\pi}_{g_i}(S)}{l}.$$
Algorithm 4 RA Set

1: Input: \( T \);
2: Select a node \( v \) from \( V \) uniformly in random;
3: Sample a realization \( g \) of \( T \);
4: Return the nodes that are reverse-adopted-reachable to \( v \) in \( g \);

and

\[
\hat{f}_{\tilde{G}_i}(S) = \frac{1}{l} \sum_{i=1}^{l} \hat{f}(g_i, S) = P \cdot \tilde{\pi}_{G_i}(S) - C \cdot |S|.
\] (9)

Now we have another estimator \( \hat{f}_{\tilde{G}_i}(S) \) that can be utilized in the framework of Alg. 2. The obtained algorithm is shown in Alg. 3 which is named as realization-based profit maximization (RPM) algorithm. According to Lemma 2 in order to obtain a good approximation one should generate sufficient number of realizations such that Eq. (8) can be satisfied. In fact, the number of total realizations required herein is the same as that of the simulations required in the RPM algorithm, shown as follows.

Lemma 6. Suppose \( l \) realizations are generated for estimating. For any \( \epsilon \in (0, \frac{1}{2}) \) and \( N > 0 \), with probability at least \( 1 - \frac{1}{N} \), \( |\hat{f}_{\tilde{G}_i}(S) - f(S)| \leq \frac{\epsilon}{l} f(V) \) holds for every \( S \) inspected by Alg. 3 if \( l \) is not less than \( \delta_0 \) defined in Eq. (7).

Proof. The proof is same as that of Lemma 4.

According to Def. 1, generating one realization costs \( O(m) \) time. For a certain realization \( g \) and a seed set \( S \), computing \( \hat{f}(g, S) \) can be done in \( O(n + m) \) time by the breadth-first search, and therefore \( \hat{f}_{\tilde{G}_i}(S) \) can be computed in \( O(l(m+n)) \). Hence, Alg. 3 totally costs \( O(lm + ln(m+n)) \). Setting \( l \) to be \( \delta_0 \) gives the running time of \( O(mn^3 \ln n) \).

Theorem 4. For any \( \epsilon \in (0, \frac{1}{2}) \) and \( N > 0 \), the RPM algorithm provides a \( \frac{1}{2} - \epsilon \) approximation with a probability at least \( 1 - 1/N \) with the running time of \( O(mn^3 \ln n) \) with respect to \( m \) and \( n \).

One can see that the running times of SPM algorithm and RPM algorithm are in the same order with respect to \( m \) and \( n \). However, the actual number of operations executed by SPM algorithm is about \( 4lmn \) while the RPM algorithm roughly conducts \( lmn \) operations. Therefore, when the graph is large, the SPM algorithm is likely to be more efficient. For practical implementation, because one realization \( g \) can be used to estimate \( f(S) \) for any seed set \( S \), we can generate the realizations in advance and use them later in real-time computation, which is another advantage of the RPM algorithm.

5 Reverse Sampling

In this section, we present two algorithms which are able to provide the approximation ratio of \( \frac{1}{2} - \epsilon \) but require significantly less running time than RPM and SPM do. Such algorithms are designed based on the technique of reverse sampling which is firstly invented by Borgs. et al. for the IM problem. In the reserve sampling framework, the key object is the reverse adopted-reachable set.

Definition 5. For a node \( v \) and a realization \( g = \{ T_1, ..., T_n \} \) of a T-C network \( T \), a node \( v^* \) is reverse-adopted-reachable to \( v \) in \( g \) if there is a sequence of nodes \( u_1, ..., u_k \) such that \( u_1 = v, u_k = v^* \) and \( u_i + 1 \in T_{u_i} \) for \( 1 \leq i \leq k - 1 \).

Definition 6. (Reverse Adopted-reachable (RA) Set) For a T-C network \( T \), an RA set is a set of nodes randomly generated by Alg. 4. As shown therein, we first select a node \( v \) in \( V \) uniformly in random and then sample a realization \( g \) of \( T \). Finally, it returns the nodes that are reverse-adopted-reachable of \( v \) in \( g \).

For a seed set \( S \subseteq V \) and a RA set \( R \), let \( x(R, S) \) be a variable such that

\[
x(S, R) = \begin{cases} 1 & \text{if } R \cap S \neq \emptyset \\ 0 & \text{otherwise} \end{cases}
\]

Since \( R \) is generated randomly, \( x(S, R) \) is a random variable. The following lemma shows that \( n \cdot x(S, R) \) is an unbiased estimate of \( \pi(S) \).

Lemma 7. For any seed set \( S \), \( E[x(S, R)] = \frac{\pi(S)}{n} \).

Proof. See Appendix B.2

Let \( \mathcal{R}_l = \{ R_1, ..., R_l \} \) be a set of \( l \) RA sets independently generated by Alg. 4 and let

\[
F(\mathcal{R}_l, S) = P \cdot n \cdot \sum_{i=1}^{l} x(R_i, S) - C \cdot |S|.
\]

The following problem plays an important role in solving the PM problem.

Problem 2 (Cost Set Cover Problem). Given a collection \( \mathcal{R}_l \) of RA sets, find a set \( S \) such that \( F(\mathcal{R}_l, S) \) is maximized.

Lemma 8. Given \( \mathcal{R}_l \), \( F(\mathcal{R}_l, S) \) is a submodular function with respect to \( S \).

Proof. The first part of \( F(\mathcal{R}_l, S) \) is exactly the set cover problem and its second part is linear to \( |S| \).

Since \( F(\mathcal{R}_l, S) \) is submodular and \( F(\mathcal{R}_l, \emptyset) + F(\mathcal{R}_l, V) = 0 + (P - C) \cdot n \geq 0 \), by Lemma 3 Alg. 4 provides an \( \frac{1}{2} \)-approximation to Problem 2.

By Lemma 7, \( F(\mathcal{R}_l, S) \) is an unbiased estimate of \( f(S) \). Hence, the set \( S \) that can maximize \( F(\mathcal{R}_l, S) \) can intuitively be a good solution to the PM problem, which is the main idea to solve the PM problem by the reverse sampling.

5.1 RA-T Algorithm

In this section, we present the first algorithm designed with the reverse sampling, as shown in Alg. 5. In this algorithm, we first decide a threshold \( l \), generate \( T \) RA sets, and then run Alg. 4 to solve Problem 2. We term this algorithm as the RA-T algorithm where RA-T stands for reverse adopted-reachable (RA) set with a threshold (T). As shown in the following, for any \( \epsilon \in (0, \frac{1}{2}) \), the RA-T algorithm produces a \( \left( \frac{1}{2} - \epsilon \right) \)-approximation to the PM problem with a high probability, as long as \( l \) is sufficiently large.

8. The RA set defined in this paper is analogous to the hyperedge in \([6]\), the reverse reachable set in \([7]\) and the random R-tuple in \([23]\).
Algorithm 5 RA-T Algorithm

1: Input: T, \(\epsilon\) and \(N_1\)
2: \(\delta_1 \leftarrow \frac{\ln N + n \ln 2}{\epsilon_1^2 r^2}\)
3: \(\delta_2 \leftarrow \frac{2 \ln N}{\epsilon_2^2 r^2}\)
4: \(l = \max(\delta_1, \delta_2)\)
5: Independently generate a collection \(\mathcal{R}_l = \{R_1, ..., R_l\}\) of \(l\) RR sets by Alg. 4.
6: \(V^* \leftarrow \) Buchbinder’s Algorithm \((F(\mathcal{R}_l, S), V)\)
7: Return \(V^*\);

Let \(\epsilon_1 > 0\) and \(\epsilon_2 > 0\) be some adjustable parameters where
\[
\epsilon_1 + \frac{1}{2} \epsilon_2 = \epsilon. \tag{10}
\]
Set that
\[
\delta_1 = \frac{\ln N + n \ln 2}{\epsilon_1^2 r^2}, \tag{11}
\]
and
\[
\delta_2 = \frac{2 \ln N}{\epsilon_2^2 r^2}. \tag{12}
\]

Lemma 9. Given a set \(\mathcal{R}_l\) of \(l\) RA sets generated independently by Alg. 4 with probability at least \(1/N\), there exists some \(S \in \mathcal{V}\) such that
\[
F(\mathcal{R}_l, S) - f(S) > \epsilon_1 \cdot f(\mathcal{V}_{opt}),
\]
if \(l \geq \delta_1\).

Proof. See Appendix B.3

Lemma 10. Given a set \(\mathcal{R}_l\) of \(l\) RA sets generated by Alg. 4 with probability at most \(1/N\)
\[
F(\mathcal{R}_l, \mathcal{V}_{opt}) - f(\mathcal{V}_{opt}) < -\epsilon_2 \cdot f(\mathcal{V}_{opt})
\]
if \(l \geq \delta_2\).

Proof. See Appendix B.4

Algorithm 6 RA-S Algorithm

1: Input: T, \(N, \epsilon\), \(k\) and \(\epsilon_3\)
2: Obtain \((\delta_1^*, \delta_2^*, \epsilon_1, \epsilon_2)\) by solving Eq System 1
3: \(l = \delta_1^*, l^* = \delta_3\) and \(\mathcal{R}_l^* = \emptyset\)
4: while \(l \leq 2 \cdot \delta_1^*\) do
5: Generate \(l\) RA sets by Alg. 4 and add them into \(\mathcal{R}_l^*\);
6: if \(l \geq \delta_1^*\) then
7: \(V^* \leftarrow\) Buchbinder’s Algorithm \((F(\mathcal{R}_l^*, S), V)\);
8: Return \(V^*\);
9: else
10: \(V^* \leftarrow\) Buchbinder’s Algorithm \((F(\mathcal{R}_l^*, S), V)\);
11: Obtain \(f_{l^*}(V^*)\) by \(l^*\) simulations;
12: if \(F(\mathcal{R}_l^*, V^*) \leq (1 + \epsilon_3) \cdot f_{l^*}(V^*)\) then
13: Return \(V^*\);
14: \(l = l + 2\)

Running time. Now let us consider the running time of Alg. 5. Given \(\mathcal{R}_l\), calculating \(F(\mathcal{R}_l, S)\) requires \(O(n^3)\) time. Therefore, line 6 takes \(O(n^3)\) time. Because generating one RA set takes \(O(m)\), the running time of Alg. 5 is \(O(m + n^3)\). Since \(l = \max(\delta_1, \delta_2)\), \(l = O(n)\). The analysis in this section is summarized as follows.

Theorem 5. Given the adjustable parameters \(\epsilon\) and \(N\), with probability at least \(1 - 2/N\), Alg. 5 returns a \(\frac{1}{2} - \epsilon\) approximation running in \(O(n^3)\) with respect to \(n\).

The success probability can be increased to \(1 - 1/N\) by scaling \(N\), which yields a \((\frac{1}{2} - \epsilon, 1 - \frac{1}{N})\)-approximation to the PM problem. Note that the running time of Alg. 5 can be further reduced. Given \(n, r, N\) and \(\epsilon\), the best choice of \(\epsilon_1\) and \(\epsilon_2\) is the one that can minimize \(\max(\delta_1, \delta_2)\). That is,
\[
(\epsilon_1, \epsilon_2) = \arg\min \max(\delta_1, \delta_2). \tag{13}
\]
In practice, there is no need to solve Eq. (13) directly, and, instead, one can enumerate \(\epsilon_1\) from 0 to 1/2 with a small search step, namely 0.01. We will adopt such a setting in experiments.

Generating RA set. It is worthy to note that, according to Def. 4, there is no need to sample the whole realization to obtain an RA set. Instead, one can obtain the adopted-reachable nodes along with the sampling of realization from the node selected in line 2 of Alg. 4 until no nodes can be further reached. Generating a RA set in this way can reduce the running time in practice but does not affect the worst-case running time of RA-T.

5.2 RA-S Algorithm

In this section, we present another algorithm. Recall that in the RA-T algorithm, the number of generated RA sets should be large enough such that \(|F(\mathcal{R}_l, V^*) - f(V^*)|\) and \(|F(\mathcal{R}_l, \mathcal{V}_{opt}) - f(\mathcal{V}_{opt})|\) are sufficiently small where \(V^*\) is obtained by solving Problem 2. In this section, we give another algorithm which generates enough RA sets such that \(|F(\mathcal{R}_l, \mathcal{V}_{opt}) - f(\mathcal{V}_{opt})|\) is small but controls the error \(|F(\mathcal{R}_4, V^*) - f(V^*)|\) by Monte Carlo simulations. We call this algorithm as the RA-S algorithm where RA and S stand for RA set and simulation, respectively.

The RA-S algorithm is formally shown in Alg. 6. In this algorithm, the parameters are \(\epsilon_1, \epsilon_2, \epsilon_3, \delta_1^*, \delta_2^*\) and \(\delta_3\), where
\(\epsilon_3\) and \(k\) are adjustable and other parameters are determined accordingly. In particular, \(\epsilon_1, \epsilon_2, \delta_1^*\) and \(\delta_2^*\) are obtained by solving the following equations,

**Equation System 1.**

\[
\begin{align*}
1 - \epsilon_2 & - \epsilon_1 = \frac{1}{2} - \epsilon \\
\delta_1^* & = \frac{(\ln N + n \ln 2)(6 + 2\epsilon_1 r)}{3\epsilon_1^2 r^2} \\
\delta_2^* & = \frac{2 \ln N}{\epsilon_2 r^2} \\
\delta_1 & = 2^k \cdot \delta_2^* \\
\epsilon_1, \epsilon_2 & > 0 \\
\delta_3 & = \frac{(2 + \epsilon_1 r) \ln N}{\epsilon_1^2 r^2},
\end{align*}
\]

In the above equation system, \(\epsilon, N, n, r, k\) and \(\epsilon_3\) are given in advance. Because \(\delta_1^*\) (resp. \(\delta_2^*\)) approaches infinity when \(\epsilon_1\) (resp. \(\epsilon_2\)) approaches zero, there must be at least one set of variables that solves the equations. As shown in Alg. 6 with such parameters, we generate a family of collections of RA sets with sizes \(\delta_2^*, 2 \cdot \delta_2^*, ..., \delta_1^* = 2^k \cdot \delta_2^*\) by a while loop from line 4 to 14. Within each iteration, we first generate a bunch of RA sets and then get a set \(V^*\) by solving Problem 2. Next, we obtain an estimate \(\tilde{f}_l(V^*)\) of \(f(V^*)\) by \(l^*\) simulations, and check if \(F(R_l, V^*)\) is smaller than \(\tilde{f}_l(V^*) = \delta_3\) by a factor of \(1 + \epsilon_3\).

### 5.2.1 Performance Guarantee

As shown in Alg. 6, the RA-S algorithm terminates either at line 8 or 13.

**Case a.** Suppose that it terminates at line 13 and let \(R_l^* = \{R_1, ..., R_l\}\) be the collection of RA sets used in the last iteration. According to line 5, half of the RA sets in \(R_l^*\) are freshly generated in the last iteration and whether they can be generated depends on the other half of the RA sets according to the condition in line 12, which means the random variables \(x(R, V_{\text{opt}})\) for \(R \in R_l^*\) are not independent. Therefore, the Chernoff bounds are not applicable. However, such dependency between the RA sets does not severely hurt the convergence of the sample mean and it turns out that the random variables \(x(R_1, V_{\text{opt}}), ..., x(R_l, V_{\text{opt}})\) in fact form a martingale. Therefore, we can alternatively use concentration inequalities for martingales as shown in Appendix A. By using martingale, the accuracy of \(F(R_l^*, V_{\text{opt}})\) can be ensured as shown in the following lemma.

**Lemma 12.** If \(l \geq \delta_2^*\)

\[
F(R_l^*, V_{\text{opt}}) - f(V_{\text{opt}}) \geq -\epsilon_2 \cdot f(V_{\text{opt}})
\]

holds with probability at least \(1 - 1/N\).

**Proof.** See Appendix B.5

Furthermore, by setting \(l^*\) as \(\delta_3, \tilde{f}_l(V^*)\) is sufficiently accurate.

**Lemma 13.** Let \(V^*\) be the set produced in line 11 and \(\tilde{f}_l(V^*)\) be the estimate of \(f(V^*)\) obtained by \(l^*\) simulations where \(l^* = \delta_3\). With probability at least \(1 - 1/N\),

\[
\tilde{f}_l(V^*) - f(V^*) \leq \epsilon_1 \cdot f(V_{\text{opt}})
\]

**Proof.** See Appendix B.6

Now we are ready to show the approximation ratio. Suppose that Eqs. 15 and 16 hold simultaneously. By Eq. 16, \((1 + \epsilon_3)(f(V^*) + \epsilon_1 \cdot f(V_{\text{opt}})) \geq (1 + \epsilon_3)\tilde{f}_l(V^*)\).

According to line 12 in Alg. 6, \((1 + \epsilon_3)\tilde{f}_l(V^*) \geq F(R_l^*, V^*)\) and, by Lemma 8, \(F(R_l^*, V_{\text{opt}}) \geq \frac{1}{2} F(R_l^*, V_{\text{opt}})\).

Therefore, \((1 + \epsilon_3)(f(V^*) + \epsilon_1 \cdot f(V_{\text{opt}})) \geq \frac{1}{2}(1 + \epsilon_3)(f(V_{\text{opt}}))\), where the section inequality follows from Eq. 15. Rearranging the above inequality yields

\[
f(V^*) \geq \frac{1 - \epsilon_2}{2(1 + \epsilon_3)} - \epsilon_1 \cdot f(V_{\text{opt}}).
\]

Since, the parameters satisfy Eq System 2, we have

\[
f(V^*) \geq \frac{1 - \epsilon_2}{2} \cdot f(V_{\text{opt}}).
\]

**Case b.** If Alg. 6 terminates at line 8, then \(l \geq \delta_1^*\) and it is sufficiently large to guarantee a \(\left(\frac{1}{2} - \epsilon, 1 - \frac{3}{N}\right)\)-approximation, as shown in the following lemma.

**Lemma 14.** If Alg. 6 terminates at line 6 returning a seed set \(V^*\), then

\[
f(V^*) \geq \frac{1}{2} \cdot f(V_{\text{opt}})
\]

holds with probability at least \(1 - 2/N\).

**Proof.** The proof of this lemma is similar to that of the analysis of the RA-T algorithm, except that we use martingale inequalities instead of Chernoff bound. Intuitively, making \(l \geq \delta_1^*\) ensures that Alg. 6 returns a good approximation in case that the stopping criterion in line 12 is never met. See Appendix B.7 for the complete proof.

By the above analysis, the RA-S algorithm has the same approximation ratio as that of the algorithms proposed in the previous sections. Because the number of RA sets is doubled in each iteration until it exceeds \(\delta_1^*\), the running of RA-S algorithm is at most larger than that of the RA-T algorithm by a constant factor. Therefore, it is still \(O(n^3)\) with respect to \(n\) and \(m\).

**Parameter setting.** As mentioned earlier, there are two free parameters, \(k\) and \(\epsilon_3\). Because \(k\) ranges from \(\delta_2^*\) to \(\delta_1^*\) and \(\delta_1^* = 2^k \cdot \delta_2^*\), \(k\) is number of iterations executed in Alg. 6 in the worst case. When \(k\) is large, the RA-S algorithm has to execute many iterations in the worst case but it also makes it possible that the RA-S algorithm can terminate at line 13 when \(k\) is much smaller than \(\delta_1^*\). On the other hand, when \(k\) is small, the number of RA sets generated in each iteration becomes larger, but the worst case requires fewer iterations. Therefore, \(k\) decides the trade-off between the load size of each iteration and the number of iterations. \(\epsilon_3\) controls the stopping criterion (line 12) of the RA-S algorithm and it is correlated with \(\epsilon_1\) and \(\epsilon_2\) according to Eq System 1. In general, a small \(\epsilon_3\) implies a strict stopping criterion but reduces the number of RA sets required. Conversely, when \(\epsilon_3\) is large, the stopping criterion become less strict but more RA sets are needed as \(\epsilon_1\) and \(\epsilon_2\) become smaller.

**An optimization.** As shown in Alg. 1, the Buchbinder’s algorithm examines the nodes one by one via a loop from line 3 to line 11, and the approximation ratio holds regardless of the examining order. Therefore, one may wonder that which examining order gives the best result. In this
paper, we sort the nodes by $\pi(v)$ in a non-increasing order in advance. In particular, we first generate a bunch of RA sets $R_l = \{R_1, ..., R_l\}$ and use $\sum_{i=1}^{l} \pi(R_i, v)$ as an estimate of $\pi(v)$ for sorting. We adopt such an examining order whenever the Buchbinder’s algorithm is used. 

**Comparison with RA-T algorithm.** The major difference between RA-T and RA-S is that they may use different numbers of RA sets. Although RA-T and RA-S algorithms have the same asymptotic running time, the RA-S algorithm can be more efficient in many cases. Recall that $l$ is the number of RA sets used for estimating. In the RA-T algorithm, even if $\delta_2$ is much smaller than $\delta_1$, $l$ has to be set as $\max(\delta_1, \delta_2) \approx \delta_1$ so that both the $F(R_l, V^*)$ and $F(R_l, V_{opt})$ are sufficiently accurate. In the RA-S algorithm, as shown in line 12 in Alg. 5, we utilize Monte Carlo simulation to check whether $F(R_l^*, V^*)$ is accurate enough, and therefore $F(R_l^*, V_{opt})$ can be a good estimate as long as $l$ is larger than $\delta_2$. As a result, if the stopping criterion is satisfied when $l$ is much smaller than $\delta_1$, RA-S returns a good solution without generating many RA sets. However, as observed in the experiments, RA-S does not dominate RA-T.

**Comparison with SPM and RPM.** According to the analysis in Sec. 4 and this section, RA-T and RA-S are theoretically efficient than SPM and RPM with respect to running time, which is also confirmed in experiments shown later in Sec. 6. However, SPM and RPM can be very fast under parallel computing systems as executing simulations and generating realizations are highly parallelizable. Furthermore, the forward sampling method designed in Sec. 4 is potentially applicable to other USM problems beyond the PM problem, while RA-T and RA-S are proposed especially for the PM problem under the triggering model.

### 6 Experiments

In this section, we present the experiments for evaluating the proposed algorithms. Besides showing the performance of the algorithms proposed in this paper, the results herein also provide useful observations on the property of the PM problem.

#### 6.1 Experimental Settings

Our experiments are performed on a server with 16 GB ram and a 3.6 GHz quadcore processor running 64-bit JAVA VM 1.6.

**6.1.1 Datasets**

The datasets selected for experiments are listed in Table 1. All the datasets are borrowed from Leskovec et al. [24]. Wiki is a who-votes-on-whom network collected from Wikipedia. HepTh and HepPh are arXiv citation graphs from the categories of high energy physics theory and high energy physics phenomenology, respectively. Youtube is a social network drawn from the Youtube video-sharing website. Pokec is a popular online social network in Slovakia. The original data of Youtube is undirected and we take each edge as a bi-directed edge to make it directed.

| Dataset | #Edges | #Nodes | Avg Degree |
|---------|--------|--------|------------|
| Wiki    | 104K   | 7K     | 14.7       |
| HepTh   | 353K   | 28K    | 12.6       |
| HepPh   | 422K   | 35K    | 12.1       |
| Youtube | 6.0M   | 1.1M   | 5.4        |
| Pokec   | 30.6M  | 1.6M   | 19.1       |

**TABLE 1: Datasets.**

6.1.2 Models.

As mentioned earlier, we adopt IC and LT models for experiments. For IC model, we consider two probability settings. One is constant probability (CP) setting where each edge has the propagation probability of 0.01 and the other one is the weighted cascade (WC) setting where $p_{(u,v)} = 1/|N_v^-|$. To distinguish those two settings, we use IC-CP to denote the former and IC-WC for the later. For LT model, following the convention [1], [7], [23], the weight of edge $(u, v)$ is set as $1/|N_v^-|$. The price $P$, coupon $C$ and intrinsics are normalized within [0,1], shown as follows. $P$ is selected from [0.2,0.6] and $C$ is set as the 90% of $P$. After setting $P$ and $C$, the intrinsic of each user is randomly generated from $[P - C, 1]$ in uniform. Under this setting, every node can be an adopter once selected as a seed user.

6.1.3 Algorithms

The algorithms considered in our experiments are shown as follows. We always set $\epsilon$ and $N$ as 0.4 and $n$, respectively.

**RA-S.** This is the algorithm proposed in Sec. 5.2. For small graphs (Wiki, HepTh and HepPh), we set that $k = 5$ and $\epsilon_3 = 0.1$, and for large graphs (Youtube and Pokec) we set that $k = 8$ and $\epsilon_3 = 0.3$. Other parameters are calculated by solving Eq System 4. Furthermore, we immediately return the seed set if the $F(R_l^*, V^*)$ calculated in two consecutive iterations decreases by less than 2%.

**RA-T.** This is the algorithm proposed in Sec. 5.1. In experiments, we set the maximum of RA sets as 5,000,000. The results under this setting are also used later for discussing how many RA sets needed in obtaining good estimates.

**SPM.** This is the algorithm proposed in Sec. 4. The parameter $l$ here controls the number of simulations used for each estimating. In the previous works regarding the IM problem, the number of simulations is typically set as 10,000 to obtain an accurate estimation. However, in this paper, we set $l$ as 2,000 for two reasons. First, it is reported in [12] that 1,000 simulations are sufficient for small graphs. Second, SPM takes more than 24 hours on Wiki even with $l = 2,000$, which means it is not a good choice even if it could provide better performance with larger $l$.

**RPM.** This is the algorithm proposed in Sec. 4.2. The number of realizations is set as 1,000 to make it terminate in a reasonable time. SPM and RPM are tested only on Wiki under IC-CP and LT as they are very time-consuming.

**MaxInf.** As the first part of our objective function (Eq. (3)), the number of adopters is important in deciding the profit. Therefore, the algorithm that maximizes the number of adopters can be a reasonable heuristic for the PM problem. Note that maximizing the number of adopters is essentially the IM problem. We adopt one of the state-of-the-art IM algorithms, D-SSA [8], as a heuristic algorithm for the PM problem. In particular, we set the size of the seed set to be $n \times i / 50$ for $i = 1, ..., 50$. For each size, we observe in the experiments, $F(R_l^*, S)$ decreases with the increase of $l$. Such a phenomenon is also observed in [12].
run the D-SSA to obtain a seed set and evaluate the profit by simulation. Finally, we select the result with the highest profit. This approach is denoted as MaxInf.

**HighDegree.** Selecting the nodes with the highest degree is a popular heuristic for the maximization problems in social networks. We first randomly decide the size of seed nodes and then select the nodes with the highest degree as the seed set. For each graph, the above process is executed for 100 times. Finally, we select the result with the highest profit.

The profits under the seed sets produced by the above algorithms are finally evaluated by 10,000 simulations. On large graphs, we show the results of RA-T and a simplified version of MaxInf where, instead of enumerating it from \( n/50 \) to \( n \), the size of the seed set of MaxInf is set as the same as that of the result produced by RA-T.

### 6.2 Discussions

The experimental results are shown Fig. 1. According to the experimental results, we make the following observations.

**RA-S and RA-T give the best overall performance for maximizing the profits.** As shown in the figures, RA-S and RA-T consistently give the highest profit on the considered datasets under different settings. While SPM and RPM are also approximation algorithms, they cannot always produce high-quality seed sets. As shown in Figs. 1a and 1b with \( l = 1,000 \) on graph Wiki, RPM performs well under IC-CP model but considerably worse under LT model. Note that the gaps between the lines of different algorithms are not very significant because the price of the product is normalized within \([0, 1]\).

**MaxInf and HighDegree are sometimes comparable to RA-S and RA-T.** As shown in Fig. 1d, MaxInf performs relatively good when the price is very low. This is because when \( P \) is small almost all the nodes can be adopters once activated due to the setting of the intrinsic, and therefore the first part \( \pi() \) dominates the objective function \( f() \). Since MaxInf aims to maximize \( \pi() \), it has a good performance in this case. When the price is very high, except the seed users, most of the users cannot be adopters and therefore the objective function is almost monotone, which makes HighDegree accidentally comparable to RA-S e.g. Fig. 1e. As an extreme case, when \( P \) approaches to 1, the seed users cannot activate any neighbor and thus the objective function is strictly monotone. In this case, \( f(S) = (P - C)|S| \) only depends on the size of the seed set and the considered problem becomes less interesting.

**RA-S and RA-T are efficient with respect to running time.** As shown in the graphs, RA-S and RA-T are even more efficient than the two heuristics, MaxInf and HighDegree, under certain circumstances. This is because the two heuristics still need to utilize Monte Carlo simulations to compare the profit under different seed sets. SPM and RPM are not scalable to large graphs as they also employ Monte Carlo simulations. For example, as shown in Figs. 1a and 1b on graph Wiki, SPM takes 38.6 and 21.2 hours under IC-CP and LT models, respectively.

**Comparison between SPM and RPM.** As shown in Fig. 1c, comparing with SPM, RPM results in a higher profit and also takes much less time on graph Wiki under IC-CP model, which gives the evidence that RPM is sometimes superior to SPM. However, RPM requires to load the realizations into memory and therefore it consumes more memory than SPM does.

**How many RA sets needed to produce a good seed set?** As shown in Sec. 5 a great deal of effort has been devoted to computing the number of RA sets needed to produce a good seed set for maximizing the profit. This is actually the key problem for many influence related problems where the reverse sampling technique is used, as shown in [7], [13], [23], [25]. While more RA sets bring a higher quality of the seed sets, the quality increases very slow when a large number RA sets have already been used, and therefore one should stop generating RA sets if the current seed set is satisfactory. We observe that the algorithms proposed in this paper can sometimes overly generate RA sets. For example, on graph HepPh under the LT model, as shown in Fig. 1c, RA-S and RA-T have the similar performance but RA-S is ten times slower than RA-T, which means the number of RA sets used in RA-S is larger than that of used in RA-T. In fact, we believe there is room for improvement of the algorithms proposed in this paper.

**Profit Vs. Price.** As observed in the experiments, the profit does not necessarily increase with the increase of price, especially under the LT model as shown in Figs. 1c and 1d. Such a phenomenon coincides with the setting of the intrinsic, because a large \( P \) narrows the set of users who are able to be adopters and therefore leads a decrease in total profit.

**Comparing to IM problem.** It is worthy to note that, in general, the PM problem considered in this paper takes more time to solve than the IM problem does. Although the SPM algorithm proposed in this paper and the CELF algorithm [10] for IM problem are both designed with the idea of forward sampling, SPM is not able to run on HepPh even with \( l = 2,000 \) while CELF terminates within 40 hours on the same graph as shown in [12]. There are several possible explanations for this phenomenon. First, the PM itself is an unconstrained optimization problem while the IM problem is a budgeted maximization problem, and thus PM requires more effort for sampling because \( \pi(V_{opt}) \) can be very large. Second, the lazy forward sampling method [10] significantly reduces the running time of the greedy algorithm for solving the IM problem without loss of the approximation guarantee. Unfortunately, currently we are not aware of any technique that can be used to boost the efficiency of the Buchbinder’s algorithm (Alg. 1).

**On large graphs.** The results on Youtube and Pokec are shown in Figs. 1h, 1i and 1j. We do not show the experimental results of HighDegree because it still performs worse than RA-T does. The results of RA-S are not included because it runs out of memory on those graphs. Note that our experiments are done on a machine with 16 GB ram and we observe that RA-S is able to run on large graphs if the size of ram is increased to 32 GB or more. According to the results, even though the number of RA sets is limited to 5,000,000, RA-T still performs much better than MaxInf does. On large graphs, MaxInf takes much less time than RA-T does, because the size of the seed set in MaxInf is given by the result of RA-T and consequently MaxInf no longer has to perform Monte Carlo simulations to compare
the profits under the seed sets with different size. As shown in Figs. [1] and [3] MaxInf can sometimes result in negative profit, which again indicates that maximizing the influence may not produce a high profit. As shown in Figs. [1h] and [1k] RA-T is able to terminate within 10 hours on Youtube and Pokec.

7 CONCLUSION AND FUTURE WORK

In this paper, we consider the problem of maximizing the profit with the usage of the coupon in social networks. By integrating coupon system into the triggering spreading model, we formulate the PM problem as a combinatorial optimization problem. Based on our formulation, several approximation algorithms are provided together with their analysis. The performance of the proposed methods is evaluated by experiments. The analysis of the PM problem is actually applicable to other influence-related problems which are non-monotone submodular, and therefore it provides an algorithmic framework.

The major future work of this paper is derived from the observations of the experiments. As discussed in Sec. [5.2] to improve the efficiency of RA-T and RA-S, one has to consider the number of RA sets generated for estimating.
On one hand we need sufficient RA sets for an accurate estimate but, on the other hand, we should avoid generating too many RA sets to reduce the running time. Essentially, it asks for a good stop criterion for generating RA sets in such kind of algorithms like RA-S and RA-T. Second, any technique that can boost the efficiency of the Buchbinder’s algorithm can definitely reduce the running time of RA-S and RA-T. Thus, finding such techniques is a promising future work. Finally, we note that the heuristics designed based on RA sets may effectively solve the PM problem while taking much less time. We leave this as part of our future work.

ACKNOWLEDGMENTS

This work was supported by National Natural Science Foundation of China under 61472272.

REFERENCES

[1] D. Kempe, J. Kleinberg, and É. Tardos, “Maximizing the spread of influence through a social network,” in Proc. of the ninth SIGKDD. ACM, 2003, pp. 137–146.
[2] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, “An analysis of approximations for maximizing submodular set functions,” Mathematical Programming, vol. 14, no. 1, pp. 265–294, 1978.
[3] S. Kalish, “A new product adoption model with price, advertising, and uncertainty,” Management science, vol. 31, no. 12, pp. 1569–1585, 1985.
[4] K. Ailawadi, S. Gupta et al., “Sales promotions,” History of Marketing Science, Singapore, pp. 463–497, 2014.
[5] N. Buchbinder, M. Feldman, J. Seffi, and R. Schwartz, “A tight linear time (1/2)-approximation for unconstrained submodular maximization,” SIAM Journal on Computing, vol. 44, no. 5, pp. 1384–1402, 2015.
[6] C. Borgs, M. Brautbar, J. Chayes, and B. Lucier, “Maximizing social influence in nearly optimal time,” in Proc. of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms. SIAM, 2014, pp. 946–957.
[7] D. Kempe, J. Kleinberg, and É. Tardos, “Maximizing the spread of influence through a social network,” in Proc. of the ninth SIGKDD. ACM, 2003, pp. 137–146.
[8] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, “An analysis of approximations for maximizing submodular set functions,” Mathematical Programming, vol. 14, no. 1, pp. 265–294, 1978.
[9] S. Kalish, “A new product adoption model with price, advertising, and uncertainty,” Management science, vol. 31, no. 12, pp. 1569–1585, 1985.
[10] K. Ailawadi, S. Gupta et al., “Sales promotions,” History of Marketing Science, Singapore, pp. 463–497, 2014.
[11] N. Buchbinder, M. Feldman, J. Seffi, and R. Schwartz, “A tight linear time (1/2)-approximation for unconstrained submodular maximization,” SIAM Journal on Computing, vol. 44, no. 5, pp. 1384–1402, 2015.
[12] C. Borgs, M. Brautbar, J. Chayes, and B. Lucier, “Maximizing social influence in nearly optimal time,” in Proc. of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms. SIAM, 2014, pp. 946–957.
[13] Z. Lu, H. Zhou, V. O. Li, and Y. Long, “Pricing game of celebrities in sponsored viral marketing in online social networks with a greedy advertising platform,” in Proc. of the 2016 ICC. IEEE, 2016, pp. 1–6.
[14] Y. Zhu, Z. Lu, Y. Bi, W. Wu, Y. Jiang, and D. Li, “Influence and profit: Two sides of the coin,” in Proc. of the 13th ICDM. IEEE, 2013, pp. 1301–1306.
[15] Y. Yang, X. Mao, J. Pei, and X. He, “Continuous influence maximization: What discounts should we offer to social network users?” in Proc. of the 2016 SIGMOD. ACM, 2016, pp. 727–741.
[16] W. Lu and L. V. Lakshmanan, “Profit maximization over social networks,” in Proc. of the 12th ICDM. IEEE, 2012, pp. 479–488.
[17] J. Tang, X. Tang, and J. Yuan, “Profit maximization for viral marketing in online social networks,” in Proc. of the 24th ICPN. IEEE, 2016, pp. 1–10.
[18] U. Feige, V. S. Mirrokni, and J. Vondrak, “Maximizing non-monotone submodular functions,” SIAM Journal on Computing, vol. 40, no. 4, pp. 1133–1153, 2011.
[19] A. 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 Proc. of the 36th INFOCOM. IEEE, 2017.
[20] J. Leskovec and A. Krevl, “SNAP Datasets: Stanford large network dataset collection,” http://snap.stanford.edu/data. Jun. 2014.
[21] W. Lu, W. Chen, and L. V. Lakshmanan, “From competition to complementarity: comparative influence diffusion and maximization,” Proc. of the VLDB Endowment, vol. 9, no. 2, pp. 60–71, 2015.
[22] R. Motwani and P. Raghavan, Randomized algorithms. Chapman & Hall/CRC, 2010.

Guang-mo Tong is a Ph.D. candidate in the Department of Computer Science at the University of Texas at Dallas. He received his BS degree in Mathematics and Applied Mathematics from Beijing Institute of Technology in 2013. His research interests include computational social systems, data communication and real-time systems. He is a student member of the IEEE.

Weili Wu (M00) is currently a Full Professor with the Department of Computer Science, the University of Texas at Dallas, Dallas, TX, USA. She received the Ph.D. and M.S. degrees from the Department of Computer Science, University of Minnesota, Minneapolis, MN, USA, in 2002 and 1998, respectively. Her research mainly deals with the general research area of data communication and data management. Her research focuses on the design and analysis of algorithms for optimization problems that occur in wireless networking environments and various database systems.

Ding-Zhu Du received the M.S. degree from the Chinese Academy of Sciences in 1982 and the Ph.D. degree from the University of California at Santa Barbara in 1985, under the supervision of Professor Ronald V. Book. Before settling at the University of Texas at Dallas, he worked as a professor in the Department of Computer Science and Engineering, University of Minnesota. He also worked at the Mathematical Sciences Research Institute, Berkeley, for one year, in the Department of Mathematics, Massachusetts Institute of Technology, for one year, and in the Department of Computer Science, Princeton University, for one and a half years. He is the editor-in-chief of the Journal of Combinatorial Optimization and is also on the editorial boards for several other journals. Forty Ph.D. students have graduated under his supervision. He is a member of the IEEE.
**APPENDIX A**

**CHERNOFF BOUND AND MARTINGALE**

Let $Z_i \in [0, 1]$ be i.i.d random variables where $E(Z_i) = \mu$. The Chernoff bounds [26] states that

$$\Pr \left[ \sum Z_i - l \cdot \mu \geq \delta \cdot l \cdot \mu \right] \leq \exp \left( -\frac{l \cdot \mu \cdot \delta^2}{2 + \delta} \right), \quad (17)$$

and

$$\Pr \left[ \sum Z_i - l \cdot \mu \leq -\delta \cdot l \cdot \mu \right] \leq \exp \left( -\frac{l \cdot \mu \cdot \delta^2}{2} \right), \quad (18)$$

for $\delta > 0$.

Martingale method provides good error estimates for a sequence of variables where the independence assumptions is not required. It has been proved by Tang et al. [21] that, for any seed set $S$ and a sequence of RA sets $R_1, \ldots, R_l$, the random variables $x(R_i, l)$, for $1 \leq i \leq l$, are martingales and they satisfy the following concentration inequalities, for any $\delta \geq 0$,

$$\Pr \left[ \sum x(R_i, S) - l \cdot \mu \geq \delta \cdot l \cdot \mu \right] \leq \exp \left( -\frac{l \cdot \mu \cdot \delta^2}{2 + \frac{\delta}{3}} \right), \quad (19)$$

and

$$\Pr \left[ \sum x(R_i, S) - l \cdot \mu \leq -\delta \cdot l \cdot \mu \right] \leq \exp \left( -\frac{l \cdot \mu \cdot \delta^2}{2} \right), \quad (20)$$

and where $\mu = E[x(R_i, l)] = \frac{\pi(S)}{n}$.

**APPENDIX B**

**PROOFS**

**B.1 Proof of Theorem 2**

Recall that, in the $i$-th iteration of the loop, node $v_i$ is considered, and $X_i$ and $Y_i$ are the subsets created in line 9 or 11. Let $S^*$ be the optimal solution and define that $S^*_i = (S^* \cup X_i) \cap Y_i$ for $1 \leq i \leq n$. Note that $S^*_0 = S^*$ and $S^*_n = X_n = Y_n$. For a set $S$ and a node $v$, define that $S + v = S \cup \{v\}$ and $S - v = S \setminus \{v\}$. Because Alg. 1 is in fact a randomized algorithm, we consider the expected value of $h(X_n)$, the following is a key lemma to prove Theorem 2.

**Lemma 15.** For each $1 \leq i \leq n$,

$$\mathbb{E}[h(S^*_{i-1}) - h(S^*_i)] \leq \frac{1}{n} (\mathbb{E}[h(X_i)] - h(X_{i-1}) + h(Y_i) - h(Y_{i-1})) + \frac{2\epsilon}{n} \cdot h(S^*)$$

**Proof.** It suffices to prove that it is true conditioned on every possible event by the $i$-th iteration. Now we fix $X_{i-1}$ and $Y_{i-1}$ and consider the $i$-th iteration. By Eq. (1), $h(X_{i-1} + v_i) - h(X_{i-1}) + h(Y_{i-1} - v_i) - h(Y_{i-1})$ is always non-negative due to the submodularity of $h()$. Thus,

$$\bar{a}_i + \bar{b}_i = \tilde{h}(X_{i-1} + v_i) - \tilde{h}(X_{i-1}) + \frac{2\epsilon}{n} \cdot L^* + \tilde{h}(Y_{i-1} - v_i) - \tilde{h}(Y_{i-1}) + \frac{2\epsilon}{n} \cdot L^* \quad \{\text{By Eq. (5)}\}$$

$$\geq h(X_{i-1} + v_i) - h(X_{i-1}) + h(Y_{i-1} - v_i) - h(Y_{i-1}) \geq 0,$$

and there are three cases to consider:

**Case 1.** ($\tilde{a}_i \geq 0$ and $\tilde{b}_i \leq 0$). In this case, $Y_i = Y_{i-1}$ and $X_i = X_{i-1} + v_i$, and therefore, $S^*_i = S^*_{i-1} + v_i$ and $h(Y_i) - h(Y_{i-1}) = 0$. If $v_i \in S^*_{i-1}$, then $h(S^*_{i-1}) - h(S^*_i) = 0$. If $v_i \notin S^*$, then by the submodularity,

$$h(S^*_{i-1}) - h(S^*_i) \leq h(Y_{i-1} - v_i) - h(Y_{i-1}) \quad \{\text{By Eq. (5)}\}$$

$$\leq \tilde{h}(Y_{i-1} - v_i) - \tilde{h}(Y_{i-1}) + \frac{2\epsilon}{n} \cdot L^*$$

Therefore, in both cases, we have $h(S^*_{i-1}) - h(S^*_i) \leq 0 \leq \frac{\epsilon}{2}$.

Since $h(Y_i) = h(Y_{i-1})$,

$$h(S^*_{i-1}) - h(S^*_i) \leq \frac{1}{2} \tilde{a}_i + \frac{1}{2} (h(Y_i) - h(Y_{i-1}))$$

$$= \frac{1}{2} (\tilde{h}(X_{i-1} + v_i) - \tilde{h}(X_{i-1}) + h(Y_i) - h(Y_{i-1})) \quad \{\text{By Eq. (5)}\}$$

$$\leq \frac{1}{2} (h(X_{i-1} + v_i) - h(X_{i-1}) + h(Y_i) - h(Y_{i-1})) + \frac{\epsilon}{n} \cdot L^*$$

$$\leq \frac{1}{2} (h(X_i) - h(X_{i-1}) + h(Y_i) - h(Y_{i-1})) + \frac{\epsilon}{n} \cdot h(S^*)$$

Thus, this case is proved.

**Case 2.** ($\tilde{a}_i < 0$ and $\tilde{b}_i \geq 0$). This case can be proved in a similar manner as that of Case 1.

**Case 3.** ($\tilde{a}_i \geq 0$ and $\tilde{b}_i > 0$). In this case, with probability $\tilde{a}_i / (\tilde{a}_i + \tilde{b}_i)$ (resp. $\tilde{b}_i / (\tilde{a}_i + \tilde{b}_i)$) that $X_i \leftarrow X_{i-1} + v_i$ and $Y_i \leftarrow Y_{i-1}$ (resp. $X_i \leftarrow X_{i-1}$ and $Y_i \leftarrow Y_{i-1} \setminus v_i$) happen.

Therefore,

$$\mathbb{E}[h(X_i) - h(X_{i-1}) + h(Y_i) - h(Y_{i-1})]$$

$$= \frac{\tilde{a}_i}{\tilde{a}_i + \tilde{b}_i} (h(X_{i-1} + v_i) - h(X_{i-1})) + \frac{\tilde{b}_i}{\tilde{a}_i + \tilde{b}_i} (h(Y_{i-1} - v_i) - h(Y_{i-1})) \quad \{\text{By Eq. (5)}\}$$

$$\geq \frac{\tilde{a}_i}{\tilde{a}_i + \tilde{b}_i} (\tilde{h}(X_{i-1} + v_i) - \tilde{h}(X_{i-1}) - \frac{2\epsilon}{n} \cdot L^*) + \frac{\tilde{b}_i}{\tilde{a}_i + \tilde{b}_i} (\tilde{h}(Y_{i-1} - v_i) - \tilde{h}(Y_{i-1}) - \frac{2\epsilon}{n} \cdot L^*)$$

$$= \frac{\tilde{a}_i^2 + \tilde{b}_i^2}{\tilde{a}_i + \tilde{b}_i} - \frac{4\epsilon}{n} \cdot L^* \geq \frac{\tilde{a}_i^2 + \tilde{b}_i^2}{\tilde{a}_i + \tilde{b}_i} - \frac{4\epsilon}{n} \cdot h(S^*) \quad (21)$$

Now let us consider $\mathbb{E}[h(S^*_{i-1}) - h(S^*_i)]$. Note that

$$\mathbb{E}[h(S^*_{i-1}) - h(S^*_i)]$$

$$= \frac{\tilde{a}_i}{\tilde{a}_i + \tilde{b}_i} (h(S^*_{i-1}) - h(S^*_{i-1} + v_i)) + \frac{\tilde{b}_i}{\tilde{a}_i + \tilde{b}_i} (h(S^*_{i-1}) - h(S^*_{i-1} - v_i))$$

$$= \frac{\tilde{a}_i}{\tilde{a}_i + \tilde{b}_i} (h(S^*_{i}) - h(S^*_{i} + v_i)) + \frac{\tilde{b}_i}{\tilde{a}_i + \tilde{b}_i} (h(S^*_{i}) - h(S^*_{i} - v_i))$$

$$\leq \frac{\tilde{a}_i^2 + \tilde{b}_i^2}{\tilde{a}_i + \tilde{b}_i} - \frac{4\epsilon}{n} \cdot h(S^*) \quad (21)$$
If $v_i \in S_{i-1}^*$ then $h(S_{i-1}^*) - h(S_{i-1}^* + v_i) = 0$ and
\[ h(S_{i-1}^*) - h(S_{i-1} - v_i) \]
By submodularity
\[ \leq h(X_{i-1} + v_i) - h(X_{i-1}) \]
By Eq. (5)
\[ \leq \hat{h}(X_{i-1} + v_i) - \hat{h}(X_{i-1}) + \frac{2\epsilon}{n} \cdot L^* = \tilde{a}_i \]
If $v_i \notin S_{i-1}^*$ then $h(S_{i-1}^* - v_i) - h(S_{i-1}^*) = 0$ and
\[ h(S_{i-1}^*) - h(S_{i-1} - v_i) \]
By submodularity
\[ \leq h(Y_{i-1} - v_i) - h(Y_{i-1}) \]
By Eq. (5)
\[ \leq \hat{h}(Y_{i-1} - v_i) - \hat{h}(Y_{i-1}) + \frac{2\epsilon}{n} \cdot L^* = \tilde{b}_i. \]

Therefore, in either case, $\mathbb{E}[h(S_{i-1}^*) - h(S_i^*)] \leq \frac{\tilde{a}_i + \tilde{b}_i}{n}$. Together with Eq. (21), $\mathbb{E}[h(S_{i-1}^*) - h(S_i^*)] \leq \frac{1}{2} \mathbb{E}[h(X_i) - h(X_{i-1}) + h(Y_i) - h(Y_{i-1})] + \frac{2\epsilon}{n} \cdot h(S_i^*)$.

Now new are ready to prove Theorem 2, as shown in the following lemma.

**Lemma 16.** $\mathbb{E}[h(X_n)] \geq \left( \frac{1}{2} - \epsilon \right) \cdot h(S^*)$.

**Proof.** Summing the inequality of Lemma 15 for $i$ from 1 to $n$ yields $\mathbb{E}[h(S_0) - h(S_n)] \leq \frac{1}{2} \mathbb{E}[h(X_n) - h(Y_n) - h(X_0) - h(Y_0)] + 2\epsilon \cdot h(S_n)$. Since $S_0 = S^*$ and $S_n = X_n = Y_n$, we have $\mathbb{E}[h(X_n)] \geq \left( \frac{1}{2} - \epsilon \right) \cdot h(S_n)$.

**B.2 Proof of Lemma 7**

Let $\pi(S, v)$ be the probability that $v$ can be an adopter under $S$ and therefore $\pi(S) = \sum_{v \in V} \pi(S, v)$ due to the linearity of expectation. Let $G^v_P$ be the set of the realizations where $v$ is an adopter under $S$ and $\Pr[g]$ be the probability that $g$ can be sampled according to the triggering distribution. By Lemma 5 $\pi(S, v) = \sum_{g \in G^v_P} \Pr[g]$. On the other hand, let $x(S, R|v)$ be the conditional variable of $x(S, R)$ when $v$ is selected in line 2 of Alg. 4. Thus, $\mathbb{E}[x(S, R)] = \sum_{v \in V} \mathbb{E}[x(S, R|v)]$. Note that $v$ is an adopter under $S$ in $g$ if and only if at least one node in $S$ in reverse-adopted-reachable to $v$. Therefore, $\mathbb{E}[x(S, R)] = \sum_{g \in G^v_P} \Pr[g] = \pi(S, v)$, which completes the proof.

**B.3 Proof of Lemma 9**

By rearrangement, $\Pr[F(R_i, S) - f(S) > \epsilon_1 \cdot f(V_{opt})]$ is equal to $\Pr[\sum_{i=1}^t x(R_i, S) - \frac{\pi(S)}{n} > \epsilon_1 \cdot f(V_{opt})]$. According to Chernoff bound, this probability is no larger than
\[ \exp\left(-\frac{1 \cdot \frac{\pi(S)}{n} \cdot (\epsilon_1 \cdot f(V_{opt})^2)}{2 \cdot 3^2 \cdot \frac{P \cdot \pi(S)}{3^2 \cdot \pi(S)}} \right). \]
By Eq. (4) and $\pi(S) \leq n$, it is further no larger than
\[ \exp\left(-\frac{3 \cdot \epsilon_1^2 \cdot P \cdot (P - C)^2}{6P^2 + 2 \epsilon_1 P \cdot (P - C)} \right). \]
Finally, by Eq. (11), this probability is at most $\frac{1}{n \cdot 2^n}$. Since there are at most $2^n$ subsets, the lemma holds.

**B.4 Proof of Lemma 10**

By rearrangement, $\Pr[F(R_i, V_{opt}) - f(V_{opt}) \leq -\epsilon_2 \cdot f(V_{opt})]$ is equal to $\Pr[\sum_{i=1}^t x(R_i, V_{opt}) - \frac{\pi(V_{opt})}{n} \leq -l \cdot \frac{\pi(V_{opt})}{n} \cdot \epsilon_2 \cdot f(V_{opt})]$. By Chernoff bound, this probability is no larger than
\[ \exp\left(-\frac{l \cdot \frac{\pi(V_{opt})}{n} \cdot (\epsilon_2 \cdot f(V_{opt})^2)}{2 \cdot \frac{P \cdot \pi(V_{opt})}{3^2 \cdot \pi(V_{opt})}} \right). \]
By Eq. (4) and $\pi(S) \leq n$, it is then no larger than
\[ \exp\left(-\frac{l \cdot \epsilon_2^2 \cdot (P-C)^2}{2P^2 + 2 \epsilon_1 P \cdot (P - C)} \right). \]
Finally, replacing $l$ with $\delta_1^2$ implies that
\[ \exp\left(-\frac{3 \cdot \epsilon_1^2 \cdot P \cdot (P - C)^2}{6P^2 + 2 \epsilon_1 P \cdot (P - C)} \right) \leq \frac{1}{N \cdot 2^n}. \]
Since there are at most $2^n$ subsets, with probability at most $1/N$, there exists some $S \subseteq V$ such that $F(R_i, S) - f(S)$
\( \epsilon_1 \cdot f(V_{opt}) \), which means, \( F(R_t, V^*) - f(V^*) \leq \epsilon_1 \cdot f(V_{opt}) \) holds with the probability at least \( 1 - 1/N \). Therefore,

\[
\begin{align*}
f(V^*) & \geq F(R_t, V^*) - \epsilon_1 \cdot f(V_{opt}) \\
& \quad \{ \text{By Lemma 3} \} \\
& \geq \frac{1}{2} F(R_t, V_{opt}) - \epsilon_1 \cdot f(V_{opt}) \\
& \quad \{ \text{By Lemma 12} \} \\
& \geq \frac{1}{2} (1 - \epsilon_2) f(V_{opt}) - \epsilon_1 \cdot f(V_{opt}) \\
& \quad \{ \text{By Eqs 1} \} \\
& \geq \left( \frac{1}{2} - \epsilon \right) \cdot f(V_{opt})
\end{align*}
\]

holds with probability at least \( 1 - 2/N \).