Maximizing Activity Profit in Social Networks

Wenguo Yang, Jing Yuan, *Student Member, IEEE*, Weili Wu, *Member, IEEE*, Jianmin Ma, and Ding-Zhu Du

***Abstract*— In the past decade, tremendous research effort has been devoted to viral marketing. Most existing works on seed selection in social networks do not take into account the scenario when a profit can be generated from group activities. Each activity has a profit that can be measured by the excitement of the participants. The excitement about one piece of information can vary significantly among different groups of people. Given a social network and a profit function, how can we select the seed users to maximize the expected total amount of profit? This problem is essentially different from the classic influence maximization problem, and existing approaches cannot be directly applied to solve the problem. In this paper, we study the problem of activity profit maximization in social networks. We first prove that the maximizing activity profit problem is nondeterministic polynomial time-hard and cannot be approximated within a constant factor by the simple greedy algorithm. Supermodular degree of a function measures the extent to which it violates submodularity. We design an algorithm that achieves an approximation ratio of $(1/(\Delta + 2))$ provided that the supermodular degree of the social graph is bounded with Δ. We then develop an exchange-based technique to further improve the quality of the solution. We also devise a randomized variation approach to overcome the computational burden of the proposed algorithms. Extensive experimental results on three real benchmark data sets demonstrate the efficacy and efficiency of our algorithms over several baseline heuristics.**

***Index Terms*— Activity profit maximization, approximation algorithms, social computing.**

I. Introduction

NOWADAYS, more and more people own personal accounts in various online social medias, such as Facebook, LinkedIn, ResearchGate, and so on. Online social networks have been booming so rapidly with users contributed data that their impact cannot be ignored in many areas, such as presidential election and viral marketing. Recent research revealed that social advertising is more effective than traditional broadcast advertising channels.

Overtime social information diffusion and influence propagation get into people's daily life more deeply and more frequently through online social media. Therefore, the influence-driven information technology and influence-based research problems have been studied extensively in the literature. Among existing works in the viral marketing, most of the efforts have been devoted into topics focusing on products used by a single person. However, some products may require more than one user. For example, co-op video games like Monaco require two or more players to beat the games' hardest levels. Other examples include online games like Texas hold 'em that requires multiple users to join before the game starts.

Wang *et al.* [1] initiated the study on products that require interaction between two users. In this paper, we would like to consider products that involve cooperation among any number of users. Consider a social network with certain information diffusion model. In the well-known influence maximization problem, a positive integer k is given and the problem is to find a set of k seeds to maximize the influence spread, i.e., the number or the expected number of active nodes. In the activity maximization problem formulated in [1], the objective function measures the total "activity strength" or say activity profit, instead of the influence spread. To consider products with any number of users, we consider an activity as an event with two properties. First, this event has two or more participants. Second, the occurrence of this event would generate a profit. For example, a group of users would jointly purchase a product or play a game together. Since each activity can be represented by a group of users, all activities form the hyperedge set of a hypergraph. Different activities may generate different profits. We are going to maximize the total profit or the expected total profit generated by active users.

As indicated in [1], the objective function of the activity maximization problem is neither submodular nor supermodular in the independent cascade (IC) model and the linear threshold (LT) model. A technique, so-called the sandwich method, is employed in their work to obtain an approximation algorithm with theoretical guaranteed performance. In this paper, we will employ a different technique to deal with our activity profit maximization problem. We will present new theoretical improvement on the technique, together with computational experiments to support our results.

A. Related Works

Influence maximization is a fundamental research problem in the study of social networks [2]–[8]. With many information diffusion models, especially the IC model and the LT model, the influence maximization is a nondeterministic polynomial time (NP)-hard problem and computing the influence spread is a #P-hard problem. However, there are randomized algorithms [9]–[11] that generates $(1 - e^{-1} - \varepsilon)$ approximation within time $O((m + n))$ with

Manuscript received September 8, 2018; accepted December 22, 2018. Date of publication January 31, 2019; date of current version February 12, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 11571015 and Grant 11571091 and in part by US NSF under Award 1747818. *(Corresponding author: Jing Yuan.)*

W. Yang is with the School of Mathematics Sciences, University of Chinese Academy of Sciences, Beijing 101408, China, and also with the Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing 100864, China.

J. Yuan, W. Wu, and D.-Z. Du are with the Department of Computer Science, University of Texas at Dallas, Richardson, TX 75080 USA (e-mail: csyuanjing@gmail.com).

J. Ma is with the College of Mathematics and Information Science, Hebei Normal University, Shijiazhuang 050041, China.

Digital Object Identifier 10.1109/TCSS.2019.2891582

probability $1 - \varepsilon$, where n is the number of nodes and m is the number of edges in the input social network. For influence maximization, Bharathi *et al.* [12] had an interesting conjecture that the influence maximization problem is NP-hard even for arborescence directed into a root. This conjecture is proven by Lu *et al.* [13] for the IC model. For the LT model, Wang *et al*. [14] proved that the influence maximization is polynomial-time solvable. This is the first time to know that the IC model and the LT model may give different computational complexities for the same problem.

The influence maximization problem has important applications in the field of viral marketing [15]–[18]. While most research efforts have been devoted to the advertisement of products to single user (i.e., one-user activities), Wang *et al.* [1] study a problem for advertising a product with two-user activities. Their difference in the definition of *activity* leads to an essential change in the property of the mathematical formulation, that is, from submodular maximization to monotone nonsubmodular maximization. Both of them belong to nonlinear combinatorial optimization. In this area, the monotone submodular maximization [19] and the nonmonotone submodular maximization [20]–[22] have been well-studied. However, the monotone nonsubmodular maximization gets ones' attention only recently [1], [23]–[26]. We are going to make a contribution in this research direction.

B. Our Contributions

In this paper, we make several contributions that are summarized as follows.

1) We propose a novel maximizing activity profit (MAP) problem to maximize the expected total profit in a social network. A unique novel feature of our problem is that our objective captures interactions among multiple active users. We are the first to explore the generalized multi-user interactions in the information diffusion process. Our problem includes the classic influence maximization problem as a special case.
2) We analyze the complexity of the MAP problem. We also prove that the objective function for the problem is neither submodular nor supermodular. We present an approximate algorithm that yields an approximation ratio of $(1/(\Delta + 2))$ provided that the supermodular degree is bounded with Δ. We also develop an exchange-based algorithm to further improve the quality of the solution.
3) We prove that computing the exact value of the objective function of MAP problem given a seed set is #P-hard. In order to tackle this challenge, we devise a randomized variation (RV) technique to overcome the computation burden of the problem. Experimental results confirm the benefits of incorporating this technique with our algorithms.
4) We conduct extensive experiments on real benchmark social network data sets to evaluate the efficacy and efficiency of our algorithms. Empirical evaluation results validate the superiority of our algorithms in both effectiveness and efficiency compared with a few baseline heuristics.

The rest of this paper is organized as follows. We formulate the problem in Section II. Complexity results and an approximation algorithm are presented in Section III. We propose an improved exchange-based algorithm and a RV technique in Section IV and V, respectively. Evaluation results are discussed in Section VI. Section VII concludes this paper.

II. Formulation

In the study of influence maximization, two most widely used influence propagation models are the IC model and the LT model. In this paper, we use a directed weighted graph $G = (V, E)$ to capture the structure of the social network and adopt the IC model to capture the diffusion dynamics in the social network. In IC model, each node has two states, either active or inactive, and each directed edge (u, v) is associated with a weight called influence probability, denoted by p_{uv}, measuring the probability that v is activated by u once u becomes active. The activation between a pair of nodes is independent of other nodes in the network. In IC model, time unfolds in discrete slots. In the beginning of the propagation process, or time slot 0, only the nodes in the seed set are active, all the other nodes are inactive. Once a node is activated in a time slot t, he/she has one chance to independently activate each of his/her inactive neighbors, with the probability of corresponding edge weight, in time slot $t + 1$. This process can be simulated as flipping a coin on the edge with a bias of its influence probability showing heads. If the coin flip shows a head, the edge is declared *live* and the activation attempt is considered a success. Otherwise, the edge is declared *blocked* and the activation attempt is considered a failure. An activated node stays active until the end of the diffusion process. The process ends when there is no new node can be activated in the network.

Since we consider activities involving multiple users, we denote an activity as a hyperedge e with a set of head nodes H_e and a set of tail nodes T_e. Note that when $|H_e| = |T_e| = 1$, e reduces to a simple edge. Since a simple edge is a special type of hyperedge, we use the term *edge* without ambiguity to unify the notion of both simple edges and hyperedges in the graph. We employ a simple extension of the original IC model to capture the information diffusion in a hypergraph (i.e., graph containing hyperedges). Given an edge e, once all nodes in its head set H_e become active in time slot t, they have one chance to independently activate each node in the tail set T_e in time slot $t + 1$. The diffusion process ends when no new activations can be made.

Since an activity involves multiple users, it can be represented by this group of users. Therefore, all activities form a hyperedge set of a hypergraph with the node set V, denoted by $\mathcal{A}$. The profit $c : \mathcal{A} \to R^+$ is a nonnegative function. For any seed set S, denote by $I(S)$, the set of active nodes at the end of the diffusion process with initial seed set S. An activity is considered *active* if all nodes in $H_e \cup H_t$ are active. We consider the profit generated by active activities. The expected profit of activities generated by active nodes would be defined as

$$f(S) = E\left[\sum_{A \subseteq I(S), A \in \mathcal{A}} c(A)\right].$$

In this paper, we study the following problem.

Definition 1 (Maximizing Activity Profit): Given a social network $G = (V, E)$ with the extended IC model, a collection of activities, $\mathcal{A}$, a profit function $c : \mathcal{A} \to R^+$, and a positive integer k, find a set S of k seeds to maximize the expected total profit of activities consisting of users activated by S through influence.

From the definition of MAP problem mentioned above, we can see that it is a generalization of many classic influence maximization derived problems, including the following.

1) *Influence Maximization Problem*: When $\mathcal{A}$ is defined as $\mathcal{A} = V$ only including all the single vertex activity and $c : \mathcal{A} \to R^+$ is defined as $c(v) = |v| = 1$, influence maximization problem is a special case of MAP.
2) *Activity Maximization Problem*: When $\mathcal{A}$ is defined as $\mathcal{A} = E$ only including all the edge set activity and $c : \mathcal{A} \to R^+$ is defined as $c(e) = w_e$, activity maximization problem is a special case of MAP.

III. Complexity and Algorithm

In this section, we first give the complexity results of MAP problem. Then, we analyze the properties of the objective function of the MAP problem. Based on these properties, we present the algorithm to solve the MAP problem.

Theorem 1: The MAP problem is NP-hard.

Proof: This is because the influence maximization problem is NP-hard and influence maximization problem is a special case of MAP. □

Theorem 2: For a given set S, the computation of

$$f(S) = E\left[\sum_{A \subseteq I(S), A \in \mathcal{A}} c(A)\right]$$

is $\#P$-hard.

Proof: This is because activity maximization problem is $\#P$-hard and activity maximization problem is a special case of MAP. □

Now, we turn to analyze the submodularity of the objective function of MAP problem. Unfortunately, the objective function of MAP problem is neither submodular nor supermodular. The definitions of submodular and supermodular are given as follows.

Definition 2 (Submodular Function): Suppose that $f : 2^V \to R^+$ is the nonnegative set value function, where V is the ground set. f is called submodular if for any subset S_1, S_2 of V with $S_1 \subseteq S_2 \subseteq V$, and any $v \in V \backslash S_2$, we have $\Delta_v f(S_1) \geq \Delta_v f(S_2)$, where $\Delta_v f(S) = f(S \cup \{v\}) - f(S)$.

Definition 3 (Supermodular Function): Suppose that $f : 2^V \to R^+$ is the nonnegative set value function, where V is the ground set. f is called supermodular if for any subset S_1, S_2 of V with $S_1 \subseteq S_2 \subseteq V$, and any $v \in V \backslash S_2$, we have $\Delta_v f(S_1) \leq \Delta_v f(S_2)$.

Theorem 3: The objective function

$$f(S) = E\left[\sum_{A \subseteq I(S), A \in \mathcal{A}} c(A)\right]$$

of MAP is nonsubmodular.

Proof: This is because the objective function of activity maximization problem is nonsubmodularity and activity maximization problem is a special case of MAP as we explained in Section II. □

Similarly, we have

Theorem 4: The objective function

$$f(S) = E\left[\sum_{A \subseteq I(S), A \in \mathcal{A}} c(A)\right]$$

of MAP is nonsupermodular.

Proof: This is because the objective function of activity maximization problem is nonsupermodularity and Activity maximization problem is a special case of MAP. □

A. Supermodular Degree Algorithm

In this section, we use the supermodular degree to measure to what degree the nonsubmodular version violates the submodularity inspired by Feldman and Izsak [24]. When the supermodular degree is bounded, denoted by Δ, an algorithm improved extendible system greedy (IESG) with consistant approximation ratio $(1/(\Delta + 2))$ is proposed. For this propose, we first give some definitions with respect to supermodularity.

Definition 4 (Supermodular Set): Given a monotone set value objective function $f(\cdot)$, the supermodular set of a node $v \in V$ is $D_f^+(v) = \{v' \in V | \Delta_v f(S \cup \{v'\}) > \Delta_v f(S), \exists S \subseteq V\}$, which includes all nodes that might increase the marginal gain of $v \in V$.

Definition 5 (Supermodular Degree): The supermodular degree, denoted by Δ, is defined as the maximum cardinality among all supermodular sets, i.e., $\Delta = \max_{v \in V} | D_f^+(v) |$.

It is obvious that when $\Delta = 0$, the function $f(\cdot)$ is submodular. As for the nonsubmodular case of influence maximization problem with bounded supermodular degree, we design an IESG algorithm based on the average marginal gain. As given in Algorithm 1, we start with an empty set S_0. The algorithm iteratively selects a set of seed nodes that mitigates the negative effect of the nonsubmodularity. Once a node v is selected as a seed node, partial nodes in $D_f^+(v)$ (denoted by D'_v) are jointly selected as seed nodes. In each iteration, a node v is selected along with nodes in its supermodular set to maximize the marginal gain of the current solution set. While the constraint on the size of the solution set, k, is satisfied, we update the seed set with newly selected nodes. The algorithm runs until k seed nodes are selected.

Compared with the naive greedy algorithm, we use a new rule in IESG to select the nodes that maximize the joint marginal gain instead of a single node with maximal marginal increment. This rule is more beneficial due to the monotone property of the objective function. The idea is that the influence of a node can be boosted by the nodes in its supermodular set, in each iteration, the algorithm always selects the set of nodes with maximal boosted influence.

IV. Analysis

In this section, we analyze the approximation ratio of the IESG algorithm. Then, we introduce the exchange improvement algorithm (EIA) to further improve the performance.

Algorithm 1 IESG

Input: a hyper-graph, G, and a constant, k.
Output: a set of seed nodes, S.

1: Initialize $i = 0$ and $S_0 = \emptyset$.
2: While $|S_i| < k$ do
3: Find out $\arg\max_{v\in V, D'_v \subseteq D^+_f(v)} [f(S_i \cup \{v\} \cup D'_v) - f(S_i \cup D'_v)]$, constrained by $|S_i \cup \{v\} \cup D'_v| < k$.
4: Update $S_{i+1} = S_i \cup \{v\} \cup D'_v$.
5: Update i to be $i+1$.
6: Return $S = S_i$ as the set of seed nodes.

A. Approximation Ratio of IESG

First, we show that the IESG algorithm designed above has $(1/(\Delta+2))$ approximation ratio. Then, we present an EIA for MAP problem to further improve the performance of the solution.

Theorem 5: Algorithm IESG has an approximation ratio of $(1/(\Delta+2))$ to the optimal solution.

Proof: Let S^* denote the optimal set of seed nodes, in terms of maximizing $f(\cdot)$. An auxiliary parameter, H_i, is used. With $H_0 = S^*$, H_i is recursively defined as an arbitrary subset of $H_{i-1} \cup S_i$, under the constraint that $S_i \subseteq H_i$ and $|H_i| = k$. Intuitively, H_i consists of S_i and a part of S^*. When i becomes larger, nodes from S_i are added to H_i, and nodes in S^* are removed from H_i, maintaining $|H_i| = k$. By definition, we have

$$|H_{i+1}| = |H_i| - |H_i \backslash H_{i+1}| + |S_{i+1} \backslash S_i|. \quad (1)$$

Since $|H_{i+1}| = |H_i| = k$, we have

$$|H_i \backslash H_{i+1}| = |S_{i+1} \backslash S_i| \le |\{v\} \cup D'_v| \le 1 + \Delta. \quad (2)$$

This is because $D'_v \subseteq D^+_f(v)$ and $\Delta = \max_v |D^+_f(v)|$. Equation (2) means that, in each greedy iteration, at most $1+\Delta$ nodes in S^* are ignored by Algorithm IESG.

We claim that the marginal gain in each greedy iteration of Algorithm IESG has a lower bound with respect to $f(H_i)$

$$f(H_i) - f(H_{i+1}) \le (1+\Delta)\big[f\big(S_i \cup \{v\} \cup D'_v\big) - f\big(S_i \cup D'_v\big)\big]. \quad (3)$$

To prove (3), let us order nodes of $H_i \backslash H_{i+1}$ in an arbitrary order (say $v_1, v_2, \ldots, v_l$), and let $H_i^j = H_i \backslash \{v_1, v_2, \ldots, v_j\}$ for $1 \le j \le l$, where $H_i^0 = H_i$ and $H_i^l \subseteq H_{i+1}$. For each j, we have

$$\begin{aligned} f\big(S_i \cup \{v_j\} \cup \big(D'_{v_j} \cap H_i^j\big)\big) &- f\big(S_i \cup \big(D'_{v_j} \cap H_i^j\big)\big) \\ &\ge f\big(S_i \cup \{v_j\} \cup H_i^j\big) - f\big(S_i \cup H_i^j\big) \\ &= f\big(S_i \cup H_i^{j-1}\big) - f\big(S_i \cup H_i^j\big). \end{aligned} \quad (4)$$

The inequality is from the definition of the modularity set, because only nodes in $D^+_f(v_j)$ can increase the marginal gain of v_j. Hence, nodes in $H_i^j \backslash D^+_f(v_j)$ might decrease the marginal gain of v_j. The equality results from the definition of H_i^j, since $\{v_j\} \cup H_i^j = H_i^{j-1}$. By accumulating (4) among j, we obtain

$$\begin{aligned} &\sum_{j=1}^{l}\big[f\big(S_i \cup \{v_j\} \cup \big(D'_{v_j} \cap H_i^j\big)\big) - f\big(S_i \cup \big(D'_{v_j} \cap H_i^j\big)\big)\big] \\ &\ge \sum_{j=1}^{l}\big[f\big(S_i \cup H_i^{j-1}\big) - f\big(S_i \cup H_i^j\big)\big] \\ &= f\big(\big(S_i \cup H_i^0\big) - f\big(S_i \cup H_i^l\big)\big) \ge f(H_i) - f(H_{i+1}). \end{aligned} \quad (5)$$

The first inequality is from (4). The last inequality is because $f(S_i \cup H_i^0) = f(H_i)$ and $f(S_i \cup H_i^l) \le f(H_{i+1})$. We have $f(S_i \cup H_i^0) = f(H_i)$, since $H_i^0 = H_i$ and $S_i \subseteq H_i$. We have $f(S_i \cup H_i^l) \le f(H_{i+1})$ by the monotonicity, since $S_i \subseteq S_{i+1} \subseteq H_{i+1}$ and $H_i \subseteq H_{i+1}$. We have

$$\begin{aligned} &(1+\Delta)\big[f(S_i \cup \{v\} \cup D'_v) - f\big(S_i \cup D'_v\big)\big] \\ &\ge \sum_{j=1}^{l}\big[f\big(S_i \cup \{v\} \cup D'_v\big) - f\big(S_i \cup D'_v\big)\big] \\ &\ge \sum_{j=1}^{l}\big[f\big(S_i \cup \{v_j\} \cup \big(D'_{v_j} \cap H_i^j\big)\big) - f\big(S_i \cup \big(D'_{v_j} \cap H_i^j\big)\big)\big] \\ &\ge f(H_i) - f(H_{i+1}). \end{aligned} \quad (6)$$

The first inequality results from (2), in which $1 \le j \le |H_i \backslash H_{i+1}| \le 1 + \Delta$. The second inequality comes from Line 3 in Algorithm IESG, which always selects the maximum average marginal gain in each greedy iteration. The third inequality comes from (5). Therefore, (3) is valid.

Since the marginal gain in each greedy iteration of Algorithm IESG has a lower bound, we can accumulate (3) among all greedy iterations (note that $S_{i+1} = S_i \cup \{v\} \cup D'_v$)

$$\begin{aligned} f(H_0) - f(H_I) &= \sum_{i=0}^{I-1}[f(H_i) - f(H_{i+1})] \\ &\le \sum_{i=0}^{I-1}(1+\Delta)\big[f\big(S_i \cup \{v\} \cup D'_v\big) - f(S_i \cup D'_v)\big] \\ &\le (1+\Delta)\sum_{i=0}^{I-1}\big[f(S_i \cup \{v\} \cup D'_v) - f(S_i)\big] \\ &\le (1+\Delta)\sum_{i=0}^{I-1}\big[f(S_{i+1}) - f(S_i)\big] \\ &\le f(S_I) - f(S_0). \end{aligned} \quad (7)$$

Here, we assume that the algorithm IESG terminates in I iterations. The first inequality is again by the monotonicity of $f(S_i \cup D'_v) \ge f(S_i)$ based on $(S_i \cup D'_v) \supseteq S_i$. Since $H_0 = S^*$, $H_I = S_I = S$ when Algorithm IESG terminates, and $S_0 = \emptyset$, we have

$$f(S^*) \le (\Delta + 2) f(S). \quad (8)$$

The theorem follows. $\square$

B. Exchange Improvement Algorithm

Although we can obtain approximation ratio for MAP, there is still a big gap between the approximate solution and the original optimal solution. In this section, we present the EIA designed based on the nice properties observed in MAP. We first discuss the optimization condition of MAP and obtain an optimum criterion. Then, we discuss M-convexity of the feasible region of MAP. At last, we present the EIA.

First, we derive the optimization condition. Although greedy algorithm IESG returns $(1/(\Delta+2))$-approximate solution to MAP, it is still a big gap between the approximate solution obtained and the optimal one in theory. How to determine whether an approximate solution can be improved or not is still a fundamental question deserve further study. To this problem, we have the following theorem.

Theorem 6: Suppose S^* is the optimum solution of the MAP problem, then

$$\min_{S\subseteq S^*} \Delta_S f(S^*\backslash S) \geq \max_{S\subseteq V\backslash S^*} \Delta_S f(S^*\backslash S_R) \tag{9}$$

where $S_R = \arg\min_{S\subseteq S^*} \Delta_S f(S^*\backslash S)$ and

$$f(S) = E\left[\sum_{A\subseteq I(S), A\in\mathcal{A}} c(A)\right].$$

Proof: Suppose the following inequation:

$$\min_{S\subseteq S^*} \Delta_S f(S^*\backslash S) \geq \max_{S\subseteq V\backslash S^*} \Delta_S f(S^*\backslash S_R)$$

is not satisfied, then there exists $S_A \subseteq V\backslash S^*$ such that $\Delta_{S_R} f(S^*\backslash S_R) < \Delta_{S_A} f(S^*\backslash S_R)$.

On the other hand

$$f(S^*) = f(S^*\backslash S_R) + \Delta_{S_R} f(S^*\backslash S_R) \tag{10}$$

$$f(S^* - S_R + S_A) = f(S^*\backslash S_R) + \Delta_{S_A} f(S^*\backslash S_R). \tag{11}$$

Therefore, $f(S^* - S_R + S_A) > f(S^*)$. This is conflict with the optimality of S^*. □

When just the singleton subset of S^* is considered, a corollary of the above optimization criterion is immediately obtained as follows.

Corollary 1: Suppose S^* is the optimum solution of the MAP problem, then

$$\min_{1\leq i\leq k} \Delta_{v_i} f(S^*\backslash\{v_i\}) \geq \max_{v\in V\backslash S^*} \Delta_v f(S^*\backslash\{v_R\}) \tag{12}$$

where $v_R = \min_{1\leq i\leq k} \Delta_{v_i} f(S^*\backslash\{v_i\})$.

Proof: The proof is similar to optimization criterion theorem above. Suppose $\min_{1\leq i\leq k} \Delta_{v_i} f(S^*\backslash\{v_i\}) \geq \max_{v\in V\backslash S^*} \Delta_v f(S^*\backslash\{v_R\})$ is not satisfied, then there exists $v_A \subseteq V\backslash S^*$ such that $\Delta_{v_R} f(S^*\backslash\{v_R\}) < \Delta_{v_A} f(S^*\backslash\{v_R\})$.

On the other hand

$$f(S^*) = f(S^*\backslash\{v_R\}) + \Delta_{v_R} f(S^*\backslash\{v_R\}) \tag{13}$$

$$f(S^* - \{v_R\} + \{v_A\}) = f(S^*\backslash\{v_R\}) + \Delta_{v_A} f(S^*\backslash\{v_R\}). \tag{14}$$

Therefore, $f(S^* - \{v_R\} + \{v_A\}) > f(S^*)$. This is conflict with the optimality of S^*. □

Now, we turn to the property of the feasible region of MAP problem. It is easy to see that the feasible region of MAP is an M-convex set [27].

Algorithm 2 EIA

Input: a solution S_0 of influence function maximization problem with $|S_0| = k$.
Output: a set of non-improvable solution $\widehat{S}$.
1: Initialize $i = 0$ and $S = S_0$
2: Find out $v_R = \arg\min_{v\in S} \Delta_v f(S\backslash\{v\})$, let $\Delta(v_R) = \Delta_{v_R} f(S\backslash\{v_R\})$
3: Find out $v_A = \arg\max_{v\in V\backslash S} \Delta_v f(S\backslash\{v_R\})$ and let $\Delta(v_A) = \Delta_{v_A} f(S\backslash\{v_R\})$
4: If $\Delta(v_R) \geq \Delta(v_A)$, then S is non-improvable, stop; otherwise $S := S - v_R + v_A$ and go to step 2.

Theorem 7: The feasible region of MAP problem $F = \{S | S \in \{0,1\}^V \bigwedge \| S \|_1 = k\}$ is M-convex set.

Proof: For $S_1, S_2 \in F$, and $u \in \text{supp}^+(S_1 - S_2)$, there exists $v \in \text{supp}^-(S_1 - S_2)$ such that $S_1 - \chi_u + \chi_v \in F$ and $S_2 + \chi_u - \chi_v \in F$ where χ_u is the indicator vector of singleton set $\{u\}$. In fact, any $v \in S_2\backslash S_1$ is candidate that meets the requirements. By the definition of M-convex set, it follows the theorem. □

Make use of the M-convexity of feasible region of influence function maximization problem, a feasible solution is reserved unchanged under exchange operations. Therefore, we have the following exchange improvement property.

Theorem 8: For any feasible solution S of MAP problem, if necessary condition for optimality (12) is not satisfied, then S can be improved through exchange operations.

Proof: According to the optimization criterion and M-convexity of feasible region, $S - \chi_{v_R} + \chi_{v_A} \in F$ outperforms S, that is, $f(S - \chi_{v_R} + \chi_{v_A}) = f(S) + \Delta_{v_A} f(S\backslash\{v_R\}) - \Delta_{v_R} f(S\backslash\{v_R\}) > f(S)$, if necessary condition for optimality (12) is not satisfied. □

Now, the definition of nonimprovable solution of activity profit maximization problem is presented.

Definition 6 (Nonimprovable Solution): S is said to be nonimprovable solution of MAP problem, if condition (12) in the necessary condition for optimality (Corollary1) is satisfied.

Inspired by the exchange improvement property mentioned above, we design an EIA as follows. The basic idea behind the EIA is to replace the node with the minimum marginal gain in the current solution with the node with the maximum marginal gain in $V\backslash S$.

V. Randomized Variation

Although the algorithm designed in the above sections have good approximate performance, there is still insurmountable difficulties to solve the MAP problem in practice due to the NP-hardness of the problem and the $\#P$-hardness of the objective function value computation. Furthermore compute the supermodular set is still NP-hardness. In this section, we use random technique design practical algorithm for the MAP problem.

In improved greedy algorithm (IESG), we should determine supermodular set $D_f^+(v)$ for each node $v \in V$ at the initial phase. Unfortunately, to decide whether a node u is in supermodular set $D_f^+(v)$ of any given node v is NP-hard [28].

Theorem 9 [28]: To determine whether a node u is in supermodular set $D_f^+(v)$ of any given node v is NP-hard.

As proposed by Theorem 2, in social network computation, the spread value of a given S is #P-hard. Traditionally, $f(S)$ is evaluated through random sampling method such as Monte Carlo simulation but how to determine the simulation number is very difficult. Therefore, it is meaningful to find effective and efficient ways to evaluate the $f(S)$ directly for any given S. In this section, we given a successive iteration update method (SIUM) to compute the $f(S)$ from the marginal gain perspective.

Definition 7 (Marginal Gain): Suppose that $f: 2^V \rightarrow R^+$ is nonnegative set value function, where V is the ground set. For any subset S of V, $\Delta_v f(S) = f(S \cup \{v\}) - f(S)$ is called marginal gain of $v \in V \backslash S$ at S. In addition, we can define $\Delta_T f(S) = f(S \cup T) - f(S)$ as marginal gain of $T \subseteq V \backslash S$ at S in the similar way.

Based on the definition of marginal gain, we have the following property in the marginal increment form.

Property 1: For a given set $S \subseteq V$ and any subset $T \subseteq S$, then we have

$$f(S) = f(S \backslash T) + \Delta_T f(S \backslash T) = f(T) + \Delta_{S \backslash T} f(T)$$

for any given set value function f.

Let $f(S) = \sum_{u \in V} q_u^S$ is the influence function value (or the spread value) of a given S, where q_u^S is the active probability (or expectation equivalently) of the node u when S is selected as seed set under the IC propagation model. Denote p_{vu} the probability of node v can activate node u along a given path. Using these notations, we have the following property.

Property 2: $f(S) = \sum_{u \in V} [p_{vu} + (1 - p_{vu}) q_u^{S \backslash v}]$ for any S and $v \in S$.

Proof: According to property 1, we have

$$f(S) = f(S \backslash v) + \Delta_v f(S \backslash v)$$

where $f(S \backslash v) = \sum_{u \in V} q_u^{S \backslash v}$ and

$$\begin{aligned} \Delta_v f(S \backslash v) &= \sum_{u \in V} \left[1 - \left(1 - q_u^{S \backslash v}\right)(1 - p_{vu}) - q_u^{S \backslash v}\right] \\ &= \sum_{u \in V} \left(1 - q_u^{S \backslash v}\right) p_{vu}\big). \end{aligned}$$

Thus, we have

$$\begin{aligned} f(S) &= \sum_{u \in V} \left[q_u^{S \backslash v} + \left(1 - q_u^{S \backslash v}\right) p_{vu}\right] \\ &= \sum_{u \in V} \left[p_{vu} + (1 - p_{vu}) q_u^{S \backslash v}\right]. \end{aligned}$$

□

Thus, we have an upper bound of

$$f(S) = E\left[\sum_{A \subseteq I(S), A \in \mathcal{A}} c(A)\right]$$

which can be reformulated as $f(S) \leq \sum_{A \subseteq I(S), A \in \mathcal{A}} (c(A) \min_{v \in A} \{q_v^S\})$. The upper bound can be further reformulated as $\overline{f}(S) = \sum_{A \subseteq I(S), A \in \mathcal{A}} (c(A)(1/|A|) \sum_{v \in A} \{q_v^S\}$ due to the inequality $\min_{v \in A} \{q_v^S\} \leq (1/|A|) \sum_{v \in A} \{q_v^S\}$.

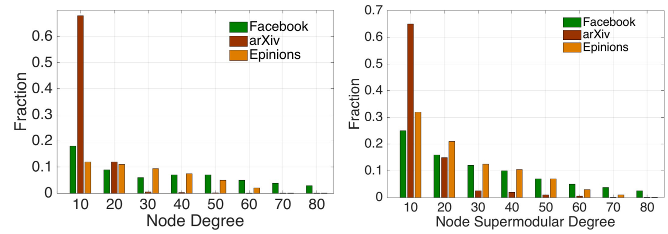


Fig. 1. Node degree and supermodular degree distribution.

Theorem 10: The upper bound objective function $\overline{f}(S)$ of MAP is submodular.

Proof: This is because the upper bound of objective function of MAP $\overline{f}(S) = \sum_{A \subseteq I(S), A \in \mathcal{A}} (c(A)(1/|A|) \sum_{v \in A} \{q_v^S\}) = \sum_{v \in V} q_v^S \overline{c}(v)$, where $\overline{c}(v) = \sum_{v \in A, A \in \mathcal{A}} (1/|A|) c(A)$. This is a weight version of influence maximization problem which is submodular. □

Considering the NP-hardness of both the MAP problem and supermodular set determine, a polling-based method inspired by Wang *et al.* [1] and based on algorithmic framework established in [29] and [10] can be generalized well to the MAP problem. For a given social hypernetwork G, IC information diffusion model, and a seed set S, let g be a "live-edge" graph instance of G and $R_g(S)$ be the set of nodes reachable from S in g. Denote by $R_{g^T}(v)$ the reverse reachable set for node v in g, where g^T is the transpose graph of $g: (u, v) \in g \iff (v, u) \in g^T$. We write $A \sim \mathcal{A}$ to indicate that we randomly pick A from $\mathcal{A}$ as a sample according to a certain distribution. Then, we have the following result.

Theorem 11: For any seed set $S \in V$

$$f(S) = C \cdot Pr_{g \sim G, A \sim \mathcal{A}}\left[\bigcap_{v \in A} (S \cap R_{g^T}(v)) \neq \emptyset\right]$$

where $C = \sum_{A \in \mathcal{A}} c(A)$.

Proof:

$$\begin{aligned} f(S) &= E\left[\sum_{A \subseteq I(S), A \in \mathcal{A}} c(A)\right] \\ &= \sum_{A \in \mathcal{A}} Pr_{g \sim G}\left[\bigcap_{v \in A} (S \cap R_{g^T}(v)) \neq \emptyset\right] c(A) \\ &= C \cdot \sum_{A \in \mathcal{A}} Pr_{g \sim G}\left[\bigcap_{v \in A} (S \cap R_{g^T}(v)) \neq \emptyset\right] \frac{c(A)}{C} \\ &= C \cdot Pr_{g \sim G, A \sim \mathcal{A}}\left[\bigcap_{v \in A} (S \cap R_{g^T}(v)) \neq \emptyset\right]. \end{aligned}$$

In the fourth equality, the activity A is randomly picked with probability $(c(A)/C)$. The theorem is obtained. □

VI. PERFORMANCE EVALUATION

In this section, we conduct extensive experiments on benchmark data sets to evaluate the effectiveness and efficiency of

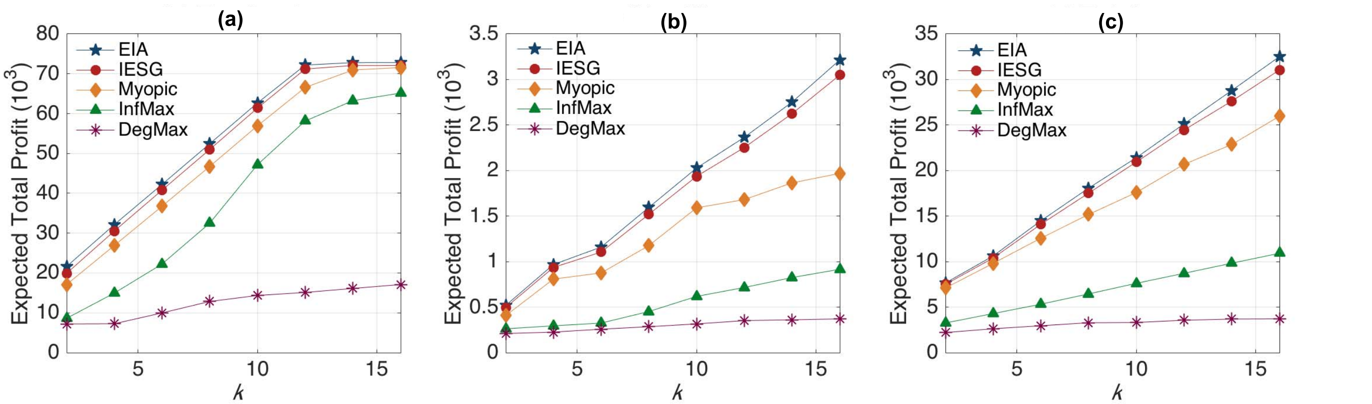


Fig. 2. Expected total profit versus seed set size produced by various algorithms under uniform profit setting. (a) Facebook. (b) arXiv. (c) Epinions.

TABLE I

CHARACTERISTICS OF SOCIAL NETWORK DATA SETS

Name	#Nodes	#Edges	Avg. degree
Facebook	899	72,821	165
arXiv	16,726	66,759	11.9
Epinions	22,166	353,546	33.5

the proposed algorithms. Table I summarizes the basic statistics of the data sets used in our experiments. Fig. 1 presents the distributions of node degree and node supermodular degree for each data set.

We compare our algorithms with three baseline algorithms. `Myopic` selects a node with the largest increment of the total profit in each iteration, until k seeds are selected. `DegMax` selects a node with the highest out-degree in each iteration, until k seeds are selected. `InfMax` selects a node with the largest increment of the influence spread in each iteration, until k seeds are selected. This algorithm only considers social influence but ignores the profit distribution.

A. Quality Evaluation

Experimental results are reported on two metrics: the expected total profit yielded and the sensitivity to varying activity profit distributions.

Figs. 2 and 3 shows the expected total profit yielded by various algorithms and baselines when the seed set size, k, ranges from 2 to 16. The x-axis holds the value of the size of the seed set and the y-axis holds the expected total profit gained from activities. Note that Fig. 2 shows the results of various algorithms under the uniform setting for activity profit distribution. Specifically, each activity is assigned a profit equal to 1. In this case, the expected total profit is exactly the number of active edges among the nodes. The expected total profit grows as k increases, since a larger k increases the chance for seeds to influence more nodes, leading to more active activities at the end of the diffusion process. Fig. 2(a)–(c) manifest that `EIA` and `IESG` outperform all the other baselines for any k under the uniform setting. We also observe that the gap between the `DegMax` algorithm and other algorithms becomes larger as k increases. `DegMax` performs poorly as it only uses the structure properties of the social network without considering social influence. `InfMax` fails to find good solutions since it only considers the diffusion process but the activity profit is not examined during seed selection. `Myopic` only selects one seed node at a time without considering the combinations of nodes that may activate many more activities via hyperedges, i.e., the marginal increase in the expected profit of a node may be boosted by the nodes in its supermodular set.

For data set *Facebook*, Fig. 2(a) shows that the expected total profit yielded by both `EIA` and `IESG` converges as k reaches 12. The underlying reason is that compared with the other two data sets, *Facebook* contains a smaller number of nodes and has a much denser network structure. Therefore, a small set of seed nodes is possible to influence almost all the nodes in the social graph. On this data set, `Myopic` has a similar performance compared with `EIA` and `IESG`. On *arXiv* and *Epinions*, the expected total profit produced by all the algorithms keeps growing as k increases. The gap between `Myopic` and our proposed algorithms becomes larger as k increases. Taking the node combinations into account, `EIA` and `IESG` obtain a better solution by taking advantage of the power of exchange and examining a node's supermodular set to incorporate the influence boosting from selecting more than one seed at a time.

Fig. 3 shows the results of various algorithms under the influence probability setting for activity profit distribution. Specifically, each activity is assigned a profit equal to the influence probability associated with the corresponding edge in the graph. As expected, the expected total profit increases as k increases. Fig. 3(a)–(c) manifest that the results are quite similar to those shown in Fig. 2(a)–(c). The proposed algorithms `EIA` and `IESG` consistently outperform all the baseline algorithms in all cases. The reason is that the baseline algorithms only consider the structure properties of the social network or the influence diffusion process but totally ignore the distribution of activity profits. This is why our proposed algorithms always have a good performance while the baseline algorithms fail in many cases.

It was worth noting that results produced by `InfMax` exhibit a better performance under the influence probability setting as measured by the expected total profit. In each iteration, `InfMax` selects a seed node that leads to the largest marginal increment of the influence spread based on current seed set until k seed nodes are selected. Therefore, `InfMax`

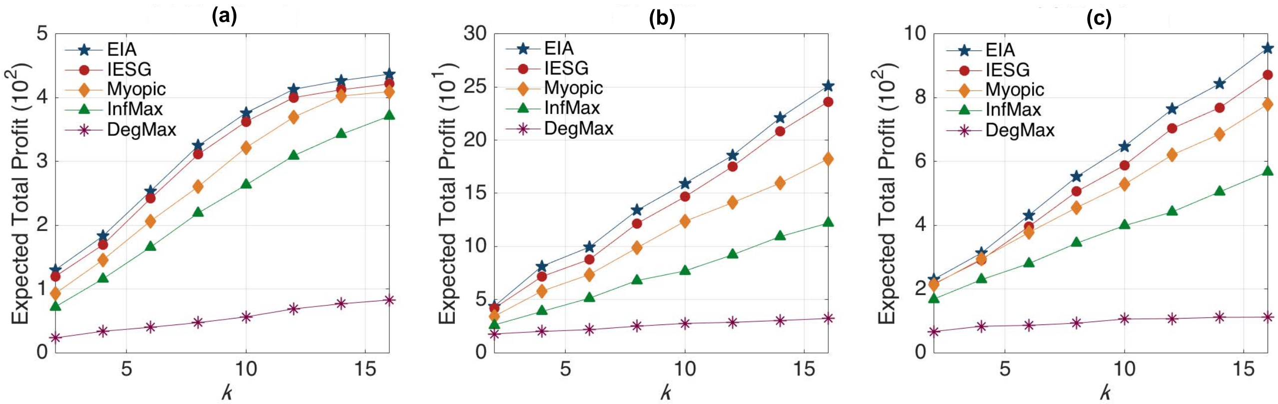


Fig. 3. Expected total profit versus seed set size produced by various algorithms under influence profit setting. (a) Facebook. (b) arXiv. (c) Epinions.

tends to select a group of highly influential nodes as the seed set. Under the influence probability setting, the activity profit for each edge is equal to the influence probability associated with the edge, and edges with higher influence probabilities have higher chance to become live and contribute to the influence diffusion. While the algorithm selects the seed set that leads to a largest influence spread, as a side effect, the edges with higher influence probability (i.e., activity profit) become active and contribute to the total profit obtained at the end of the diffusion process.

To understand the effectiveness of various algorithms, we take a closer look at the results across different data sets. We plot the distributions of node degree and node supermodular degree for each data set, and the result is illustrated in Fig. 1. We observe that the fraction of nodes decreases rapidly as the node degree increases. The fraction of nodes with higher degree is much smaller than the fraction of nodes with lower degree. These results empirically illustrate the phenomenon that the node degree distribution follows a power-law distribution for each of the three tested data sets. Recall that as we discussed in Section III, the supermodular degree Δ is defined as the size of the largest supermodular set of the nodes in the social graph. It measures the degree to which the objective function violates the submodularity. As illustrated in Fig. 1, the fraction of nodes with supermodular set of size over 80 is less than 2% for data set *Facebook*, and it is less than 0.1% for data sets *arXiv* and *Epinions*. These results empirically demonstrate that for each of the three data sets, the supermodular degree of the social graph is very small compared with the number of nodes $|V|$ in the graph. This intuitively suggests that the degree to which the objective function violates the submodularity is limited. Therefore, the improved greedy algorithm has a good performance on each of all three data sets. Taking the node combinations into account, `IESG` fully utilizes the power of boosted influence from a node's modularity set and selects a high-quality set of seed nodes in each iteration, leading to a comprehensive solution set as a result.

B. Scalability Evaluation

We further evaluate the efficiency and scalability of our algorithms. In our experiments, we adopt two synthetic profit settings to evaluate our proposed algorithms. Since the activity profit settings do not affect the running time, we only report the result of running time on the three data sets.

Fig. 4 shows the running time produced by our algorithms and the baselines with varying seed set sizes. The seed set size k varies from 2 to 16. The x-axis holds the value of k and the y-axis holds the running time (measured in seconds). As can be seen, our algorithms show scalability and efficiency and finish within 2 min for *Epinions*. The running time of the algorithms grows when the size of seed set k increases. The growth rate is almost linear except `DegMax`. Taking the source combinations into account, `EIA` and `IESG` examine the combinations of node set and obtain a better solution by spending more time since the number of node combinations is much higher than the number of nodes. `DegMax` has the shortest running time, as it only considers the node degree when selecting the seed set. `InfMax` and `Myopic` are also faster than our algorithms. Since they only consider the influence diffusion and marginal increment of single nodes but totally ignore the distribution of activity profits and node combinations, their performance is substantially weaker than our algorithms. Therefore, it is worthwhile to take a bit more time to produce comprehensive solution sets with higher quality.

We further evaluate the efficiency of the RV as discussed in Section V. The simple greedy algorithm for estimating the expected influence spread is computationally prohibitive, as computing the marginal gain for each node is #P-hard [4], and is typically approximated by a sufficiently large number of Monte Carlo simulations. In our problem, we have proved in Section III that computing the expected profit of activities given a seed set is also #P-hard. Next, we incorporate the RV technique to evaluate the benefit of the method in terms of performance and efficiency.

Fig. 5 shows the performance and running time produced by `IESG` and `IESG-RV`, improved greedy with RV. This set of experiments is conducted on the data set *Epinions*. Fig. 5(b) shows the running time of both algorithms. The results are reported with the varying seed set sizes. The seed set size k varies from 2 to 16. The x-axis holds the value of k and the y-axis holds the running time (measured in seconds). As we

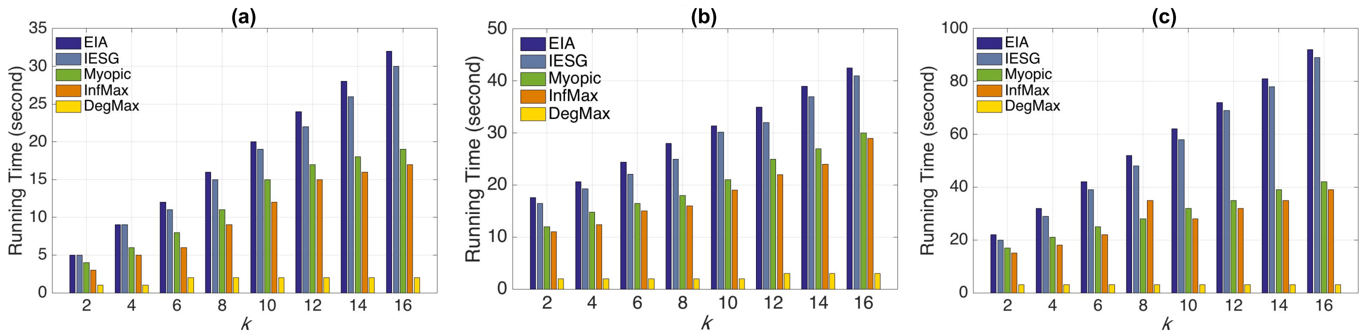


Fig. 4. Running time versus seed set size produced by various algorithms. (a) Facebook. (b) arXiv. (c) Epinions.

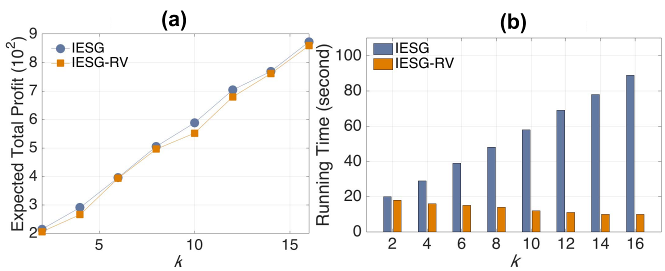


Fig. 5. Performance versus running time for random variation. (a) Performance. (b) Running time.

discussed earlier, the running time of `IESG` grows when the size of seed set k increases, and the growth rate is almost linear. In contrast, the running time of *IG-RV* decreases when k increases. This is because the time cost in the RV depends on the number of sampled edges. As we discussed in Section V, the expected number of samples is inversely proportional to the probability of the event $\bigcap_{v \in A}(S \cap R_{g^T}(v)) \neq \emptyset$. The latter increases when k increases. The improved greedy with RV is faster than the original improved greedy algorithm on *Epinions* in all cases. Specifically, when $k = 16$, the running time is reduced by 89%. This empirically suggests that the RV proposed in Section V provides a promising technique to speed up the improved greedy algorithm, especially on large data sets.

Fig. 5(a) shows the performance of the two algorithms as measured by the expected total profit on *Epinions*. The parameter setting is the same as above. We observe that the performance of the two algorithms is quite similar with varying size of the seed set. Taking both effectiveness and efficiency into account, it demonstrates that by incorporating with the RV technique, the running time of the improved greedy algorithm is reduced significantly while the good performance is preserved.

VII. Conclusion

In this paper, we study a novel and important MAP problem. We are the first to explore the generalized multiuser interactions in information diffusion. Our problem includes several classic influence maximization related problems as special cases. We present an approximate algorithm with approximate ratio of $\frac{1}{\Delta+2}$ provided that the supermodular degree is bounded with Δ. We design an exchange-based algorithm to further improve the quality of the solution. We develop a RV technique to reduce the computation burden of the MAP problem.

References

[1] Z. Wang, Y. Yang, J. Pei, and E. Chen. (2016). "Activity maximization by effective information diffusion in social networks." [Online]. Available: https://ieeexplore.ieee.org/document/8010858

[2] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proc. 9th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2003, pp. 137–146.

[3] W. Xu, Z. Lu, W. Wu, and Z. Chen, "A novel approach to online social influence maximization," *Social Netw. Anal. Mining*, vol. 4, no. 1, p. 153, 2014.

[4] W. Chen, C. Wang, and Y. Wang, "Scalable influence maximization for prevalent viral marketing in large-scale social networks," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2010, pp. 1029–1038.

[5] W. Chen, Y. Yuan, and L. Zhang, "Scalable influence maximization in social networks under the linear threshold model," in *Proc. IEEE 10th Int. Conf. Data Mining (ICDM)*, Dec. 2010, pp. 88–97.

[6] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2009, pp. 199–208.

[7] Y. Zhang, Q. Gu, J. Zheng, and D. Chen, "Estimate on expectation for influence maximization in social networks," in *Proc. Adv. Knowl. Discovery Data Mining*, pp. 99–106, 2010.

[8] A. Goyal, F. Bonchi, and L. V. S. Lakshmanan, "A data-based approach to social influence maximization," *Proc. VLDB Endowment*, vol. 5, no. 1, pp. 73–84, Sep. 2011.

[9] Y. Tang, X. Xiao, and Y. Shi, "Influence maximization: Near-optimal time complexity meets practical efficiency," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2014, pp. 75–86.

[10] Y. Tang, Y. Shi, and X. Xiao, "Influence maximization in near-linear time: A martingale approach," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2015, pp. 1539–1554.

[11] H. T. Nguyen, M. T. Thai, and T. N. Dinh, "Stop-and-stare: Optimal sampling algorithms for viral marketing in billion-scale networks," in *Proc. Int. Conf. Manage. Data*, 2016, pp. 695–710.

[12] S. Bharathi, D. Kempe, and M. Salek, "Competitive influence maximization in social networks," in *Internet and Network Economics*. Berlin, Germany: Springer, 2007, pp. 306–311.

[13] Z. Lu, Z. Zhang, and W. Wu, "Solution of Bharathi–Kempe–Salek conjecture for influence maximization on arborescence," *J. Combinat. Optim.*, vol. 33, no. 2, pp. 803–808, 2017.

[14] A. Wang, W. Wu, and L. Cui, "On Bharathi–Kempe–Salek conjecture for influence maximization on arborescence," *J. Combinat. Optim.*, vol. 31, no. 4, pp. 1678–1684, 2016.

[15] M. Richardson and P. Domingos, "Mining knowledge-sharing sites for viral marketing," in *Proc. 8th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2002, pp. 61–70.

[16] J. Leskovec, L. A. Adamic, and B. A. Huberman, "The dynamics of viral marketing," *ACM Trans. Web*, vol. 1, no. 1, p. 5, 2007.

[17] P. Domingos and M. Richardson, "Mining the network value of customers," in *Proc. 7th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2001, pp. 57–66.

[18] F. Bonchi, C. Castillo, A. Gionis, and A. Jaimes, "Social network analysis and mining for business applications," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, p. 22, Apr. 2011.

[19] G. L. Nemhauser and L. A. Wolsey, "Best algorithms for approximating the maximum of a submodular set function," *Math. Oper. Res.*, vol. 3, no. 3, pp. 177–188, 1978.
[20] U. Feige, V. S. Mirrokni, and J. Vondrák, "Maximizing non-monotone submodular functions," *SIAM J. Comput.*, vol. 40, no. 4, pp. 1133–1153, 2011.
[21] J. Lee, V. S. Mirrokni, V. Nagarajan, and M. Sviridenko, "Non-monotone submodular maximization under matroid and knapsack constraints," in *Proc. 41st Annu. ACM Symp. Theory Comput.*, 2009, pp. 323–332.
[22] M. Feldman, J. Naor, and R. Schwartz, "A unified continuous greedy algorithm for submodular maximization," in *Proc. IEEE 52nd Annu. Symp. Found. Comput. Sci. (FOCS)*, Oct. 2011, pp. 570–579.
[23] U. Feige and R. Izsak, "Welfare maximization and the supermodular degree," in *Proc. 4th Conf. Innov. Theor. Comput. Sci.*, 2013, pp. 247–256.
[24] M. Feldman and R. Izsak. (2014). "Constrained monotone function maximization and the supermodular degree." [Online]. Available: https://arxiv.org/abs/1407.6328
[25] W. Lu, W. Chen, and L. V. Lakshmanan, "From competition to complementarity: Comparative influence diffusion and maximization," *VLDB Endowment*, vol. 9, no. 2, pp. 60–71, 2015.
[26] W. Chen, T. Lin, Z. Tan, M. Zhao, and X. Zhou, "Robust influence maximization," in *Proc. 9th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 795–804.
[27] K. Murota, *Discrete Convex Analysis*. Philadelphia, PA, USA: SIAM, 2003.
[28] G. Tong, private communication, Dec. 2017.
[29] C. Borgs, M. Brautbar, J. Chayes, and B. Lucier, "Maximizing social influence in nearly optimal time," in *Proc. 25th Annu. ACM-SIAM Symp. Discrete Algorithms*, 2014, pp. 946–957 s.

Wenguo Yang received the M.A. degree in operation research and control theory from Beijing Jiaotong University, Beijing, China, and the Ph.D. degree in operation research and control theory from the Graduate University of the Chinese Academy of Sciences, Beijing, in 2006.

He is currently a Professor with the School of Mathematics Sciences, University of Chinese Academy of Sciences, Beijing, China. He has authored or co-authored more than 50 research papers in reputed refereed academic journals and conference proceedings, such as *Journal of Combinatorial Optimization, Optimization*, *International Journal of Computer Mathematics*, and *Journal of Systems Science and Complexity*. He has presided over two national science foundation programs and several practical applications. His current research interests include social network, robust optimization, nonlinear combinatorial optimization, emergency management, and telecommunication network optimization. He has supervised many M.Sc. and Ph.D. students in these areas.

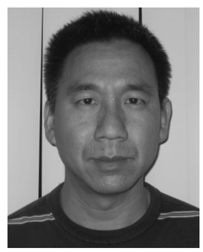
Jing Yuan (S'15) received the B.S. degree in computer science from Nanjing University, Nanjing, China. She is currently pursuing the Ph.D. degree with the University of Texas at Dallas, Richardson, TX, USA.

Her current research interests include span social computing, ecommerce, combinatorial optimization and algorithms.

Ms. Yuan is a Student Member of ACM.

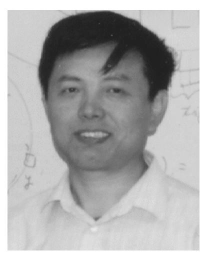
Weili Wu (M'00) received the M.S. and Ph.D. degrees from the Department of Computer Science, University of Minnesota, Minneapolis, MN, USA, in 1998 and 2002, respectively.

She is currently a Full Professor with the Department of Computer Science, University of Texas, Dallas, TX, USA. She is involved in the design and analysis of algorithms for optimization problems that occur in wireless networking environments and various database systems. Her current research interests include data communication and data management.



Jianmin Ma received the Ph.D. degree in mathematics from Colorado State University, Fort Collins, CO, USA.

He is currently a Professor of mathematics with the College of Mathematics and Information Science, Hebei Normal University, Shijiazhuang, China. His current research interests include machine learning, data mining, social networks, and discrete mathematics.



Ding-Zhu Du received the M.S. degree from the Chinese Academy of Sciences, China, in 1982, and the Ph.D. degree from the University of California at Santa Barbara, Santa Barbara, CA, USA, in 1985, under the supervision of Prof. Ronald V. Book.

He was a Professor with the Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, USA. He was with the University of Texas at Dallas, Richardson, TX, USA, Mathematical Sciences Research Institute, Berkeley, CA, USA, and the Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA. He was also with the Department of Computer Science, Princeton University Princeton, NJ, USA. Forty Ph.D. students have graduated under his supervision.

Dr. Du is the Editor-In-Chief of the *Journal of Combinatorial Optimization* and also on the editorial boards for several other journals.