Viral Marketing for Digital Goods in Social Networks

Yu Qiao, Jun Wu, Lei Zhang, and Chongjun Wang *

State Key Laboratory for Novel Software Technology,
Nanjing University, Nanjing 210023, China
yuqiao@smail.nju.edu.cn, wujun@nju.edu.cn, zhangl@nju.edu.cn,
chjwang@nju.edu.cn

Abstract. Influence maximization is a problem of finding a small set of highly influential individuals in social networks to maximize the spread of influence. However, the distinction between the spread of influence and profit is neglected. The problem of profit maximization in social network extends the influence maximization problems to a realistic setting aiming to gain maximum profit in social networks. In this paper, we consider how to sell the digital goods (near zero marginal cost) by viral marketing in social network. The question can be modeled as a profit maximization problem. We show the problem is an unconstrained submodular maximization and adopt two efficient algorithms from two approaches. One is a famous algorithm from theoretical computer science and that can achieve a tight linear time (1/2) approximation. The second is to propose a profit discount heuristic which improves the efficiency. Through our extensive experiments, we demonstrate the efficiency and quality of the algorithms we applied. Based on results of our research, we also provide some advice for practical viral marketing.

Keywords: influence maximization, viral marketing, social network, profit maximization, digital goods

1 Introduction

A combination of recent economic and computational trends, such as negligible cost of duplicating digital goods and, more importantly, the emergence of the online social network as one of the most important arenas for marketing, has created a number of new pricing and marketing problems [3, 2]. In 2001, Domingos and Richardson first introduced the concept of viral marketing to the social network [6]. With the in-depth research and development of social network such as Facebook and Google+, successful online social network (OSN) is creating a media environment for viral marketing. The spread of awareness about a specific product in social network through "word of mouth" is a trend of advertising.

Influence maximization problem is a hot topic for social network analysis which is first defined as an optimization problem by Kempe et al. [1] : A social network is modeled as an undirected graph $G = (V, E)$, with vertices in V

* Corresponding author

modeling the users in the network and weighted edges E reflecting the influence between users, and a positive number k. The goal is to find a seed set A, including k nodes, such that with a given propagation model, the information propagation range of A is maximized. Influence is propagated in the network according to a stochastic cascade model. Three cascade models, namely the Independent Cascade(IC) model , the Weight Cascade(WC) model, and the Linear Threshold(LT) model are extracted from mathematical sociology and considered in [1]. While computer scientists are trying to use mathematic theories and computing devices to understand the diffusion process in online social network (OSN), numerous researchers concern themselves with the efficient approximation algorithms and applications of influence maximization problem [1, 5, 4] (see section 2).

In a real-world scenario, developers and application makers will give away a paid software for free temporarily in *App Store*. Then, if the user has a good experience, they will recommend it to their friends, and by that time the application will likely come back to its regular price, which should help bring in some profit. This is a typical case of viral marketing on the Internet. Digital goods such as softwares and songs are a suitable target for marketing online, since their duplicating and delivery costs are minimal. Digital goods is a popular topic of theoretical computer science [2, 7].

Nowadays most researchers who study viral market are focusing on the spread of influence rather than the original goal of viral marketing: profit. To address the aforementioned limitation, we introduce the concept of free samples and profit achieved by activated users to the influence maximization problem and propose the profit maximization problem for digital goods (PMDG) in social networks. The study is focused on selecting a set of users and giving them free samples to promote the product in order to gain the maximum profit by the diffusion process under the LT model and IC model. The problem is a non-negative, non-monotone and submodular optimization problem without knapsack, which is more complex than the problem of spread of influence. More sophisticated algorithms are needed to solve the problem. Thus, we adopt two efficient algorithms to obtain the maximum profit. One is a famous algorithm from theoretical computer science which can achieve a tight linear time (1/2)-approximation. The other is a highly efficient profit discount heuristic. Through our extensive experiments, we show both algorithms are effective and efficient. Based on results of our research, we also provide some advice for practical viral marketing.

Roadmap. The rest of the paper is organized as follows. Section 2 discusses some related work. Section 3 formalizes the viral marketing problem and discusses properties of our proposed problem. Section 4 presents the efficient algorithms. Data sets and experimental results are discussed in section 5. The last section provides conclusion, possible future work and some practical advice.

2 Related Work

Influence Maximization: Kempe et al. [1] are the first to propose a greedy algorithm for influence maximization. Due to a nice property called submodularity, the greedy algorithm produces near-optimal solution with a theoretical guarantee. However, it suffers from scalability, since the large number of Monte-Carlo simulations are necessary for accuracy. The typical algorithm Cost-Effective Lazy Forward(CELF) utilizes the submodularity to reduce the number of necessary influence estimations with the same performance as the original greedy algorithm [5]. Unfortunately, these improved greedy algorithm is still too inefficient because of the heavy Monte-Carlo simulation. The degree discount heuristics [8] are likely to the scalable solutions to the influence maximization for large scale real-life social networks with their influence spread getting close to that of the greedy algorithm and their extremely fast speed. Recently, Lu et al. [4] proposed a method which scales well to the large scale networks by breaking down the big network to small subgraphs.

Digital Goods : In the domain of marketing for digital goods, there has also been lots of work by theoretical computer scientists who study the pricing strategies for digital goods. For instance, [3] and [2] studied the competitive auction for digital goods aiming to maximize revenue. Hartline et al. [9] studied the marketing strategies for digital goods and proposed the influence-and-exploit(IE) framework. The paper mainly made a study of optimal pricing strategy for digital goods in social networks and the valuation of a user is determined by the buyers who own the product.

Profit Maximization : Lu et al. [10] addressed the difference between influence and profit in their Linear Threshold model with user valuations (LT-V). The LT-V model is focusing on the pricing strategies and the additional adopting process. In contrast, our work is focused on "unbudgeted" seeds selection in social networks. The "unbudgeted" greedy (U-Greedy) algorithm is proposed for seeds selection in the profit maximization, starting with an empty set A, and then add the node v with maximum marginal profit in each round, i.e., $v = argmax_{v \in V \setminus A} R(A \cup \{v\}) - R(A)$, into A until the profit starting decreasing. $R(A)$ is the profit (see Section 3). The U-Greedy algorithm can not give a tight bound and its performance is unstable in some conditions. It also suffers the scalability due to the heavy Monte-Carlo computation in large scale networks.

Submodular Maximization : As we mentioned in section 1, in our setting, the problem for promoting digital goods is an unconstrained submodular maximization, which is different from the influence maximization problem. From a pure algorithm perspective, without a cardinality constraint, maximizing a submodular function is well known to be NP-Hard and an (2/5) approximation can be achieved by local search [11]. Moreover, a linear time algorithm has been proposed for unconstrained submodular maximization, which can achieve an (1/2) approximation [12].

3 Preliminaries

In Section 3, we discuss the proposed model for viral marketing problem and define the objective function. Then we will discuss the properties of the problem. Table 1 gives the important notations used in this paper.

Table 1. Notations

Notations	Descriptions
$G = (V, E)$	An undirected network with nodes set V and edge set E
M	Number of the edges in G
N	Number of the nodes in G
P_v	Expected profit earned from node v
A	Seed set
W_{vu}	Influence weight of node v to u in LT model
Q_{vu}	Probability that node v will influence u in IC model
$R(A)$	Total expected profit achieved by seed set A
$\gamma(A)$	Set of active nodes at the end of propagation

3.1 Problem Statement

We consider a marketer who wants to sell the digital goods to a group of users in communities or social networks. The goal of marketer is to maximize the revenue rather than spread of awareness the products. Since the delivery cost and manufacturing cost of the digital goods are near zero, the supply is unlimited. When the supply of goods is unlimited, the marketer could give lots of free samples to the selected users aiming to promoting the product. In other words, the free samples are given to the seed set, the selected users will help to influence their neighbors to buy the product. We formalize the adopting mechanism in the setting of LT model [1, 16] as an example. The mechanism for IC model can be reached analogously, so we omit here. Given a graph $G = (V, E)$, each edge $(v, u) \in E$ is associated with an influence weight for each $W_{v,u}$. $Nei_a(v)$ means the set of v's active neighbors. The diffusion process begins with an initial set A of active users who are given free samples at round $t = 0$. At each round i, an inactive (who do not own the product) node v will be activated if the sum of influence weight $\sum_{u \in Nei_a(v)} W_{u,v}$ from its active neighbor $Nei_a(v)$ is larger than the influence threshold θ_v. Once the user is activated, he will buy the product and stay active. The propagation terminates when no more nodes can be activated.

Definition 1. *The profit maximization problem for digital goods (PMDG) is defined as: given a graph $G = (V, E)$, the goal is to find out an unconstrained* *set A, $A \subseteq V$, such that profit function $R(A)$ is maximizing. The profit function $R(A)$, is defined as:*

$$R(A) = \sum_{v\in\gamma(A)} P_v - \sum_{v\in A} P_v\,, \tag{1}$$

where $\gamma(A)$ is the set of active nodes at the end of propagation and P_v means the profit earned by single node v. The goal is to maximizing $R(A)$.

The first term in the above definition is the sum of profit achieved by activated user, the second term is the sum of users' profit who are given free samples for marketing.

3.2 Properties

In the Definition 1, $\sum_{v\in\gamma(A)} P_v$ is a typical influence maximization problem of weighted nodes. However, PMDG remains the submodurity which is a natural "diminishing returns" property [19]: the marginal gain from adding an element to a set A is at least as high as the marginal gain form adding the same element to a superset of A. Formally, a submodular function satisfies:

$$f(A \cup \{v\}) - f(A) \geq f(T \cup \{v\}) - f(T)$$

for all elements v and all pairs of sets $A \subseteq T$. However, unlike influence maximization, PMDG is not monotone. In view of the above discussion, PMDG will be a consequence of the follows:

Theorem 1. *The profit function $R(A)$ is a non-negative, non-monotone and submodular function.*

Proof. (Non-negativity): In (1), the first term of $R(A)$ is always larger than the second one, since $A \subseteq \gamma(A)$. Thus, $R(A)$ will be a positive number.

(Submodularity): Observe that $\sum_{v\in\gamma(A)} P_v$ is a influence function and its submodularity has been proven in [1]. $(-\sum_{v\in A} P_v)$ is a linear function and it's a special submodular function. $R(A)$ can be written as $\sum_{i\in\gamma(A)} P_v + (-\sum_{v\in A} P_v)$ and it's a sum of two submodular function. A positive linear combination of submodular functions is still a submodular function [19]. Therefore, $R(A)$ remains submodularity.

(Non-monotonicity): Obviously, $R(\emptyset)$ and $R(V)$ are both zero. Without loss of generality, we assume there exists a node v who can influence at least one of her neighbors. For the set $\{v\}$ which only have one element, $\sum_{v\in\gamma(\{v\})} P_v$ is larger than $\sum_{v\in\{v\}} P_v$ and $R(\{v\})$ is positive. When adding the elements to the set A from null set to full set, $R(A)$ will be from zero to be positive and then from a positive number to zero again. Thus, Non-monotonicity holds.

This completes the proof. □

Next, we show the hardness of PMDG.

Theorem 2. *The problem of profit maximization for digital goods is NP-Hard for both Independent Cascade model and Linear Threshold model.*

Proof. (Independent Cascade model) Firstly, we give the proof in the condition of IC model. Consider an instance of the NP-complete *Set Cover* problem, defined by a collection of subset $\{S_1, S_2...S_k\}$ of a ground set $V = \{v_1, v_2...v_n\}$; the problem is to identify the smallest of sub-collection of S whose union equals the universe. This can be viewed as a special case of PMDG.

Given an arbitrary instance of Set Cover problem, we can model the problem as maximizing $\{|N| - |S|\}$, where $|N|$ is the cardinality of the ground set V and m is the cardinality of feasible solution. In [1], Kempe et al. propose the snapshot simulation based on "flip coins". The propagation of each node is determined in advance and the result can be viewed as a subgraph. So the subgraphs for each node are equivalent to sub-collection S in Set Cover problem. The reduced profit for free sample can viewed as the $|S|$the the cardinality of feasible solution in Set Cover problem. Thus we can reduce NP-Hard Set Cover problem to our problem.

(Linear Threshold model.) In an instance of *Minimum Vertex Cover* problem, defined by an undirected n-node graph $G = (V, E)$, there exists a vertex set V' which is subset of V such that $uv \in E \Rightarrow u \in V' \vee v \in V'$. It means that the set will cover the edges of G. Next we will show that Minimum Vertex Cover problem can be viewed as a special case of our problem.

Given an instance of *Minimum Vertex Cover* problem and a graph G. Minimum Vertex Cover problem also can be modeled as optimization problem, max $\{|M| - |V'|\}$. $|M|$ is the number of edges in the graph. The special case of profit maximization problem can be regarded as using fewest free samples to influence all users. If we want to activated all user, all edges need to be activated. Thus, the problem can be modeled as max $\{|M| - |A|\}$. This matches the optimization objective of minimum vertex cover problem.

Combining the proofs of two settings gives Theorem 2. □

4 Algorithms for PMDG

In this section, we attempt to find good strategies for our problem. What set A of seeds, should we initially give the samples for free so the expected profit is maximizing? It means that we want to find a set A which maximizes $R(A)$. As aforementioned, we have shown that our problem is *NP-Hard* and it's computational intractable, but PTAS is available. Though we do not compute the optimal set A, we compute an A that gives a good approximation. Unlike for influence maximization, the function is non-monotone and "unbudgeted", thus standard greedy hill-climbing strategy [19] is not applicable here. In [10], U-Greedy is proposed for approximation and it's a standard greedy algorithm with terminating condition. It achieves an approximation better than $(1-1/e) \cdot R(A^*) - \Theta(max\{|A_g|, |A^*|\})$, where A_g is set returned by U-Greedy and A^* is optimal solution. As for digital goods setting, since 10%-15% of nodes may be selected as seeds, the performance of U-Greedy is unstable and computational complexity is quite high. Motivated by this, we make use of the recently developed non-monotone submodular maximization technic in [12].

4.1 Approximation Algorithm

There are two algorithms in [12] to maximize the non-monotone submodular function, *Deterministic Algorithm* and *Randomized Algorithm*. The deterministic algorithm is a linear time method that can compute the seed set A and can achieve at least a (1/3)-fraction of optimal profit. Furthermore, the randomized algorithm can achieve an (1/2) approximation and the result can also be achieved in a linear time. Now we will show how to apply those technics to our problem.

Algorithm 1 :Deterministic Algorithm

Input: R, $G=(V,E)$

$i=0$
$X_0 \leftarrow \emptyset$ and $Y_0 \leftarrow V$
sort V in an arbitrary order $v_1, v_2, .., v_N$
while $i < |V|$ **do**
  $a_i \leftarrow R(X_{i-1} \cup \{v_i\}) - R(X_{i-1})$
  $b_i \leftarrow R(Y_{i-1} \backslash \{v_i\}) - R(Y_{i-1})$
  if $a_i \geq b_i$ **then**
    $X_i \leftarrow X_{i-1} \cup \{v_i\}$ and $Y_i \leftarrow \{Y_{i-1}\}$
  else
    $X_i \leftarrow \{X_{i-1}\}$ and $Y_i \leftarrow Y_{i-1} \backslash \{v_i\}$
  end if
end while

Output: X_n or Y_n

The deterministic algorithm is presented in Algorithm 1. Two sets are used in the algorithm. X is initially an empty set and Y equals the node set V. The node set V need to be sorted in an arbitrary order, $v_1, v_2, .., v_N$. The algorithm will proceed in N rounds, where N is the number of nodes in the graph. In each round i, the node v_i will be added to X_{i-1} or removed from Y_{i-1} by comparing the marginal contribution of node v_i for X_{i-1} and Y_{i-1}. At the end of the iterations, X_n equals Y_n and the two set are both solutions for our problem. Obviously, Algorithm 1 runs in a linear time.

The Algorithm 2 implements the randomized algorithm. The initial setting of Algorithm 2 is the same as Algorithm 1. The adding or removing decision of v_i for Algorithm 1 is deterministic, but Algorithm 2 makes a decision with uncertainty,which is also based on the a_i and b_i. However, it provides a guarantee of (1/2) approximation to our problem.

Complexity Analysis. The time complexity of standard greedy algorithm for influence maximization is $O(N^2M)$, where N is the number of nodes in the graph and M is the number of edges. In contrast, the complexities of both deterministic algorithm and randomized algorithm are $O(NM)$. In addition, both methods provide a tight approximation guarantee [12].

Algorithm 2 :Randomized Algorithm

Input: *R*, *G=(V,E)*
i=0
$X_0 \leftarrow \emptyset$ and $Y_0 \leftarrow V$
sort V in an arbitrary order $v_1, v_2, .., v_N$
while $i < |V|$ **do**
  $a_i \leftarrow R(X_{i-1} \cup \{v_i\}) - R(X_{i-1})$
  $b_i \leftarrow R(Y_{i-1} \backslash \{v_i\}) - R(Y_{i-1})$
  $a_i{}' \leftarrow max\{a_i, 0\}$
  $b_i{}' \leftarrow max\{b_i, 0\}$
  with probability $a_i{}'/(a_i{}' + b_i{}')$ **do:**
  $X_i \leftarrow X_{i-1} \cup \{v_i\}$ and $Y_i \leftarrow \{Y_{i-1}\}$
  else with compliment probability do:
  $X_i \leftarrow \{X_{i-1}\}$ and $Y_i \leftarrow Y_{i-1} \backslash \{v_i\}$
end while
Output: X_n or Y_n

4.2 Profit Discount

Inspired by [8], we propose a heuristic named *Profit Discount* for IC model. Although the algorithms discussed in Section 4.1 improve the efficiency dramatically, they're still suffered from heavy Monte-Carlo simulation. Thus, we are motivated to propose a fast heuristic based on degree. In sociology literature, degree and other centrality-based heuristics are commonly used to estimate the influence nodes in social networks [8, 20]. In the setting of our problem, we regard the degree as the expected profit gained by single point. The expected profit achieved by single point v is measured by $\sum_{u \in Nei(v)} Q_{vu} \cdot P_u$, where $Nei(v)$ means neighbors of v. P_u and Q_{vu} mean the expected profit from u and influencing probability in IC model.

Algorithm 3 :Profit Discount

Input: *A*=$\emptyset$,*G=(V,E)*,*R*
$i = 0$
while *true* **do**
  while *each node v in* $V \backslash A$ **do**
    compute *expected profit* $D(v)$ of v
  end while
  $u_i = argmax_{v \in V \backslash A} D(v)$
  if $R(A) - R(A \cup \{u_i\}) \geq 0$ **then**
    $A = A \cup \{u_i\}$
  else
    break
  end if
end while
Output: A

We adopt the thought of "discount" to our problem, which implies that the influence of v's seed neighbor should be taken into consideration when deciding whether to select v as a seed. Firstly, we assume the probability of propagation is small, thus the indirect influence of v will be neglected. Because v is a neighbor of seed u, with probability at least Q_{uv}, v will influenced by u. Thus, at probability Q_{uv}, we do not need to take v into the initial set. Furthermore, selecting the node v will reduce the expected profit of its neighbor in the seed set. The heuristic will select the node with maximum expected profit in each round until the total expected profit begins decreasing. This is the main idea of the profit discount heuristic.

Let $Nei_a(v)$ be the set of v's seed neighbors. The expected profit earned by node v, donated as $D(v)$, in IC model is defined as:

$$D(v) = \prod_{t \in Nei_a(v)} (1 - Q_{tv}) \cdot \sum_{u \in Nei(v) \setminus Nei_a(v)} (P_u \cdot Q_{vu}) - \sum_{t \in Nei_a(v)} (P_v \cdot Q_{tv}) \quad (2)$$

where $\prod_{t \in Nei_a(v)} (1 - Q_{tv})$ means that probability that the node v will not be influenced by any of the already selected seeds. $\sum_{u \in Nei(v) \setminus Nei_a(v)} (P_u \cdot Q_{vu})$ are the profit earned by v from its neighbors which is not seeds. The last term in (2) presents the reduced expected profit when free sample is provided to v. The selecting procedure for seeds will terminate when the total profit begin to decrease.

The profit discount heuristic is formulized in Algorithm 3. The speed of the algorithm is much faster than U-Greedy, deterministic algorithm and randomized algorithm, since mass Monte-Carlo simulations are unnecessary. In section 5, we will show that the performance of profit discount heuristic approaches to the results achieved by U-Greedy.

5 Empirical Evaluations

We conduct the experiments on three real-life networks to evaluate the proposed model and algorithms. All algorithms are implemented in C++ and experiments are conducted on a server with 1.8GHz eight-core Intel E5-2428 CPU, 32GB for memory, and running Windows Server 2010 operating system. CELF algorithm is applied to accelerating the U-Greedy. We run 10,000 simulations and take the average of profit achieved. In order to reduce the uncertainty of propagation and obtain accuracy in IC model, the technic of snapshot is adopted in simulations.

5.1 Datasets

We use collaboration network datasets from real-life for empirical valuation. Three networks are undirected graphs. The number of vertices, edges and the other detail information are summarized in Table 2. Detailed description of datasets is as follows.

- **ca-HepPh.** This data set is from the e-print Arxiv High Energy Physics category. If the author v and u published a paper together, the network contains an edge connecting nodes v to u.

- **ca-CondMat.** The network is from the e-print Arxiv and covers scientific collaborations between authors' papers submitted to Condense Matter category. If v and u co-authored a paper, the edge of v and u will be built.
- **ca-AstroPh.** This data set is a collaboration network of Arxiv Astro Physics category. If v and u are co-author of a paper, the network will contain an edge connecting node v to u.

Table 2. Statistic of Network Data

Name	Node	Edge	Average degree	Maximum degree
ca-HepPh	15233	58831	7.7	314
ca-CondMat	23133	93497	8.1	560
ca-AstroPh	18772	198110	21.1	504

5.2 Parameters Sets

- **Independent Cascade model.** The probability of each edge is selected uniformly at random from {0.01, 0.03, 0.05}.
- **Linear Threshold model.** The threshold for every node is uniformly and randomly chosen from 0 to 1. For trivalency, the edge weight is selected uniformly at random from {0.03, 0.06, 0.09}.
- **Profit distribution.** Following [10], we set the profit subject to normal distribution $N(\mu, \sigma^2)$ with $\mu = 0.53$ and $\sigma = 0.13$.

5.3 Baseline Methods

We compare our applied algorithms with three algorithms in both IC and LT models on three networks.

- **U-Greedy.** This is a modified standard greedy algorithm. In each round, the maximum of marginal contributor will be selected. When the total profit begins decreasing, the algorithm will terminate.
- **Random.** As a baseline comparison, simply select a random user in the graph. Terminate when the profit achieved is less than last round.
- **Degree.** As a comparison, a simple heuristic that selects a node with the maximum degree will terminate, when the total profit begin to decrease

5.4 Experiments Results

We will show the decreasing curve of profit achieved by the U-Greedy and profit discount algorithm when number of seeds grows larger, though both algorithms will terminate when the profit starts decreasing. Deterministic algorithm and randomized algorithm just return the seed set and the selecting procedure is not progressive thus two straights are used to present the results of deterministic algorithm and randomized algorithm.

Discussion on results. The results are shown in Fig. 1 to Fig. 6. From the figures, we can observe that our adopted algorithms perform well in all settings. The profit achieved by the heuristic approaches the performance of U-Greedy. From Table 2, we could know that the connectivity of ca-AtroPh is stronger than the others. Thus, the relative size of solutions for ca-AtroPh is smaller than the two others'. The relative size of the optimal solution set A is inversely proportional to the connectivity.

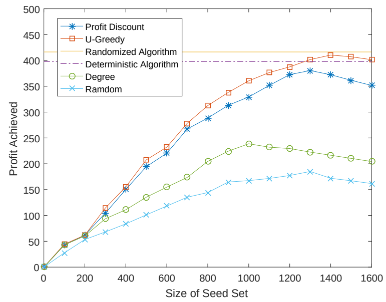

Fig. 1. Profit on ca-HepPh(IC)

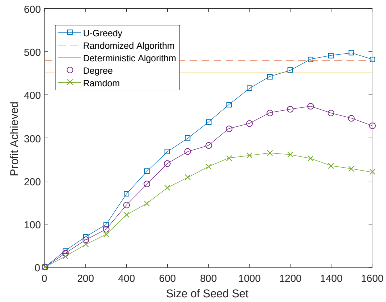

Fig. 2. Profit on ca-HepPh(LT)

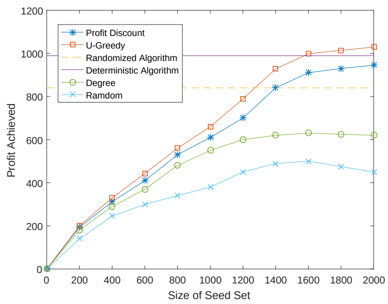

Fig. 3. Profit on ca-CondMat(IC)

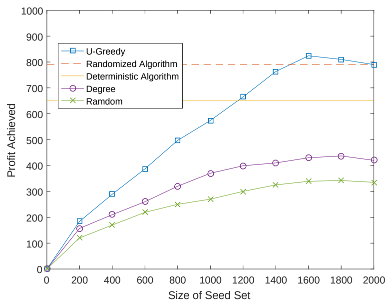

Fig. 4. Profit on ca-CondMat(LT)

Running time. The Table 2 shows the running time of each algorithm in the network datasets. The speed of profit discount heuristic in IC model is impressive. It takes just few minutes and achieves a reasonable results. The two approximation algorithms are also faster than the modified standard greedy algorithm. Both algorithms finish in a linear time.

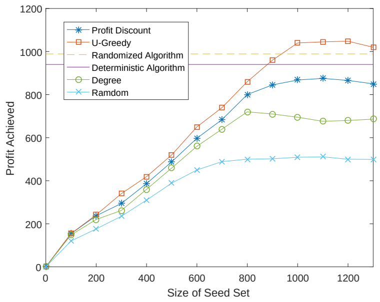

Fig. 5. Profit on ca-AstroPh(IC)

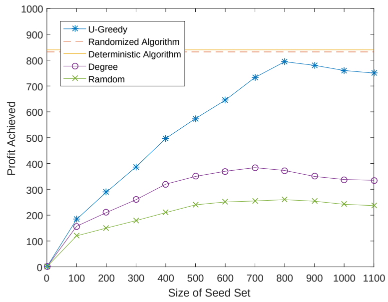

Fig. 6. Profit on ca-AstroPh(LT)

Table 3. Running time in hours

	ca-HepPh		ca-CondMat		ca-AstroPh	
	IC model	LT model	IC model	LT model	IC model	LT model
Profit discount	0.1	-	0.1	-	0.1	-
Deterministic	5.4	6.1	8.4	9.2	7.8	8.7
Randomized	6.1	4.7	7.3	9.0	6.4	7.9
U-Greedy	18.8	22.7	28.2	37.5	26.7	36.7

5.5 Practical Advice on Viral Marketing

According to the empirical valuation, we give some advice on viral marketing for digital goods. First of all, strong connected network is better choice for viral marketing. Once the network is given, an important problem is to find the number of seed users. The number of initial set is key point to achieve high profit. In the real life, the company should investigate the connectivity of the targeted networks and make a “connectivity/number of optimal seeds” growth curve to help make decisions. When the connectivity is relatively high, less users will be selected into seed set. In contrast, more seed users are necessary in the setting of low connectivity.

6 Conclusion

In this paper, we make a study of the profit maximization problem for digital goods in social networks and extend the influence maximization problem to P-MDG in both IC and LT model. We also show the properties of PMDG, such as submodularity and non-monotonicity. In addition, famous approximation algorithms are applied to PMDG. We also propose a highly efficient heuristic. Our experimental results show that our applied algorithms outperform the baseline in both efficiency and effectiveness.

As for future work, *mechanism design* is one popular topic for both fields of theoretical computer science and economics [17, 18]. How to let agents tell the truth is the focus of the research. In [15], the technic of algorithmic game theory is applied to the problem of social network analysis. It is interesting to research whether the algorithmic game theory can be applied to our setting for self-interested users.

Influence maximization in heterogeneous social networks is also an interesting topic in social network analysis [13, 14]. Another future direction is for this research. We plan to introduce the concept of money into heterogeneous social networks and study the viral marketing in multi-platform setting.

7 Acknowledgments

This paper is supported by the National Key Research and Development Program of China (Grant No.2016YFB1001102) and the National Natural Science Foundation of China(Grant No.61375069, 61403156, 61502227), this research is supported by the Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing University.

References

1. Kempe, D. and Kleinberg, J. and Tardos, E.: Maximizing the spread of influence through a social network. International Conference on Knowledge Discovery and Data Mining. pp, 137-146 (2003)
2. Goldberg A V, Hartline J D, Wright A.: Competitive auctions and digital goods Twelfth Acm-siam Symposium on Discrete Algorithms. pp, 735–744 (2001)
3. Chen, N. and Gravin, N. and Lu, P.: Optimal competitive auctions. ACM Symposium on Theory of Computing. pp, 253-262 (2014)
4. Lu, W. X., Zhang, P., Zhou, C., Liu, C., Gao, L.: Influence maximization in big networks: an incremental algorithm for streaming subgraph influence spread estimation. International Joint Conference on Artificial Intelligence. pp, 2076-2082 (2015)
5. Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., Vanbriesen, J. M., Glance, N.: Cost-effective outbreak detection in networks. Knowledge Discovery and Data Mining. pp, 420-429 (2007)
6. Domingos P, Richardson M.: Mining the network value of customers. Knowledge Discovery and Data Mining. pp, 57-66 (2001)
7. Alaei, S., Malekian, A., Srinivasan, A.: On random sampling auctions for digital goods. Electronic Commerce. pp, 187-196 (2009)
8. Chen W., Wang Y., Yang S.: Efficient influence maximization in social networks. Knowledge Discovery and Data Mining. pp, 199-208 (2009)
9. Hartline, J. D., Mirrokni, V., Sundararajan, M.: Optimal marketing strategies over social networks. International World Wide Web Conferences. pp, 189-298 (2008)
10. Lu W., Lakshmanan L V.: Profit Maximization over Social Networks. International Conference on Data Mining. pp, 189-298 (2008)
11. Feige, U., Mirrokni, V., Vondrak, J.: Maximizing non-monotone submodular functions. Foundations of Computer Science. pp, 461-471 (2007)

12. Buchbinder, N., Feldman, M., Naor, J., Schwartz, R. : A tight linear time (1/2)-approximation for unconstrained submodular maximization. Foundations of Computer Science. pp, 649-658 (2012)
13. Wang, Y., Huang, H., Feng, C., Yang, X.: A co-ranking framework to select optimal seed set for influence maximization in heterogeneous network. Web Technologies and Applications. pp, 141-153. (2015)
14. Zhan, Q., Yang, H., Wang, C., Xie, J. A solution to influence maximization problem under cost control International Conference on Tools with Artificial Intelligence. pp, 849-856 (2013)
15. Singer Y. : How to win friends and influence people, truthfully: influence maximization mechanisms for social networks. Web Search and Data Mining. pp, 733-742 (2012)
16. Granovetter M.: Threshold Models of Collective Behavior American Journal of Sociology,. pp, 1420-1443 (2015).
17. Zhang, L., Chen, H., Wu, J., Wang, C. Xie, J.: False-name-proof mechanisms for path auctions in social networks. European Conference on Artificial Intelligence. pp, 323-332 (2016).
18. Nisan, N., Roughgarden, T., Tardos, E., Vazirani, V. V.: Algorithmic game theory. Cambridge University Press, (2007)
19. Nemhauser, G., Wolsey, L. A., Fisher, M. L.: An analysis of approximations for maximizing submodular set functionsI. Mathematical Programming. pp, 265-294 (1978)
20. Wasserman S, Faust K.: Social network analysis :methods and applications: methods and applications. Cambridge University Press, (1994)