2013 IEEE 13th International Conference on Data Mining

Influence and Profit: Two Sides of the Coin

Yuqing Zhu*, Zaixin Lu†, Yuanjun Bi*, Weili Wu*, Yiwei Jiang‡, and Deying Li§¶
*Department of Computer Science, University of Texas at Dallas. Email: {yuqing.zhu, yuanjun.bi, weiliwu}@utdallas.edu
†NSF Center for Research on Complex Networks, Texas Southern University. Email: luz@tsu.edu
‡Department of Mathematics, Zhejiang Sci-Tech University. Email: ywjiang@zstu.edu.cn
§School of Information, Renmin University of China. Email: deyingli@ruc.edu.cn
¶Corresponding Author: Deying Li.

***Abstract*—Influence maximization problem is to find a set of seeds in social networks such that the cascade influence is maximized. Traditional models assume all nodes are willing to spread the influence once they are influenced, and they ignore the disparity between influence and profit of a product. In this paper by considering the role that price plays in viral marketing, we propose price related (PR) frame that contains PR-I and PR-L models for classic IC and LT models respectively, which is a pioneer work. We find that influence and profit are like two sides of the coin, high price hinders the influence propagation and to enlarge the influence some sacrifice on profit is inevitable. We propose Balanced Influence and Profit (BIP) maximization problem. We prove the NP-hardness of BIP maximization under PR-I and PR-L model. Unlike influence maximization, the BIP objective function is not monotone. Despite the non-monotony, we show BIP objective function is submodular under certain conditions. Two unbudgeted greedy algorithms separately are devised. We conduct simulations on real-world datasets and evaluate the superiority of our algorithms over existing ones.**

***Index Terms*—Influence maximization, profit maximization, social networks, IC model, LT model.**

I. Introduction

In social networks, viral marketing [1] [2] is an act of prompting products by other customers spreading to their friends and colleagues, has become one of the most effective marketing strategy. In viral marketing, influence maximization problem is to choose a group of infusive people within one social network to conduct the largest cascade of positive influence, and it has become one of the fundamental problems. In [3], Kempe et al. studied this issue as a problem in discrete optimization. The authors proposed two popular models: *Independent Cascade* (IC) model and *Linear Threshold* (LT) model, which are the basis of many successive variant models.

It goes without saying that price plays a significant role in people's buying behaviors. When a product hits the market, whether a person purchases it should depend on the person's evaluation and the product's price. Price affects influence propagation in reality. To understand this, think one question, when one of your friends shares you a product's information, will you spread it immediately to others? No. You may think it is attractive, check the product's price and consider to buy it. After you have the product and the experience is good indeed, you then propagate its information to others. Thus, in this paper, we consider the price's effect in influence propagation.

Influence is the "fame", yet it is not enough because the ultimate goal of a product is making profit. However, influence and profit are like two sides of the coin, both of which can not be maximized at the same time, and people have to make choices between them. For instances, in price war, the company reduces the price to attract customers. The price may be lower than the manufacturing cost to cater the buyers, so that more individuals will buy the product and influence more people. In this occasion, the company values the influence better than the profit. But in luxury goods market where the profit is gigantic, the situation reverses. Very few people own the goods and spread the influence, thus the propagated influence is small. The impossibility of maximizing influence and profit at the same time is frustrating, nevertheless, we can employ weights to indicate the preference of a company upon influence and profit in decision making. The manager should adopt the strategy that maximizes the weighted combination of influence and profit.

On the whole, in this paper, taking note of the impact of price on influence and profit in social networks, we propose a propagation rule that when one individual is influenced, he or she will spread the influence to the friends only when the price is lower than his or her evaluation. Under this rule, for an individual, the enhancement of the influence leads to the decreasing of the profit. By assigning weights to influence and profit we obtain their linear combination, which is the optimization objective function. Our ultimate goal is to select the seeds and assign the prices to maximize the objective function.

Our contributions are as follows: First, to better approach the reality, we adopt price into influence spread models, and propose Price Related propagation (PR) frame. PR frame contains PR-I and PR-L models which extend classic IC and LT models respectively. We are the first one introducing price into IC model. Distinct from existing works, an influenced node does not always propagate influence. Molding by price, our PR frame considers every individual's willingness in influence spread, which makes it more realistic.

Second, in PR frame high price hinders the nodes from turning active and propagating influence. If people seek to increase the influence of a product they must give up some profit, and vice versa. Motivating by this observation, we consider the preference and propose Balanced Influence and Profit (BIP) maximization problem. We prove BIP is NP-hard. BIP objective function is not monotone, but it is submodular

1550-4786/13 $31.00 © 2013 IEEE
DOI 10.1109/ICDM.2013.40

given certain conditions. Unbudgeted greedy algorithms are designed for both PR-I or PR-L model.

Last but not the least, we conduct extensive simulations on real-world datasets. We show that the preference is valid on controlling the influence and profit. Moreover, an algorithm exists for one special case in BIP maximization, and we confirm that our algorithm outperforms the existing algorithm.

II. Related Work

Influence maximization has been studied extensively. That the influence propagates from customers to customers recursively was first noted by Domingos and Richardson, who proposed the problem of mining the influence within a social network instead of a set of independent entities for the first time. In [1] Domingos and Richardson modeled the influence as a Markov random field and devised algorithms to mine social network models from databases. Kempe, Kleiberg et al. proposed influence maximization as finding a set of k seeds that maximizes the influence spread firstly in [3]. They proposed two basic diffusion models: *Independent Cascade* (IC) model and *Linear Threshold* (LT) model and proved this problem is $\mathbf{NP}$-hard. Computing exact influence under both IC model and LT model are #P-hard, and they have been proved by Chen in [4] and [5] respectively.

Some works on profit(revenue) maximization also emerged. Kleinberg and Leighton [6] considered maximizing the seller's revenue in online posted-price auction mechanism. They defined the *additive* regret of a pricing strategy as the difference between the strategy's expected revenue and the revenue derived from the optimal fixed-price strategy, and considered buyer's evaluations are: 1) equal to one same number p, 2) sampled from one probability distribution, 3) chosen by an oblivious adversary. Both upper and lower bounds on regret was given for each case. In [7], Hartline et al. identified the *influence-and-exploit strategies* which *influences* the population by giving free samples to chosen buyers initially, and employs greedy pricing strategy to extract revenue from the remaining buyers. Arthur et al. [8] proposed a model in which whether a buyer buys a product is influenced by friends who own the product and the product's price. The authors proved maximizing the expected revenue under their model is $\mathbf{NP}$-hard, and proposed an *influence-and-exploit* based strategy that generates revenue guaranteed to be within a constant factor of the optimal strategy.

In [9], Lu and Lakshmanan argued the difference between adoption and influence in social networks. By adding the product's price into LT model their model named LT-V was proposed. The authors proved the profit maximization problem under LT-V model is submodular but non-monotone, and presented an unbudgeted greedy framework for the problem.

III. Propagation Model and Marketing Strategy

We propose new models that better suit the commercial environment. Our models extend both classic IC and LT models. A new easy-to-implemented marketing strategy along with a powerful but costly-to-implemented one are also designed.

A. IC and LT Models With Price Related Propagation

Our new model considers the monetary factor in propagation, thus we name it *Price Related propagation* frame and abbreviate it as PR frame. PR frame contains PR-I model based on the IC model, and PR-L model based on the LT model. The difference between PR frame and the classic IC and LT models is that every node u in PR-I model has three stages: *neutral*, *influenced*, and *active*. Node u being neutral means u has no idea or positive attitude of the product. When u becomes influenced, u starts to hold positive attitude to the product. However, only active nodes will purchase the product, and propagate the influence by telling the network neighbors. In PR frame, holding positive attitude and propagating attitude no longer equal, which makes our model more natural. The individuals in social networks are independent human beings, they are influenced by the people around, but they also have their judgments. When it comes to information propagation, surely people will evaluate the information before they spread it.

Each node u in PR frame has a *evaluation* for the product, which is the highest price this node thinks the product worths. Only when the offered price is lower than u's evaluation, u will deem this product worthy to adopt, turn *active* and propagate the influence by informing others about the product.

LT-V model in [9] only modifies the LT model, while our PR frame adds price's effect into both the IC and LT models.

B. Marketing Strategy

Considering the implementing difficulty, we design two marketing strategies. One is *BinarY priCing* (BYC), and the other is *PAnoramic Pricing* (PAP).

Definition 1 (BYC). *All chosen seed individuals are given samples for free, and all non-seed nodes are charged the same non-zero price.*

Definition 2 (PAP). *Prices for Individuals are unconstrained different values, the seeds are offered discounts if needed.*

IV. The Choice between Influence and Profit: Two Sides of The Coin

For an instance of PR frame where $|V| = n$, the prices of all nodes form a vector $\vec{\boldsymbol{p}} = (p_1, ..., p_i, ...p_n) \in [0,1]^n$, and denote the profit and influence by $\mathcal{R}$ and $\mathcal{I}$, respectively. Let S be the set of seeds, α and β respectively be the number of active nodes and number of influenced nodes. All $\mathcal{R}$, $\mathcal{I}$, α and β are related to S and $\vec{\boldsymbol{p}}$. Denote the constant manufacturing cost as $\hat{c} \in [0,1]$, we have

$$\begin{aligned} \mathcal{I}(S,\vec{\boldsymbol{p}}) &= \alpha(S,\vec{\boldsymbol{p}}) + \beta(S,\vec{\boldsymbol{p}}) \ , \\ \mathcal{R}(S,\vec{\boldsymbol{p}}) &= \sum_{v_i \in V \text{ is active}} (p_i - \hat{c}) \ . \end{aligned} \tag{1}$$

Using a parameter $\hat{\lambda} \in [0,1)$ to denote this preference, the following problem is defined:

$$\max_{(S,\vec{\boldsymbol{p}})} : \ \mathcal{B}(S,\vec{\boldsymbol{p}}) = \hat{\lambda} \cdot \mathcal{I}(S,\vec{\boldsymbol{p}}) + (1-\hat{\lambda}) \cdot \mathcal{R}(S,\vec{\boldsymbol{p}}) \tag{2}$$

We call $\mathcal{B}(S, \vec{\boldsymbol{p}})$ the *Balanced Influence and Profit* (BIP), and optimization problem (2) is named BIP maximization.

A. UniActive: A Special Scenario of PR Frame

Generally speaking, BIP maximization problem is hard to optimize. To understand its complexity, we check a special scenario named *UniActive*, where every individual's evaluation is 1 and the strategy is BYC. For each element p_i in $\vec{\boldsymbol{p}}$, if v_i is a seed, $p_i = 0$, otherwise $p_i = \hat{p}$ where $\hat{p} \in [\hat{c}, 1]$ is a constant. Under UniActive all influenced nodes will automatically switch to active since $\hat{p} \leq 1$, i.e. the evaluation. Hence, influenced and active are equivalent, $\beta(S, \vec{\boldsymbol{p}}) = 0$, $\mathcal{I}(S, \vec{\boldsymbol{p}}) = \mathcal{I}(S) = \alpha(S)$. Further, $\mathcal{R}(S, \vec{\boldsymbol{p}}) = \mathcal{R}(S) = (\hat{p} - \hat{c})(\mathcal{I}(S) - |S|) - \hat{c} \cdot |S|$. Combine it with (2) and define two new constants:

$$\begin{cases} \hat{x} \coloneqq (1-\hat{\lambda})(\hat{p}-\hat{c}) + \hat{\lambda} \\ \hat{y} \coloneqq (1-\hat{\lambda})\hat{p} \end{cases} \tag{3}$$

($\hat{x}, \hat{y} \geq 0$), the BIP maximization under UniActive scenario is:

$$\max_S \mathcal{B}(S) = \hat{x} \cdot \mathcal{I}(S) - \hat{y} \cdot |S| \tag{4}$$

Theorem 1. *$\mathcal{B}(S)$ in (4) is non-monotone when $\hat{y} \neq 0$.*

Combine all the possible cases of different $\hat{x}$ and $\hat{y}$ together, we denote the problem family under PR-I model as $\Omega_{\mathcal{B}}^{I}$, and that under PR-L model as $\Omega_{\mathcal{B}}^{L}$. Because whether $\hat{x} = 0$ or $\hat{y} = 0$ optimization problem (4) turns trivial, in theorems 2 and 3, $\hat{x}, \hat{y} > 0$.

Theorem 2. *Under PR-I model, when $\hat{x} < \hat{y}$, BIP maximization (4) is $\mathbf{NP}$-hard.*

Theorem 3. *Under PR-L model, for arbitrary $\hat{x}$ and $\hat{y}$, BIP maximization (4) is $\mathbf{NP}$-hard.*

For PR-I model, Theorem 2 shows $\Omega_{\mathcal{B}}^{I}$ is $\mathbf{NP}$-hard as a problem family. A direct result from Theorem 2 is that merely maximizing profit ($\hat{\lambda} = 0$) under PR-I model is still $\mathbf{NP}$-hard. For PR-L model, [9] showed $\Omega_{\mathcal{B}}^{L}$ is $\mathbf{NP}$-hard, while Theorem 3 gives a more general proof, and shows a much stronger conclusion that all non-trivial subcases in $\Omega_{\mathcal{B}}^{L}$ are $\mathbf{NP}$-hard.

B. Submodularity in General PR Frame

Deducing from the previous theorems, BIP maximization in general PR frame is $\mathbf{NP}$-hard, and general BIP objective function is non-monotone.

Theorem 4. *Given a PR-I instance and a fix $\vec{\boldsymbol{p}}$ that $p_i \geq \hat{c}$, $\mathcal{B}(S, \vec{\boldsymbol{p}})$ is submodular w.r.t. S.*

Theorem 5. *Given a PR-L instance and a fix $\vec{\boldsymbol{p}}$ that $p_i \geq \hat{c}$, $\mathcal{B}(S, \vec{\boldsymbol{p}})$ is submodular w.r.t. S.*

V. Algorithms

To give a pricing method for BIP problem, we take both the manufacturing cost and local influence into consideration. Let $d^{out}(v_i)$ denotes the outdegree of v_i and

$$s(v_i, u) \coloneqq \begin{cases} w(v_i, u) & \text{for PR-I model,} \\ b(v_i, u) & \text{for PR-L model.} \end{cases}$$

The *optimal price* p_i' for v_i is calculated as follows:

$$\begin{aligned} \widetilde{\mathcal{B}}_i(p) \quad &\coloneqq \quad \hat{\lambda}(1 - F_i(p)) \sum \tfrac{s(v_i,u)}{d^{out}(v_i)} \\ &\quad\ + (1-\hat{\lambda})(1-F_i(p))(p-\hat{c}), \qquad (5) \\ p_i' \quad &= \quad \operatorname*{arg\,min}_{p \in [0,1]} \widetilde{\mathcal{B}}_i(p). \qquad (6) \end{aligned}$$

(5) is so far the only formula that considers the manufacturing cost and the network structure. It has the same computation expense as the existing optimal price. Nevertheless, the price calculated by (6) is still myopic.

A. Distributions of Nodes' Evaluations

Suppose user v_i's evaluation is a random variable X_i, we study two distributions:

Normal Distribution:
$X_i \sim \mathrm{N}(\mu, \sigma^2)$, the CDF $F_i(x_i) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{t} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \, dt$.
Uniform Distribution:
$X_i \sim \mathrm{U}(0, 1)$, the CDF $F_i(x_i) = x_i$.

Note that given the price p_i offered to v_i, the probability v_i turning active is $\mathrm{Prob}(x_i \geq p_i) = 1 - F_i(p_i)$.

B. Determine the Seeds and Prices under BYC

BYC is the strategy that takes the least implementation expense for a company, and here we design ABYC: an algorithm for BYC. ABYC contains two stages: First, to offer every individual a same price. Second, to choose the seeds whom free samples are given, which means, to choose the seeds and change their prices to 0.

(6) is not used in the first stage since the obtained p_i' may vary from v_i. Instead we calculate the *universal optimal price* as follows: $p_{_U}' = \operatorname*{arg\,min}_{p \in [0,1]} \sum_{i}^{n} \widetilde{\mathcal{B}}_i(p)$.

In the second stage, we adopt greedy policy. Every round for each non-seed node u we compute the marginal BIP gain of picking u as a seed, and choose the node that provides the highest marginal gain. When no non-seed node brings positive marginal BIP gain, ABYC stops.

For a vector $\vec{\boldsymbol{p}} = (p_1, ...p_n)$, let $(\vec{\boldsymbol{p}}_{-i}, q)$ be the vector obtained by altering p_i, the i-th element of $\vec{\boldsymbol{p}}$ to q, i.e. $(\vec{\boldsymbol{p}}_{-i}, q) = (p_1, ..., p_{i-1}, q, p_{i+1}, ..., p_n)$.

Algorithm 1 ABYC: the Algorithm for BYC

```
S ← ∅, p ← 0 ;
for ∀v_i ∈ V do
    p_i ← p'_U ;
end for
while true do
    u ← argmax_{v_i ∈ V\S} {B(S ∪ {v_i}, (p_{-i}, 0)) − B(S, p)};
    if B(S ∪ {u}, (p_{-i}, 0)) − B(S, p) > 0 then
        S ← S ∪ {u};  p ← (p_{-i}, 0);
    else  break;
    end if
end while
output (S, p);
```

C. Determine the Seeds and Prices under PAP

Suppose v_i's new price is q, the marginal BIP gain of adding set v_i is: $\mathcal{B}(S \cup \{v_i\}, (\vec{\boldsymbol{p}}_{-i}, q)) - \mathcal{B}(S, \vec{\boldsymbol{p}})$. Since $\mathcal{B}(S, \vec{\boldsymbol{p}})$ is a constant w.r.t. q, we only have to maximize $\mathcal{B}(S \cup \{v_i\}, (\vec{\boldsymbol{p}}_{-i}, q))$. When v_i is offered price q, there are only two outcomes in the sample space, outcome ω_1 that v_i accepts the price and turns active, outcome ω_0 that v_i rejects the price, stays influenced and never spreads the influence. For ω_1, the influence collected from v_i is 1 and the profit collected from v_i is $q - \hat{c}$, suppose the influence from other nodes is I_1 and the profit from other nodes is R_1, the BIP is $g_i(q) = \hat{\lambda}(I_1 + 1) + (1 - \hat{\lambda})(q - \hat{c} + R_1)$, which is a linear function w.r.t. q. For ω_0, the influence gain collected from v_i is 1 and the profit gain collected from v_i is 0, suppose the influence from other nodes is I_0 and the profit from other nodes is R_0, the BIP is $h_i = \hat{\lambda}(I_0+1)+(1-\hat{\lambda})R_0$, a constant independent of q. Prob$(\omega_1) = 1-F_i(q)$ and Prob$(\omega_0) = F_i(q)$. Hence the expected BIP is:

$$\delta_i(q) = g_i(q) \cdot (1 - F_i(q)) + h_i \cdot F_i(q) \tag{7}$$

Algorithm 2 APAP: the Algorithm for PAP

$S \leftarrow \emptyset$, $\vec{\boldsymbol{p}} \leftarrow \vec{0}$;
for $\forall v_i \in V$ **do**
  $p_i \leftarrow p_i' = \arg\min_{p \in [0,1]} \widetilde{\mathcal{B}}_i(p)$;
end for
while true **do**
  for $\forall v_i \in V \setminus S$ **do**
    $p_i^* \leftarrow \arg\max_{q \in [0,1]} \delta_i(q)$;
  end for
  $u \leftarrow \arg\max_{v_i \in V \setminus S} \{\mathcal{B}(S \cup \{v_i\}, (\vec{\boldsymbol{p}}_{-i}, p_i^*)) - \mathcal{B}(S, \vec{\boldsymbol{p}})\}$;
  if $\mathcal{B}(S \cup \{u\}, (\vec{\boldsymbol{p}}_{-i}, p_i^*)) - \mathcal{B}(S, \vec{\boldsymbol{p}}) > 0$ **then**
    $S \leftarrow S \cup \{u\}$; $\vec{\boldsymbol{p}} \leftarrow (\vec{\boldsymbol{p}}_{-i}, p_i^*)$;
  else $break$;
  end if
end while
output $(S, \vec{\boldsymbol{p}})$;

VI. Experiments

As same as all the existing greedy algorithms for influence or profit maximization on estimating the marginal gain, we use Monte Carlo simulation with 10000 generated samples. This simulation is costly and in order to finish our experiments in a reasonable time, the sizes of our operated datasets are restricted.

Table I. Dataset Statistics

	Enron	NetHept	Epinion
#Nodes	4.7K	5.2K	5.1K
#Edges	47.6K	30.7K	113.4K
Avg. outdegree	10.11	5.89	22.34
Max outdegree	1023	64	968
Max component size	2671	3165	3324

A. Simulation Setup

Table I presents the statistics of our employed three datasets: Enron[11], NetHept[12], and Epinions[13]. The raw data of NetHept, Enron and Epinion is bigger, and we use the partition software METIS[14] to collect our datasets from the original graphs.

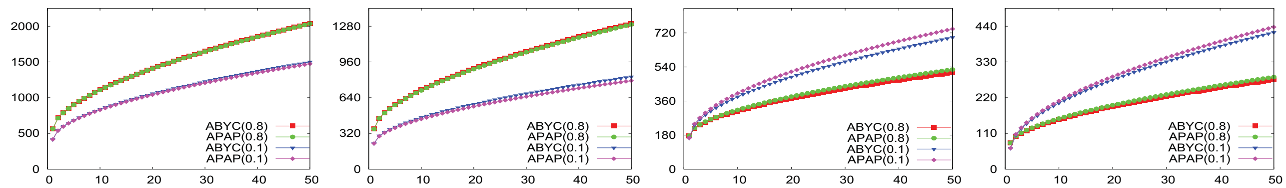
In PR-I model, we use *Weighted Cascade* method to assign edge weights. For an edge from node u to v, the probability $w(u, v)$ that u influences v is $\beta/d^{in}(v)$, where $\beta \in (0, 1]$ is an adjustable parameter and we set $\beta = 0.4$. In PR-L model, we use *Trivalency* method. The weight $b(u, v)$ of every edge is randomly chosen from $\{0.001, 0.01, 0.1\}$, normalization on the weights may be used to guarantee that for each node v, the sum of its in-edges' weights is no greater than 1. In fact neither PR-I nor PR-L model has preference on either weight assignment method, due to space limitation we only present two combinations.

In our experiments we set the Normal distribution of user evaluation is related to the cost, i.e. most customers have evaluations higher than the manufacturing cost $\hat{c}$. *3-sigma rule* tells us if $X\sim$N(μ, σ^2), then Prob$\{\mu - 3\sigma < X < \mu + 3\sigma\} = 99.74\%$. Hence we let $\mu - 3\sigma = \hat{c}$ and $\mu + 3\sigma = 1$, it follows that $\mu = (1 + \hat{c})/2$ and $\sigma = (1 - \hat{c})/6$.

B. Simulation Results

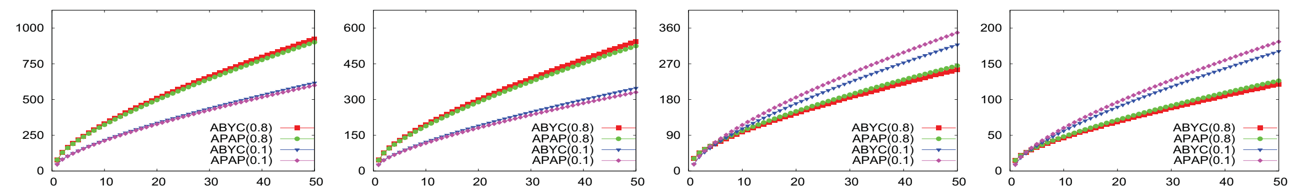
In all test cases we increasingly pick seeds until 50 seeds are chosen. We use $X\sim\mathbf{N}$ to denote that the individuals' evaluations are normal distributed, while $X\sim\mathbf{U}$ means the evaluations are uniform distributed. Through Fig. 1 to Fig. 6, the abscissa is the size of seed set, and the ordinate is influence or profit according to the subfigure's caption.

- ***Preference*** $\hat{\lambda}$: We set $\hat{\lambda}$ at two values, 0.8 and 0.1. APAP(0.8) and APAP(0.1) represent APAP with $\hat{\lambda} = 0.8$ and APAP with $\hat{\lambda} = 0.1$, respectively. Fig. 1(a) and Fig. 1(c) plot the influence and profit in a same graph respectively. From Fig. 1(a), we see the influence achieved by APAP(0.8) is about 45% higher than that achieved by APAP(0.1), while from Fig. 1(c), we see the profit achieved by APAP(0.8) is about 55% lower than that achieved by APAP(0.1). Note that higher $\hat{\lambda}$ represents higher preference on influence, we have testified the effectiveness of $\hat{\lambda}$ in APAP. Its effectiveness in ABYC can be easily evaluated too. Similar phenomena can be found from Fig. 2 to Fig. 6.
- ***ABYC and APAP***: Note that both APAP and ABYC are designed to maximize BIP, the weighted combination of influence and profit, but we show influence and profit separately to better explain the algorithms' performances. Fig. 1 to Fig. 3 plot the results of weighted cascade PR-I graph on Enron, NetHept, and Epinion respectively. From these figures, the expected profit of APAP are consistently higher than that of ABYC. The gaps between APAP and ABYC are tiny yet, since in weighted cascade graph, the nodes have higher probabilities to impact others. Hence offering price 0 to seeds can bring nice potential profit. The profit achieved when $\hat{\lambda} = 0.1$ should be higher than the profit achieved when $\hat{\lambda} = 0.8$, however when



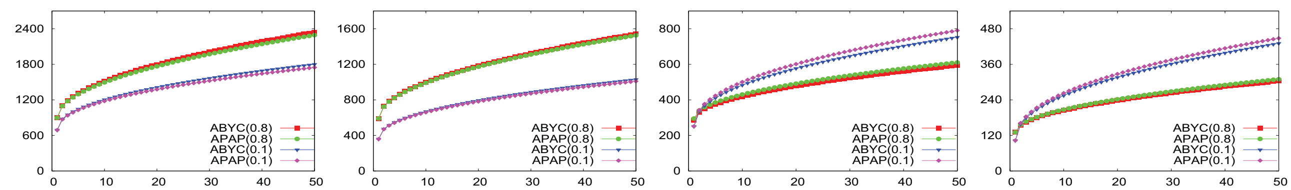
(a) Influence with $X \sim \mathbf{N}$ (b) Influence with $X \sim \mathbf{U}$ (c) Profit with $X \sim \mathbf{N}$ (d) Profit with $X \sim \mathbf{U}$

Figure 1: Influence and profit of weighted cascade PR-I on Enron



(a) Influence with $X \sim \mathbf{N}$ (b) Influence with $X \sim \mathbf{U}$ (c) Profit with $X \sim \mathbf{N}$ (d) Profit with $X \sim \mathbf{U}$

Figure 2: Influence and profit of weighted cascade PR-I on NetHept



(a) Influence with $X \sim \mathbf{N}$ (b) Influence with $X \sim \mathbf{U}$ (c) Profit with $X \sim \mathbf{N}$ (d) Profit with $X \sim \mathbf{U}$

Figure 3: Influence and profit of weighted cascade PR-I on Epinion

the number of seeds are very small we do not see that happens since the scale is too tiny to represent the general phenomenon.

From Fig. 4 to Fig. 6 which show the results of the trivalency PR-L graphs, APAP still achieves higher profit than ABYC, and the gaps of profit grow bigger. For instance, we see APAP achieves 15% higher than ABYC on profit when $\hat{\lambda} = 0.1$. The greater gaps come from the lower impact between nodes in trivalency method, and offering a seed price 0 cannot bring enough future profit to compensate the profit lost on this seed.

Through the results we see that APAP achieves lower influence than ABYC. The reason is that ABYC that blindly offers price 0 to every seed maximizes the influence solely gained from that seed, APAP may also offer price 0 to a seed if necessary but it also has to consider the profit. Though ABYC works better than APAP on maximizing the influence, the gaps are always tiny. Note that it is the BIP, i.e. the balanced influence and profit which APAP and ABYC both try to maximize, and APAP works indubitably better than ABYC on BIP Maximization.

Table II. Profit comparison of APAP and PAGE

Alg.	Enron		NetHept		Epinion	
	$\hat{c}: 0.2$	$\hat{c}: 0.3$	$\hat{c}: 0.2$	$\hat{c}: 0.3$	$\hat{c}: 0.2$	$\hat{c}: 0.3$
APAP	148.6	125.3	78.5	68.9	166.8	147.3
PAGE	125.8	89.0	68.3	50.3	143.5	103.1

• ***The performance of APAP on a special case***: PAGE is an existing algorithm for profit maximization in PR-L model, i.e. the special case where $\hat{\lambda} = 0$ and the propagation model is PR-L. We compare the achieved profit of APAP and PAGE in trivalency PR-L graph under the condition $X \sim \mathbf{N}$. Table II plots the profits APAP and PAGE gain by 50 seeds. APAP makes a better consideration of the manufacturing cost $\hat{c}$, thus it works better than PAGE. The gap grows as $\hat{c}$ increases.

• ***Price Variation***: When users' evaluations are Normal distributed, we plot the prices for seeds obtained by APAP on weighted cascade PR-I Epinion graph in Fig. 7. When $\hat{\lambda} = 0.8$, the price barely grows, for in this case influence is valued more and high price will hinder influence propagation. When $\hat{\lambda} = 0.1$, the price increases faster, since the profit is valued more.

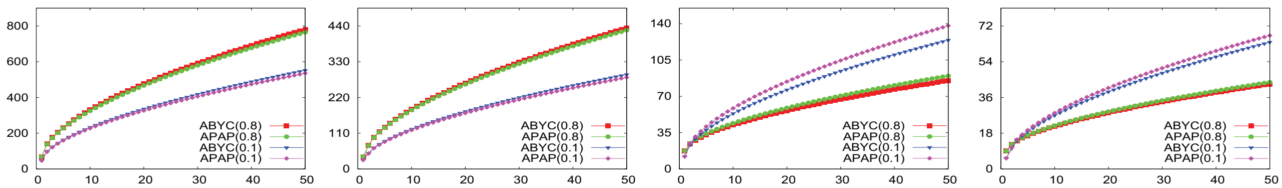
• ***Running Time***:

Table III. Running time on weighted cascade PR-I graph (hour)

Alg.	Enron		NetHept		Epinion	
	N	U	N	U	N	U
ABYC	1.2	0.3	0.3	0.2	2.5	0.7
APAP	1.6	0.4	0.5	0.3	3.1	1.0

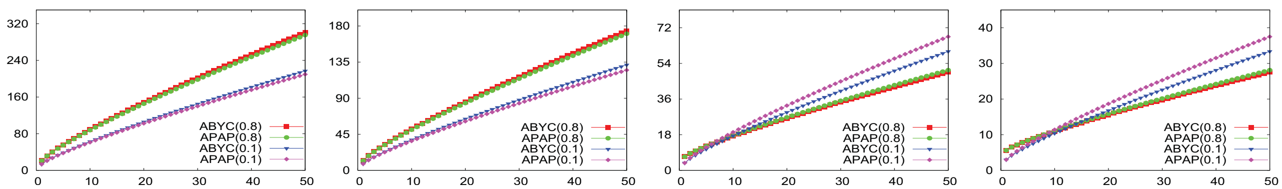
VII. Conclusion

In this paper we distinguished influence and profit in social networks, and proposed the PR frame, in which PR-I and PR-L models extend classic IC and LT models respectively. Observed that maximum influence and profit do not exist at the same time, we proposed BIP Maximization to cater the decision maker's different preference towards influence and profit. We proved that BIP Maximization is $\mathbf{NP}$-hard, and the objective function is submodular under certain conditions,



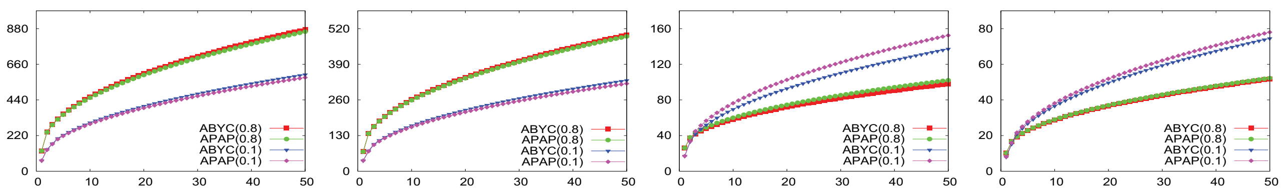
(a) Influence with $X{\sim}\mathbf{N}$ (b) Influence with $X{\sim}\mathbf{U}$ (c) Profit with $X{\sim}\mathbf{N}$ (d) Profit with $X{\sim}\mathbf{U}$

Figure 4: Influence and profit of trivalency PR-L on Enron



(a) Influence with $X{\sim}\mathbf{N}$ (b) Influence with $X{\sim}\mathbf{U}$ (c) Profit with $X{\sim}\mathbf{N}$ (d) Profit with $X{\sim}\mathbf{U}$

Figure 5: Influence and profit of trivalency PR-L on NetHept



(a) Influence with $X{\sim}\mathbf{N}$ (b) Influence with $X{\sim}\mathbf{U}$ (c) Profit with $X{\sim}\mathbf{N}$ (d) Profit with $X{\sim}\mathbf{U}$

Figure 6: Influence and profit of trivalency PR-L on Epinion

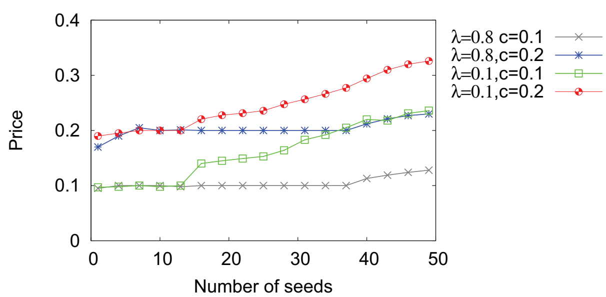


Figure 7: Seed prices of weighted cascade PR-I on Epinion

although it is not monotone. Two algorithms ABYC and APAP have been devised. Through the simulation, the parameter we set to control the preference on influence and profit was confirmed to be effective. We saw APAP has the best performance on maximizing the BIP objective function. For a special case in PR-L model, an algorithm exists and APAP also outperforms it.

Acknowledgement

This research was supported in part by National Science Foundation of USA under grants CNS-1016320, CCF-0829993, NSF 1137732 and NSF 1241626, and National Natural Science Foundation of China under grants 61070191, 91124001, 11001242, and 11071220.

References

[1] P. Domingos and M. Richardson. Mining the network value of customers. In *Proceedings of the 7th International Conference on Knowledge Discovery and Data Mining*, KDD '01, pages 57–66. ACM Press, 2001.

[2] M. Richardson and P. Domingos. Mining knowledge-sharing sites for viral marketing. In *Proceedings of the 8th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '02, pages 61–70. ACM, 2002.

[3] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '03, pages 137–146. ACM, 2003.

[4] W. Chen, C. Wang, and Y. Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '10, pages 1029–1038, New York, NY, USA, 2010. ACM.

[5] W. Chen, Y. Yuan, and L. Zhang. Scalable influence maximization in social networks under the linear threshold model. In *Proceedings of the 10th IEEE International Conference on Data Mining*, ICDM '10, 2010.

[6] R. Kleinberg and T. Leighton. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '03, pages 594–628. IEEE Computer Society, 2003.

[7] J. Hartline, V. Mirrokni, and M. Sundararajan. Optimal marketing strategies over social networks. In *Proceedings of the 17th international conference on World Wide Web*, WWW '08, pages 189–198. ACM, 2008.

[8] D. Arthur, R. Motwani, A. Sharma, and Y. Xu. Pricing strategies for viral marketing on social networks. *CoRR*, abs/0902.3485, 2009.

[9] W. Lu and L. V. S. Lakshmanan. Profit maximization over social networks. In *Proceeding of the 12th IEEE International Conference on Data Mining*, ICDM '12, 2012.

[10] L. Kiefer. Sequential minimax search for a maximum. *Proceedings of the American Mathematical Society*, pages 502–206, Apr 1953.

[11] J. Leskovec, A. Dasgupta K. Lang, and M. Mahoney. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics*, pages 29–123, 6(1) 2009.

[12] http://research.microsoft.com/en-us/people/weic/graphdata.zip.

[13] P. Massa and P. Avesani. Trust-aware bootstrapping of recommender systems. In *Proceedings of ECAI 2006 Workshop on Recommender Systems*, pages 29–33.

[14] http://glaros.dtc.umn.edu/gkhome/views/metis.