Competitive Opinion Maximization in Social Networks

by

Jianjun Luo

A Thesis

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Master of Science

in

Data Science

by

April 2018

APPROVED:

Professor Xiangnan Kong, Advisor

Professor Yanhua Li, Reader

Abstract

Influence maximization in social networks has been intensively studied in recent years, where the goal is to find a small set of seed nodes in a social network that maximizes the spread of influence according to a diffusion model. Recent research on influence maximization mainly focuses on incorporating either user opinions or competitive settings in the influence diffusion model. In many real-world applications, however, the influence diffusion process can often involve both real-valued opinions from users and multiple parties that are competing with each other. In this paper, I present the problem of competitive opinion maximization (COM), where the game of influence diffusion includes multiple competing products and the goal is to maximize the total opinions of activated users by each product. This problem is very challenging because it is #P-hard and no longer keeps the property of submodularity. I propose a novel model, called ICOM (Iterative Competitive Opinion Maximization), that can effectively and efficiently maximize the total opinions in competitive games by taking user opinions as well as the competitor's strategy into account. Different from existing influence maximization methods, I inhibit the spread of negative opinions and search for the optimal response to opponents' choices of seed nodes. I apply iterative inference based on a greedy algorithm to reduce the computational complexity. Empirical studies on real-world datasets demonstrate that comparing with several baseline methods, the ICOM approach can effectively and efficiently improve the total opinions achieved by the promoted product in the competitive network.

Acknowledgements

I would like to express my deepest appreciation to Professor Xiangnan Kong for his support as my advisor. I am deeply indebted to him for critiquing my research, sharing his insights, reading over drafts of my thesis, and contributing to my growth in research. I absolutely could not have done this without his help and encouragement.

I thank Professors Yanhua Li for his generosity with his time to be my thesis reader, giving valuable critiques and insights within a limited time frame.

I would like to thank Xinyue, a senior student for sharing his insights and experience. In addition I would like to express my gratitude to the members of DSRG for advice and encouragement.

Lastly, I express my warmest gratitude to my family, for their unconditional support.

Contents

List of Figures

List of Tables

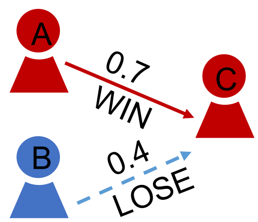
Chapter 1

Introduction

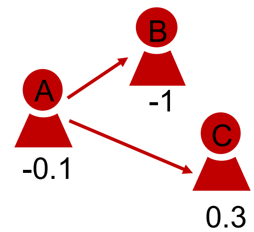
As John Donne wrote, "No man is an island." Individuals are linked together by different relationships: family, friends, *etc.* These interactions determine a social network, where information propagates and influence spreads among the members. A lot of studies investigate how people are affected by their "neighbors" in this underlying network and the diffusion processes of "word-of-mouth" effects, especially for the application in viral marketing [33]. Motivated by the increasing applications (such as viral marketing, rumor control, public opinion formation, and personalized recommendation), the problem of influence maximization has been studied intensively in recent years.

1.1 Overview

The influence maximization problem intends to identify the influential users in a social network, so as to maximize the coverage of an item (*e.g.*, a product or a political view) by the propagation. More formally, given a social network and a diffusion model, the influence maximization problem aims at selecting a small set of seed nodes that maximizes the spread of influence in the social network. This



(a) Competition: Party *Red* wins.



(b) Total opinions: $\mathcal{O} = -0.8$

Figure 1.1: The problem of competitive opinion maximization in social networks has two major features: (a) Two parties compete in the diffusion. A party with larger influence weight (relationship strength) wins. (b) Infected users have their own opinions. Total opinions of a party are the sum of its infected users' opinions.

problem is especially important in viral marketing. For example, with a limited budget of product promotion, a company may want to selectively choose a small set of users to distribute free samples, hoping they would recommend the product to their friends or followers, consequently increases the product sales or brand awareness by word-of-mouth marketing.

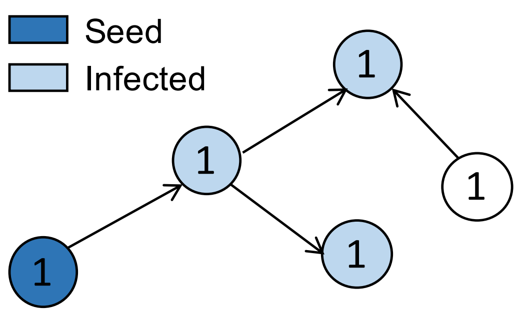
Influence maximization (IM) problem was first formalized by Domingos *et al.* in [13], and Kempe *et al.* then proposed two popular diffusion models, *i.e.*, the independent cascade (IC) model and the linear threshold (LT) model [22]. Conventional influence maximization problem assumes that the influence propagating in the network is positive, *i.e.*, the more users get exposed to the target item, the better the goal of promotion will be achieved, *e.g.*, the higher the sales or reputation of the item will have. Later, lots of the works focus on studying the diffusion of a single item in a social network. However, in many real-world applications, neither of these two assumptions is true (as shown in Figure 1.1). On the one hand, it is often the case that multiple parties start promotion in a social network simultaneously, and their target items are of the same type, *i.e.*, with same features and competing with each other. A user will choose the item which is widely adopted by its neighbors or by its close friends. On the other hand, if users who get exposed to the target item express negative opinions towards it, its reputation will be harmed. The opinions depend on users' preference, so the promoters should try to cover more users with positive opinions as the target audience. Thus the promoters should take user opinions as well as the competitor's strategy into account.

In this thesis, I study the problem of competitive opinion maximization (COM) under a competitive linear threshold (CLT) model. In CLT model, each party propagates the same way as it does in the linear threshold model, but an activated node cannot be influenced again by another party. An inactive node which receives influence from different parties at the same time will be activated by the one who sends the highest influence weight. Formally, the competitive opinion maximization problem corresponds to selecting a small set of seed users as the optimal response to the observed or assumed opponent's choices of seeds. The objective of selection is to maximize the total opinions gained after a competitive diffusion. Any two parties of the competition can be divided as a first mover and a second mover. A second mover can simply make its selection based on known opponent's selection, but the first mover needs to search for the optimal choices to maximize the total opinions under the disadvantage of being observed by its opponent. The problem of opinion maximization in CLT diffusion model has not been studied in this context so far.

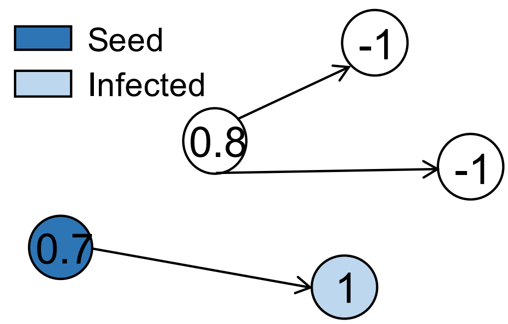
1.2 Challenge

The major research challenges on competitive opinion maximization can be summarized as follows:

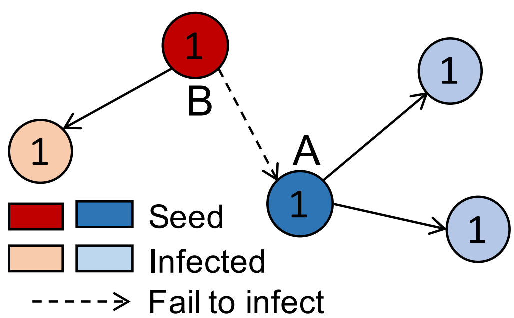
• **Users' Opinions:** Conventional influence maximization problems and competitive influence maximization (CIM) problems [4, 5, 6, 21], a natural extension of influence maximization assuming multiple items to be propagated in a social net-



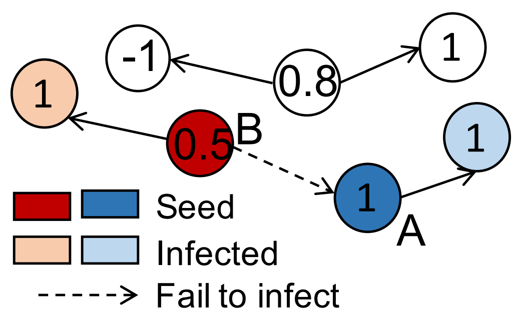
(a) Influence Maximization [22]. Diffusion result:$|\mathcal{I}| = 4$



(b) Opinion Maximization [31, 34]: $\mathcal{O} = 1.7$



(c) Competitive Influence Maximization [3]: $|\mathcal{I}_{\text{blue first}}| = 3$, $|\mathcal{I}_{\text{red second}}| = 2$



(d) Competitive Opinion Maximization (this thesis): $\mathcal{O}_{\text{blue first}} = 2$, $\mathcal{O}_{\text{red second}} = 1.5$

Figure 1.2: Comparison of four related problems and the results of their influence diffusion. Here $\mathcal{I}$ represents the set of nodes infected in the diffusion. $\mathcal{O}$ represents the total opinions of the infected users. $\mathcal{I}_{\text{blue first}}$ represents the users being infected by the *blue* party (first mover) in the competitive diffusion.

work, all assume the users who are influenced will adopt or like the item (as shown in Figures 1.2(a) and 1.2(c)). However, there are unpopular items that most of the users in the social network will express negative opinions. In such case, the promotion without considering opinions will lead to a bad overall reputation. From a long-term perspective, a promoter hopes most of the expressed opinions in the social network are positive.

The importance of users' opinions can also be explained by market segmentation or niche product. For example, literary film, experimental music or anime culture is well accepted by certain customer groups, but disliked by other groups. As shown in Figure 1.2(b), the best seed node can only infect one node ($|\mathcal{I}| = 2 < 3$), but its total opinions are maximized ($\mathcal{O} = 1.7 > 0.8$). The same rule applies to the seed selection in Figure 1.2(d).

I assume opinions will not influence the adoption of other nodes, but the infection of unwanted nodes damages the effect of promoting. It is important to target the appropriate customers because maximizing the spread of influence during diffusion is no longer the optimal solution.

• **Rational Competitors:** Another challenge comes from the fact that multiple parties are competing in the market. If a user has already adopted an item, he or she will not accept another of the same type. It is possible that the user adopt both of the items in a very special case that receives same influence at the same time, then I assume the volume of his or her opinion will be weaken to half of the original. We assume all the parties in the market are rational, which means they are likely to choose similar ideal customers if ignoring the competitors. The failure in a competition can block the diffusion of influence and significantly lower the total opinions.

For example, as shown in Figure 1.2(d), the first mover, the *blue* party will choose node B as a seed if it ignores the *red* party, expecting $\mathcal{O}_{\text{blue}} = 3.5$. But in that case, node A will be selected by *red* party and block *blue* party's diffusion, $\mathcal{O}_{\text{blue}} = 1.5$. It is worth noting that knowing node B is chosen, the *red* party will not select B again to get half of the opinions of four nodes, $\mathcal{O}_{\text{red}} = 1.75 < 2$. So the best seed node for party *blue* is node A, $\mathcal{O}_{\text{blue}} = 2$. Similarly, the *blue* party will select node A instead of B because $\mathcal{I}_{\text{blue}} = 3 > 2$.

A wise decision should take opponents' choices into consideration and estimate the outcome based on the possible failure.

• **Scalability:** Different from the problems that maximize the influence, our COM problem is neither submodular or monotone, which means a naive greedy algorithm has no guarantee to the approximation ratio in [4]. Still, if the opponents' choices of seeds are observed, we can use a greedy algorithm as a heuristics to search through the network. However, in some cases, the promoter can only make decisions based on the known opponent's strategy and will be observed by competitors as the first mover. Due to the passive position, a naive greedy algorithm for a first mover requires a lot of simulations, which is very slow for large networks. To address the problem, I utilize the inference from iterative simulation and design an efficient and effective heuristic algorithm.

1.3 Contribution

The main contributions of this thesis can be summarized as follows:

• I present a novel problem of Competitive Opinion Maximization (COM) about how to effectively select a set of seeds to maximize opinions under competition in a social network.

• I propose a novel solution, called ICOM (Iterative Competitive Opinion Maximization) method, to solve COM problem. By explicitly exploiting the users' opinions and opponents' information, our ICOM method can effectively find a set of seeds to compete with other parties for total opinions with an iterative inference procedure.

- Empirical studies on real-world tasks demonstrate that the proposed iterative opinion maximization approach can significantly boost the performances in terms of total opinions achieved by the promoter with competitors in real-world data sets.

The rest of the thesis is organized as follows. Chapter 3 introduces the notation and basic theory used throughout the thesis. Chapter 4 presents the ICOM algorithm based on greedy method. Chapter 5 presents the experimental setup and reports the experiment results over three real-world data sets. Chapter 6 concludes the thesis.

Chapter 2

Related Works

To the best of our knowledge, this paper is the first work on competitive opinion maximization. This work is related to influence maximization, opinion maximization and competitive influence maximization.

2.1 Influence Maximization

Influence maximization is to study how to choose a small set of seed nodes in a network, which has the best opportunity to influence the most number of nodes through a given diffusion model. Domingos *et al.* [13] first proposed to model the customer's network value that derives from his or her influence on other customers for viral marketing. Richardson *et al.* [33] extended the work by proposing a less computational cost model and applied it to the data from a knowledge-sharing site. In [22] Kempe *et al.* obtained the first provable approximation guarantee for the two basic stochastic influence cascade models they proposed, the independent cascade (IC) model and the linear threshold (LT) model. The objective function under these two models has nice properties of monotonicity and submodularity.

Later, many works have been proposed to tackle the limitation of the simple greedy algorithm. Chen *et al.* designed a scalable algorithm for the IC model that can handle large-scale social networks with more than a million edges, and proposed the first scalable influence maximization algorithm tailored for the LT model [9, 10, 11]. In [17], an alternative algorithm is proposed to compute the spread under LT model, which outperforms LDAG heuristic in [9]. While using heuristics to estimate the spread is one way to improve the efficiency, an efficient algorithm called CELF proposed by Leskovec *et al.* [27] exploits submodularity and dramatically improves the efficiency of the greedy algorithm. Based on that, Goyal *et al.* [16] introduces a further optimized approach CELF++ for influence maximization in social networks.

Several extensions, like distinguishing specific users from others and differentiating product adoption from influence, have been proposed to describe the real world situation more accurately [2, 24]. An alternative data-based approach was proposed by Goyal *et al.* to directly estimate influence spread by exploiting historical data, *i.e.*, traces of past action propagation.

2.2 Opinion Maximization

Traditional influence maximization problem has an assumption that people always hold positive opinions towards the promoted products, which is not satisfied in reality. Therefore, it is meaningful to study the extended problem, opinion maximization, where infected people can hold negative opinions. Instead of selecting the most influential nodes, the aim of opinion maximization is to make the item favorable and get more positive opinions.

In [8], Chen *et al.* proposed a model that incorporates the emergence and propagation of negative opinions into the IC model to maximize the expected number of positive activated nodes. Zhang *et al.* [34] considered the negative and neutral opinions, proposed an adapted IC model to maximize the total opinions of activated users. While the users' opinions are generated randomly in [34], Gionis *et al.* [15] took the process of opinion formation into consideration. They assumed the opinions of individuals get formed dynamically by the mutual influence of internal opinions and the neighbors' opinions. The goal is to find a set of users whose positive opinions about an item will maximize the overall opinions for the item in a social network.

Some studies [25, 26, 28] exploit opinions in a different way. They assume there are both positive and negative relationships in the network, thus people are more likely to adopt their friends' opinions and do the opposite to the foe's opinions. Such works extend influence maximization problem for the signed networks, and are different from the opinion maximization problem discussed in this thesis.

2.3 Competitive Influence Maximization

The problem is to study the simultaneous propagation of multiple items in a social network. The solutions for competitive influence maximization can basically be classified into two types: opponent strategy known or opponent strategy unknown.

Given the opponent's selected seed nodes, a popular solution is to minimize the opponent's influence, *i.e.*, Influence Blocking Maximization [4, 5, 21]. Borodin *et al.* and He *et al.* proposed the greedy algorithms based on the LT model, while Budak *et al.* provided a greedy algorithm for the IC model. Carnes *et al.* [6] studied the competitive influence maximization problem from a follower's perspective, *i.e.*, finding a best response to the first mover's selection. For competitive influence maximization problems with known and fixed competitor's strategy but not seed set, to some extend we can observe or predict the opponent's seed selection and then search for the best response directly.

For the problem where the opponents' strategy and choices are not known, Bharathi *et al.* [3] proposed a natural generalization of the IC model and used game theory to study the diffusion with multiple competing items. In [7], Chen *et al.* proposed a data-driven approach to study the multi-player influence maximization and proposed a game to collect the picking strategies from human or AI to analysis. In [29], Lin *et al.* proposed a learning-based framework using reinforcement learning and game theory to address the multi-round competitive influence maximization problem. Zhang *et al.* studied the maximization problem of multiple competing or complementary products in a social network at the same time in[35].

Chapter 3

Problem Formulation

In this section, we will define important notations, the competitive linear threshold (CLT) diffusion model and competitive opinion maximization (COM) problem separately.

3.1 Concept Definitions

We first define some important concepts, which will be used throughout this paper.

Definition 1 (Social Network): An online social network can be represented as $G(\mathcal{V}, \mathcal{E}, \mathcal{W})$, where nodes $\mathcal{V} = \{u_1, \cdots, u_n\}$ is the set of users, $\mathcal{E}$ is the set of social links among users in $\mathcal{V}$ and $\mathcal{W} = \{w_{ij} | i = 1, \cdots, n; j = 1, \cdots, n\}$ is the set of link weights. In the network, user (node) u_i can be influenced by its neighbor u_j according to the weight w_{ij}. The set of items (products) associated with G is denoted by $\mathcal{I} = \{i_1, \cdots, i_m\}$. Each item i_t has a rating vector $\mathbf{r}_t = (r_{1t}, \cdots, r_{nt})$, where r_{it} denotes the rating user u_i gives product i_t.

Definition 2 (Opinions): $\mathbf{o}_t = (o_{1t}, o_{2t}, \cdots, o_{nt})$ is the opinion vector of item t, where $o_{it} \in [-1, 1]$ represents the opinion of user u_i towards item i_t. The opinion is mapped from the rating r_{it} using minmax normalization, $o_{it} = 2 \cdot \frac{r_{it} - r^t_{\min}}{r^t_{\max} - r^t_{\min}} - 1$, where

Table 3.1: Important Notations.

Symbol	Definition
$\mathcal{V} = \{u_1, \cdots, u_n\}$	the set of nodes
$\mathcal{E} = \{e_i \in \mathcal{V} \times \mathcal{V}\}$	the set of edges or links
$\mathcal{I} = \{i_1, \cdots, i_m\}$	the set of items (products)
$\mathcal{P} = \{p_1, \cdots, p_c\}$	the set of parties competing in the network
$\mathcal{S} = (\mathcal{S}^1, \mathcal{S}^2, \cdots, \mathcal{S}^c)$	the set of seed node selection of c parties
u_i	user (node) i
w_{ij}	influence weight from u_i to u_j
i_t	item (product) t
p_v	party (promoter) v
r_{it}	the rating of i_t assigned by u_i
$o_{it} \in [-1, 1]$	u_i's opinion towards i_t
$s_i^v \in [0, 1]$	the active status of u_i promoted by p_v
$\mathcal{S}^v$	the seed user selection of p_v
$\mathbf{r}_t = (r_{1t}, \cdots, r_{nt})$	the vector of ratings on item i_t assigned by all users in $\mathcal{V}$
$\mathbf{o}_t = (o_{1t}, o_{2t}, \cdots, o_{nt})$	the vector of opinions on item i_t hold by all users in $\mathcal{V}$
$\mathbf{s}_v = (s_1^v, s_2^v, \cdots, s_n^v)$	the vector of status of users activated by p_v

$r^t_{\max}$ and $r^t_{\min}$ are the minimum and maximum rating of product i_t.

Definition 3 (Party): There are multiple parties $\mathcal{P} = \{p_1, \cdots, p_c\}$ promoting competing products in the network. These competing products are identical, thus users will express the same opinions (preferences) on the competing items from different parties. In other words, these parties share the same opinion vector when they promote competing items.

Definition 4 (Active User Vector): Users who are influenced by Party v to adopt the promoted item are defined to be activated by p_v, while others are inactive. Active user vector $\mathbf{s}_v = (s_1^v, s_2^v, \cdots, s_n^v)$ represents the activated status of all users in the network by party p_v. Given the target item i_t, when user u_i is activated by party p_v and adopt the product with opinion o_{it}, $s_i^v = 1$, otherwise $s_i^v = 0$ (inactivated).

In special cases $s_i^v \in (0,1)$ when u_i is activated by more than one party.

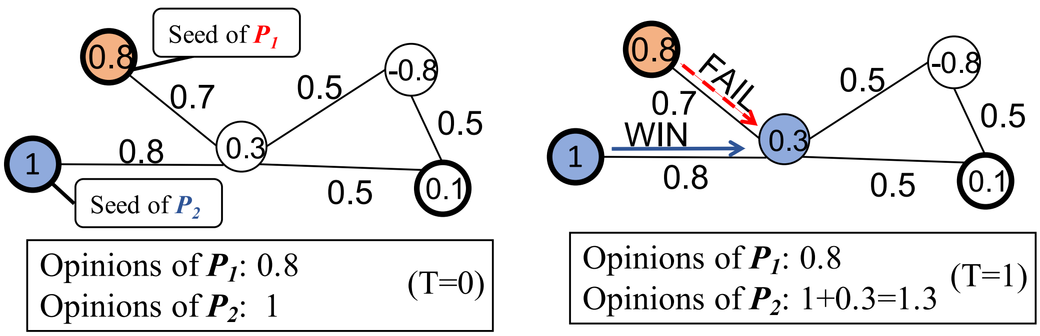
Definition 5 (Seed Set): The decisions made under the marketing strategies of c competing parties can be represented as seed user set list $\mathcal{S} = (\mathcal{S}^1, \mathcal{S}^2, \cdots, \mathcal{S}^c)$, while $\mathcal{S}^{-v} = (\mathcal{S}^1, \cdots, \mathcal{S}^{v-1}, \mathcal{S}^{v+1}, \cdots, \mathcal{S}^c)$ defines the known or predicted seed sets selected by all parties except p_v.

3.2 Problem Definition

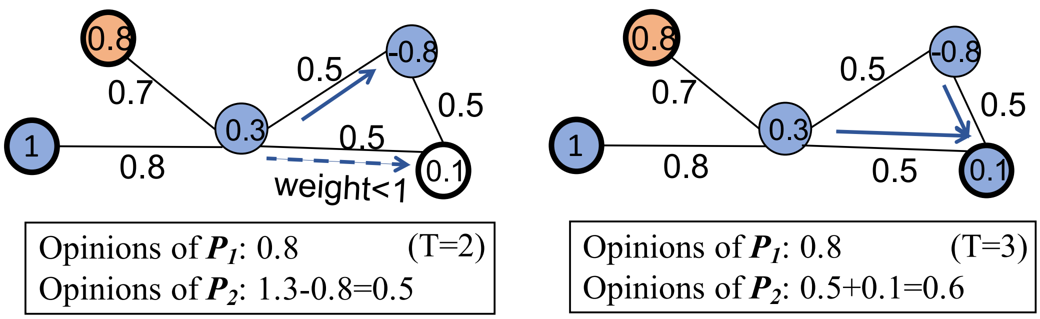
Definition 6 (Competitive Linear Threshold): Competitive Linear Threshold (CLT) model is similar to the naive LT model, but with multiple participants. Each node u_i in the network has an activation threshold θ_i and is possibly influenced by each neighbor j with weight w_{ij}. The influences of different parties start to propagate at the same time. At any timestamp T, if the total received weight of an inactive node i from party p_v's activated neighbors is higher than that from other parties and θ_i, then u_i is activated by p_v at timestamp $T+1$ (as shown in Figure 3.1(a)). Once a node is activated by one party, it cannot be activated by any other party. When more than one party selects the same node as seed node or is eligible to activate the same node at the same timestamp T, *e.g.*, p_1 and p_2, having $\sum_{i=0}^{n} w_{ij} s_i^1 = \sum_{i=0}^{n} w_{ij} s_i^2 > \theta_j$, node j will be equally activated by both parties p_1 and p_2 at $T+1$, denoted as $s_i^1 = s_i^2 = \frac{1}{2}$. The influence process repeats until no new node becomes active by any party (as shown in Figure 3.1(b)).

Then we define the competitive opinion maximization problem with CLT model as follows.

Definition 7 (Competitive Opinion Maximization): Given the network G, an item $i_t \in \mathcal{I}$, corresponding opinion $\mathbf{o}_t$, the budget of seeds k and the CLT diffusion model, the goal of each party p_v in COM is to select a marketing strategy, *i.e.*, a set of seed nodes $\mathcal{S}^v (|\mathcal{S}^v| = k)$ among $\mathcal{V}$ to propagate the influence to maximize the total



(a) Two parties $\mathcal{P}_1, \mathcal{P}_2$ compete for a node, the one with higher influence weight wins.



(b) Parties continue to propagate until no new nodes can be activated.

Figure 3.1: An example of CLT diffusion process. In this undirected network, the threshold of bold circle nodes are 1, while the light ones are 0.5. Weights of influence are on the edge, and opinions are in the node.

opinions towards item i_t achieved at the end of the diffusion, *i.e.*, $\mathbf{s}_v = I(\mathcal{S}^v|\mathcal{S}^{-v})$,

$$\mathcal{S}^v = \arg\max_{\mathcal{S}^v} \sum_{s_i^v > 0} s_i^v o_{it}$$
$$\text{s.t.} \quad |\mathcal{S}^v| \leq k$$

where $I(\mathcal{S}^v|\mathcal{S}^{-v})$ denotes the influence function of party p_v on item i_t given the known seed user sets $\mathcal{S}^{-v}$.

In other words, for p_v, the objective is to select seed set $\mathcal{S}^v$ to maximize the total opinions of eventually-influenced users $\sigma(\mathcal{S}^v)$.

For simplicity, we set $|\mathcal{P}| = 2$ in the following method description and experiment. We assume the opinions towards the items promoted by different parties are given and the same as $\mathbf{o}_t$, if the items and i_t are competing products. This is consistent with the general case that users have relatively fixed preferences towards the same kind of products. The parties can be classified into two types: the first-mover who only knows the strategy of the opponents, by which it can estimate $\mathcal{S}^{-v}$, and the second-mover who also knows the competitors' exact choices of seed users $\mathcal{S}^{-v}$. This thesis focuses on the competitive LT model, but the proposed model shall be able to be adapted to competitive IC model.

Computing the conventional influence is #P-hard, but the influence function $I(\mathcal{S})$ is monotone and submodular[22], where $\mathcal{S}$ is the initial seed set. However, for COM problem, the activated users with negative opinions can harm the final performance and lower the total opinions achieved, which makes the problem no longer keeps the property of submodularity or monotonicity.

Chapter 4

Method

In this section, the proposed method is introduced in details. We first propose an adapted greedy algorithm on competitive opinion maximization (GCOM) by incorporating the competitive setting and opinions objective into greedy algorithm for influence maximization. Based on GCOM, we propose an iterative method (ICOM) to efficiently and effectively maximize the total opinions under CLT model. Finally we compare the time complexity of ICOM and GCOM.

4.1 Preliminary

In [22], Kempe *et al.* proposed the greedy algorithm which is a $(1-1/e)$ approximation to solve the influence maximization problem. It is a simple strategy for influence maximization (IM) without considering competition or users' opinions. The objective function is the expected number of activated users at the end. However, it can be easily adapted to chase opinion maximization (OM) by modifying the objection function from $\sigma(\mathcal{S}) = |I(\mathcal{S})|$ to $\sigma(\mathcal{S}) = I(\mathcal{S}) \cdot \mathbf{o}_t$. The seed selection made by this strategy should outperform IM in terms of total opinions. The pseudo-code of the OM algorithm is available in Algorithm 1.

Algorithm 1 Greedy Alogorithm for OM

Input: social network G, opinion vector $\mathbf{o}_t$, seed set size k

Output: the seed user selection $\mathcal{S}$

initialize $\mathcal{S} \leftarrow \emptyset$
while $\mathcal{V} \setminus \mathcal{S} \neq \emptyset \wedge |\mathcal{S}| \neq k$ **do**
    $O_{\max} \leftarrow -\infty$
    for `each` u `in` $\mathcal{V} \setminus \mathcal{S}$ **do**
        $\mathcal{S} \leftarrow \mathcal{S} \cup \{u\}$
        $\mathbf{s} \leftarrow I(\mathcal{S})$
        $O \leftarrow \mathbf{s} \times \mathbf{o}_t$
        if $O > O_{\max}$ **then**
            $O_{\max} \leftarrow O$, $u_{\text{best}} \leftarrow u$
    $\mathcal{S} \leftarrow \mathcal{S} \cup \{u_{\text{best}}\}$
Return $\mathcal{S}$

It is worth noting that there are many improved algorithm to solve influence maximization problem [1, 12, 14, 19, 20, 32]. Algorithms like CELF++[16] and LDAG[9] are able to dramatically improve the efficiency under LT model. However, in this thesis, I do not focus on improving the efficiency of seed selection algorithm in influence maximization as these studies. Instead, I propose a strategy that can solve COM problem effectively with opinions and opponent's information better than conventional IM strategies without additional consideration.

4.2 Greedy Algorithm on COM

Although two parties start to propagate in the network at the same time, they can be classified into the first mover and the second mover based on the accessible information. A second mover in the competition knows the competitor's exact choices of seed users $\mathcal{S}^{-v}$, while a first mover only knows the competitor's seed user selection strategy M.

We first use an analogous greedy algorithm GCOM to search the optimal response to the competitors' observed or simulated choices directly (Figure 4.1).

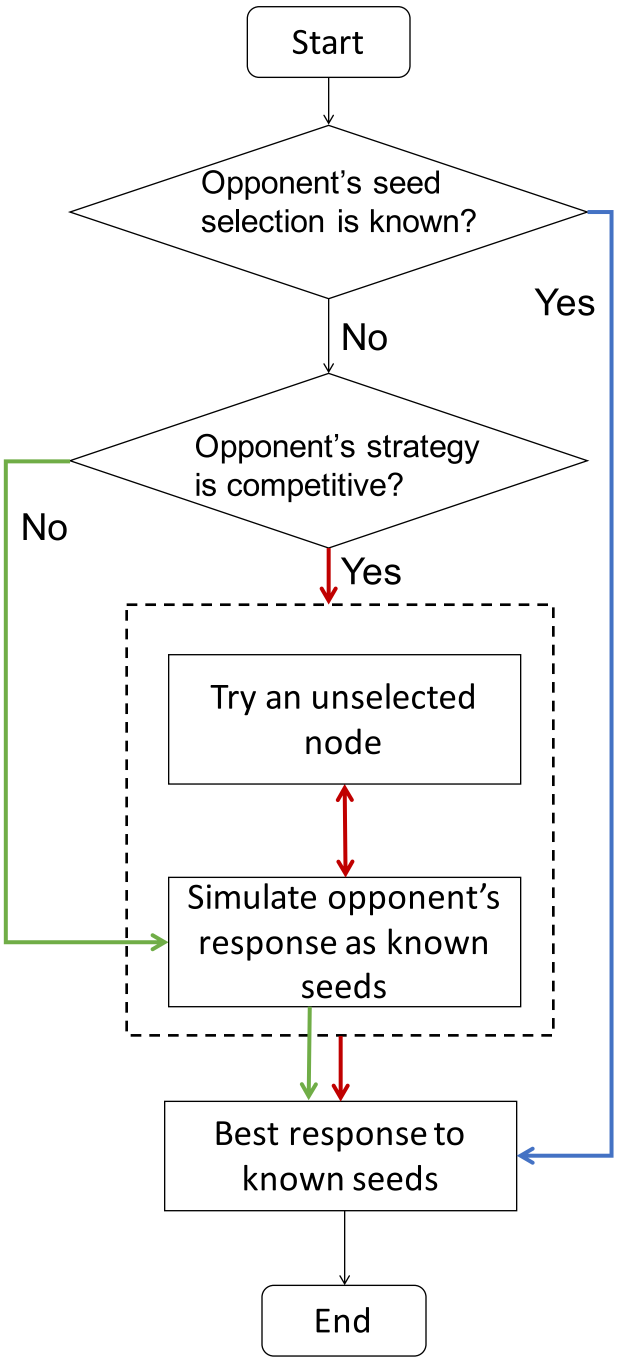


Figure 4.1: Flow chart of greedy competitive opinion maximization strategy: A second mover knows competitor's selection (blue), while a first mover only knows competitor's strategy (red or green). The first mover tries to simulate all opponent's possible selection to make optimal response.

Algorithm 2 Greedy Alogorithm for COM

Input: social network G, opinion vector $\mathbf{o}_t$, seed user size k, competitors' strategy(algorithm) M, competitors' choices of seed users $\mathcal{S}^{-v}$ (optional)

Output: the seed user selection $\mathcal{S}^v$

initialize $\mathcal{S}^v \leftarrow \emptyset$
if $\mathcal{S}^{-v} = \emptyset \wedge M$ ignores competition **then**
  $\mathcal{S}^{-v} \leftarrow M(G, \mathbf{o}_t, k)$
if $\mathcal{S}^{-v} \neq \emptyset$ **then**
  while $\mathcal{V} \setminus \mathcal{S}^v \neq \emptyset \wedge |\mathcal{S}^v| \neq k$ **do**
    $\sigma(\mathcal{S}^v) \leftarrow I(\mathcal{S}^v|\mathcal{S}^{-v}) \cdot \mathbf{o}_t$
    $u_{\text{best}} \leftarrow \arg\max_{u \in \mathcal{V} \setminus \mathcal{S}^v} \sigma(\mathcal{S}^v \cup \{u\}) - \sigma(\mathcal{S}^v)$
    $\mathcal{S}^v \leftarrow \mathcal{S}^v \cup \{u_{\text{best}}\}$
else
  while $\mathcal{V} \setminus \mathcal{S}^v \neq \emptyset \wedge |\mathcal{S}^v| \neq k$ **do**
    $O_{\max} \leftarrow -\infty$
    for `each` u `in` $\mathcal{V} \setminus \mathcal{S}^v$ **do**
      $\mathcal{S} \leftarrow \mathcal{S}^v \cup \{u\}$
      $\mathcal{S}^{-v} \leftarrow M(G, \mathbf{o}_t, k, \mathcal{S})$
      $\mathbf{s}^v \leftarrow I(\mathcal{S}|\mathcal{S}^{-v})$
      $O \leftarrow \mathbf{s}^v \cdot \mathbf{o}_t$
      if $O > O_{\max}$ **then**
        $O_{\max} \leftarrow O$, $u_{\text{best}} \leftarrow u$
    $\mathcal{S}^v \leftarrow \mathcal{S}^v \cup \{u_{\text{best}}\}$
Return $\mathcal{S}^v$

If p_v is a second mover who has observed the competitor's choices of seed users $\mathcal{S}^{-v}$, to form its own seed set, the promoter will search through all the nodes in $\mathcal{V}$ to greedily select the seeds. Before the diffusion, Every node in the network is inactive because the propagation of different parties starts at the same time. The party v selects the seed nodes one by one until the budget of k is achieved. When searching for the qth ($q < k$) seed user, it tries every node that is not selected in the previous $q - 1$ seeds to form different temporary $\mathcal{S}^v$. The final result O simulated by the temporary $\mathcal{S}^v$ and the fixed $\mathcal{S}^{-v}$ will be recorded. Let the qth seed of p_v be the available node with the maximum simulated total opinion $\sigma(\mathcal{S}^v \cup \{u\}$.

The selecting process of a first mover who only knows the opponent's strategy

M is similar. If the opponent's strategy M does not consider competition, knowing M is the same as knowing $\mathcal{S}^{-v}$. The first mover can estimate competitor's seed set applying M. However, if the opponent's strategy takes competition into account, every time to find a node with maximum marginal gain, the promoter needs to simulate the $\mathcal{S}^{-v}$ based on the temporary $\mathcal{S}^{v}$ repeatedly. The pseudo-code of the algorithm is available in Algorithm 2.

However, it will be impractical and not scalable for the first mover when the opponent's seed set depends on its seed set and can not be predicted before selection due to the extra cost of simulating second movers' choices. So I adapt GCOM by using iterative inference and propose an efficient and effective method to address the COM problem.

4.3 Iterative Algorithm on COM

Now that I propose the iterative competitive opinion maximization (ICOM) method (Algorithm 3) to improve the GCOM algorithm by simplifying the work flow for first movers. As discussed in section 4.2, reducing the simulation times required by a first mover whose opponent seed set is not fixed is a key point to tackle the limitation of simple greedy method. Inspired by the tit for tat strategy in game theory, ICOM is a heuristic method based on a small batch of multi-round simulation to approximate the strategic dominance.

Instead of considering every available node for qth seed to form possible temporary seed sets and then simulating the possible responses from second movers, ICOM randomly chooses k nodes as the very first seed set of first mover $\hat{\mathcal{S}}^{v}$. Given the opponent's strategy M, we can easily infer the possible response $\mathcal{S}^{-v}$ from the competitor to the initial $\hat{\mathcal{S}}^{v}$. Then the iterative selection begins, and it includes

Algorithm 3 Iterative Greedy Alogorithm for COM

Input: social network G, opinion vector $\mathbf{o}_t$, seed user size k, competitors' strategy(algorithm) M, competitors' choices of seed users $\mathcal{S}^{-v}$ (optional), maximum round of iteration r

Output: the seed user selection $\mathcal{S}^v$

initialize $\mathcal{S}^v \leftarrow \emptyset$
if $\mathcal{S}^{-v} = \emptyset$ and M is a competitive model **then**
    Randomly generate an initial $\hat{\mathcal{S}}^v$
    $\mathcal{S}^{-v} \leftarrow M(G, \mathbf{o}_t, k, \hat{\mathcal{S}}^v)$
    $O_{\max} \leftarrow -\infty$
    for *round* in $1, \cdots, r$ **do**
        $\mathcal{S}^v \leftarrow GCOM(G, \mathbf{o}_t, k, M, \mathcal{S}^{-v})$
        $\mathcal{S}^{-v} \leftarrow M(G, \mathbf{o}_t, k, \mathcal{S}^v)$
        $\mathbf{s}^v \leftarrow I(\mathcal{S}|\mathcal{S}^{-v})$
        $O \leftarrow \mathbf{s}^v \cdot \mathbf{o}_t$
        if $O > O_{\max}$ **then**
            $O_{\max} \leftarrow O$, $\mathcal{S}^v_{\text{best}} \leftarrow \mathcal{S}^v$
    $\mathcal{S}^v \leftarrow \mathcal{S}^v_{\text{best}}$
else
    if $\mathcal{S}^{-v} = \emptyset$ **then**
        $\mathcal{S}^{-v} \leftarrow M(G, \mathbf{o}_t, k)$
    $\mathcal{S}^v \leftarrow GCOM(G, \mathbf{o}_t, k, M, \mathcal{S}^{-v})$
Return $\mathcal{S}^v$

following steps: First, a new temporary $\mathcal{S}^v$ can be selected based on GCOM and the inferred $\mathcal{S}^{-v}$, *i.e.*, first mover p_v can search for a optimal response towards $\mathcal{S}^{-v}$ as a second mover. Then a new inferred $\mathcal{S}^{-v}$ can be predicted based on the new temporary $\mathcal{S}^v$ and fixed strategy M model. The final active users of p_v diffused by $\mathcal{S}^v$ after competing with $\mathcal{S}^{-v}$ will be $\mathbf{s}^v$, which is used to estimate the total opinions O achieved in this round. Repeat such iterative inference until the maximum round r is reached. We choose the $\mathcal{S}^v$ with the highest total opinion $O_{\max}$ in all these rounds as the seed set of p_v.

4.4 Complexity Analysis

In this section, we simply compare the time complexity of ICOM with GCOM to validate the advantage of iterative inferences. We assume two parties are competing in the market using GCOM or ICOM. ICOM and GCOM use the same CLT model for propagation, which we use $O(|A|)$ to denote its time complexity. For the second mover, it is not hard to realize that both methods have the same work flow and therefore the same complexity $O(k|\mathcal{V}||A|)$, since $O(|\mathcal{V}|) < O(|A|)$. However, for the first mover, GCOM has to simulate the second mover's decision process repeatedly, which leads to a complexity of $O(k^2|\mathcal{V}|^2|A|)$. Using ICOM, we do not need to search all inactive nodes for selecting a seed to run the second mover simulation repeatedly, but just iterate several times to find the optimal solution. The complexity of ICOM for a first mover is $O(rk|\mathcal{V}||A|)$.

Chapter 5

Experiments

To test the effectiveness of ICOM in addressing opinion maximization problem under the CLT model, I conduct experiments on three real-world datasets. In this chapter, I first introduce the data sets as well as the experiment setup. After that I present the results of different strategy combination under competitive propagation model to show the effectiveness of ICOM. At last we discuss and analyze the influence of parameter.

5.1 Data Collection

In order to evaluate the performance of the proposed approach for opinion maximization under competitive environment, I tested the approach on three real-world networks with ratings (Summarized in Table 5.1).

CiaoDVD: Ciao is a website for product reviews and price comparison. The dataset is crawled from the entire category of DVDs from the website (`http://dvd.ciao.co.uk`) by Guo *et al.* in [18]. The trust relationships between users are presented by the directed edges of the networks. The ratings of the product have a scale of one to five. Items with less than 5 ratings are filtered from the graph.

Table 5.1: Summary of experimental datasets.

	Data Sets		
Characteristics	CiaoDVD	Flixster	Filmtrust
# Nodes	2,740	5,320	1,642
# Links	20.8k	44.3k	44.6k
# Items	13.1k	3,470	2,071
# Ratings	34.3k	110k	35.5k
Link type	directed	undirected	directed

Flixster: Flixster is a website and a mobile app for movie information and ratings. The data set is consist of 137,925 nodes and 1,269,373 edges, with ratings of 48,794 items given by the users. It is hard to perform experiments on the large size of networks, so I used the smaller dataset sampled by Graclus [2]. The users of Flixster link with each other in the form of "friendship", which are shown as undirected links in the network. The movies are rated on a scale from 0.5 to 5. I also remove the items with less than 5 ratings from the users in the subgraph.

Filmtrust: FilmTrust is a small dataset crawled from the entire FilmTrust website in June, 2011. The website integrates social networks with movie ratings and reviews. Similar to CiaoDVD, this is a directed network that users links others with "trust". The ratings of movies are given on a scale of 0.5 to 4. Items with less than 5 ratings are filtered from the graph.

5.2 Experiment Setting

To make the link type of Flixster consistent with other datasets, we replace the undirected links with directed links. For example, friendship between u_i and u_j is represented by two directed links $u_i \rightarrow u_j$ and $u_i \leftarrow u_j$.

In CLT model, a user u_i can influence neighbors with certain weights. The weight of a directed link e_{ij} measures the influence from u_i to u_j. We calculate the weight of e_{ij} using Jaccard coefficient, which is widely used in social influence analysis. The strength of relationship, *i.e.*, the weight of link is defined as $\frac{\Gamma(u_i)\cap\Gamma(u_j)}{\Gamma(u_i)\cup\Gamma(u_j)}$. The threshold of users $[\theta_1,...,\theta_m]$, is randomly generated from a uniform distribution within $[0, 1]$.

Complete ratings of an item from all the users can be estimated by the incomplete rating vector in datasets. To predict unknown ratings based on observed ratings, I use the matrix factorization method for collaborative filtering following [23, 30, 31]. The rating of user u_i towards item i_t can be approximated by the inner product of user profile and item profile, which can be learned given the observed ratings. The approximated ratings which exceed the scope are replaced by the highest or lowest rating allowed in corresponding data set.

I then convert the ratings to opinions using minmax normalization. The opinion $o_{it}=2\cdot\frac{r_{it}-r^t_{\min}}{r^t_{\max}-r^t_{\min}}-1$, where $r^t_{\max}$ and $r^t_{\min}$ are the minimum and maximum rating of product i_t. The entire range of ratings of i_t are mapped to the range -1 to 1, which distinguishes the influence of positive or negative opinion towards the item. The reason to convert the ratings to opinions is that ratings in real-world dataset are always positive. It cannot model the damage of negative opinions.

Given the network represented by the weighted adjacency matrix, and the opinions converted from ratings, I randomly choose three items from each data set to perform the experiments. For each experiment, there are two parties in the market: a first mover and a second mover. There are five strategies for each party to take, *i.e.*, for an item in a network, there are 25 kinds of setting to simulate the possible result in the competitive market. The experiments were conducted on a Linux server with 8GB of RAM at the Turing cluster.

Table 5.2: Summary of compared methods.

Method	**Competition**	**Objective**	**Publication**
Random	No	None	[22]
IM	No	Influence	[22]
OM	No	Opinion	[31, 34]
ICIM	Yes	Influence	this thesis
ICOM	Yes	Opinion	this thesis

5.3 Compared Methods

In order to demonstrate the effectiveness of the iteratively inferred competitive opinion maximization approach, I test with following methods.

• **ICOM**: ICOM is the proposed method based on CLT diffusion model.

• **ICIM**: ICIM is a degenerative version of ICOM, which neglects the preference of users on target item, aiming to get wider spread. It takes the number of infected nodes as optimal objective instead of total opinions.

• **Random**: Method Random is a baseline method that selects inactive nodes as seeds randomly.

• **IM**: Method IM is a greedy method for influence maximization problem under linear threshold model.

• **OM**: Method OM is an adapted version of IM, which also greedily selects seeds. Instead of selecting the node with largest marginal spread, it selects the one that can gain highest opinion.

All methods except Random include a simple greedy method to select the next seed with largest marginal objective, so IM, OM, ICIM are sensible to be the baseline of ICOM. I can improve the efficiency of these algorithms by adapting that common step, the greedy algorithm in the future study. However, in this thesis I do not put emphasis on that.

The comparison shows the advantage of combining opinions and competitors information into the strategy for COM problem. All of the experiments are evaluated under competitive environment, even though Method Random, IM or OM selects seeds as if it is the only promoter in the diffusion. In different settings, I use corresponding strategies to get the seed nodes of the first mover and the second mover. I then use CLT diffusion model to propagate the influence at same time and output the total opinions of each mover at the end as measurement. For models with randomness, their performances are measured by the 5-time average results.

5.4 Performances on ICOM

I first study the effectiveness of the proposed ICOM method on competitive opinion maximization. I present the total opinions achieved by the first mover in Table 5.3. The performance is grouped by the opponent's strategy. Experiments are conducted over 9 randomly chosen items from 3 different networks. The budget of seed sets is 10, while the inferring methods will only do 5 rounds inference. Performance ranks of each model within the group are also listed. We use the average rank to compare the general performance of the model on items with different popularity and in various networks. The performance of second mover are shown in Table 5.4. Similarly, the result is grouped by the first mover's strategy.

The first observation Table 5.3 and Table 5.4 is as follows: almost all the methods that explicitly exploit the opinions of the item can achieve higher total opinions than the baseline Random, and the corresponding degenerate method IM or CIM which only chases for the influence spread. These results can support the importance of considering users' opinions and seed selection jointly. Maximizing the number of activated users is not guaranteeing the total opinions gained for the item. In

Table 5.3: Results of different strategies as first movers. The budget of seed set $k = 10$, and the maximum inference round $r = 5$. The results are reported as "average opinions + (rank)".

		Ciao			Flixster			Filmtrust			Avg.
Competitor	Methods	Item C_1	Item C_2	Item C_3	Item X_1	Item X_2	Item X_3	Item T_1	Item T_2	Item T_3	Rank
Random+	Random	310.1 (5)	210.8 (5)	197.6 (5)	67.3 (3)	49.7(4)	-90.0(4)	83.9 (5)	76.6 (5)	18.3 (5)	(4.6)
	IM	583.4 (3)	372.0 (3)	355.6 (2)	28.0 (5)	18.7(5)	-51.4(3)	130.5 (4)	120.7 (4)	28.3 (4)	(3.7)
	OM	595.6 (1)	389.4 (1)	378.5 (1)	66.3 (4)	56.9(3)	15.5(1)	137.7 (3)	126.5 (3)	36.9 (3)	(2.2)
	ICIM	564.0 (4)	353.7 (4)	341.0 (4)	100.1 (2)	91.2(2)	-200.7(5)	166.9 (2)	152.9 (2)	37.9 (2)	(3)
	ICOM	586.4 (2)	382.8 (2)	352.5 (3)	117.4 (1)	121.9(1)	-51.3(2)	173.6 (1)	222.1 (1)	39.4 (1)	(1.6)
IM+	Random	76.8 (5)	51.5 (5)	48.3 (5)	163.9 (4)	132.0 (4)	-276.9 (4)	261.0 (3)	238.5 (3)	58.7 (3)	(4)
	IM	255.5 (4)	170.2 (4)	164.5 (4)	5.9 (5)	4.3 (5)	-6.7 (3)	171.1 (5)	157.1 (5)	37.2 (5)	(4.4)
	OM	287.2 (3)	299.0 (3)	248.8 (3)	185.6 (3)	157.7 (3)	15.7 (2)	178.1 (4)	164.1 (4)	46.0 (4)	(3.2)
	ICIM	542.8 (2)	369.2 (2)	353.8 (2)	211.8 (2)	162.9 (2)	-348.8 (5)	398.1 (2)	364.4 (2)	89.6 (2)	(2.3)
	ICOM	553.2 (1)	378.5 (1)	364.2 (1)	226.6 (1)	179.7 (1)	20.4 (1)	402.4 (1)	368.9 (1)	96.3 (1)	(1)
OM+	Random	75.1 (5)	48.0 (5)	46.0 (5)	140.1 (3)	112.9 (4)	-343.5 (5)	261.0 (3)	238.3 (3)	58.7 (3)	(4)
	IM	347.2 (3)	125.5 (4)	154.8 (4)	20.9 (4)	10.0 (5)	-354.4 (3)	171.0 (5)	158.1 (5)	37.4 (5)	(4.2)
	OM	254.8 (4)	174.8 (3)	169.0 (3)	3.1 (5)	50.6 (3)	3.3 (2)	174.6 (4)	160.1 (4)	41.5 (4)	(3.6)
	ICIM	565.2 (2)	370.8 (2)	353.2 (2)	207.5 (2)	160.7 (2)	-354.4 (3)	395.0 (2)	362.7 (2)	88.4 (2)	(2.1)
	ICOM	558.5 (1)	380.2 (1)	358.2 (1)	226.1 (1)	184.9 (1)	12.9 (1)	402.4 (1)	368.9 (1)	96.3 (1)	(1)
ICIM+	Random	11.6 (5)	6.2 (5)	6.4 (5)	0.8 (4)	1.3 (4)	-1.4 (4)	75.3 (3)	68.8 (3)	15.8 (3)	(4)
	IM	92.2 (3)	37.2 (4)	38.2 (4)	-1.6 (5)	-1.6 (5)	-0.8 (3)	1.4 (5)	2.3 (5)	-0.3 (5)	(4.3)
	OM	78.2 (4)	55.8 (3)	60.2 (3)	5.6 (3)	10.1 (3)	15.7 (1)	8.6 (4)	8.3 (4)	8.3 (4)	(3.2)
	ICIM	306.0 (1)	210.8 (2)	196.0 (1)	80.4 (1)	59.1 (2)	-132.0 (5)	369.4 (1)	338.6 (2)	78.6 (2)	(1.9)
	ICOM	300.8 (2)	222.0 (1)	191.4 (2)	73.7 (2)	70.4 (1)	14.2 (2)	231.3 (2)	338.9 (1)	84.5 (1)	(1.6)
ICOM+	Random	9.8 (5)	3.6 (5)	4.7 (5)	-2.9 (4)	-0.3 (4)	-339.6 (3)	73.7 (3)	4.2 (4)	0.0 (4)	(4.1)
	IM	83.2 (4)	35.5 (4)	26.8 (4)	-4.6 (5)	-5.1 (5)	-356.7 (4)	1.4 (5)	2.1 (5)	-0.7 (5)	(4.6)
	OM	94.5 (3)	40.8 (3)	53.2 (3)	4.2 (3)	3.9 (3)	14.7 (1)	8.6 (4)	8.1 (3)	7.9 (2)	(2.8)
	ICIM	291.6 (1)	184.3 (2)	173.1 (2)	118.6 (1)	89.1 (1)	-358.1 (5)	362.8 (1)	74.2 (2)	7.6 (3)	(2)
	ICOM	287.6 (2)	195.5 (1)	184.4 (1)	105.2 (2)	81.4 (2)	13.8 (2)	294.3 (2)	334.5 (1)	67.5 (1)	(1.6)

some cases, *e.g.*, when diffusing an item with distinct feature that makes opinions from different people in stark contrast, or an unpopular item that many people host negative opinions, the attempt to maximize the spread can even lead to a result worse than to select seed nodes randomly. The strategies like IM or ICIM will choose some nodes that have a high infecting ability, but also belong to a social circle that it and its neighbors have negative opinion towards the item. For example, for the item X_3, whatever strategy the opponent takes, ICOM outperforms ICIM and Random by exploiting the opinion to this item from every users in the network. This also explains the reason why OM outperforms IM and Random.

It can also be observed that in many situations, the performance of models considering competition have a significant improvement compared with the ones that ignore. These results support the claim that in a multi-player propagation, taking

Table 5.4: Results of different strategies as second movers. The budget of seed set $k = 10$, and the maximum inference round $r = 5$. The results are reported as "average opinions + (rank)".

		Ciao			Flixster			Filmtrust			Avg.
Competitor	Methods	Item C_1	Item C_2	Item C_3	Item X_1	Item X_2	Item X_3	Item T_1	Item T_2	Item T_3	Rank
Random+	Random	273.9 (5)	185.5 (5)	181.6 (5)	121.4 (4)	98.9 (4)	-230.1 (4)	315.3 (3)	287.1 (3)	69.0 (3)	(4)
	IM	583.4 (4)	372.0 (4)	355.6 (4)	28.0 (5)	18.7 (5)	-51.4 (3)	130.5 (5)	120.7 (5)	28.3 (5)	(4.4)
	OM	595.6 (3)	389.4 (3)	378.4 (3)	66.2 (3)	56.9 (3)	15.5 (2)	137.7 (4)	126.5 (4)	36.9 (4)	(3.2)
	ICIM	643.2 (2)	412.6 (2)	394.4 (2)	210.0 (2)	162.3 (2)	-345.1 (5)	321.9 (2)	294.7 (2)	72.3 (2)	(2.3)
	ICOM	655.8 (1)	429.4 (1)	416.3 (1)	229.2 (1)	182.5 (1)	20.6 (1)	328.8 (1)	364.2 (1)	93.0 (1)	(1)
IM+	Random	76.8 (5)	51.5 (5)	48.3 (5)	163.9 (4)	132.0 (4)	-276.9 (4)	261.0 (3)	238.5 (3)	58.7 (3)	(4)
	IM	255.5 (4)	170.2 (4)	164.5 (4)	5.9 (5)	4.3 (5)	-6.7 (3)	171.1 (5)	157.1 (5)	37.2 (5)	(4.4)
	OM	287.2 (3)	299.0 (3)	248.8 (3)	185.6 (3)	157.7 (3)	15.7 (2)	178.1 (4)	164.1 (4)	46.0 (4)	(3.2)
	ICIM	542.8 (2)	369.2 (2)	353.8 (2)	211.8 (2)	162.9 (2)	-348.8 (5)	398.1 (2)	364.4 (2)	89.6 (2)	(2.3)
	ICOM	553.2 (1)	378.5 (1)	364.2 (1)	226.6 (1)	179.7 (1)	20.4 (1)	402.4 (1)	368.9 (1)	96.3 (1)	(1)
OM+	Random	75.1 (5)	48.0 (5)	46.0 (5)	140.1 (3)	112.9 (4)	-343.5 (5)	261.0 (3)	238.3 (3)	58.7 (3)	(4)
	IM	347.2 (3)	125.5 (4)	154.8 (4)	20.9 (4)	10.0 (5)	-354.4 (3)	171.0 (5)	158.1 (5)	37.4 (5)	(4.2)
	OM	254.8 (4)	174.8 (3)	169.0 (3)	3.1 (5)	50.6 (3)	3.3 (2)	174.6 (4)	160.1 (4)	41.5 (4)	(3.6)
	ICIM	565.2 (2)	370.8 (2)	353.2 (2)	207.5 (2)	160.7 (2)	-354.4 (3)	395.0 (2)	362.7 (2)	88.4 (2)	(2.1)
	ICOM	558.5 (1)	380.2 (1)	358.2 (1)	226.1 (1)	184.9 (1)	12.9 (1)	402.4 (1)	368.9 (1)	96.3 (1)	(1)
ICIM+	Random	87.5 (4)	64.1 (3)	58.2 (4)	84.8 (2)	55.7 (3)	-124.7 (4)	230.1 (1)	210.0 (2)	49.7 (2)	(2.8)
	IM	92.2 (3)	37.2 (5)	38.2 (5)	-1.6 (5)	-1.6 (5)	-0.8 (3)	1.4 (5)	2.3 (5)	-0.3 (5)	(4.6)
	OM	78.2 (5)	55.8 (4)	60.2 (3)	5.6 (4)	10.1 (4)	15.7 (2)	8.6 (4)	8.3 (4)	8.3 (4)	(3.8)
	ICIM	299.1 (2)	193.4 (2)	188.0 (2)	101.0 (1)	83.8 (1)	-172.8 (5)	24.6 (3)	21.4 (3)	8.5 (3)	(2.4)
	ICOM	314.8 (1)	215.6 (1)	210.2 (1)	76.5 (3)	63.2 (2)	20.6 (1)	31.8 (2)	286.7 (1)	81.9 (1)	(1.4)
ICOM+	Random	73.6 (5)	49.2 (3)	62.4 (3)	90.7 (2)	42.3 (3)	-264.7 (3)	228.5 (1)	145.4 (1)	53.2 (1)	(2.4)
	IM	83.2 (4)	35.5 (5)	26.8 (5)	-4.6 (5)	-5.1 (5)	-356.7 (5)	1.4 (5)	2.1 (5)	-0.7 (5)	(4.9)
	OM	94.5 (3)	40.8 (4)	53.2 (4)	4.2 (4)	3.9 (4)	14.7 (1)	8.6 (4)	8.1 (4)	7.9 (3)	(3.4)
	ICIM	302.4 (2)	178.4 (2)	192.6 (2)	105.0 (1)	75.4 (1)	-353.6 (4)	162.3 (2)	22.7 (3)	4.4 (4)	(2.3)
	ICOM	310.3 (1)	208.4 (1)	198.3 (1)	83.6 (3)	70.9 (2)	14.3 (2)	100.7 (3)	26.7 (2)	19.4 (2)	(1.9)

opponent's strategy into consideration usually enhance competitiveness. Making decision based on the known or simulated seeds of opponents can avoid the possible failure in the competition or weaken performances brought by sharing the same customers with opponents. For example, when the opponent takes OM as the model, ICOM outperforms OM by exploiting the mechanism of competition. If both of the players ignore the competitor, greedily choosing the users that can achieve highest opinion after diffusion, they will choose the same seed users. As a result, they have to share the same activated users in the end, and it is a waste of seed budget. Thus, a market with multiple players should take competition into account and apply a more flexible and effective strategy to choose its seed users.

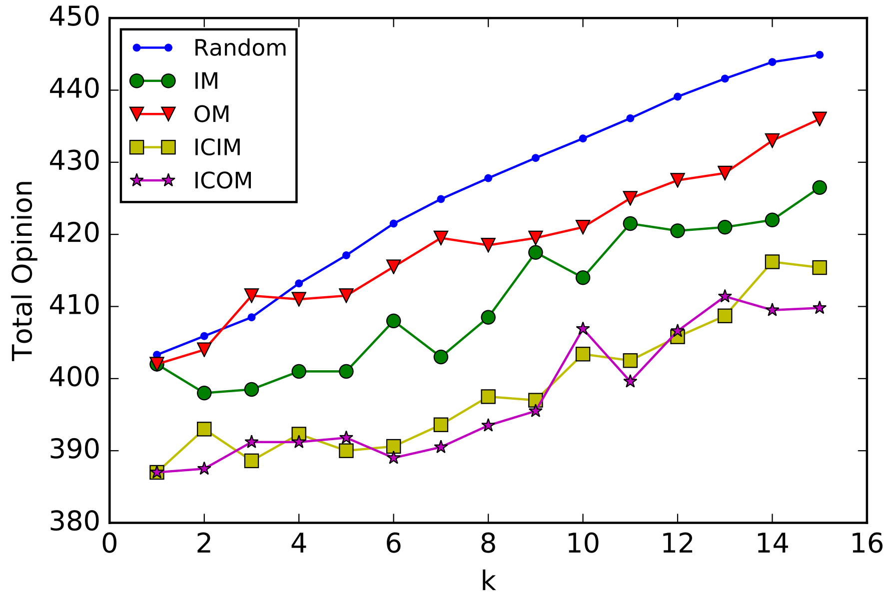
These two main observations indicate the necessity and superiority of ICOM to achieve maximum opinions in a competitive environment. The overall results showed that no matter as a first mover or a second mover and no matter what strategy the opponent takes, ICOM strategy, which exploits both opinion and competition information, is better than other baseline methods.

5.5 Discussion of Parameters

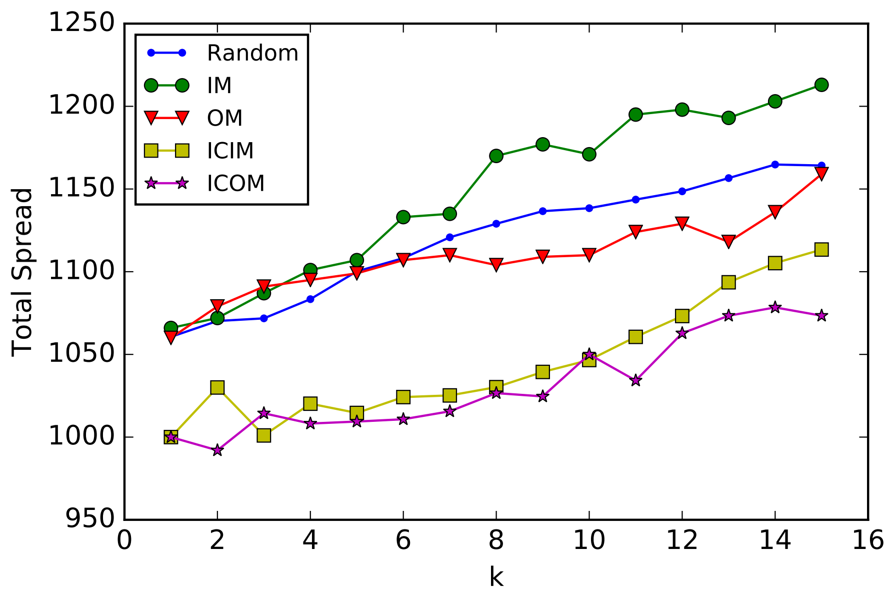
In this section, I first compare the total outcomes from both parties of 5 different setting in the market with various seed user budget k. Then I present the influence of different maximum number of round r on the performance of ICOM model.

To have a better understand in how the competitive strategy works in the market, I fix the second mover strategy and change the first mover strategy, to see how the overall market grows as the budget k grows. Figure 5.1 presents the performances from both participants in terms of opinions and influence spread. One can see that as the budget k grows, the total gain over opinions or spread got by the two parties increases, which supports the intuition that increasing the budget can improve the final performances. However, the choice of first mover model makes the outcome different, while the second mover's strategy is ICOM. As discussed in previous section, the first mover will achieve more opinions if it takes competitive strategy. But from the figure, A competitive strategy will significantly decreases the second mover's gain and the overall gain of both parties. Therefore when both parties try to take a competitive strategy, maximizing its own gain, the overall market in terms of opinions or spread would decline.

Figure 5.2 shows the convergence of the proposed ICOM by testing the performance after each iteration step. I randomly select an item from each of the dataset to run an experiment that both players are using ICOM, and each setting repeats 5 times to measure the performance. The budget k is 10 and the maximum tested



(a) Total opinions in the market

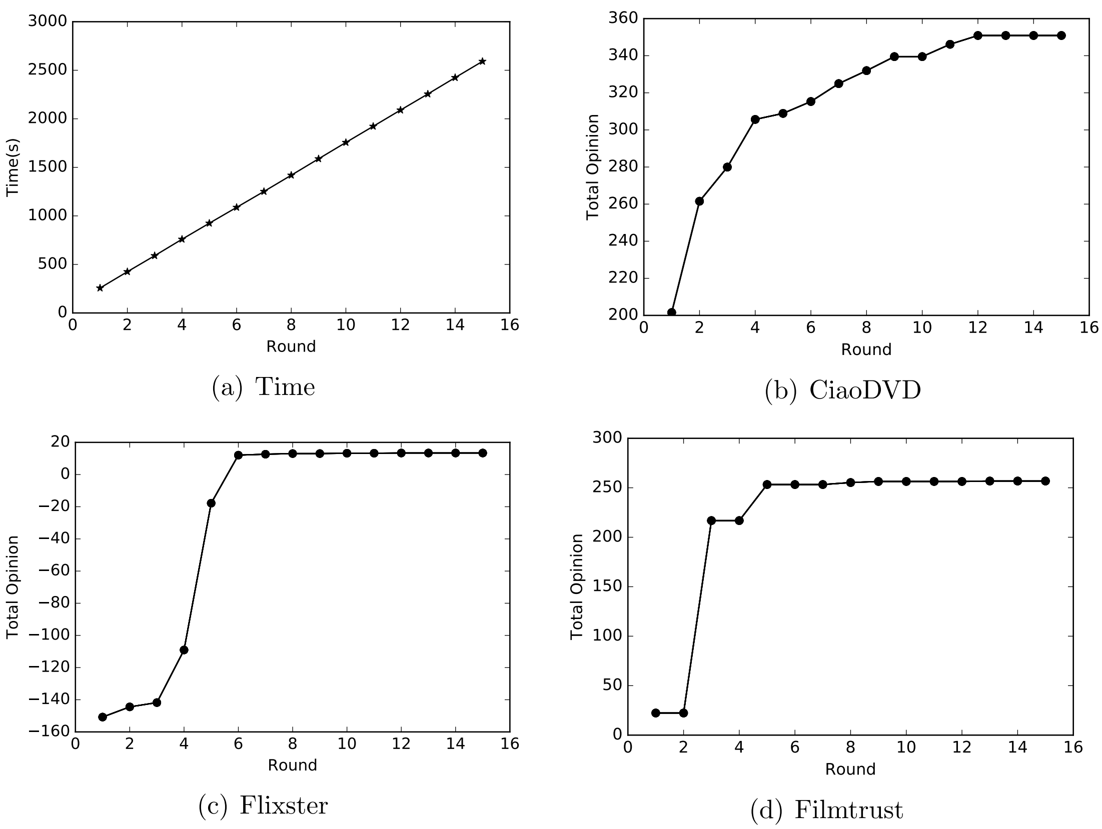


(b) Total spread in the market

Figure 5.1: Performances of two parties in Ciao network when the second mover uses ICOM and the first mover chooses different strategies ($r = 5$).

round r is 15. In Figure 5.2(a) I demonstrate the relationship between r and the cost of time in the experiment of Ciao dataset, while other two figures are similar and omitted. It indicates that the time cost of ICOM linearly grows along with the parameter r.

Figure 5.2(b) to 5.2(d) shows the total opinions achieved along with different r. It shows that the iterative inference of ICOM converges after six iterations in Flixster dataset and five iterations in Filmtrust dataset. For Ciao dataset, the performance slow down the growth after five rounds and converges after 12 iterations. From these results, we can see that the performance of ICOM converges very fast after the first few iterations, and is not sensitive to the maximum number of rounds as long as r is assigned with a modest number. Thus in previous experiment, we use 5 as the default maximum number of rounds. This also supports our intuition that using inference makes the cost of time controllable and exploiting competitor's selection is important and effective for competitive opinion maximization.



(a) Time
(b) CiaoDVD
(c) Flixster
(d) Filmtrust

Figure 5.2: Influence of parameter round r

Chapter 6

Conclusion

In this thesis, I study the competitive opinion maximization (COM) problem and defines a competitive propagation model based on LT model, where different parties compete to activate nodes. Previous work has either applied competitive setting or taken expected opinions as the objective function, but this is the first attempt to conclude both setting.

To solve this problem, I first introduce a simple greedy method and then propose an adapted method ICOM to improve the efficiency. ICOM estimates the users' opinions towards the target item and optimizes the seed selection collectively by exploiting the information from competitors. It also utilizes the iterative inferences to improve the performance of opinions and reduce time complexity when competitors' seed selections are unknown. A brief theoretical analysis shows it is able to reduce the time complexity significantly comparing to the simple greedy method.

Based on three real-world social networks, the experimental results show that the strategies consider opinions outperforms those do not, and taking competition into consideration leads to a better performance than fixed seed selection. The results validate the effectiveness and efficiency of the proposed model ICOM, which is a smart marketing strategy combining both competition and opinion information.

As mentioned in the thesis, there are some state-of-art methods that are able to improve the efficiency of algorithm for influence maximization. Therefore it could be interesting to apply these improvement to the proposed iterative framework. Another future work direction could be generalizing the COM problem under different diffusion model.

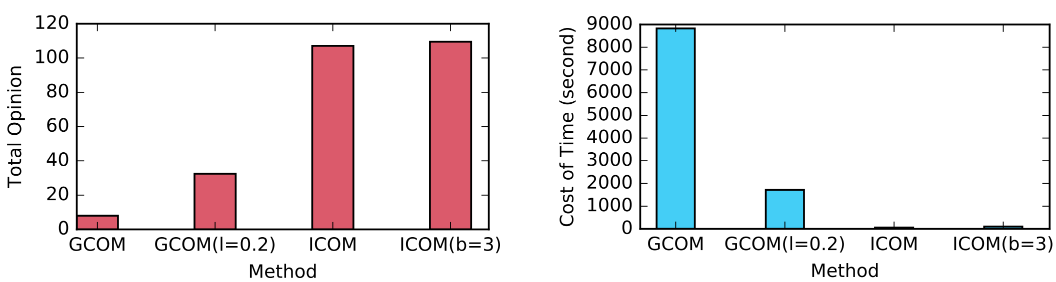
Appendix A

Algorithm Experiment

A.1 Space limitation and Beam Search

I run experiments on a sampled small dataset to validate the conclusion we get in section 4.4. Figure A.1 shows that ICOMs significantly outperform GCOMs and cost less time. Besides the basic GCOM and ICOM, I try another two simple settings for GCOM and ICOM.

First, to reduce the time complexity, I introduce a parameter l to control the search area of a seed for the first mover, where $l = 0.2$ indicates the search area is



(a) Total opinions achieved

(b) Cost of time

Figure A.1: Performances for (adapted) GCOM and ICOM in a sampled 400-node Ciao network ($k = 5, r = 10$).

20% of the original one. The results show that such "pruning" setting improves the performance of GCOM, although it is still not as well as ICOM.

Then to improve the performance in terms of opinions, I try to apply beam search in ICOM. After a few preliminary experiments for parameter setting, I use beam $b = 3$ to test the effect. In this pilot test, the beam search setting slightly increases the opinions achieved, but also the cost of time. Thus such setting doesn't bring great improvement.

Bibliography

[1] Eyal Ackerman, Oren Ben-Zwi, and Guy Wolfovitz. Combinatorial model and bounds for target set selection. *Theoretical Computer Science*, 411(44-46):4017–4022, 2010.

[2] Smriti Bhagat, Amit Goyal, and Laks VS Lakshmanan. Maximizing product adoption in social networks. In *WSDM*, pages 603–612, 2012.

[3] Shishir Bharathi, David Kempe, and Mahyar Salek. Competitive influence maximization in social networks. *Internet and Network Economics*, pages 306–311, 2007.

[4] Allan Borodin, Yuval Filmus, and Joel Oren. Threshold models for competitive influence in social networks. In *WINE*, volume 6484, pages 539–550, 2010.

[5] Ceren Budak, Divyakant Agrawal, and Amr El Abbadi. Limiting the spread of misinformation in social networks. In *WWW*, pages 665–674, 2011.

[6] Tim Carnes, Chandrashekhar Nagarajan, Stefan M Wild, and Anke Van Zuylen. Maximizing influence in a competitive social network: a follower's perspective. In *ICEC*, pages 351–360, 2007.

[7] Hung-Hsuan Chen, Yan-Bin Ciou, and Shou-De Lin. Information propagation game: a tool to acquire humanplaying data for multiplayer influence maximization on social networks. In *KDD*, pages 1524–1527, 2012.

[8] Wei Chen, Alex Collins, Rachel Cummings, Te Ke, Zhenming Liu, David Rincon, Xiaorui Sun, Yajun Wang, Wei Wei, and Yifei Yuan. Influence maximization in social networks when negative opinions may emerge and propagate. In *SDM*, pages 379–390, 2011.

[9] Wei Chen, Chi Wang, and Yajun Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *KDD*, pages 1029–1038, 2010.

[10] Wei Chen, Yajun Wang, and Siyu Yang. Efficient influence maximization in social networks. In *KDD*, pages 199–208, 2009.

[11] Wei Chen, Yifei Yuan, and Li Zhang. Scalable influence maximization in social networks under the linear threshold model. In *ICDM*, pages 88–97, 2010.

[12] Chun-Ying Chiang, Liang-Hao Huang, Bo-Jr Li, Jiaojiao Wu, and Hong-Gwa Yeh. Some results on the target set selection problem. *Journal of Combinatorial Optimization*, 25(4):702–715, 2013.

[13] Pedro Domingos and Matt Richardson. Mining the network value of customers. In *KDD*, pages 57–66, 2001.

[14] Aram Galstyan, Vahe Musoyan, and Paul Cohen. Maximizing influence propagation in networks with community structure. *Physical Review E*, 79(5):056102, 2009.

[15] Aristides Gionis, Evimaria Terzi, and Panayiotis Tsaparas. Opinion maximization in social networks. In *SDM*, pages 387–395, 2013.

[16] Amit Goyal, Wei Lu, and Laks VS Lakshmanan. Celf++: optimizing the greedy algorithm for influence maximization in social networks. In *Proceedings of the 20th international conference companion on World wide web*, pages 47–48. ACM, 2011.

[17] Amit Goyal, Wei Lu, and Laks VS Lakshmanan. Simpath: An efficient algorithm for influence maximization under the linear threshold model. In *Data Mining (ICDM), 2011 IEEE 11th International Conference on*, pages 211–220. IEEE, 2011.

[18] Guibing Guo, Jie Zhang, Daniel Thalmann, and Neil Yorke-Smith. Etaf: An extended trust antecedents framework for trust prediction. In *ASONAM*, pages 540–547, 2014.

[19] Jia-Lin He, Yan Fu, and Duan-Bing Chen. A novel top-k strategy for influence maximization in complex networks with community structure. *PloS one*, 10(12):e0145283, 2015.

[20] Xinran He and David Kempe. Robust influence maximization. In *KDD*, 2016.

[21] Xinran He, Guojie Song, Wei Chen, and Qingye Jiang. Influence blocking maximization in social networks under the competitive linear threshold model. In *SDM*, pages 463–474, 2012.

[22] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *KDD*, pages 137–146, 2003.

[23] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8), 2009.

[24] Jong-Ryul Lee and Chin-Wan Chung. A query approach for influence maximization on specific users in social networks. *TKDE*, 27(2):340–353, 2015.

[25] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. Predicting positive and negative links in online social networks. In *Proceedings of the 19th international conference on World wide web*, pages 641–650. ACM, 2010.

[26] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. Signed networks in social media. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1361–1370. ACM, 2010.

[27] Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie Glance. Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 420–429. ACM, 2007.

[28] Yanhua Li, Wei Chen, Yajun Wang, and Zhi-Li Zhang. Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 657–666. ACM, 2013.

[29] Su-Chen Lin, Shou-De Lin, and Ming-Syan Chen. A learning-based framework to handle multi-round multi-party influence maximization on social networks. In *KDD*, pages 695–704, 2015.

[30] Xinyue Liu, Chara Aggarwal, Yu-Feng Li, Xiaugnan Kong, Xinyuan Sun, and Saket Sathe. Kernelized matrix factorization for collaborative filtering. In *SDM*, pages 378–386. SIAM, 2017.

[31] Xinyue Liu and Xiangnan Kong. Active opinion maximization on social networks. 2017.

[32] Flaviano Morone and Hernán A Makse. Influence maximization in complex networks through optimal percolation. *Nature*, 524(7563):65, 2015.

[33] Matthew Richardson and Pedro Domingos. Mining knowledge-sharing sites for viral marketing. In *KDD*, pages 61–70, 2002.

[34] Huiyuan Zhang, Thang N Dinh, and My T Thai. Maximizing the spread of positive influence in online social networks. In *ICDCS*, pages 317–326, 2013.

[35] Jiawei Zhang, Senzhang Wang, Qianyi Zhan, and S Yu Philip. Intertwined viral marketing in social networks. In *ASONAM*, pages 239–246, 2016.