GetReal: Towards Realistic Selection of Influence Maximization Strategies in Competitive Networks

Hui Li
School of Cyber Engineering
Xidian University, China
hli@xidian.edu.cn

Sourav S Bhowmick
School of Computer Engineering
Nanyang Technological University, Singapore
assourav@ntu.edu.sg

Jiangtao Cui
School of Cyber Engineering
Xidian University, China
cuijt@xidian.edu.cn

Yunjun Gao
College of Computer Science
Zhejiang University, China
gaoyj@cs.zju.edu.cn

Jianfeng Ma
School of Cyber Engineering
Xidian University, China
jfma@mail.xidian.edu.cn

ABSTRACT

State-of-the-art classical *influence maximization* (IM) techniques are "competition-unaware" as they assume that a group (company) finds seeds (users) in a network independent of other groups who are also simultaneously interested in finding such seeds in the same network. However, in reality several groups often compete for the same market (*e.g.,* *Samsung*, *HTC*, and *Apple* for the smart phone market) and hence may attempt to select seeds in the same network. This has led to increasing body of research in devising IM techniques for *competitive* networks. Despite the considerable progress made by these efforts toward finding seeds in a more realistic settings, unfortunately, they still make several unrealistic assumptions (*e.g.,* a new company being aware of a rival's strategy, *alternate* seed selection, etc.) making their deployment impractical in real-world networks. In this paper, we propose a novel framework based on game theory to provide a more realistic solution to the IM problem in competitive networks by jettisoning these unrealistic assumptions. Specifically, we seek to find the "best" IM *strategy* (an algorithm or a mixture of algorithms) a group should adopt in the presence of rivals so that it can maximize its influence. As each group adopts some strategy, we model the problem as a game with each group as competitors and the expected influences under the strategies as payoffs. We propose a novel algorithm called GetReal to find each group's best solution by leveraging the competition between different groups. Specifically, it seeks to find whether there exist a *Nash Equilibrium* (NE) in a game, which guarantees that there exist an "optimal" strategy for each group. Our experimental study on real-world networks demonstrates the superiority of our solution in a more realistic environment.

SIGMOD'15, May 31–June 4, 2015, Melbourne, Victoria, Australia.

http://dx.doi.org/10.1145/2723372.2723710.

Categories and Subject Descriptors

F.2.2 [**Analysis of Algorithms and Problem Complexity**]: Non-numerical Algorithms and Problems

Keywords

Influence Maximization; competitive network; game theory; nash equilibrium; pure and mixed strategies

1. INTRODUCTION

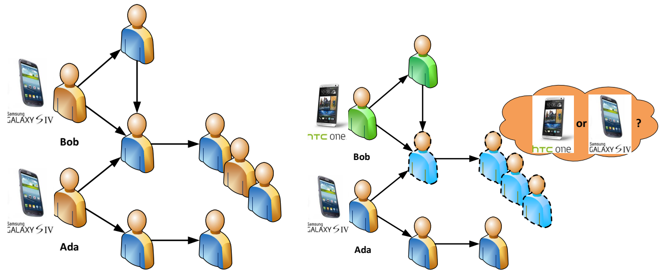
Given a social network as well as an *influence propagation* (or *cascade*) model, the problem of *influence maximization* (IM) is to find a set of initial users of size k (referred to as *seeds*) so that they eventually influence the largest number of individuals (referred to as *influence spread*) in the network [18]. Kempe et al. [18] proved that this problem is NP-hard and presented a greedy approximate algorithm and guarantee that the influence spread is within $(1 - 1/e)$ of the optimal influence spread. Since then a long stream of greedy and heuristic-based techniques [8, 12, 13, 19–21] have been proposed to improve efficiency and scalability of the IM problem.

Despite the significant progress made by state-of-the-art IM approaches, they inevitably suffer from a key drawback primarily due to the unrealistic assumption of the social network to be *non-competitive* in nature. Specifically, these techniques are effective for a network where only *one* company (*i.e.,* group)[1] is maximizing the influence of a product irrespective of the absence or presence of other competitors. Unfortunately, this scenario rarely occurs in the real world as competition for market between rivals is ubiquitous. Given a market of a particular product category (*e.g.,* mobile phone market), there exist several rival groups (*e.g.,* *Apple*, *Samsung*, *HTC*) attempting to maximize their influences[2]. A common concern of these groups is the scenario where their rivals may take precedence in the market.

[1] In the sequel, we shall use the terms "company" and "group" interchangeably.

[2] In this paper, we do not consider a market which is heavily monopolized by a certain group where users have no other choice but to buy their product(s). For such monopoly market the most important problem is not IM but price setting.

As a result, competitive new products from rival companies are often promoted during the same period[3]. For example, *HTC* and *Samsung* promoted *HTC One* and *Samsung Galaxy S4*, respectively, almost simultaneously. Consequently, these companies also desire to maximize the influence of their products in the market at the same time. Unfortunately, the aforementioned IM techniques generate unsatisfactory results when deployed over such *competitive* networks (detailed in Section 6). Several groups in the competition may adopt the same or similar IM algorithms, leading to poor estimations of influence spreads.



(a) Non-competitive (b) Competitive

Figure 1: [Best viewed in color] Influence propagation in different networks.

1.1 Motivating Example

Consider the social network in Figure 1(a). In order to maximize the influence of a new mobile phone model, assume that *Samsung* adopts a state-of-the-art IM algorithm to select two users ($k = 2$) whom it wishes to provide free samples of the phone. Suppose *Samsung* decided to choose *Ada* and *Bob* as candidate seeds based on the output of the algorithm. These users eventually influence nine persons in the network. Now consider Figure 1(b). Assume that *Samsung* has a rival in the mobile phone market, namely *HTC*. Similar to *Samsung*, *HTC* also wants to maximize the influence of their new mobile phone which is released almost simultaneously. Thus, *HTC* may also adopt a state-of-the-art IM algorithm to select the candidate seeds. It is indeed possible for *HTC* to select the same candidate seeds as *Samsung* (*Ada* and *Bob*) based on the output of the chosen IM algorithm. However, in reality, *Ada* or *Bob* may adopt *only one* mobile phone model (which they prefer) to recommend to their friends. For instance, *Ada* may adopt *Samsung* whereas *Bob* may adopt *HTC*. Note that users who are influenced by both *Ada* and *Bob* (persons in brown color in Figure 1(b)), will rarely buy both phones. Instead, they may need to make a choice between these two phones. Consequently, in reality *both companies may not eventually influence a large number of users as expected*, especially when both of them adopt IM algorithms that recommend similar candidate seeds. Clearly, this will lead to inferior influence spread quality.

Now consider the case where we assume that *Samsung* and *HTC* adopt different IM algorithms. In this case, it is possible to have lesser degree of overlap between the candidate seed sets. However, existing IM techniques still estimate the seed set of one group *disregarding* the impact of the influence spread of the other group. *In this paper, we propose a novel framework to study the* IM *problem under competitive networks where more than one competing groups want to maximize their influence.*

1.2 IM Research in Competitive Networks

Due to the aforementioned limitations, there have been increasing research efforts to address the IM problem in competitive networks [1–3, 5, 15, 29, 31] (detailed in Section 2). Although these efforts assume the existence of competing groups in a network, they suffer from two key limitations that may impede their adoption in real-world settings. First, they assume that a company A, who has already selected the seeds, is unaware of the existence of their new competitor B (*e.g., Samsung* is unaware of *HTC*). However, this is highly unlikely in reality as a company is typically aware of their potential rivals. Note that due to this unrealistic assumption, a company A may fail to adapt their IM *strategy* to one that is suitable for a competitive network (in the presence of the rival B) in order to maximize influence. In this paper, an IM *strategy* refers to an IM algorithm or a mixture of IM algorithms that a group adopts to maximize their influence. Second, they assume that the new company B is aware of its rival A's strategy and selected seeds (target users) to whom free samples of a product have been provided. Clearly, this is highly unlikely to happen as majority of these companies who are promoting their products in social networks (*e.g., Twitter*) do not own these networks to track rivals' strategies or seeds.

More recently, game theory is leveraged to investigate the IM problem in competitive networks [10, 11, 26, 32]. Specifically, the efforts in [10, 26] assume that each individual is able to alternate their choices at arbitrary time step, which deviates from traditional settings of the IM problem. In real applications, especially in viral marketing, whenever a user buys a product she is viewed as being influenced. Consequently, as the user is already influenced by a product, it is highly unlikely that she is going to be influenced by any other products. In [11] and [32], the authors investigate the IM problem between non-cooperative companies who select the seeds *alternately*. However, the assumption of "alternate seed selection" is unrealistic in may real-world applications as it is extremely difficult for companies to select targets alternatively (like playing chess).

1.3 Overview

In this paper, we take an important step towards reconsidering the IM problem in a competitive network under *more* realistic settings. Specifically, we jettison the aforementioned assumptions (*e.g.,* a new company being aware of a rival's strategy, competing companies may select seeds alternatively) made by state-of-the-art IM techniques. Instead, we make the following more realistic assumptions to address the IM problem in competitive networks.

- Firstly, given r different groups $\Psi = \{p_1, \dots, p_r\}$ who are maximizing their influences in a competitive network G, each group p_i may *independently* select k seeds under some strategy (algorithm). Note that different groups may or may not select seeds simultaneously.
- Secondly, we assume that each group p_i is aware of the existence of their rivals $p_1, \dots, p_{i-1}, p_{i+1}, \dots, p_r$ but is unaware of the IM strategies adopted by them.
- Thirdly, as motivated in Section 1.1, we assume that during influence propagation once a node in G is influ-

[3]For simplicity, we assume that the companies promote their own products without seeking any aid from a third-party for promotion.

enced by some group p_i, it cannot be affected by any other groups.

Based on the above assumptions, we seek to find the "best" IM strategy (strategy for brevity) for each group in Ψ either by maximizing its influence or by minimizing its rivals', or both. Specifically, we provide an answer to the following question: *which existing* IM *strategy a company should adopt in the presence of rivals so that it can maximize its influence?*

Towards this goal, we model the IM problem in a competitive network as a game with each group as competitors and the expected influences under the strategies as payoffs. Consequently, we reformulate the IM task in such networks as the *optimal* IM *strategy selection* problem in a game (detailed in Section 3). By leveraging game theory, we first show that *pure* and *mixed* strategies may achieve *Nash Equilibrium* (NE) [16] (*i.e.,* a state in which none of the groups in the competition can influence more users by changing their strategies) and then discuss how to select the "best" IM strategy by finding the existence of NE (Section 4). To realize this solution, in Section 5 we propose a novel algorithm called GetReal[4] (**G**am**E** **T**heory-based **RE**alistic M**A**ximization of Inf**L**uence) that seeks to find whether there exist a NE and returns both *pure* and *mixed* strategies for NE in a competitive network with r players and z strategies. Note that finding the existence of NE is particularly important as it guarantees that there exist an "optimal" strategy (strategy that exhibits the most expected influence in comparison with other strategies for all scenarios) for each group despite it being oblivious to their competitors' seeds. We experimentally demonstrate that GetReal enables us to obtain superior quality influence spread in a competitive network compared to state-of-the-art IM techniques. We also show that this is primarily due to the fact that it can find the best strategy for a group in a competitive network attempting to maximize its influence against its rivals. In summary, we make the following contributions in this paper.

- We reformulate the IM problem in a competitive network where several companies are competing to influence the users over the same product almost simultaneously under more realistic settings compared to the state-of-the-art.
- We conduct a series of game theoretic analysis over the combinations of IM strategies by different companies in the competition. Based on our analysis, we propose a novel and generic algorithm called GetReal to generate optimal solutions for these companies by finding the *Nash Equilibrium* (NE) under different conditions. Note that our solution ensures that GetReal is not tightly coupled to any specific influence propagation model. This enhances its generality as well as portability as it can be easily realized over several popular models such as the *independent cascade* (IC) model and the *linear threshold* (LT) model.
- By applying GetReal to real-world datasets under more realistic assumptions, we show its effectiveness and superiority over state-of-the-art methods.

[4]The name is inspired by an award-winning television programme in Singapore called *Get Rea!* (pronounced as "get real") that takes a "fresh and hard look" at real-world issues related to life and living in Singapore. Similarly, in this paper we take a fresh and hard look at the real-world assumptions made by state-of-the-art IM techniques.

Table 1: Key notations used in this paper.

Symbol	*Definition*
$G(V,E)$	A social network graph
n	number of vertices in G
m	number of edges in G
$\Psi = \{p_1, \dots, p_r\}$	r different groups competing for influence
$\Phi = \{\phi_1, \dots, \phi_z\}$	z different IM algorithms each group may choose from
ϕ^i_j	p_i adopts ϕ_j
S_i	the initial seeds selected by p_i
A^j_i	the nodes eventually activated by p_i in j-th round, thus A^0_i is the initial seeds that are eventually activated by p_i
$\sigma(\cdot)$	expected number of influenced node
$\sigma^0(\cdot)$	expected number of influenced node in singleton form, without competition
$\sigma_i(\phi_{t1}, \dots, \phi_{tr})$	the expected influence of p_i with $p_1, \dots, p_r$ adopting strategies $\phi_{t1}, \dots, \phi_{tr}$, respectively
$\Sigma(\Psi^r, \Phi^r)$	expected influence in competitive network with groups Ψ, each of which can choose a strategy from Φ
R	number of rounds of simulation
$E_{\phi_{t1} \dots \phi_{tr}}(\sigma(\cdot))$	expected $\sigma(\cdot)$ when p_i adopts ϕ_{ti}, $\forall i = 1, \dots, r$

The rest of this paper is organized as follows. We review related work in Section 2. In Section 3, we formally define a more realistic setting of the IM problem in competitive networks. We propose a game theory-driven method to solve the problem of selecting best IM strategies for groups in a competitive network in Section 4. In Section 5, we present the GetReal algorithm to identify the "best" IM strategies involving r groups and z strategies. We present the experimental results in Section 6. Finally, the last section concludes this paper. The key notations used in this paper are given in Table 1.

2. RELATED WORK

2.1 IM in Non-competitive Networks

Kempe et al. [18] proved that the IM problem is NP-hard. Hence, they proposed an approximate greedy algorithm based on the fact that if a greedy maximization algorithm of a *submodular* function f returns the result A_{greedy}, then the following holds $f(A_{greedy}) \geq (1 - 1/e) \max_{|A| \leq k} f(A)$ [24]. That is, a greedy algorithm can give near optimal solution to the problem of maximization of a submodular function. Since then a large body of greedy [8,12,18,20–22] and heuristic-based [6–9,13,17,19] IM techniques are reported in the literature to improve efficiency, scalability, and influence spread quality. However, as discussed in Section 1, these techniques are suitable for maximizing influence in a non-competitive network. In reality, competition is ubiquitous. Consequently, if all competitors in a particular product market adopt the aforementioned approaches to influence the most number of users, then the initial seeds determined by these approaches may be shared by many competitors. Given that a user may eventually be influenced by a single competitor's product among many alternative options, existing IM approaches naturally fail to influence the expected number of users in a competitive network.

2.2 IM in Competitive Networks

Carnes et al. [5] and Bharathi et al. [1] are among the first to investigate the IM problem in competitive networks.

They proved that maximizing the influence of a competing group with the prior knowledge of rival side's initial seeds is NP-hard and submodular. Thus, they proposed a pair of hill-climbing algorithms to maximize the influence with prior knowledge of rival side's choice. Borodin et al. [2] studied similar problem under the LT model and further extended the solution to general threshold model [18]. The efforts in [3,15,31] investigate the problem of limiting the spread of a given campaign, which shares similarity with the IM problem in competitive networks. More recently, [29] proposed to solve the same problem with a minimal number of seeds selected by a latter group, who is maximizing its influence, given that another group have already influenced a number of users. However, as remarked in Section 1, these approaches suffer from several limitations stemming primarily from unrealistic assumptions. This limits their deployment in competitive networks.

Clark and Poovendran [10] and Pathak et al. [26] assumed that each individual is able to alternate their choice at arbitrary time step in a competitive influence maximization environment. However, in viral marketing, whenever a user buys a product she is viewed as being influenced. Consequently, once a user is influenced, she may not be influenced by other products subsequently. Recently, Fazeli et al. [11] and Tzoumas et al. [32] have investigated the IM problem between non-cooperative companies who select seeds alternately using game theory. They assume that a pair of companies select the seeds alternately in sequence, until both companies have selected k seeds. Unfortunately, in the real world it is highly unlikely that a competitor of a company is aware (or informed) every time the latter selects a seed. Furthermore, these techniques attempt to maximize the overall influence of both companies by leveraging a third party association. However, such scenario may not happen frequently in practice as many companies tend to maximize their own influence while minimizing their competitors'. They typically adopt their own strategies without knowing their competitors'. In contrast, GetReal assumes that each individual company is unaware of the choices of their competitors.

Goyal and Kearns [14] jettisoned the aforementioned unrealistic assumptions and similar to our proposed technique assume that each group is unaware of their competitors' choices. However, this approach has the following drawbacks. Firstly, for the initial seeds which are simultaneously selected by both groups, namely p_1 and p_2, they assume that each node is infected by p_1 and p_2 with a probability proportional to the number of initial *adopters* of p_1 and p_2. However, this strategy may run into a dilemma as the number of initial adopters of p_1 and p_2 cannot be determined in reality *before* all the initial seeds for each group p_1 and p_2 are identified. In contrast, in our work a seed is activated by both groups with equal chance. Secondly, they defined a novel cascade model where the probability of a node to be activated is proportional to the fraction of active nodes in its neighborhood. However, in popular cascade models for the IM problem (*e.g.,* IC, WC) the probability of a node to be activated is affected by the exponential of the number of activated neighbors. Hence, the aforementioned technique has a limited applicability whereas, as we shall see later, our method can be applied to any cascade models. Thirdly, according to their model, the neighbors of a node v attempt to activate v at the same time. Hence, if v is not infected by its neighbors, it cannot be considered as a candidate seed. However, in IC, WC and LT models, if a node v is not affected by some neighbor, it is indeed possible to be affected by other neighbors at some step later.

Additionally, the work in [11,14,32] theoretically analyzed the existence of NE and the expected payoffs of competitors within the game without proposing any method to search for the NE. Furthermore, no simulation or experimental study has been undertaken on real-world networks to validate the approaches. In contrast, we not only analyze the existence of NE but also propose an algorithm that searches for NE given the *group space* and *strategy space*. Furthermore, we demonstrate its effectiveness with simulation on several real-world networks.

In [27], Prakash et al. theoretically showed winner takes all in the context of *epidemic propagation problem*, where a node i can be infected by a particular virus with probability β and cured with probability δ. Then when $\beta/\delta \geq \tau = \lambda_{1,A}$ (*i.e.,* the largest eigenvalue of adjacency matrix A) the epidemic survives, otherwise it dies out. IM problem is completely different as it only assumes that a node can transmit from inactive to active mode but not vice versa. Hence, influence cannot die out as long as it activates some nodes. Furthermore, the authors argued that even if two viruses both satisfy the *epidemic threshold* τ, one will dominate over the other. This is because a node can transmit from "infected by virus v_1" to "cured" and then to "infected by virus v_2". Such scenario does not happen in the IM problem setting.

3. PROBLEM DESCRIPTION

In this section, we formally define the problem of IM *strategy selection* in competitive networks. We begin by formally defining the notion of IM *strategy*. To facilitate our discussion, we focus on the *Independent Cascade* (IC) and *Weighted Cascade* (WC) models as these are the most popular cascade models [7, 8, 18–20]. However, our proposed technique is orthogonal to any specific cascade model as it provides a general solution to the proposed problem.

3.1 IM Strategy

We model a social network as directed graph $G = (V, E)$, where nodes in V represent individuals in the network and edges in E represent relationships between them. In contrast to classical IM techniques, influence diffusion in a competitive network does not happen in a "singleton" form. Instead, multiple pieces of influence diffuse simultaneously from competitive groups in the network. Each competing group aims to maximize their own influence considering the fact that some nodes may not be influenced if they have already been influenced by another group. Moreover, each group may adopt different IM *strategies*. Formally, an IM *strategy* is defined as follows.

Definition 1. *Given a network $G(V, E)$ and a cascade model C, a* ***strategy*** *is an algorithm or a mixture of algorithms that a group adopts to maximize their influence subject to a budget k. All algorithms (*e.g., *Hill-climb algorithms, Heuristic algorithms, etc.) that a group may adopt form a strategy space and is denoted by $\Phi = \{\phi_1, \ldots, \phi_z\}$. If a single algorithm is adopted, it is called a* ***pure strategy****. Otherwise, we refer to it as a* ***mixed strategy****. We denote a mixed strategy as $\phi^* = \{\rho_1\phi_1, \ldots, \rho_z\phi_z\}$ where $\sum_{i=1}^{z} \rho_i = 1$. That is, ϕ^* is a mixture of Φ, namely, each player selecting seeds using ϕ_i with a probability ρ_i.* ∎

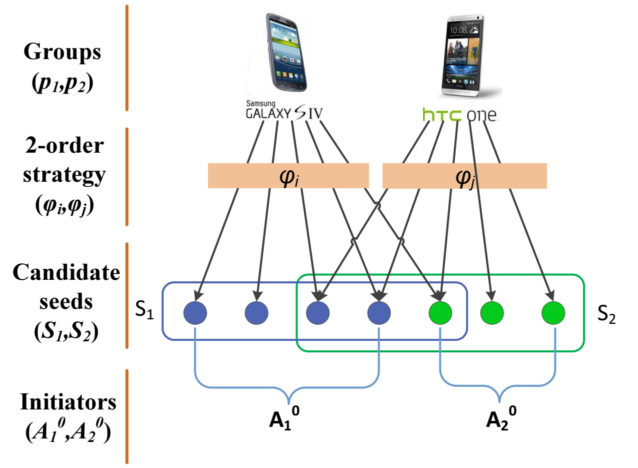


Figure 2: The seeds in a competitive network ($r = 2$,$z = 2$).

3.2 IM Strategy Selection Problem

Clearly, IM algorithms for non-competitive networks are unsuitable to provide superior solution to the IM problem in competitive networks. On the other hand, existing IM algorithms for competitive networks make several unrealistic assumptions. Hence, we revisit the IM problem in competitive networks under relatively more realistic settings and cast it to the problem of *optimal* selection of IM strategies in a competitive network. In the sequel, we refer to this reformulation of the IM problem in competitive networks as *influence maximization with competition* (IMC), which is formally defined below.

DEFINITION 2. *Let $\Psi = \{p_1, p_2, \ldots, p_r\}$ be the set of groups, each of which is maximizing their influence in a network $G(V, E)$. Let $\Phi = \{\phi_1, \ldots, \phi_z\}$ be a set of* IM *strategies that $p_i \in \Psi$ can choose from. Assume that each $v \in V$ can be activated by at most one $p_i \in \Psi$ and p_i is unaware of the strategies and seeds of other groups in Ψ. Given a specific cascade model C and a budget number k, the* ***influence maximization with competition (IMC) problem*** *is to choose $\phi_j \in \Phi$ for each $p_i \in \Psi$ that finds a seed set S_i in G where $|S_i| = k$ such that the expected number of nodes that are influenced by S_i, $\sigma(S_i)$, is the largest. That is,*

$$S_i = \underset{S_i' \subseteq V, |S_i'|=k}{\arg\max}\ \sigma(S_i').$$

■

Remark. It is worth noting that the goal of the above problem is not to propose yet another new IM technique that leverages local adoption dynamics and competition during the diffusion process to generate superior influence spreads. Instead, it *selects the best existing* IM *strategy from Φ for each competing group in Ψ in a competitive network*[5]. Since there are already an array of IM algorithms for non-competitive networks, in this paper we explore how a group can leverage them by choosing the "best" algorithm(s) (strategy) from this existing collection to maximize influence in a competitive network.

[5]Similar to [14], we assume groups have the same budget k. Our technique can be easily extended to arbitrary budgets for different groups.

Recall that a common concern of each group in the competition is that their rivals may take precedence in the market. As a result, competitive new products from rival companies are often promoted at similar times. To this end, in our model each group in a competitive network simultaneously selects the seed set and triggers off the diffusion process. Consider the simplest case where there is only one group. Then the selected seed nodes are eventually initiated by the group, which we refer to as *initiators* of the group. That is, the initiators and seeds are the same in a non-competitive IM problem. However, this may not be the case when there are multiple groups. In this case, each group may select their own seed set, which may overlap with one another. Without loss of generality, let S_i be the seeds selected by p_i and A_i^0 be the set of initiators of p_i. As depicted in Figure 2, S_1 and S_2 represent the seeds selected by p_1 and p_2, respectively. Observe that some of the seeds are selected by both p_1 and p_2. However, in reality they can only act as seeds of only one group (p_1 or p_2). Hence we can consider the nodes that eventually act as seeds of p_1 as *initiators* (denoted as A_1^0). Obviously, A_1^0 comprises of two parts: the nodes in $S_1 \setminus S_2$, which definitely act as seeds of p_1; the nodes in $S_1 \cap S_2$, which act as seeds of p_1 with *certain probabilities*.

Let $\mathcal{S} = \{S_1, S_2, \ldots S_r\}$ be the seed sets of $\Psi = \{p_1, p_2, \ldots, p_r\}$ in $G(V, E)$. A node selected by p_i can also be selected as a seed by other groups. Hence, we can use a bitmap $B = \{0, 1\}^r$ to denote whether a seed of p_i is also selected by other groups. Specifically, if a seed of p_i is also a seed of p_j, we set $b_j = 1$. Hence, all seeds of p_i can be matched to 2^{r-1} different bitmaps where $b_i = 1$. These seeds can be classified into two groups according to their bitmaps. The first group of seeds corresponds to the sequence $B = b_1 \ldots b_r$ where $b_i = 1, b_j \neq 1$ for $\forall j \neq i$. Note that these seeds are selected by p_i exclusively and definitely belong to the initiators of p_i, A_i^0. The second group of seeds corresponds to the sequences $B = b_1 \ldots b_r$ where $b_i = 1$ and $\exists j1, \ldots, js \neq i$ such that $b_{jh} = 1$ for $h = 1, \ldots, s$ (*i.e.,* these seeds are selected by not only p_i but also some other groups $p_{j1}, \ldots, p_{js}$). These seeds will be eventually initiating the influence propagation of only one group with equal probability $\frac{1}{s+1}$. Obviously, there are $2^{r-1} - 1$ bitmaps belonging to the second group. Seeds corresponding to each of these $2^{r-1} - 1$ bitmaps form a fuzzy set Q, for each seed $v \in Q$, it belongs to A_i^0 with probability $\frac{1}{|\{j|b_j^v=1\}|}$ where b_j^v is j-th bit in v's bitmap $B^v = b_1^v b_2^v \ldots b_{i-1}^v 1 b_{i+1}^v \ldots b_r^v$.

Therefore, based on the pigeonhole principle, the expected size of A_i^0 should be less than or equal to k.

The example depicted in Figure 2 illustrates the relationship between A_i^0 and S_i. Let p_1 (*e.g., Samsung*) adopts a strategy ϕ_i which returns the seeds S_1; p_2 (*e.g., HTC*) adopts a strategy ϕ_j which returns the seeds S_2 ($|S_1| = |S_2| = k = 5$). Observe that there are three nodes in $S_1 \cap S_2$, which are eventually influenced by one group, either p_1 or p_2. Let us assume that two out of these three nodes in $S_1 \cap S_2$ are eventually influenced by p_1. Hence, these two nodes along with the other two nodes $v \in S_1 \setminus S_2$ are the initiators of p_1, denoted as A_1^0.

Extension of Cascade Models. We now extend the IC and WC models to competitive networks describing how multiple influences diffuse simultaneously in a given network. Note that according to the IC model, if there are t neighbors of v that are in A^i, then $v \in A^{i+1}$ with probability

		p_2	
		ϕ_1	ϕ_2
p_1	ϕ_1	$(\lambda g, \lambda g)$	$(\alpha g, \beta h)$
	ϕ_2	$(\beta h, \alpha g)$	$(\gamma h, \gamma h)$

Table 2: Expected influence under $\Psi \times \Phi$.

$1-(1-p)^t$. Hence, IC model in competitive networks can be defined as follows.

Let A_j^i be the set of nodes that are influenced by p_j in the i-th round and $|A_j^0| \leq k$. If there are t_j neighbors of v that are in A_j^i, then $v \in A_j^{i+1}$ with probability:

$$\frac{t_j}{\sum_{j=1}^{r} t_j}(1-(1-p)^{\sum_{j=1}^{r} t_j}).$$

Similarly, in the WC model for competitive networks, v belongs to A_j^{i+1} with the following probability:

$$\frac{t_j}{\sum_{j=1}^{r} t_j}(1-(1-1/v.degree)^{\sum_{j=1}^{r} t_j}).$$

In the following section, we propose a game theoretic strategy to investigate the solutions for groups maximizing their influences in a competitive network under these two models.

4. FINDING BEST STRATEGIES

As discussed in the preceding section, the goal of IMC problem is to select the best IM strategy from Φ for each competing group in Ψ in a competitive network such that the influence of each group is maximized under a specified cascade model. To this end, in this section we propose a novel method using game theory to find the best solution for groups that are competing for influence scope with a set of rivals. We first describe the strategies that each group may adopt. As remarked earlier, each strategy is correlated to seed selection according to a specific IM algorithm or even a mixture of algorithms. With each group adopting some strategy, a game can be formed with each group as competitors and the expected influence under the strategies as payoffs. Hence, we can seek NE within the game, from which the best strategy for each group is obtained.

To simplify our discussion, we assume the competitive network involves two groups and two strategies. That is, $r = 2$ (*i.e.,* $\Psi = \{p_1, p_2\}$) and $z = 2$ (*i.e.,* $\Phi = \{\phi_1, \phi_2\}$) (*e.g.,* Figure 2). It is worth noting that although it is widely acknowledged that game between 2 or 3 competitive players is prevalent in many real markets (*e.g., Samsung* vs. *Apple* in smartphone; *AMD* vs. *Intel* in desktop CPUs; *Verizon*, *Sprint* and *AT&T* in US mobile service market) [25], the solution proposed in this paper is not limited to this setting. Particularly, it is applicable to more complex competitive networks with more than two groups and two strategies.

Let g and h be the expected influence using strategies ϕ_1 and ϕ_2, respectively. Clearly, in a competitive network with $\Psi = \{p_1, p_2\}$, the expected influence of p_1 and p_2 both adopting the same strategy ϕ_1 (*resp.,* ϕ_2) should be equal to each other. Let us denote it as λg (*resp.,* γh). In the case where p_1 adopts ϕ_1 and p_2 adopts ϕ_2, the expected influence of p_1 and p_2 can be represented as αg and βh, respectively. Similarly, in the symmetric case where p_1 adopts ϕ_2 while p_2 adopts ϕ_1, the expected influence of p_1 and p_2 are βh and αg, respectively. Thus, we have the matrix in Table 2 showing the expected influences under all combinations in $\Psi \times \Phi$.

DEFINITION 3. *Let Ψ and Φ be the group space and strategy space, respectively. Then a **r-order strategy** can be denoted as Φ^r, and an instance of r-order strategy $\Phi_j^r = (\phi_{j1}, \ldots, \phi_{jr}) \in \Phi^r$ refers to the case where p_i adopts ϕ_{ji} for $i = 1, \ldots, r$, respectively.* ∎

DEFINITION 4. *Let Ψ and Φ be the group space and strategy space, respectively. The **expected influence over all groups for a r-order strategy** Φ^r is defined as $\Sigma : \Phi^r \mapsto \mathbb{R}^r$. Let $\sigma_i(\Phi_j^r)$ be the expected influence of a r-order strategy Φ_j^r. Then the expected influence over all r-order strategies is given as follows:*

$$\begin{aligned}\Sigma(\Phi^r) \;=\; & ((\sigma_1(\phi_1, \ldots, \phi_1), \ldots, \sigma_r(\phi_1, \ldots, \phi_1)),\\ & (\sigma_1(\phi_1, \ldots, \phi_1), \ldots, \sigma_{r-1}(\phi_1, \ldots, \phi_1),\\ & \quad \sigma_r(\phi_1, \ldots, \phi_1, \phi_2)),\\ & \vdots\\ & (\sigma_1(\phi_{z-1}, \phi_z, \ldots, \phi_z), \sigma_2(\phi_{z-1}, \phi_z, \ldots, \phi_z),\\ & \quad \ldots, \sigma_r(\phi_{z-1}, \phi_z, \ldots, \phi_z)),\\ & (\sigma_1(\phi_z, \ldots, \phi_z), \ldots, \sigma_r(\phi_z, \ldots, \phi_z))).\end{aligned}$$

Obviously, the number of element $|\Sigma(\Phi)| = z^r$, each of which is a vector in $\mathbb{R}^r$ space. ∎

Table 2 shows the expected influences of 2-order strategies in a competitive network with $r = 2, z = 2$. By transferring the expected influence for 2-order strategies into such a table, the IMC problem can be reformulated as a Nash Equilibrium (NE) existence problem in the field of game theory. For the sake of completeness, we first briefly introduce the notion of Nash Equilibrium.

Nash Equilibrium (NE). In game theory, the Nash Equilibrium is a solution concept of a non-cooperative game involving two or more players, in which each player is assumed to know the equilibrium strategies of the other players, and no player has anything to gain by changing only his own strategy unilaterally [16, 23]. If each player has chosen a strategy and no player can benefit by changing strategies while the other players keep theirs unchanged, then the current set of strategy choices and the corresponding payoffs constitute a NE. Stated simply, p_1 and p_2 are in NE if p_1 is making the best decision she can, taking into account p_2's decision, and p_2 is making the best decision he can, taking into account p_1's decision. Likewise, a group of players are in NE if each one is making the best decision that he or she can, taking into account the decisions of the others. In the following, we discuss the solution to the NE from the parameters in Table 2. In particular, in contrast to [11, 14, 32], we present a generalized technique, which is applicable to *any* existing cascade model, for finding the existence of NE.

4.1 The Values of γ, λ, α and β

Solving the NE within the expected influence of r-order strategies shown in Table 2 requires detailed comparison over the adjacent entries in the table. We now discuss the entries in the expected influence of r-order strategy table. Through the discussion, we are able to show how the expected influence of strategies affect the existence of NE. In the next section, we shall propose an algorithm for solving the problem in more general case.

THEOREM 1. *Let g, h be the expected influence for strategies ϕ_1, ϕ_2 with seeds set size k, respectively, in a network without competition. The expectation for the expected influence of 2-order strategy when both groups adopt the same strategy in a competitive network (*i.e., $(\sigma_1(\phi_1,\phi_1), \sigma_2(\phi_1,\phi_1))$ *and* $(\sigma_1(\phi_2,\phi_2), \sigma_2(\phi_2,\phi_2))$*) can be represented as $(\lambda g, \lambda g)$ and $(\gamma h, \gamma h)$, respectively. Moreover, the following holds $\lambda \in [\frac{1}{2}, 1-\frac{\epsilon_1}{2g}]$ and $\gamma \in [\frac{1}{2}, 1-\frac{\epsilon_2}{2h}]$ where $\epsilon_i = E_{\phi_i}(\sigma(S_1 \cap S_2))$.*[6]

PROOF. Let $\sigma^0(S)$ be the expected number of influenced nodes with initial seeds set S in a network without competition. As S is generated from a randomized algorithm ϕ_i, we use $E_{\phi_i}(\sigma^0(S))$ to denote the expected $\sigma^0(S)$ that is computed over ϕ_i. Then $E_{\phi_1}(\sigma^0(S)) = g$ when we adopt ϕ_1 strategy. Hence, $E_{\phi_1}(\sigma^0(S_1)) = E_{\phi_1}(\sigma^0(S_2)) = g$.

On one hand, as $\sigma^0(\cdot)$ is monotonically increasing [18], the following holds: $g = E_{\phi_1}(\sigma^0(S_1)) \leq E_{\phi_1}(\sigma^0(S_1 \cup S_2))$. Also, $S_1 \cup S_2 = A_1^0 \cup A_2^0$ and $A_1^0 \cap A_2^0 = \emptyset$. Thus,

$$E_{\phi_1}(\sigma^0(S_1 \cup S_2)) = E_{\phi_1}(\sigma(S_1 \cup S_2)) = E_{\phi_1}(\sigma(A_1^0 \cup A_2^0)).$$

As there is no overlap between A_1^∞ and A_2^∞, then the following holds:

$$\begin{aligned} E_{\phi_1}(\sigma(A_1^0 \cup A_2^0)) &= E_{\phi_1}(|A_1^\infty|) + E_{\phi_1}(|A_2^\infty|) \\ &= \lambda g + \lambda g = 2\lambda g. \end{aligned}$$

Thus, $g \leq 2\lambda g$ which indicates $\lambda \geq \frac{1}{2}$.

Similarly, $h \leq 2\gamma h$ under strategy ϕ_2. Thus, $\gamma \geq \frac{1}{2}$.

On the other hand, as $\sigma^0(\cdot)$ is submodular [18][7], the following holds under strategy ϕ_1:

$$\begin{aligned} \sigma^0(S_1 \cup S_2) - \sigma^0(S_1) &= \sigma^0(S_1 \cup (S_2 \setminus S_1)) - \sigma^0(S_1) \\ &\leq \sigma^0((S_1 \cap S_2) \cup (S_2 \setminus S_1)) \\ &\quad -\sigma^0(S_1 \cap S_2). \end{aligned}$$

In fact, $E_{\phi_1}(\sigma^0(S_1 \cup S_2)) = E(|A_1^\infty|) + E(|A_2^\infty|) = 2\lambda g$ and

$$\begin{aligned} E_{\phi_1}(\sigma^0((S_1 \cap S_2) \cup (S_2 \setminus S_1))) &= E_{\phi_1}(\sigma^0(S_2)) \\ &= E_{\phi_1}(\sigma^0(S_1)) = g. \end{aligned}$$

If we represent $E_{\phi_1}(\sigma^0(S_1 \cap S_2))$ as ϵ_1, then $2\lambda g - g \leq g - \epsilon_1$ which indicates $\lambda \leq 1 - \frac{\epsilon_1}{2g}$.

Similarly, $2\gamma h - h \leq h - \epsilon$. Hence, $\gamma \leq 1 - \frac{\epsilon_2}{2h}$ where $\epsilon_2 = E_{\phi_2}(\sigma^0(S_1 \cap S_2))$. □

COROLLARY 1. *Let g, h be the expected influences for strategies ϕ_1, ϕ_2, respectively, in a network without competition. The expectation for the expected influence of 2-order strategy where both groups adopt different strategies in a competitive network (*i.e., $(\sigma_1(\phi_1,\phi_2)$, $\sigma_2(\phi_1,\phi_2))$ *and* $(\sigma_1(\phi_2,\phi_1)$, $\sigma_2(\phi_2,\phi_1))$*) can be represented as $(\alpha g, \beta h)$ and $(\beta h, \alpha g)$, respectively. Moreover, the following holds $\alpha+\beta \in [1, 1+\frac{g-\epsilon}{h}]$ where*

$$\epsilon = E_{\phi_1\phi_2}(\sigma(S_1 \cap S_2)) = E_{\phi_2\phi_1}(\sigma(S_1 \cap S_2)).$$

PROOF. Similar to the proof for Theorem 1, the following holds:

$$g + h - \epsilon \geq \alpha g + \beta h \geq g.$$

On the other hand, according to the result in [8], $g \geq h$. Thus,

$$(\alpha+\beta)h \leq \alpha g + \beta h \leq (\alpha+\beta)g.$$

Combining both sets of inequalities together, we can generate the following inequalities:

$$\begin{cases} (\alpha+\beta)h \leq g + h - \epsilon \\ (\alpha+\beta)g \geq g \end{cases}$$

Hence,

$$\begin{cases} \alpha+\beta \leq 1 + \dfrac{g-\epsilon}{h} \\ \alpha+\beta \geq 1 \end{cases}$$

□

In fact, the values of γ, λ, α and β are affected by the degree of overlap between the seeds generated from different algorithms, which is eventually decided by the topological characteristics of the network and the adopted influence cascade model. For instance, if a network always generates the same initial seeds according to a specific algorithm, the values of λ and γ are $\frac{1}{2}$. Otherwise, if a network generates different initial seeds according to the same algorithm, the values of λ and γ are close to the upper bound. Note that α and β are also affected by the topological characteristics of networks in the same way. Experimental results in Section 6 shall justify our conclusion.

4.2 Pure Strategy Nash Equilibrium

Next we show in detail the NE that can be found from the expected influence table of 2-order strategy. We begin by formally defining NE in the context of the IMC problem.

DEFINITION 5. *A r-order strategy $\Phi_j^r = (\phi_{t1}, \ldots, \phi_{tr})$ is an **equilibrium** for p_s if $\forall \phi'_{ts} \in \Phi$ and $\phi'_{ts} \neq \phi_{ts}$, the following holds: $\sigma_s(\phi_{t1}, \ldots, \phi_{t(s-1)}, \phi_{ts}, \phi_{t(s+1)}, \ldots, \phi_{tr}) \geq \sigma_s(\phi_{t1}, \ldots, \phi_{t(s-1)}, \phi'_{ts}, \phi_{t(s+1)}, \ldots, \phi_{tr})$. If a r-order strategy Φ_j^r is an equilibrium for all groups $p_i \in \Psi$, then it is a **Nash Equilibrium**[8].* ∎

Let (ϕ_i, ϕ_j) be the 2-order strategy where p_1 adopts ϕ_i while p_2 adopts ϕ_j. We iteratively examine each 2-order strategies in Table 2. If a 2-order strategy dominates the others with respect to the expected influence for both groups p_1 and p_2, it is a *pure strategy* NE.

If $\lambda g \geq \beta h$ and $\alpha g \geq \gamma h$, the 2-order strategy (ϕ_1, ϕ_1) is the NE. It can be explained as follows. In case p_2 adopts ϕ_1, ϕ_1 is superior to ϕ_2 for p_1 as the following holds

$$\sigma_1(\phi_1, \phi_1) = \lambda g \geq \beta h = \sigma_1(\phi_2, \phi_1).$$

In the other case when p_2 adopts ϕ_2, ϕ_1 is also superior to ϕ_2 for p_1, as

$$\sigma_1(\phi_1, \phi_2) = \lambda g \geq \beta h = \sigma_1(\phi_2, \phi_2).$$

Thus, no matter whichever strategy p_2 adopts, p_1 has to adopt ϕ_1. Similarly, it is easy to see that p_2 has to adopt ϕ_1 in both cases (*i.e.,* p_1 adopts ϕ_1 or ϕ_2).

[6] Note that as all greedy IM algorithms are in fact based on sampling method, the seeds selected by the same algorithm may not be exactly the same.

[7] Note that $\sigma^0(S)$ is the expected number of influenced nodes with initial seeds set S in a network without competition and therefore satisfies submodularity. Hence, we do not need to know whether $\sigma(S)$, the competitive version, is submodular or not.

[8] We only consider dominant strategy nash equilibrium. The strategy of every player (*i.e.,* p_1, p_2, ...) in this equilibrium always dominates other strategies no matter how other players choose their strategies. This is consistent with our problem settings where companies are unaware of each other's strategy and aim to maximize their own benefits while not dominated by any other groups.

Similarly, If $\lambda g \leq \beta h$ and $\alpha g \leq \gamma h$, the 2-order strategy (ϕ_2, ϕ_2) is the NE. Particularly, Nash [23] has proved that in a symmetric game where the players are commutative with each other, if there is a NE, it exists when all the players adopt the same strategy (*i.e.,* (ϕ_1, ϕ_1) or (ϕ_2, ϕ_2))[9].

4.3 Mixed Strategy Nash Equilibrium

Observe that when none of the 2-order strategy dominates all the other ones, there does not exist a NE based on pure strategy. However, there exists another type of solution, namely *mixed strategy*, where players choose a probability distribution over possible strategies. Nash [23] has proved for a game with finite set of actions (*i.e.,* strategies), at least one mixed strategy NE must exist in such a game.

DEFINITION 6. *A r-order **mixed strategy** is a r-order strategy $\Phi_j^r = (\phi_1^*, \ldots, \phi_r^*)$ where ϕ_i^* is a probabilistic strategy that p_i adopts $\phi_1, \ldots, \phi_z$ with probability $\rho_{i1}, \ldots, \rho_{iz}$ ($\sum_{j=1}^{z} \rho_{ij} = 1$), respectively.* ∎

Hence, in a network with $r = 2, z = 2$, the expectation for p_1 adopting ϕ_1 can be given as follows.

$$E[\rho\sigma_1(\phi_1, \phi_1) + (1-\rho)\sigma_1(\phi_1, \phi_2)] = \rho\lambda g + (1-\rho)\alpha g. \quad (1)$$

On the other hand, the expectation for p_1 adopting ϕ_2 is given as:

$$E[\rho\sigma_1(\phi_2, \phi_1) + (1-\rho)\sigma_1(\phi_2, \phi_2)] = \rho\beta h + (1-\rho)\gamma h. \quad (2)$$

Nash [23] proved that the mixed strategy exists when both equations equal to each other. Specifically, all the players in a symmetric game where the players are commutative with each other, will *adopt the same mixed strategy*: $\phi_1^* = \cdots = \phi_r^*$. Hence, let $\rho\lambda g + (1-\rho)\alpha g = \rho\beta h + (1-\rho)\gamma h$. Then a mixed strategy NE can be found when the following holds.

$$\rho = \frac{\gamma h - \alpha g}{\gamma h - \alpha g + \lambda g - \beta h} \quad (3)$$

That is, in a problem where pure strategy NE does not exist, the best solution for each group (*i.e.,* p_1, p_2) is to adopt a mixed strategy by adopting ϕ_1, ϕ_2 with the following probabilities, respectively.

$$\frac{\gamma h - \alpha g}{\gamma h - \alpha g + \lambda g - \beta h} \quad (4)$$

$$\frac{\lambda g - \beta h}{\gamma h - \alpha g + \lambda g - \beta h} \quad (5)$$

Next, we propose a framework to solve the IMC problem by exploiting NE in the expected influence for r-order strategy.

5. THE GETREAL ALGORITHM

In the preceding section, we have formally defined r-order strategy as well as the corresponding expected influence (Definitions 3 and 4). Using the r-order strategy, we reformulate the IMC problem as a NE existence problem. By investigating the simplest case, namely the existence of pure NE in IMC problem with $r = 2, z = 2$ (Theorem 1 and Corollary 1), we describe a method to find the pure NE in general IMC problem with $r \geq 2, z \geq 2$ in Section 4.2. In case there is no pure strategy, we additionally show how to find the mixed NE in general IMC problem with $r \geq 2, z \geq 2$ in Section 4.3. Hence we now have all the machinery in place to present the GETREAL algorithm.

Algorithm 1: The GETREAL Algorithm.

Input: network $G(V, E)$, group space $\Psi = \{p_1, \ldots, p_r\}$, seeds sets $S_1, \ldots, S_z$ generated by strategies $\phi_1, \ldots, \phi_z$, respectively; the expected influence function $\sigma_i(S_{t1}, \ldots, S_{tr})$ under a given influence cascade model.
Output: the best strategy $\phi^* = \{\rho_1\phi_1, \ldots, \rho_z\phi_z\}$ for each group, where $\sum_{i=1}^{z} \rho_i = 1$.

```
begin
    foreach ψ = (φ_t1, ..., φ_tr) ∈ ψ(Φ^r) do
        foreach p_i ∈ Ψ do
            compute σ_i(ψ) = σ_i(S_t1, ..., S_tr);
    foreach φ_i ∈ Φ do
        if ∀p_s ∈ Ψ and φ_j^s ≠ φ_i, σ_s(φ_i, ..., φ_i) ≥
        σ_s(φ_i^1, ..., φ_i^{s-1}, φ_j^s, φ_i^{s+1}, ..., φ_i^r) then
            return φ* = {φ_i} /* the pure strategy */;
    Solve the equation set
    Σ_{ℓ∈L} ρ⃗ ⊙ ℓ⃗ σ_1(φ_1, φ_{-1}(ℓ⃗)) = Σ_{ℓ∈L} ρ⃗ ⊙ ℓ⃗ σ_1(φ_2, φ_{-1}(ℓ⃗));
    = ··· = Σ_{ℓ∈L} ρ⃗ ⊙ ℓ⃗ σ_1(φ_z, φ_{-1}(ℓ⃗));
    return φ* = {ρ_1φ_1, ..., ρ_zφ_z} where (ρ_1, ..., ρ_z) = ρ⃗
    /* the mixed strategy */;
```

The main idea of the algorithm is as follows. Given a competitive network $G(V, E)$ with group space $\Psi = \{p_1, \ldots, p_r\}$ and strategy space $\Phi = \{\phi_1, \ldots, \phi_z\}$, let $f : \Psi \mapsto \Phi$ be a function that maps every element of Ψ to an element in Φ. Hence, each function f corresponds to a specific case in the game where each group adopts a specific strategy. We compute the expected influence of each group under each function. If there exist any strategy which exhibits the most expected influence under all functions, it is output as a pure strategy that is correlated to a pure NE. Otherwise, we need to find the mixed strategy by searching for mixed NE solution as follows. Let $\sigma_i(\phi_j, \phi_{-i})$ be the expected influence of p_i while p_i adopts ϕ_j and other groups adopt a combination over all the strategies, namely ϕ_{-i}. Clearly, ϕ_{-i} contains z^{r-1} cases. For instance, ϕ_{-1} contains the cases $(\phi_2^1, \ldots, \phi_r^1), (\phi_2^1, \ldots, \phi_{r-1}^1, \phi_r^2), \ldots, (\phi_2^z, \ldots, \phi_r^z)$. That is, the number of remaining $r-1$ users adopting an arbitrary strategy ϕ_i follows a distribution $L = (\ell_1, \ldots, \ell_z)$ subject to $\sum_{i=1}^{z} \ell_i = r - 1$, where ℓ_i indicates the number of players adopting ϕ_i. Let $\phi_{-i}(L)$ denote the case where players other than ψ_i adopt strategies according to the distribution L. Hence, ϕ_{-1} contains all the possible distributions of L which contains z^{r-1} cases. Generally, given that ϕ_{-1} follows a distribution $L = (\ell_1, \ldots, \ell_z)$ and for each player ϕ_i occurs with probability ρ_i, then L occurs with a probability $\prod_{i=1}^{r} \rho_i^{\ell_i}$. Formally, we denote $\prod_{i=1}^{r} \rho_i^{\ell_i}$ as $\vec{\rho} \odot \vec{\ell}$ where $\vec{\rho} = (\rho_1, \ldots, \rho_r)$ and $\vec{\ell} = (\ell_1, \ldots, \ell_r)$. We now present the GETREAL algorithm to find the best solution.

Algorithm 1 outlines the GETREAL algorithm. Firstly, for each instance of r-order strategy in Φ^r, the algorithm com-

[9] We do not consider the asymmetric nash equilibrium, where an equilibrium is formed by groups adopting different pure strategy. In this case, there is at least one group dominated by the others in any nash equilibrium, and thus that group will reject the nash. In our algorithm, we tackle this using a randomized mixed strategy NE.

Table 3: Description of real-world networks.

Network	*No. of nodes*	*No. of edges*
Phy	37,154	231,584
Hep	15,233	58,891
Wiki-talk	2,394,385	5,021,410

putes the expected influence for each group $p_i \in \Psi$ using $\sigma(\Phi^r)$ (Lines 2-4). Obviously, there are z^r instances in r-order strategy set Φ^r. For each instance, the algorithm computes the expected influence for r different groups. Next, it tests whether there exist any pure strategy NE as described in Definition 5 (Lines 5-7). Note that we have also shown in Section 4.3, if there is a NE in a symmetric game where the players are commutative with each other, it exist when all players adopt the same strategy. Hence, we examine only z instances in the r-order strategy space, namely, $(\phi_i, \ldots, \phi_i)$ for all $i = 1, \ldots, z$ (Lines 5). For each of these instances, namely $(\phi_i, \ldots, \phi_i)$, the algorithm checks whether there is an increase in the expected influence of p_s for all $s = 1, \ldots, r$ when p_s changes its strategy (Lines 6). If there is any increase, the r-order strategy $(\phi_i, \ldots, \phi_i)$ is rejected and we turn on to $(\phi_{i+1}, \ldots, \phi_{i+1})$. Otherwise, the pure strategy NE happens when each group adopts ϕ_i (Lines 7). Recall that we only aim to find symmetric pure strategy where NE happens when all groups adopt the same strategy. We do not consider the asymmetric NE where an equilibrium is formed by groups adopting different pure strategy. For this case, the best strategy is to adopt a randomized mixed strategy NE which is guaranteed to exist in finite action games [23]. Specifically, it exists when all the z expectations of the payoff (*i.e.,* expected influence) for p_s adopting each strategy equal to each other. Thus, GetReal finds the solution (*i.e.,* $\rho_1, \ldots, \rho_z$) towards the equation set with z equations (Lines 8-10) based on our discussion in Section 4.3. Note that it has been proven that finding the mixed equilibrium in a game with r players (*i.e.,* groups) and z actions (*i.e.,* strategies) is NP-complete [28]. Hence, it works effectively in games where $z, r \leq 3$, which represents many real-world scenarios [25]. Observe that GetReal can easily be extended (by replacing Lines 8-10 with a state-of-the-art *approximate* approach in game theory [4, 28]) to handle more complex networks (*i.e.,* $z, r \geq 4$) by solving NE in complex games[10].

6. PERFORMANCE STUDY

GetReal is implemented in C++. We ran all experiments on 3.2GHz Quad-Core Intel i5 machines with 8GB RAM, running Windows 7. Note that there is no existing IM algorithm for competitive networks which we can compare with GetReal. This is because, in addition to the unrealistic assumptions, the efforts in [11, 14, 32] do not provide solutions for finding NE as they focus on theoretical study of the existence of NE.

6.1 Experimental Setup

Table 3 summarizes the three real-world social network graphs used in our experiments. *Phy* and *Hep* are two academic collaboration networks[11] which are also used in several prior studies such as [7, 8, 13, 18, 20]. The *Wiki-talk*[12] is a large network containing millions of nodes representing all the users and discussions in Wikipedia from its inception to January 2008. Nodes in the network represent Wikipedia users and edges represent talk page editing relationship.

Recall that the goal of this work is to choose an optimal existing IM strategy for each group so that each group can maximize influence. Hence, we report the following experiments under different cascade models. Recall from Section 4, 2-player game is the most classic game in many real applications [25]. Hence, we focus our experimental study on 2-order strategies. Note that the qualitative results on 3-order strategies are similar and we do not report them here due to space constraints (requires 27 graphs for a single dataset).

- Firstly, we show the optimal 2-order strategy that is selected by our framework and compare it with the other two 2-order strategies under each model (IC and WC). Hence, there are six different 2-order strategies as listed below.
- Secondly, we show the impact of our realistic assumptions on the IM problem by comparing GetReal with some of the state-of-the-art algorithms designed for non-competitive networks.
- Thirdly, for the case when pure NE does not exist, we show the 2-order mixed strategy found by our framework and compare it with other 2-order strategies as well as the random strategy.

We denote the strategies involved in these experiments as follows:

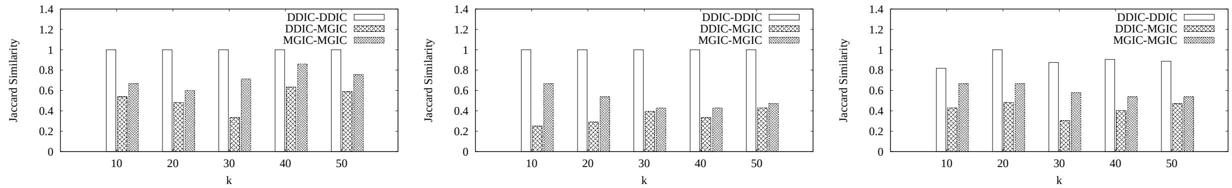
- "DDIC-MGIC": 2-order strategy (*DegreeDiscountIC*, *MixGreedyIC* [8]).
- "MGIC-MGIC": 2-order strategy (*MixGreedyIC*, *MixGreedyIC*).
- "DDIC-DDIC": 2-order strategy (*DegreeDiscountIC*, *DegreeDiscountIC*).
- "SDWC-MGWC": 2-order strategy (*SingleDiscount*, *MixGreedyWC* [8]).
- "MGWC-MGWC": 2-order strategy (*MixGreedyWC*, *MixGreedyWC*).
- "SDWC-SDWC": 2-order strategy (*SingleDiscount*, *SingleDiscount*).
- "S-MGIC": running *MixGreedyIC* in non-competitive network.
- "S-MGWC": running *MixGreedyWC* in non-competitive network.
- "*Mixed Str.*": 2-order strategy ($\phi^* = \rho\phi_1, (1-\rho)\phi_2$) with both p_1 and p_2 adopt ϕ_1, ϕ_2 with probability $\rho, 1-\rho$, respectively.
- "*Random*": both p_1 and p_2 adopt ϕ_1, ϕ_2 randomly.

Observe that we chose the popular IM techniques in [8] as representative strategies. However, our strategy space is not limited to these selected algorithms. Particularly, GetReal is orthogonal to any specific choice of IM techniques. Other IM techniques such as [7, 19] can be chosen as well.

[10] Pure NE can be found in polynomial time and is not the main challenge in such complex game.

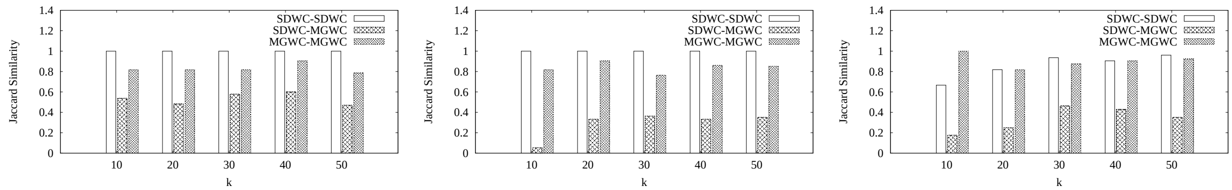
[11] Hep and Phy are downloaded from `http://research.microsoft.com/enus/people/weic/graphdata.zip`.

[12] Downloaded from `http://snap.stanford.edu/data/wiki-Talk.html`.



(a) *Hep* (b) *Phy* (c) *Wiki*

Figure 3: Overlap between S_1 and S_2 (IC model).



(a) *Hep* (b) *Phy* (c) *Wiki*

Figure 4: Overlap between S_1 and S_2 (WC model).

6.2 Experimental Results

Similarity Between the Selected Seeds Sets. We ran a series of 2-order strategies with $\phi_1 =$*MixGreedy* (IC) and $\phi_2 =$*DegreeDiscount* (IC). Before comparing the influence spread, we compare the selected seeds in the three different 2-order strategies, namely, "DDIC-MGIC", "MGIC-MGIC" and "DDIC-DDIC". We compute the jaccard similarity between S_1 and S_2 under each 2-order strategy, which is computed as follows:

$$Sim(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|}.$$

Note that it can also be computed as follows.

$$Sim(S_1, S_2) = \frac{2k - (|A_1^0| + |A_2^0|)}{|A_1^0| + |A_2^0|}.$$

Figure 3 plots $Sim(S_1, S_2)$ by varying $k = 10, \ldots, 50$ for the three datasets. Obviously, the 2-order strategies "DDIC-DDIC" and "MGIC-MGIC" exhibit higher similarity than the strategy "MGIC-DDIC". Specifically, the seeds selected by the same IM algorithms have larger overlap and is consistent with our discussion in Section 1. We also evaluated the jaccard similarities when $\phi_1 =$*MixGreedy* (WC) and $\phi_2 =$*SingleDiscount*. The results are reported in Figures 3(b), 3(c), and 4. All these cases show similar behavior as Figure 3(a).

The Pure Strategy Solutions. The aim of this set of experiments is to find the best pure strategy a group should choose, given all the possible pure strategies their competitors may choose from. Based on the seeds selected by different algorithms under three r-order strategies under IC model, we compute the expected influence for groups p_1 and p_2. Based on Algorithm 1, we find the pure strategy NE solution towards *Hep* dataset as "MGIC-MGIC". That is, the best solution for p_1 to maximize their influence is to choose *MixGreedy* (IC model) to select the seeds, no matter which strategy p_2 adopts (*MixGreedy* or *DegreeDiscount*). In fact, p_2 will also select *MixGreedy* (IC) for the same reason as p_1. Figures 5(a) and 5(b) justify our conclusion. In both figures we plot the expected influence of p_1 by fixing the strategy of p_2. The curve "MGIC" (*resp.,* "DDIC") means p_1 adopts *MixGreedy* (*resp., DegreeDiscount*). In Figure 5(a) we fix p_2 to adopt "MGIC" and vary the approaches for p_1. Clearly, "MGIC" dominates "DDIC". That is, p_1 should adopt "MGIC" when p_2 adopts "MGIC". In Figure 5(b) we fix p_2 to adopt "DDIC" and vary the approaches for p_1. Again "MGIC" dominates "DDIC". That is, p_1 should adopt "MGIC" when p_2 adopts "DDIC". Thus, p_1 should adopt "MGIC" no matter whichever strategy p_2 adopts ("MGIC" or "DDIC"). Similarly, p_2 should also select "MGIC" as p_2 and p_1 is symmetric. Likewise, we can also find pure strategy NE in the other four scenarios, namely, *Phy* and *Wiki* datasets under both IC and WC models. These are reported in Figures 6 and 7, respectively.

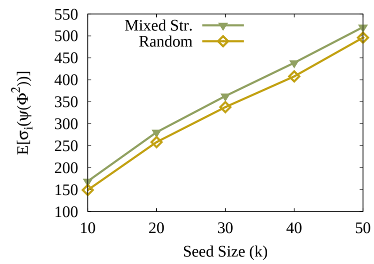
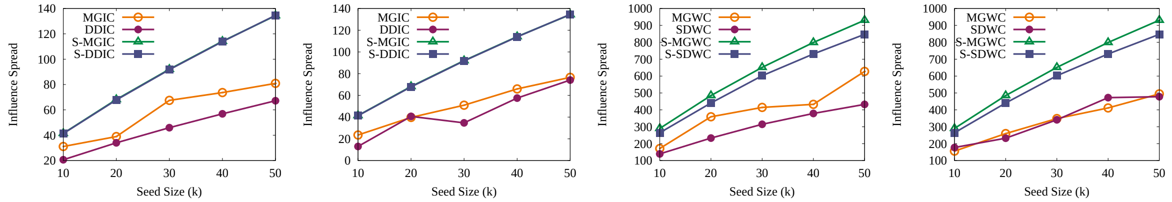


Figure 8: Comparison between mixed ($\rho = 0.582$) and random strategies (*Hep* under WC model).

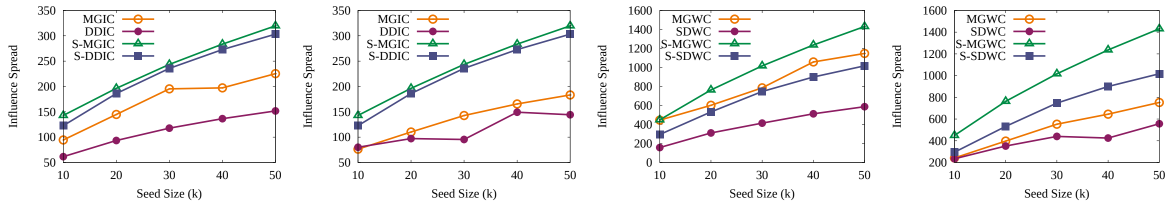
The Mixed Strategy Solutions. In all the datasets and cascade scenarios discussed above, there exist one case where we cannot find a pure strategy solution: *Hep* under WC model (Figures 5(c) and 5(d)). Hence, we search for mixed strategy (Lines 8-10 in Algorithm 1). Executing GETREAL on this scenario computes the value of ρ as 0.582. That is, the mixed strategy for p_1 (*resp., p_2*) is to adopt *MixGreedy* and *SingleDiscount* with probabilities 0.582 and 0.418, respectively.

In order to test the performance of mixed strategy NE, we simulate the influence propagation for $R = 50$ rounds by choosing each 2-order strategy pair according to ρ. Specifically, "MGWC-SDWC" and "SDWC-MGWC" are adopted with probability $0.582 \times 0.418 = 0.243$, respectively. "MGWC-MGWC" and "SDWC-SDWC" are adopted with probabilities $0.582^2 = 0.339$ and $0.418^2 = 0.175$, respectively. The his-



(a) p_2 adopts "MGIC" (IC) (b) p_2 adopts "DDIC" (IC) (c) p_2 adopts "MGWC"(WC) (d) p_2 adopts "SDWC" (WC)

Figure 5: [Best viewed in color] Influence spread in competitive network ($z = 2, r = 2$, *Hep* dataset).



(a) p_2 adopts "MGIC" (IC) (b) p_2 adopts "DDIC" (IC) (c) p_2 adopts "MGWC"(WC) (d) p_2 adopts "SDWC" (WC)

Figure 6: [Best viewed in color] Influence spread in a competitive network ($z = 2, r = 2$, *Phy* dataset).

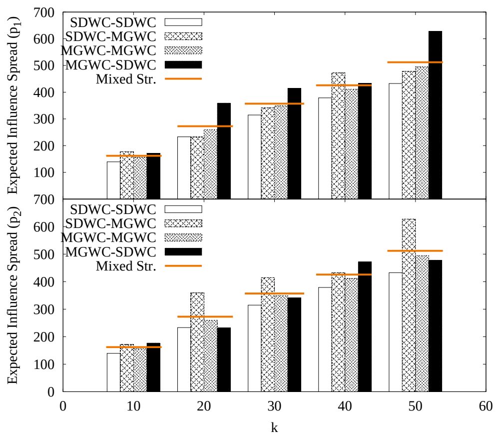


Figure 9: Average influence under different 2-order strategies (*Hep* under WC model).

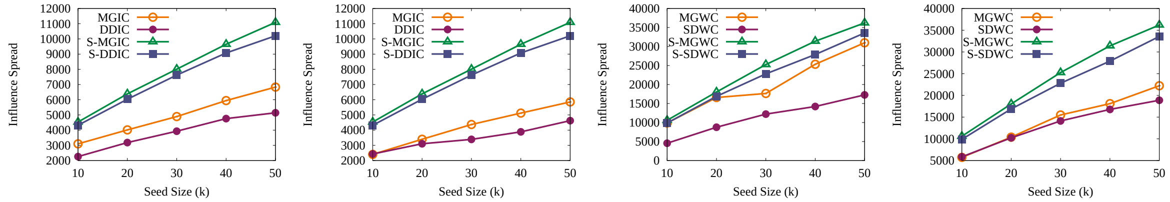
tograms in Figure 9 show the average influence spread of different 2-order pure strategies for p_1 and p_2. The horizontal lines in the figure represent the average number of eventually influenced nodes under 2-order mixed strategy with $\rho = 0.582$ over all 50 rounds for each group. Obviously, there does not exist a histogram, namely 2-order strategy, which dominates others for both p_1 and p_2. Thus, there does not exist a pure strategy NE. In this case, p_1 or p_2 has no idea which strategy they should adopt. Hence, we propose to use mixed strategy, the performance of which is shown by the horizontal lines. The optimal mixed strategy with $\rho = 0.582$ outperforms the influence spread of "MGWC-MGWC" and "SDWC-SDWC" by 20% and 9%, respectively. On the other hand, the area under each horizontal line is larger than the area covered by all the histograms for each k. In other words, the expectation of the mixed strategy at $\rho = 0.582$ is superior to the strategy which randomly select a 2-order strategy with equal probability 0.25. This result is shown in Figure 8. For both p_1 and p_2, mixed strategy is better than random selection by 7%. Observe that the degree of improvement over random selection depends on the difference between the expected influence of pure strategy, as the expected influence under mixed strategy is bounded by the minimum and maximum of that under a particular pure strategy within the strategy space. In the above case, the difference between the expected influences of p_1 adopting MGWC and SDWC is small. Hence, the expected influence of mixed strategy of p_1 is limited. In fact, regardless of the existence of pure strategy NE, randomly selecting seeds over strategy space is not the optimal solution.

Table 4: Response Times.

network	Hep		Phy		Wiki-talk	
model	IC	WC	IC	WC	IC	WC
$(r = z = 2)$	0.022s	0.034s	0.024s	0.024s	0.023s	0.03s
$(r = z = 3)$	0.043s	0.083s	0.044s	0.092s	0.098s	0.44s

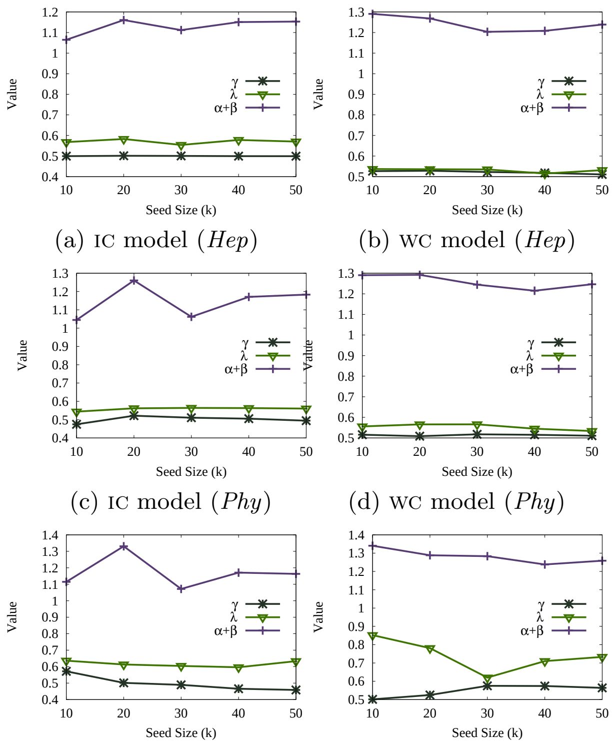
Response times. Table 4 reports the response time of finding NE (Lines 5-11 in Algorithm 1) when there are 2 (resp. 3) players and 2 (resp. 3) strategies. Observe that, given the expected influence of different strategies for each group, it is able to return the NE less in than a second.

Values of $\gamma, \lambda, \alpha, \beta$. Finally, we evaluate the values of $\gamma, \lambda, \alpha, \beta$. For both IC and WC models, we compute the values from the results generated under three 2-order strategies. Figures 10(a)-(b) plot the values for different $k = 10, \ldots, 50$ in *Hep*. In both figures, λ and γ are within the range [0.5,0.59], which supports Theorem 1. Moreover, both λ and γ are close to 0.5, which indicates that the summation over the influence of p_1 and p_2 is close to that of the influence in regular form without competition. In other words, it can be interpreted as follows. By adopting the same strategy (*i.e., MixGreedy* or *DegreeDiscount*), p_1 and p_2 select almost the same seeds (*i.e.,* $|\frac{|S^1 \cap S^2|}{|S^1 \cup S^2|} - 1| < \epsilon$). As p_1 and p_2 are symmetric in the competition, the expectation over the seeds initially activated by them is the same (*i.e.,* $E(|A_1^0|) = E(|A_2^0|)$). The summation over them equals to the total number of seeds activated, which in this case is close to k. Hence, both of $E(|A_1^0|)$ and $E(|A_2^0|)$ are close to $k/2$. In both figures, $(\alpha+\beta)$



(a) p_2 adopts "MGIC" (IC) (b) p_2 adopts "DDIC" (IC) (c) p_2 adopts "MGWC"(WC) (d) p_2 adopts "SDWC" (WC)

Figure 7: [Best viewed in color] Influence spread in a competitive network ($z = 2, r = 2$, *Wiki* dataset).



(a) IC model (*Hep*) (b) WC model (*Hep*)

(c) IC model (*Phy*) (d) WC model (*Phy*)

(e) IC model (*Wiki*) (f) WC model (*Wiki*)

Figure 10: [Best viewed in color] Values of $\gamma, \lambda, \alpha, \beta$.

values are within the range [1.08,1.16] (*resp.,* [1.2,1.29]) under IC model (*resp.,* WC model), which justifies our discussion in Corollary 1.

Observe that the values of γ, λ, α, and β do not vary significantly. However, they are different in the two models. For instance, λ ranges from 0.56 to 0.59 in the IC model. However, it is between 0.51 and 0.52 in the WC model. This phenomenon indicates that γ, λ, α, and β are unaffected by k. Instead they are influenced by the cascade models and network topology. This is consistent with our discussion in Section 4.1. The values of γ, λ, α, and β for *Wiki* and *Phy* under these two models are reported in Figures 10(c)-(f) and also show similar characteristics as above.

7. CONCLUSIONS & FUTURE WORK

State-of-the-art IM techniques for non-competitive networks can seldom be deployed in real-world applications as they ignore competitions between rival companies to maximize influence of a product. That is, they are not suitable for the case where multiple groups are maximizing their influence simultaneously in a network. Our work reported here contributes towards the goal of enhancing usefulness of IM techniques by designing a pragmatic IM solution that is competition-aware and is grounded on several realistic assumptions. In particular, we study the setting where each group is aware of the existence of their rivals but is unaware of the strategies adopted by their rivals. We propose a framework that finds the best solution for each group, who are maximizing their influences, based on game theory. To this end, we propose a novel algorithm called GetReal that seeks to find whether there exists a Nash Equilibrium (NE) (*i.e.,* whether there exists an optimal strategy that will exhibit the most influence spread in comparison with other strategies for all scenarios). Specifically, it can find either pure strategy NE or mixed strategy NE in a network with competition. Experimental results demonstrate that it can find the optimal strategy for each group no matter which strategy their rivals adopt.

Despite considering more realistic assumptions compared to the state-of-the-art techniques for the IM problem, complex social behaviors of users and groups in a competitive network may engender some special scenarios where some of these assumptions may not hold. For instance, a user may buy a product of a company p_1 and then change her mind to buy another product from a rival p_2. In this case, the assumption that once a user is influenced by some group, it cannot be affected by any other groups, does not hold. Furthermore, under certain scenarios the game may become a *biased* one. For example, two groups p_1 and p_2 may collude with each other secretly to compete with p_3. In this case, p_1 and p_2 not only know each other's strategy, but also can design a *cooperative* strategy to minimize the influence of p_3. Therefore, p_1 and p_2 can be viewed as a special company which selects $2k$ seeds in the game. It may also happen that p_1 and p_2 may explicitly form an alliance (*e.g.,* buying either product leads to customer services from both companies). In this case the influence probability for p_1 and p_2 may increase. Hence, as part of future work, we plan to investigate such *biased competition* where groups exhibit different cascade models and explore efficient methods to solve the IMC problem in a more complex competitive networks.

Acknowledgment

Hui Li, Sourav S Bhowmick, Jiangtao Cui and Jianfeng Ma are supported by National Nature Science Foundation of China (No. 61202179, 61173089, 61472298 and U1135002), National High Technology Research and Development Program (863 Program) (No. 2015AA011704), SRF for ROCS, SEM and the Fundamental Research Funds for the Central Universities.

8. REFERENCES

[1] S. Bharathi, D. Kempe, and M. Salek. Competitive influence maximization in social networks. In *WINE*, 2007.

[2] A. Borodin, Y. Filmus, and J. Oren. Threshold models for competitive influence in social networks. In *WINE*, 2010.

[3] C. Budak, D. Agrawal, and A. El Abbadi. Limiting the spread of misinformation in social networks. In *ACM WWW*, 2011.

[4] I. Caragiannis, A. Fanelli, N. Gravin, and A. Skopalik. Approximate pure nash equilibria in weighted congestion games: existence, efficient computation, and structure. In *ACM ICEC*, 2012.

[5] T. Carnes, C. Nagarajan, S. M. Wild, and A. van Zuylen. Maximizing influence in a competitive social network: a follower's perspective. In *ACM ICEC*, 2007.

[6] W. Chen, A. Collins, R. Cummings, T. Ke, Z. Liu, D. Rincón, X. Sun, Y. Wang, W. Wei, and Y. Yuan Influence maximization in social networks when negative opinions may emerge and propagate. In *SDM*, 2011.

[7] W. Chen, C. Wang, and Y. Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *KDD*, 2010.

[8] W. Chen, Y. Wang, and S. Yang. Efficient influence maximization in social networks. In *KDD*, 2009.

[9] W. Chen, Y. Yuan, and L. Zhang. Scalable influence maximization in social networks under the linear threshold model. In *ICDM*, 2010.

[10] A. Clark and R. Poovendran. Maximizing influence in competitive environments: a game-theoretic approach. In *GameSec*, 2011.

[11] A. Fazeli and A. Jadbabaie. Game theoretic analysis of a strategic model of competitive contagion and product adoption in social networks. In *CDC*, 2012.

[12] A. Goyal, F. Bonchi, and L. V. S. Lakshmanan. A data-based approach to social influence maximization. *PVLDB*, 2011.

[13] A. Goyal, W. Lu, and L. V. S. Lakshmanan. Simpath: An efficient algorithm for influence maximization under the linear threshold model. In *ICDM*, 2011.

[14] S. Goyal and M. Kearns. Competitive contagion in networks. In *STOC*, 2012.

[15] X. He, G. Song, W. Chen, and Q. Jiang. Influence blocking maximization in social networks under the competitive linear threshold model. In *SDM*, 2012.

[16] O. M. J and A. Rubinstein. A course in game theory. 1994.

[17] K. Jung, W. Heo, and W. Chen. IRIE: Scalable and Robust Influence Maximization in Social Networks. In *ICDM*, 2012.

[18] D. Kempe, J. M. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *KDD*, 2003.

[19] J. Kim, S. Kim, and H. Yu. Scalable and parallelizable processing of influence maximization for large-scale social networks. In *ICDE*, 2013.

[20] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. M. VanBriesen, and N. S. Glance. Cost-effective outbreak detection in networks. In *KDD*, 2007.

[21] H. Li, S. S. Bhowmick, and A. Sun. Cinema: conformity-aware greedy algorithm for influence maximization in online social networks. In *EDBT*, 2013.

[22] H. Ma, H. Yang, M. R. Lyu and I. King. Mining social networks using heat diffusion processes for marketing candidates selection. In *CIKM*, 2008.

[23] J. Nash. Non-cooperative games. *The Annals of Mathematics*, 54(2):286–295, Sept. 1951.

[24] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. An analysis of approximations for maximizing submodular set functions. *Math. Programming*, 14(3):265–294, 1978.

[25] J. von Neumann and O. Morgenstern. Theory of Games and Economic Behavior. *Princeton University Press*, 2007.

[26] N. Pathak, A. Banerjee, and J. Srivastava. A generalized linear threshold model for multiple cascades. In *ICDM*, 2010.

[27] B. A. Prakash, A. Beutel, R. Rosenfeld, and C. Faloutsos. Winner takes all: competing viruses or ideas on fair-play networks. In *WWW*, 2012.

[28] G. R. Schoenebeck, and S. Vadhan. The computational complexity of nash equilibria in concisely represented games. *ACM Trans. Comput. Theory*, 4(2):4:1–4:50, May 2012.

[29] S. Shirazipourazad, B. Bogard, H. Vachhani, A. Sen, and P. Horn. Influence propagation in adversarial setting: how to defeat competition with least amount of investment. In *CIKM*, 2012.

[30] Y. Tang, X. Xiao, and Y. Shi. Influence Maximization: Near-Optimal Time Complexity Meets Practical Efficiency. *In SIGMOD*, 2014,

[31] J. Tsai, T. H. Nguyen, and M. Tambe. Security games for controlling contagion. In *AAAI*. 2012.

[32] V. Tzoumas, C. Amanatidis, and E. Markakis. A game-theoretic analysis of a competitive diffusion process over social networks. In *WINE*, 2012.