

A Game-Theoretic Approach to Competitive Viral Marketing in Social Networks

Student No. 3

April 24, 2013

1 Introduction

Recently, the rapidly increasing popularity of online social networking sites such as Facebook, Twitter, and Google+ opens up great opportunities for large-scale viral marketing campaigns. *Viral marketing*, first introduced to the data mining community by Domingos and Richardson [6], is a cost-effective marketing strategy that promotes products by giving free or discounted items to a selected group of highly influential individuals (also called *seeds*), in the hope that through word-of-mouth effects, a large number of adoption will occur.

Motivated by viral marketing, *influence maximization* has emerged as a key computational problem concerning the propagation of innovations and technology adoptions through social networks. In a seminal paper, Kempe et al. [12] formulated influence maximization as follows. Given a directed graph $G = (V, E)$ representing a social network, and a positive integer k , find k users, such that by targeting them, the expected spread of influence is maximized under a certain propagation model. Here, “targeting” means giving a buyer free or discounted samples so that she adopts the product, and k represents the company’s budget, i.e., the maximum number of samples it is willing to give away.

The bulk majority of research in this field assumes that there is only one company promoting one product, and there is no competition. However, in real world, there are often multiple companies providing comparable products and fight for market share, e.g., Apple vs. Samsung in smartphones, Canon vs. Nikon in digital SLR cameras, and Sony vs. Microsoft in video game platforms. Since those consumer technologies are usually not free to adopt, it is unlikely that one person will purchase from more than one company in a short period of time. Therefore, competitions naturally exist, and recognizing this, there has been recent work proposing competitive influence propagation models and tackle influence maximization problems under such models [1, 5, 4, 11, 16, 3, 2, 13].

However, there are two missing pieces in most previous work. First, they focus on the optimizing an objective for a single company and do not fully consider strategic interactions among companies. Second, it is often assumed that a marketing company has complete knowledge of the social network graph and can directly mine influential seeds from it. This may not be the case in

reality, as a network graph is owned as critical and proprietary data by the social network operator, e.g., Facebook, Twitter, or Google, and hence may not be readily available to “seed seekers”.

In this work, we address the above issues by proposing a game-theoretic framework and study companies’ behaviours in the proposed game. We model a competitive viral marketing campaign as a strategic game, called the *the competitive viral marketing game*, hosted by the social network owner which we refer to as the *host* hereafter. The host designs a mechanism for the game such that two goals are achieved: First, *the collective spread of influence over all companies is maximized*, so that it maintains good reputation for future business. Second, *fairness is ensured*, in the sense that the more a company invest in the campaign, the higher return (utility) it can expect from the game. Each competing company is modelled as an agent participating in the game. We then provide a sufficient condition for the existence of pure-strategy Nash equilibria.

The rest of the paper is organized as follows. Section 2 introduce necessary background knowledge and discuss representative related work. Section 3 describes the competitive viral marketing game and its properties. Finally, Section 4 concludes the paper and discusses possible directions for future work.

2 Background and Related Work

2.1 Influence Maximization

Kempe et al. [12] introduce influence maximization and study it on two propagation models: *Independent Cascade* (IC) and *Linear Threshold* (LT) originated in mathematical sociology [7, 10]. In their framework, a social network is modelled as a directed graph $G = (V, E)$, with nodes representing individuals and edges representing social relationship between nodes. Each edge $(u, v) \in E$ labeled by an influence weight $p_{u,v}$, capturing the extent to which v will be influenced by u once u adopts a new technology or product. The dynamics of an influence diffusion process proceed in discrete time steps. At any given time step, each node is either active (i.e., being an adopter) or inactive. Initially, all nodes are inactive. At time 0, a set S of seeds are activated and new activations start spreading over the network in a probabilistic manner based on the propagation model. The process quiesces when no more nodes can be activated, and once activated, nodes remain active.

The *expected spread of influence* of a set S , denoted by $\sigma(S)$, is defined as the expected number of active nodes by the end of diffusion. The *influence maximization problem*, given as input $G = (V, E)$ and k , asks for a set $S \subseteq V$, $|S| \leq k$, such that $\sigma(S)$ is maximized. Under both IC and LT models, the problem is NP-hard [12]. Kempe et al., however, show that the influence function $\sigma(S)$ is *monotone* (i.e., $\sigma(S) \leq \sigma(T)$ whenever $S \subseteq T$) and *submodular* (i.e., $\sigma(S \cup \{w\}) - \sigma(S) \geq \sigma(T \cup \{w\}) - \sigma(T)$ for any $S \subseteq T$ and $w \in V \setminus T$). With those two desired properties, a simple greedy algorithm which in each iteration

Algorithm 1: GREEDY ($G = (V, E), k, \sigma$)

```
1  $S \leftarrow \emptyset$ ;  
2 for  $i = 1 \rightarrow k$  do  
3    $w \leftarrow \arg \max_{u \in V \setminus S} [\sigma(S \cup \{u\}) - \sigma(S)]$ ;  
4    $S \leftarrow S \cup \{w\}$ ;  
5 Output  $S$ ;
```

greedily extends the seed set S with the node w providing the largest marginal influence $\sigma(S \cup \{w\}) - \sigma(S)$ gives a $(1 - 1/e)$ -approximation to the optimum [14, 12] (see Algorithm 1).

2.2 Competitive Viral Marketing

There have been some recent studies on competitive viral marketing, by extending the IC or the LT model. A common theme is that they all focus on the client perspective, and some of them are restricted to just two companies.

Bharathi et al. [1] proposed a natural extension to the IC model to incorporate competitions and show that the last player to select the seed set can apply the greedy algorithm to obtain a $(1 - 1/e)$ -approximation to the optimal strategy. Carnes et al. [5] study the problem from the “follower’s perspective”. The follower is the player trying to introduce a new product into an environment where a competing product already exists, and the follower knows the early adopters of its competitors. Under their models, the influence function is monotone and submodular, and thus the greedy algorithm can be applied to provide approximation guarantees. Later, Budak et al. [4] and Chen et al. [11] studied the problem of influence blocking maximization, where one entity tries to block the influence propagation of its competitor as much as possible, under extended IC and LT models respectively. Both of their models involve only two entities (clients). Pathak et al. [16] propose an extension of the voter model to study multiple cascades. Borodin et al. [3] propose extensions to the LT model to deal with competing products.

More recently, Borodin et al. [2] and Lakshmanan et al. [13] studied competitive viral marketing from the host’s perspective: [2] designed strategy-proof mechanisms whose goal is to maximize the overall influence spread and avoid the situation where companies can achieve a higher utility by declaring a lower budget to the host. [13] focuses on models under which the overall influence spread can be approximated within a factor of $1 - 1/e$ by the aforementioned greedy algorithm, and studies the problem of fairly allocating the greedily chosen global seed set to participating companies, based on reported budgets. Interestingly, the fair allocation mechanism proposed in [13] guarantees that underreporting budgets is suboptimal for companies, but it fails to address a potential issue that companies might still behave strategically and hence undermine the entire campaign. Therefore, in this paper we attempt to close this gap by making

explicit game-theoretic connections to [13].

3 The Competitive Viral Marketing Game and Its Properties

3.1 Problem Setting

In this section, we first introduce necessary notations and definitions, and then describe the assumptions on influence propagation models applicable to our framework.

Notations. Let $N = \{1, 2, \dots, n\}$ be the set of n competing companies. We denote by k_i the budget of company i , $\forall i \in N$. For each company i , the host will allocate to it a seed set $S_i \subseteq V$ with $|S_i| = k_i$. For simplification, we assume that each targeted seed will only receive a sample from one company. Hence, all seed sets are pairwise disjoint, i.e., $S_i \cap S_j = \emptyset$, whenever $i \neq j$. This can be justified, when, for example, the host does not want its users to be disturbed too much by receiving multiple samples from different advertisers.

We call $S = S_1 \cup S_2 \cup \dots \cup S_n$ – the union of all seed sets – the *global* seed set. On the other hand, reversely, given any set $S \subseteq V$ such that $|S| = \sum_i k_i$, we say that (S_1, S_2, \dots, S_n) is a *n-partition* of S if and only if S_i and S_j are disjoint, $\forall i \neq j, i \in N, j \in N$, and $|S_i| = k_i, \forall i \in N$. Also, we denote by S_{-i} the vector comprising all individual seed sets excluding S_i , and k_{-i} the budget profile comprising the budgets of all companies excluding k_i . Finally, we use $\sigma(S)$ to represent the expected collective spread of influence over all companies when S is the global seed set, and $\sigma(S_i, S_{-i})$ to denote the expected influence spread resulted for company i when S_i is allocated to i .

Model Assumptions. Notice that our framework does not necessarily assume a particular competitive influence propagation model to work with, but rather we make the following assumptions in terms of the properties that we expect any applicable model to have.

1. The collective influence spread is equal to the sum of influence spread of all companies'. That is, given k_1, k_2, \dots, k_n , for any $S \subseteq V$, $|S| = \sum_i k_i$, and any n -partition of S , it holds that $\sigma(S) = \sum_i \sigma_i(S_i, S_{-i})$.
2. The collective influence function $\sigma(S)$ is monotone and submodular w.r.t. the global seed set S .
3. The individual influence function $\sigma_i(S_i, S_{-i})$ is monotone and submodular w.r.t. i 's own seed set S_i , $\forall i \in N$.
4. Given a global seed set S and an arbitrary $S_i \subseteq S$ with $|S_i| = k_i$ the influence spread for company i is independent of how other seeds get allocated. That is, $\sigma_i(S_i, S_{-i})$ remains the same regardless of the exact shape of S_{-i} provided that S_i is held constant.

As long as a propagation model satisfies all the above properties, our analysis will apply. Examples of such models include the K -LT model proposed by Lakshmanan et al. [13] and the OR model proposed by Borodin et al. [3]. We argue that the properties are abstract but reasonably natural. Besides, they are desired partially due to technical concerns such as computational tractability. We omit the detailed model definitions and justifications (which can be found in [13] and [3]), since it is not the focus of this work.

3.2 Game Formulation

The competitive viral marketing game $\mathcal{G} = \langle N, A, u \rangle$ is formulated as follows. The set of agents is exactly the set of all participating companies, and hereafter we use agent and company interchangeably to refer to an competing entity. For each agent i , its set A_i of available actions consists of all budget numbers i can choose to report to the host. Since the social network has $|V|$ nodes in total, $A_i = \{0, 1, 2, \dots, |V|\}$, $\forall i \in N$. Thus, $A = A_1 \times A_2 \times \dots \times A_n$. Notice that some action profile in A will render the seed selection and partition impossible. For example, suppose all declared budgets are greater than $|V|/n$; this leads to $\sum_i k_i \geq |V|$, making it impossible for the host to proceed. As we shall see later, due to (1) the mechanism designed for this game, (2) expected influence spread is typically low in real-world social networks (compared to the size of the network), and (3) agent's budgetary concerns, the above situation can be avoided in the equilibrium of the game.

Recall that the host has a two-fold goal: influence maximization and fair allocation. First, given budgets reported by agents, it wants to maximize the collective influence spread over all agents, so that it fetches as much revenue as possible and in the mean time maintains good reputations for future business. Second, it wants to ensure that the allocation of seeds is as fair as possible, such that agents invest more in the campaign (reporting higher budgets) will receive higher expected return than those who invest less.

Lakshmanan et al. [13] proposed a seed allocation scheme that not only guarantees a $(1 - 1/e)$ -approximation to the influence maximization objective but also optimizes the fairness objective. First, each agent i reports its budget k_i to the host. An agent's budget is privately held information and is not known to any other agent. Then, the host selects a total number of $\sum_i k_i$ seeds from the network, which forms the global seed set S . It does so by running the greedy approximation algorithm (cf. Algorithm 1), and due to submodularity of $\sigma(S)$, the $(1 - 1/e)$ approximation can be achieved for the influence maximization objective. After that, the host partitions S in such a way that the influence spread for each agent is proportional to their reported budgets. In other words, $\sigma_i(S_i, S_{-i})$ is as close to $\frac{k_i}{\sum_j k_j} \sigma(S)$ as possible.

As mentioned in Section 2, [13] ignores game-theoretic implications in this fair allocation process. In particular, it is not clear that how agents should come up with their budget demand, i.e., what dictates their choices? Also, although Borodin et al. [2] also formulates a game-theoretic framework, too, they assume

that each agent will magically have a “truthful” budget demand in mind. Notice that agents are rational and self-interested, seeking to maximize utility (profit, if we speak of companies), and hence an agent’s demand for seeds should be a strategic move made based on their knowledge of the game.

Therefore, in formulating our game, we adopt the fair allocation mechanism in [13]. Next, to complete the game, we define utility functions $u = (u_1, u_2, \dots, u_n)$. In a nutshell, an agent’s utility consists of two parts: a gain function $g_i(\cdot)$ representing its expected profit from the campaign and a loss function $\ell_i(\cdot)$ representing its costs by participating.

We assume that each agent i sets a fixed price and thus fetches a fixed profit p_i from each adopted customer in the network. Therefore, we have

$$\begin{aligned} g_i(k_i, S_i, S_{-i}) &= p_i \cdot [\sigma_i(S_i, S_{-i}) - |S_i|] \\ &= \frac{p_i k_i}{\sum_j k_j} \sigma(S_i \cup S_{-i}) - p_i k_i. \end{aligned} \quad (1)$$

Notice that the host is known a priori to run the greedy algorithm to select the global seed set S , and thus the collective influence spread can be re-written as a function of the total budget, instead of a set function. That is, given a budget profile (k_1, k_2, \dots, k_n) , let $\sigma^g(k_i, k_{-i})$ be the expected total spread achieved by a greedily chosen set with $\sum_i k_i$ elements. Similarly, $\sigma_i^g(k_i, k_{-i})$ is defined to be $\frac{k_i}{\sum_j k_j} \sigma^g(k_i, k_{-i})$. Hence, we can re-write agent i ’s gain function (1) as a function of the budget profile (k_i, k_{-i}) :

$$g_i(k_i, k_{-i}) = p_i \cdot [\sigma_i^g(k_i, k_{-i}) - k_i] = \frac{p_i k_i}{K} \sigma^g(k_i, k_{-i}) - p_i k_i, \quad (2)$$

where we have used K as the shorthand for $\sum_j k_j$. It can be easily seen that i ’s profit depends on not only its own budget, but also all agents’.

Agent i ’s loss function consists of two types of costs, both are charges by the host as its business model. More specifically, the host charges α for each seed requested and β for per one expected adopter. Both α and β are assumed constant and the same for all agents. The per-seed charge α is easily justified, since the host is providing services to agents. On the other hand, the per-adopter charge β can be thought of as a surcharge preventing agents to abuse the system (mechanism) by reporting arbitrarily high demand for seeds. Thus, we can write the loss function as

$$\ell_i(k_i, k_{-i}) = \alpha k_i + \beta \cdot [\sigma_i^g(k_i, k_{-i}) - k_i] \quad (3)$$

The utility of agent i is simply the different between its gain and loss, and thus combining (2) and (3), we have:

$$\begin{aligned} u_i(k_i, k_{-i}) &= g_i(k_i, k_{-i}) - \ell_i(k_i, k_{-i}) \\ &= (p_i - \beta) \cdot \sigma_i^g(k_i, k_{-i}) - (p_i + \alpha - \beta) \cdot k_i. \end{aligned} \quad (4)$$

We also assume that $p_i + \alpha > \beta$, which simply says that the product is reasonably profitable, meaning p_i is at least higher than $\beta - \alpha$ (which can be negative if the host is not too greedy!).

3.3 Nash Equilibria of the Game

As the competitive viral marketing game \mathcal{G} is finite (both N and A are finite), mixed-strategy Nash equilibria are guaranteed to exist [15]. However, in the context of a marketing campaign, pure-strategy Nash equilibria are more interesting as agents will need to submit deterministic budgets to the host.

There are many works in the game theory literature studying the existence of pure-strategy Nash equilibria. In general, determining whether a game has a pure-strategy Nash equilibrium is NP-hard [8]. One way to show that pure-strategy Nash equilibria exist is to use fixed point theorems as sufficient conditions, e.g., the Kakutani’s fixed point theorem [15]. However, this approach does not apply here as the actions sets are finite, and hence not convex in our case. Recently, Gottlob et al. [8] showed that if a game has “small neighborhood” and its corresponding “dependency hypergraph” has bounded treewidth, then finding pure Nash equilibria for this game is feasible in polynomial time. Small neighborhood means that for each agent i , the number of i ’s “neighbors” (other agents whose actions will affect i ’s utility) is logarithmic w.r.t. $|N|$. The dependency hypergraph is a graph with agents as nodes, and its hyperedges are drawn based on neighboring relationships of agents. Clearly, their result cannot be used in our case, either, as the competitive viral marketing game does not satisfy the small neighborhood condition – in fact, all agents affect each other’s utility in the game.

In view of the above, we explore the existence of pure-strategy Nash equilibria by directly observing the structure and properties of the game. Here we provide a sufficient condition for the existence of pure Nash equilibrium for the case of $|N| = 2$. We can think of the strategic moves of the two agents as a “line-sweeping” process where they increment their budgets by one at each step. For example, initially, they start with $k_1 = k_2 = 0$. Then, they simultaneously increment k_1 and k_2 by one at a time, until one agent sees its utility starting to decrease, which we say the agent “hits” its “critical point”. In fact, the line-sweeping process can start with any arbitrary small positive integers, and they may lead to different critical points and hence a different equilibrium.

Definition 1 (CRITICAL POINT). *Agent i is said to reach its critical point k_i^* in the line-sweeping process with an arbitrary but fixed start point if and only if*

1. $u_i(k_i^*, k_{-i}) > u_i(k_i^* + 1, k_{-i})$ and $u_i(k_i^*, k_{-i}) > u_i(k_i^* - 1, k_{-i})$;
2. $u_i(k_i^*, k_{-i}) > u_i(k_i^* + 1, k_{-i} + 1)$;
3. $u_i(k_i^*, k_{-i}) > u_i(k_i^* - 1, k_{-i} - 1)$.

Without loss of generality, assume an arbitrary start point for line-sweeping, and that agent 2 reaches critical point k_2^* first in the “line-sweeping” budget incrementing process. From that point on, agent 1 can still possibly go on, so we will proceed by case analysis.

Case 1. Suppose that agent 1 also reaches the critical point at the same time. Then, by Definition 1, neither agent has profitable deviations from the current budgets, and thus they are already in an equilibrium.

Case 2. Suppose otherwise, that agent 1 has not reached its critical point and hence will continue to increment k_1 . Consider the very next step of “line-sweeping”, in which k_1 has gone up to $k_1 + 1$. The budget profile now becomes $(k_1 + 1, k_2^*)$. First, there is no need to consider the case in which agent 2 also increments k_2^* by one, since this contradicts with Definition 1. Second, if agent 2 stays with k_2^* rather than decrementing to $k_2^* - 1$, we can let k_1 keep going up until either agent 1 also reaches critical point or agent 2 does want to go down. Therefore, it suffices to consider the case where agent 2 has a profitable deviation by going down to $k_2^* - 1$. This may happen when the cost saved is more than the marginal profit brought to agent 2 by the newest seed of agent 1.

At this point (budget profile being $(k_1 + 1, k_2^* - 1)$), we can show by contradiction that agent 1 cannot have a profitable deviation to $(k_1, k_2^* - 1)$, otherwise it would not have gone from k_1 to $k_1 + 1$ in the first place (for simplicity and space considerations, we omit the full proof for this claim here). Thus, the sufficient condition for a pure-strategy Nash equilibrium is that agent 1 also does not have a profitable deviation from $(k_1 + 1, k_2^* - 1)$ to $(k_1 + 2, k_2^* - 1)$, i.e., $u_1(k_1 + 1, k_2^* - 1) \geq u_1(k_1 + 2, k_2^* - 1)$, which implies that $(k_1 + 1, k_2^* - 1)$ is an equilibrium. For this condition to hold, a simple calculation yields that the marginal gain in global influence spread brought by the $(k_1 + k_2 + 1)^{\text{th}}$ seed in the greedy algorithm, denoted by z , should satisfy:

$$z \leq \frac{p_1 + \alpha - \beta}{p_1 - \beta} \cdot \frac{k_1 + k_2^* + 1}{k_1 + 2} - \frac{k_2^* - 1}{(k_1 + k_2^*)(k_1 + 2)} \cdot \sigma^g(k_1 + k_2^*).$$

This condition is instance-dependent in the sense that it involves parameters of a specific game instance, e.g., p_1 , α , β , and also the graph structure and influence probabilities which determine influence spread values.

4 Conclusion and Discussions

In this work, we propose a game-theoretic approach to address the problem of competitive viral marketing campaign design from the host’s perspective. In particular, unlike previous work, in our framework competing companies are able to make strategic choices on the budgets to report. We provide a sufficient condition for the existence of pure-strategy Nash equilibria.

There are a couple of possible directions for future work: First, one can look for a more general result on pure-strategy Nash equilibria, and fully characterize them, especially for the game instances with more than two players. Second, it is also interesting to investigate other mechanisms for allocating seeds.

References

- [1] S. Bharathi, D. Kempe, and M. Salek. Competitive influence maximization in social networks. In *WINE*, pages 306–311, 2007.

- [2] A. Borodin, M. Braverman, B. Lucier, and J. Oren. Strategyproof mechanisms for competitive influence in networks. In *WWW*, 2013.
- [3] A. Borodin, Y. Filmus, and J. Oren. Threshold models for competitive influence in social networks. In *WINE*, pages 539–550, 2010.
- [4] C. Budak, D. Agrawal, and A. El Abbadi. Limiting the spread of misinformation in social networks. In *WWW*, pages 665–674, 2011.
- [5] T. Carnes, C. Nagarajan, S. M. Wild, and A. van Zuylen. Maximizing influence in a competitive social network: a follower’s perspective. In *ICEC*, pages 351–360, 2007.
- [6] P. Domingos and M. Richardson. Mining the network value of customers. In *KDD*, pages 57–66, 2001.
- [7] J. Goldenberg, B. Libai, and E. Muller. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters*, 12(3):211–223, 2001.
- [8] G. Gottlob, G. Greco, and F. Scarcello. Pure nash equilibria: Hard and easy games. *J. Artif. Intell. Res. (JAIR)*, 24:357–406, 2005.
- [9] A. Goyal, F. Bonchi, and L. V. S. Lakshmanan. Learning influence probabilities in social networks. In *WSDM*, pages 241–250, 2010.
- [10] M. Granovetter. Threshold models of collective behavior. *American Journal of Sociology*, 83(6):1420–1443, 1978.
- [11] X. He, G. Song, W. Chen, and Q. Jiang. Influence blocking maximization in social networks under the competitive linear threshold model. In *SDM*, pages 463–474, 2012.
- [12] D. Kempe, J. M. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *KDD*, pages 137–146, 2003.
- [13] L. Lakshmanan and et al. Competitive viral marketing: In the eyes of a campaign host. In *KDD*, 2013 (submitted).
- [14] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. An analysis of approximations for maximizing submodular set functions - i. *Mathematical Programming*, 14(1):265–294, 1978.
- [15] M. J. Osborne and A. Rubinstein. *A Course in Game Theory*. The MIT Press, Cambridge, Massachusetts, USA, 1994.
- [16] N. Pathak, A. Banerjee, and J. Srivastava. A generalized linear threshold model for multiple cascades. In *ICDM*, pages 965–970, 2010.

- [17] K. Saito, R. Nakano, and M. Kimura. Prediction of information diffusion probabilities for independent cascade model. In *Proc. of the 12th Int. Conf. on Knowledge-Based Intelligent Information and Engineering Systems (KES'08)*, 2008.