The Thirty-Second AAAI Conference
on Artificial Intelligence (AAAI-18)

When Social Advertising Meets Viral Marketing: Sequencing Social Advertisements for Influence Maximization

Shaojie Tang
Naveen Jindal School of Management
University of Texas at Dallas
shaojie.tang@utdallas.edu

Abstract

Recent studies reveal that social advertising is more effective than conventional online advertising. This is mainly because conventional advertising targets at individual's interest while social advertising is able to produce a large cascade of further exposures to other users via social influence. This motivates us to study the optimal social advertising problem from platform's perspective, and our objective is to find the best ad sequence for each user in order to maximize the expected revenue. Although there is rich body of work that has been devoted to ad sequencing, the network value of each customer is largely ignored in existing algorithm design. To fill this gap, we propose to integrate viral marketing into existing ad sequencing model, and develop both non-adaptive and adaptive ad sequencing policies that can maximize the viral marketing efficiency.

Introduction

Social advertising has been proven to be more effective than conventional online advertising due to the rapid growth of social networking sites. Different from conventional online advertising, a typical social advertisement can propagate across the social network through "reposts" or "shares". For example, Facebook allows advertisers to promote their product through promoted posts and boost posts, which could propagate to other users through a sequence of re-shares. This motivates us to study the ad sequencing problem in the context of social advertising, e.g., finding the best ad sequence for each user in order to maximize the expected revenue from all influenced users.

Although there is rich body of work that has been devoted to ad sequencing, the network value of each customer is largely ignored in existing algorithm design. Most existing work (Craswell et al. 2008; Kempe and Mahdian 2008; Tang 2017) in the field of ad sequencing adopts a simple and practical model to capture user's reaction upon reading a sequence of ads: the user scans through slots from top to bottom, she clicks an ad i with (user-specific) probability q_i and continues to read the next ad with (user-specific) probability c_i. Under the pay-per-click model, each ad has a pay-per-click revenue and advertiser pays such fee each time when one of their ads is clicked. Then given a list of candidate ads, they mainly focus on finding the best ad sequence for a single user so as to maximize the expected revenue. However, in the context of social advertising, users are not isolated and one click on an ad could trigger a large cascade of further exposures to other users. Therefore, we must take into account the network value of each user when determining the ad sequencing for her. We notice that influence maximization has been extensively studied in the literature (Kempe, Kleinberg, and Tardos 2003; Tang et al. 2011; Tong et al. 2017), but none of them considered the sequencing problem on each individual user, and they simply assume that the user will click all ads allocated to her (Tang and Yuan 2016). To fill this gap, we, for the first time, integrate viral marketing into existing ad sequencing model and develop both *non-adaptive* and *adaptive* ad sequencing policies that can maximize the viral marketing efficiency.

The contributions of this paper can be summarized as follows:

- We are the first to formulate and study the social ad sequencing problem. Our objective is to determine the best ad sequence for each arrival user so as to maximize the expected revenue generated from all influenced users;
- We first study our problem under non-adaptive setting, where we can not observe the actual influence generated from the previously allocated ads. Under this model, we propose a simple greedy algorithm that can achieve 1/2-competitive ratio;
- We then extend our model to adaptive setting, where we can observe the full or partial influence from previously allocated ads, we propose a simple greedy policy that can achieve a bounded competitive ratio. Although we focus on ad sequencing in this paper, our results apply to a broad range of optimization problems that can be formulated as an adaptive submodular maximization problem.

Related Work

Due to the rapid increase in internet users, internet advertising has attracted much attention these days (e.g., (Yang et al. 2015; Edelman, Ostrovsky, and Schwarz 2005; Lahaie et al. 2007; Varian 2007)). (Aggarwal et al. 2008; Kempe and Mahdian 2008; Tang 2017) study the ad sequencing problem under the cascade model. Majority of existing works assume that users are isolated from each other, thus it suffice to compute the best ad sequencing for each individual user. However, this assumption does not hold in the context of social advertising, this is because users are connected with each other in the social network, thus one user's decision could affect the others via social influence. To fill this gap, we are the first to study the *Social Ad Sequencing problem*, taking into account the network value of each user when determining the best sequence for her. We present two ad sequencing policies that can maximize the viral marketing efficiency under both non-adaptive and adaptive settings.

Ad Sequencing Model and Propagation Model

Advertisements

The input is a group of ads, denoted by $\mathcal{A} = \{1, 2, \cdots, |\mathcal{A}|\}$. Assume each advertiser i is willing to pay r_i per engagement, typical engagement includes *share* and *like*. In this paper we use share as a representative.

Extension: In the model studied in (Tang and Yuan 2016), each advertiser i also has a finite budget B_i, representing the maximum amount of advertising fee she is willing to pay. Thus, the actual payment made by i is the minimum one between B_i and r_i times the number of engagements. Fortunately, all results derived in this paper can be easily extended to this model.

Ad Scanning Process

Users are arriving sequentially, upon the arrival of a new user, we must decide immediately the best ad sequencing for her. Given an user u and a sequence of ads allocated to her $\sigma_u \in \Sigma_u$, where Σ_u is the strategy space of u, i.e., Σ_u contains all candidate ad sequences that can be allocated to u, we assume that u will view the ads sequentially (Craswell et al. 2008). After examining an ad, say i, in the sequence, the user shares i with her friends with probability $q_i(u)$. This probability is decided by the intrinsic quality or relevance of ad i. Independently of whether ad i was shared or not, the user continues to examine the next ad with probability $c_i(u)$; otherwise, terminates the scanning process. We use $\sigma_u(t)$ to denote the ad placed in slot t by σ_u, u will see a particular slot k with probability $\prod_{t=1}^{k-1} c_{\sigma_u(t)}(u)$. Thus u will share $\sigma_u(k)$ with the following probability:

$$p_{\sigma_u(k)} = q_{\sigma_u(k)}(u) \prod_{t=1}^{k-1} c_{\sigma_u(t)}(u) \qquad (1)$$

In the rest of this paper, we say a user u is seeded by i if and only if i is shared by u. Possible extensions to the previous ad scanning model can be found in (Kempe and Mahdian 2008; Tang 2017).

Propagation Model

Given that ad i has been shared by some users S, we adopt (enhanced) *Independent Cascade Model* (IC), which is investigated recently in the context of marketing (Goldenberg, Libai, and Muller 2001a; 2001b; Kempe, Kleinberg, and Tardos 2003; Tang and Yuan 2016; Yuan and Tang 2017a; 2017b), to capture the dynamics of cascade of i. Under IC model, we use $G_i = (U, p_i(E))$ to denote the diffusion graph under ad $i \in \mathcal{A}$, where U represent the set of all users in the network, $p_i(u, v)$ is the diffusion probability between u and v for ad i. The cascade process runs in discrete steps, in each timestep, when a user u re-shares an ad i, it has one chance of influencing each inactive neighbor v and the success depends on the diffusion probability $p_i(u, v)$. The expected revenue gained from S, denoted by $I_i(S)$, can be calculated as $I_i(S) = r_i C_i(S)$ where $C_i(S)$ is the expected number of influenced users given seed set S. We add a set of edges $\{(u', u) | u \in U\}$ with diffusion probabilities $\{p_i(u', u) = q_i(u) | u \in U\}$ to G_i, these additional edges are used to capture the uncertainty from ad scanning process.

Problem Statement

Let $V \subseteq U$ denote the set of all arriving users. Notice that since users are arriving in an online manner, the complete information about V is only available at the end of the campaign. Given an allocation $\mathcal{S} = \{\sigma_u\}_{u \in V}$, let $p_u^i(\sigma_u)$ denote the probability that u shares i under $\mathcal{S}$, and $p_u^i(\sigma_u)$ can be computed according to (1), the probability that a subset of users $Z \subseteq V$ successfully become the seed set of i is

$$\Pr(Z; \mathcal{S}; i) = \prod_{u \in Z} p_u^i(\sigma_u) \prod_{u \in V \setminus Z} (1 - p_u^i(\sigma_u))$$

We use $f_i(\mathcal{S})$ to denote the expected revenue of i under $\mathcal{S}$, then

$$f_i(\mathcal{S}) = \sum_{Z \in 2^V} \Pr(Z; \mathcal{S}; i) I_i(Z)$$

It follows that the expected revenue under allocation $\mathcal{S}$ is

$$f(\mathcal{S}) = \sum_{i=1}^{N} f_i(\mathcal{S})$$

We study our problem from platform's perspective and our objective is to identify and allocate a sequence of ads to each arriving user so as to maximize the expected revenue. We study this problem under both adaptive and non-adaptive settings.

Non-Adaptive Ad Sequencing: Under the non-adaptive setting, we aim at computing the best ad sequence for each arriving user, without observing the resulting cascade from previously allocated ads. Our problem can be formulated as **P.A**, where $\forall u \in V : |\mathcal{S} \cap \Sigma_u| \leq 1$ specifies that only one ad sequence can be selected for each user.

P.A $\max f(\mathcal{S})$
subject to: $\forall u \in V : |\mathcal{S} \cap \Sigma_u| \leq 1$

Adaptive Ad Sequencing: Under the adaptive setting, we can observe partial or even full realization of the resulting influence from previously allocated ads. Therefore, the decision made at each stage depends on the actual cascading happens in previous stages. Similar to (Golovin and Krause 2011b), we first define full diffusion realization as follows.

Definition 1 (Full Diffusion Realization) *For each $i \in \mathcal{A}$, the state of every edge (u, v) in G_i is either "live" or 'blocked", indicating whether the propagation of i through (u, v) is a success or not. The state of the diffusion of i can be represented using function $\psi_i : E \to [0, 1]$, called* full diffusion realization *or* diffusion realization *of i in short.*

Let $\Psi = \{\psi_i | i \in \mathcal{A}\}$ denote the diffusion realization of all ads upon the arrival of u. We define our adaptive policy $\pi : \Psi \to \sigma_u$, which is a function from the current "observation" Ψ to an ad sequence σ_u, specifying which ad sequence should be allocated to u given the resulting cascade Ψ from previously allocated ads. Our objective is to identify the best policy that maximizes the expected revenue.

P.B: *Maximize* $f(\pi)$

Non-Adaptive Ad Sequencing

We first study the non-adaptive ad sequencing problem. The performance of our online sequencing algorithm is evaluated using competitive analysis. Given a set of arriving users, we say our online algorithm achieves β competitive ratio if the expected revenue of our algorithm is at least β times the expected revenue of the offline optimal solution.

Algorithm Design

We follow a simple greedy idea to design our algorithm. Upon the arrival of a user, we select the sequence with the largest expected incremental marginal gain.

Assume u is the newly arrived user and the existing allocation is $\mathcal{S}$, let $\Delta_{\mathcal{S}}(\sigma_u)$ denote the expected incremental marginal gain by adding σ_u.

$$\Delta_{\mathcal{S}}(\sigma_u) = \sum_{i \in \mathcal{A}} \sum_{Z \in 2^V} \Pr(Z; \mathcal{S} \cup \{\sigma_u\}; i) I_i(Z) - \sum_{i \in \mathcal{A}} \sum_{Z \in 2^V} \Pr(Z; \mathcal{S}; i) I_i(Z)$$

Upon the arrival of u, we aim to compute the best sequence by solving the following optimization problem.

P.1 $\max \Delta_{\mathcal{S}}(\sigma_u)$
subject to: $\sigma_u \in \Sigma_u$

We introduce a dynamic programming based method for computing such sequence in polynomial time. Our approach is inspired by the one proposed for traditional ad sequencing problem (Kempe and Mahdian 2008).

Let $\Delta^i_{\mathcal{S}}(u)$ denote the expected marginal profit brought by i being shared by u. Although calculating the exact value of $\Delta^i_{\mathcal{S}}(u)$ is #P-hard (Chen, Wang, and Wang 2010), we can estimate its value using Monte Carlo simulation.

$$\Delta^i_{\mathcal{S}}(u) = \sum_{Z \in 2^V} (\Pr(Z; \mathcal{S}; i) I_i(Z \cup \{v\}) - \Pr(Z; \mathcal{S}; i) I_i(Z))$$

Our greedy algorithm (Algorithm 1) works as follows: We first calculate $\Delta^i_{\mathcal{S}}(u)$ for each $i \in \mathcal{A}$, then sort all ads in non-decreasing order of $\frac{\Delta^i_{\mathcal{S}}(u) q_i(u)}{1 - c_i(u)}$. Assume $\mathcal{A}$ has been sorted, then we adopt dynamic programming to find the optimal sequencing:

$$\sigma[i, t] = \max\{\Delta^i_{\mathcal{S}}(v) q_i(u) + c_i(u) \sigma[i-1, t-1], \sigma[i-1, t]\}$$

In the above recursion function, $\sigma[i, t]$ stores the optimum value that can be obtained from ads $i, \cdots, |\mathcal{A}|$ in slots $t, \cdots, T$, where T is maximum number of ads that can be allocated to u.

Algorithm 1 Non-Adaptive Ad Sequencing

1: Upon the arrival of a new user u, calculate $\Delta^i_{\mathcal{S}}(u)$ for each $i \in \mathcal{A}$;
2: Sort $\mathcal{A}$ in non-decreasing order of $\frac{\Delta^i_{\mathcal{S}}(u) q_i(u)}{1 - c_i(u)}$;
3: Adopt dynamic programming to find the optimal sequencing $\sigma[i, t] = \max\{\Delta^i_{\mathcal{S}}(u) q_i(u) + c_i(u) \sigma[i-1, t-1], \sigma[i-1, t]\}$;

Lemma 1 *Line 3 in Algorithm 1 returns an optimal ad sequencing to* ***P.1***.

Proof: Given a sequence σ_u, we have $\Delta_{\mathcal{S}}(\sigma_u) = \sum_{i \in \mathcal{A}} p^i_u(\sigma_u) \Delta^i_{\mathcal{S}}(u)$. Thus, solving **P.1** is equivalent to finding a σ_u that maximizes $\sum_{i \in \mathcal{A}} p^i_u(\sigma_u) \Delta^i_{\mathcal{S}}(u)$ where $p^i_u(\sigma_u)$ is defined in (1). Notice that for a given $\mathcal{S}$, $\Delta^i_{\mathcal{S}}(u)$ is a fixed value for each i, thus we can treat $\Delta^i_{\mathcal{S}}(u)$ as the per-engagement value of ad i as defined in traditional ad sequencing problem (Kempe and Mahdian 2008), and this enables us to find the optimal solution via dynamic programming. □

We next prove that by solving **P.1** optimally for each arriving user, our greedy policy can achieve a constant competitive ratio.

Theorem 1 *Algorithm 1 achieves competitive ratio* $1/2$.

Proof: Notice that the original utility function $f(\cdot)$ is defined on $\mathcal{S}$ where $\forall u \in V : |\mathcal{S} \cap \Sigma_u| \leq 1$. To facilitate our analysis, we generalize its definition to any subset of $\bigcup_{v \in U} \Sigma_v$ by allowing multiple sequences to be allocated to the same user: $f : 2^{\bigcup_{v \in U} \Sigma_v} \to \mathbb{R}$. Intuitively, we assume that if multiple sequences are allocated to the same user, she will scan them one by one independently. This generalized model allows multiple attempts to share one ad, however, we assume that the same ad can only be shared at most once.

We next give a formal description of this generalization. For any subset of sequences $\mathcal{S} \in \bigcup_{v \in U} \Sigma_v$, let $\mathcal{S}_u = \mathcal{S} \cap \Sigma_u$ denote the set of sequences in $\mathcal{S}$ that are allocated to u, thus $\mathcal{S} = \bigcup_{v \in U} \mathcal{S}_v$. Given a $\mathcal{S}$, we define the probability of i being shared by u as

$$p^i_u(\mathcal{S}) = 1 - \prod_{\sigma \in \mathcal{S}_u} (1 - p^i_u(\sigma)) \tag{2}$$

The probability that a subset of users $Z \subseteq V$ successfully become the seed set of i is

$$\Pr(Z; \mathcal{S}; i) = \prod_{u \in Z} p^i_u(\mathcal{S}) \prod_{u \in V \setminus Z} (1 - p^i_u(\mathcal{S}))$$

Define $f_i(\mathcal{S}) = \sum_{Z \in 2^V} \Pr(Z; \mathcal{S}; i) I_i(Z)$. It follows that the expected cascade under allocation $\mathcal{S}$, which is not necessarily to be feasible, is $f(\mathcal{S}) = \sum_{i=1}^{N} f_i(\mathcal{S})$.

We next prove that $f(\cdot)$ is a monotone and submodular function. It is easy to prove that $f(\cdot)$ is monotone. Eq. (2) implies that adding a new sequence to an existing allocation does not decrease the click probability of any ad. Therefore $f(\cdot)$ is a non-decreasing function. We next prove the submodularity of $f(\cdot)$. Since $f(\cdot)$ is a linear combination of $f_i(\cdot)$, it suffice to prove the submodularity of $f_i(\cdot)$. Consider two arbitrary allocations $\mathcal{S}', \mathcal{S}'' \subseteq \Sigma$ such that $\mathcal{S}' \subseteq \mathcal{S}''$, and an arbitrary sequence $\sigma_u \in \Sigma_u \setminus \mathcal{S}''$. Because $\forall i \in \mathcal{A}, \forall u \in V : p_u^i(\mathcal{S}) = 1 - \prod_{\sigma \in \mathcal{S}_u}(1 - p_u^i(\sigma))$ and $\mathcal{S}' \subseteq \mathcal{S}''$, we have $\forall i \in \mathcal{A}, \forall u \in V : p_u^i(\mathcal{S}') \leq p_u^i(\mathcal{S}'')$. We next prove that $f_i(\mathcal{S}_A \cup \{\sigma_u\}) - f_i(\mathcal{S}_A) \geq f_i(\mathcal{S}_B \cup \{\sigma_u\}) - f_i(\mathcal{S}_B)$. Our proof is inspired by (Yuan and Tang 2017a). To facilitate our analysis, we next take an alternative view at the propagation process generated by any allocation $\mathcal{S}$. For each user v and each ad i, we select a random number g_v uniformly from $[0, 1]$: if $p_v^i(\mathcal{S}) > g_v$, then v is declared to be live, otherwise, v is blocked. Similarly, we select a random number g_e uniformly from $[0, 1]$ for each edge e, then an edge is declared to be live if and only $p_e > g_e$. It is easy to verify that a user is influenced by i if it can be reached from some live user through a path consisting of live edges. Since $\forall i \in \mathcal{A}, \forall v \in V : p_v^i(\mathcal{S}') \leq p_v^i(\mathcal{S}'')$, for every node v with $p_v^i(\mathcal{S}') > g_v$, we have $p_v^i(\mathcal{S}'') > g_v$. Moreover, the set of live edges under $\mathcal{S}'$ are identical to $\mathcal{S}''$. This implies that if a user is influenced by i under $\mathcal{S}''$, it must be influenced by i under $\mathcal{S}''$. This is sufficient to prove that $f_i(\mathcal{S}_A \cup \{\sigma_u\}) - f_i(\mathcal{S}_A) \geq f_i(\mathcal{S}_B \cup \{\sigma_u\}) - f_i(\mathcal{S}_B)$, thus $f_i(\cdot)$ is submodular.

On the other hand, $\forall u \in V : |\mathcal{S} \cap \Sigma_u| \leq 1$ is a partition matroid constraint. Therefore, an online greedy policy achieves $1/2$-competitive ratio. $\square$

Adaptive Ad Sequencing

Adaptive Sequencing with Full Feedback

Algorithm 2 Adaptive Ad Sequencing

1: Upon the arrival of a new user u, calculate $\Delta_{\mathcal{S}}^i(u|\Psi)$ for each $i \in \mathcal{A}$;

2: Sort all posts in non-decreasing order of $\frac{\Delta_{\mathcal{S}}^i(u|\Psi) q_i(u)}{1 - c_i(u)}$;

3: Adopt dynamic programming to find the optimal sequencing $\sigma[i, t] = \max\{\Delta_{\mathcal{S}}^i(u|\Psi) q_i(u) + c_i(u)\sigma[i-1, t-1], \sigma[i-1, t]\}$;

4: Update the diffusion realization Ψ;

We first study the case when full feedback is available. We assume that the diffusion realization of all previously allocated ads is available before the arrival of the next user, it allows us to adaptively decide the best ad sequencing for a newly arrived user after observing the actual cascade resulting from previously allocated ads.

Our policy (Algorithm 2) is performed in a sequential greedy manner as follows: After observing the diffusion realization Ψ, we aim to identify the best sequence for u by solving the following optimization problem:

P.2 $\max \Delta_\Psi(\sigma_u)$
subject to: $\sigma_u \in \Sigma_u$

In the above formulation, $\Delta_\Psi(\sigma_u)$ represents the expected marginal profit of σ_u under Ψ and $\Delta_\Psi^i(\sigma_u)$ represents the expected marginal profit from i, then $\Delta_\Psi(\sigma_u)$ can be calculated as $\sum_{i \in \mathcal{A}} \Delta_\Psi^i(\sigma_u)$. By solving **P.2**, we are able to obtain a sequence that maximizes the expected marginal benefit under Ψ. Similar to the proof of Lemma 1, we can prove that Line 3 in Algorithm 2 returns an optimal sequence to **P.2**.

To bound the competitive ratio of Algorithm 2, we first prove that $f(\cdot)$ is monotone and adaptive submodular. A complete description of adaptive submodular optimization can be found in (Golovin and Krause 2011b).

Lemma 2 *$f(\cdot)$ is monotone and adaptive submodular.*

Proof: Since $f(\cdot)$ is a linear combination of $f_i(\cdot)$, it suffice to prove that $f_i(\cdot)$ is monotone adaptive submodular. In the rest of this paper, we use $f_i(\mathcal{S}_\Psi \cup \sigma_u)$ to represent $f_i(\mathcal{S}_\Psi \cup \{\sigma_u\})$.

It is easy to prove that $f_i(\cdot)$ is an non-decreasing function, we next focus on proving $\Delta_\Psi^i(\sigma_u) \geq \Delta_{\Psi'}^i(\sigma_u)$ for all $\Psi \subseteq \Psi'$, where $\Delta_\Psi^i(\sigma_u) = \mathbb{E}_\Phi[f_i(\mathcal{S}_\Psi \cup \sigma_u), \Phi) - f_i(\mathcal{S}_\Psi), \Phi)|\Phi \sim \Psi]$ and Φ a global realization that is consistent with Ψ. Consider any node $v \in U \setminus \{u\}$ and a realization Ψ, the increased utility from v brought by σ_u being added can be calculated as follows:

$$\left(p_u^i(\mathcal{S}_\Psi \cup \sigma_u|\Psi) - p_u^i(\mathcal{S}_\Psi|\Psi)\right) \Delta^i(u \rightarrow v|\Psi) \tag{3}$$

where $p_u^i(\mathcal{S}_\Psi \cup \sigma_u|\Psi) - p_u^i(\mathcal{S}_\Psi|\Psi)$ is the increased probability of i being shared by u after adding σ_u and $\Delta^i(u \rightarrow v|\Psi)$ is the increased utility from v after i has been shared by u. It follows that $\Delta_\Psi^i(\sigma_u) = \sum_{v \in U}(3)$. Similarly, this marginal utility under Ψ' can be calculated as

$$\left(p_u^i(\mathcal{S}_{\Psi'} \cup \sigma_u|\Psi') - p_u^i(\mathcal{S}_{\Psi'}|\Psi')\right) \Delta^i(u \rightarrow v|\Psi') \tag{4}$$

We next prove that $(3) \geq (4)$. It has been proved that $\Delta^i(u \rightarrow v|\Psi) \geq \Delta^i(u \rightarrow v|\Psi')$ in (Golovin and Krause 2011b), in order to prove $(3) \geq (4)$, it suffice to prove

$$p_u^i(\mathcal{S}_\Psi \cup \sigma_u|\Psi) - p_u^i(\mathcal{S}_\Psi|\Psi) \geq p_u^i(\mathcal{S}_{\Psi'} \cup \sigma_u|\Psi') - p_u^i(\mathcal{S}_{\Psi'}|\Psi') \tag{5}$$

We prove (5) under two cases:

- u is seeded by i under Ψ': we have $p_u^i(\mathcal{S}_{\Psi'} \cup \sigma_u|\Psi') - p_u^i(\mathcal{S}'_\Psi|\Psi') = 0$, thus (5) holds;
- u is not seeded by i under Ψ': because $\Psi \subseteq \Psi'$, u is not seeded under Ψ. Thus $p_u^i(\mathcal{S}_\Psi \cup \sigma_u|\Psi) - p_u^i(\mathcal{S}_\Psi|\Psi) = p_u^i(\mathcal{S}_{\Psi'} \cup \sigma_u|\Psi') - p_u^i(\mathcal{S}'_\Psi|\Psi') = p_u^i(\sigma_u)$.

It follows that $(3) \geq (4)$, thus $\sum_{v \in U}(3) \geq \sum_{v \in U}(4)$. Then we have $\Delta_\Psi^i(\sigma_u) \geq \Delta_{\Psi'}^i(\sigma_u)$. $\square$

We next show that Algorithm 2 achieves a constant competitive ratio.

Theorem 2 *Algorithm 2 achieves $1/2$-competitive ratio.*

This theorem can be proved based on Lemma 2 and Theorem 7 in (Golovin and Krause 2011a).

Adaptive Sequencing with Partial Feedback

We next study the case when only partial feedback is available. Unfortunately, the utility function $f(\cdot)$ under partial feedback model is not adaptive submodular (Golovin and Krause 2011b). To tackle this challenge, we first introduce the following definition.

Definition 2 *Given the existing allocation $\mathcal{S}$, we define partial feedback realization (p-realization) Ψ_p as the realization that can be observed upon the arrival of the next user, and we define the full feedback realization (f-realization) $\Psi_f \in \mathbf{\Psi}_f$ as the complete realization of $\mathcal{S}$, where $\mathbf{\Psi}_f$ is the set of all possible f-realizations that are consistent with Ψ_p, thus $\Psi_p \subseteq \Psi_f$. Notice that we have $\Psi_p = \Psi_f$ under full feedback model.*

Let Ψ_f^* denote the f-realization such that $f(\Psi_p) = f(\Psi_f^*)$, intuitively, Ψ_f^* is the most *pessimistic* realization under which no additional users, other than those who have been influenced under Ψ_p, can be influenced by $\mathcal{S}$. In the rest of this paper, we assume Ψ_f^* happens with probability α_u upon the arrival of u, and α_u can be readily calculated with the propagation model and ad scanning model. Let $\alpha = \min_{u\in V} \alpha_u$, we next prove that the greedy policy described in Algorithm 2 achieves $\alpha/(\alpha+1)$-competitive ratio under partial feedback model. Notice that this result subsumes the full feedback model as our special case, e.g., $\alpha = 1$ under full feedback model.

Theorem 3 *Algorithm 2 achieves $\alpha/(\alpha+1)$-competitive ratio.*

Proof: Consider any node $v \in U \setminus \{u\}$ and a p-realization Ψ_p, the increased utility from v brought by an arbitrary sequence σ_u being added under Ψ_p is at least:

$$\alpha \left(p_u^i(\mathcal{S}\cup\sigma_u|\Psi_f^*) - p_u^i(\mathcal{S}|\Psi_f^*)\right)\Delta^i(u\rightarrow v|\Psi_f^*) \tag{6}$$

Consider an arbitrary f-realization Ψ_f, the increased utility from v brought by an arbitrary sequence σ_u being added is:

$$\left(p_u^i(\mathcal{S}\cup\sigma_u|\Psi_f) - p_u^i(\mathcal{S}|\Psi_f)\right)\Delta^i(u\rightarrow v|\Psi_f) \tag{7}$$

We next prove that $\forall u \in V \forall v \in U \forall \Psi_f \in \mathbf{\Psi}_f : \frac{1}{\alpha}(6) \geq (7)$. Since Ψ_f^* is the most pessimistic realization, we make the following three observations:

- if v is influenced under Ψ_f^*, it must be influenced under Ψ_f, then $\Delta^i(u\rightarrow v|\Psi_f^*) = \Delta^i(u\rightarrow v|\Psi_f) = 0$;
- if v is not influenced under Ψ_f^* but it is influenced under Ψ_f, then $\Delta^i(u\rightarrow v|\Psi_f^*) \geq \Delta^i(u\rightarrow v|\Psi_f) = 0$;
- if v is not influenced under Ψ_f, then any live path that connects u and v can not contain any influenced nodes under Ψ_f, thus this path also exists in Ψ_f^* (under the same global realization for the remaining nodes), it follows that $\Delta^i(u\rightarrow v|\Psi_f^*) \geq \Delta^i(u\rightarrow v|\Psi_f)$.

It follows that $\Delta^i(u\rightarrow v|\Psi_f^*) \geq \Delta^i(u\rightarrow v|\Psi_f)$. On the other hand, if u is seeded by i under Ψ_f^*, it must be seeded by i under Ψ_f, therefore

$$p_u^i(\mathcal{S}\cup\sigma_u|\Psi_f^*) - p_u^i(\mathcal{S}|\Psi_f^*) \geq p_u^i(\mathcal{S}\cup\sigma_u|\Psi_f) - p_u^i(\mathcal{S}|\Psi_f)$$

It follows that

$$\forall u \in V \forall v \in U \forall \Psi_f \in \mathbf{\Psi}_f : \frac{1}{\alpha}(6) \geq (7)$$

Thus, we have $\mathbb{E}_\Phi[f_i(\mathcal{S}\cup\sigma_u,\Phi) - f_i(\mathcal{S},\Phi)|\Phi\sim\Psi_p] \geq \alpha\mathbb{E}_\Phi[f_i(\mathcal{S}\cup\sigma_u,\Phi) - f_i(\mathcal{S},\Phi)|\Phi\sim\Psi_f^*] \geq \alpha\mathbb{E}_\Phi[f_i(\mathcal{S}\cup\sigma_u,\Phi) - f_i(\mathcal{S},\Phi)|\Phi\sim\Psi_f]$. Assume σ_u^g is the sequence added by our greedy algorithm, for every $\sigma_u \in \Sigma_u$, we have

$$\begin{aligned}&\mathbb{E}_\Phi[f_i(\mathcal{S}\cup\sigma_u^g,\Phi) - f_i(\mathcal{S},\Phi)|\Phi\sim\Psi_p]\\ &\geq \mathbb{E}_\Phi[f_i(\mathcal{S}\cup\sigma_u,\Phi) - f_i(\mathcal{S},\Phi)|\Phi\sim\Psi_p]\\ &\geq \alpha\mathbb{E}_\Phi[f_i(\mathcal{S}\cup\sigma_u,\Phi) - f_i(\mathcal{S},\Phi)|\Phi\sim\Psi_f]\end{aligned} \tag{8}$$

Let $\mathcal{O}$ denote the optimal solution before u's arrival and σ_u^o is the optimal sequence assigned to u. Due to $f(\cdot)$ is adaptive submodular under full feedback model, we have

$$\begin{aligned}&\mathbb{E}_\Phi[f_i(\mathcal{S}\cup\mathcal{O}\cup\sigma_u^o\cup\sigma_u^g,\Phi) - f_i(\mathcal{S}\cup\mathcal{O}\cup\sigma_u^g,\Phi)|\Phi\sim\Psi_f]\\ &\leq \mathbb{E}_\Phi[f_i(\mathcal{S}\cup\sigma_u^o,\Phi) - f_i(\mathcal{S},\Phi)|\Phi\sim\Psi_f].\end{aligned}$$

It follows that $\mathbb{E}_{\Psi_f}[\mathbb{E}_\Phi[f_i(\mathcal{S}\cup\mathcal{O}\cup\sigma_u^o\cup\sigma_u^g,\Phi) - f_i(\mathcal{S}\cup\mathcal{O}\cup\sigma_u^g,\Phi)|\Phi\sim\Psi_f]|\Psi_f\sim\mathbf{\Psi}_f] \leq \mathbb{E}_{\Psi_f}[\mathbb{E}_\Phi[f_i(\mathcal{S}\cup\sigma_u^o,\Phi) - f_i(\mathcal{S},\Phi)|\Phi\sim\Psi_f|\Psi_f\sim\mathbf{\Psi}_f]$. According to (8), we have

$$\begin{aligned}&\frac{1}{\alpha}\mathbb{E}_\Phi[f_i(\mathcal{S}\cup\sigma_u^g,\Phi) - f_i(\mathcal{S},\Phi)|\Phi\sim\Psi_p]\\ &\geq \mathbb{E}_\Phi[f_i(\mathcal{S}\cup\sigma_u^o,\Phi) - f_i(\mathcal{S},\Phi)|\Phi\sim\Psi_f]\end{aligned} \tag{9}$$

Therefore, (9) is lower bounded by $(9) \geq \mathbb{E}_{\Psi_f}[\mathbb{E}_\Phi[f_i(\mathcal{S}\cup\mathcal{O}\cup\sigma_u^o\cup\sigma_u^g,\Phi) - f_i(\mathcal{S}\cup\mathcal{O}\cup\sigma_u^g,\Phi)|\Phi\sim\Psi_f]|\Psi_f\sim\mathbf{\Psi}_f]$. Thus Algorithm 2 achieves $\alpha/(\alpha+1)$-competitive ratio. □

Performance Evaluation

In this section, we evaluate the effectiveness and efficiency of our proposed ad sequencing strategies on three benchmark social network datasets.

Experimental Setup

Datasets. We conduct extensive experiments on three real-world benchmark social networks in the literature of viral marketing: *Epinions*, *Slashdot*, and *Pokec* to examine the effectiveness and efficiency of the proposed algorithms. Basic statistics of the datasets are summarized in Table 1, where n denotes the number of nodes and m denotes the number of edges in the social graph. *Epinions* is a who-trusts-whom network that is taken from a social consumer review website (http://www.epinions.com/). *Slashdot* is a social graph that exhibits friend/foe relationships in a user community interested in technology oriented news (http://slashdot.org/). *Pokec* is the most popular online social network in Slovakia (http://pokec.azet.sk/). The popularity of *Pokec* has not changed even after the emergence of Facebook. In this graph, nodes represent authors and the edges are directed since friendships in *Pokec* are oriented. All datasets used in our experiments are publicly available at (Leskovec and Krevl 2014).

Influence Model. The influence diffusion model governs the way that information spreads in the social network driven by social influence. In this work, we adopt the standard Independent Cascade (*IC*) model as the influence model,

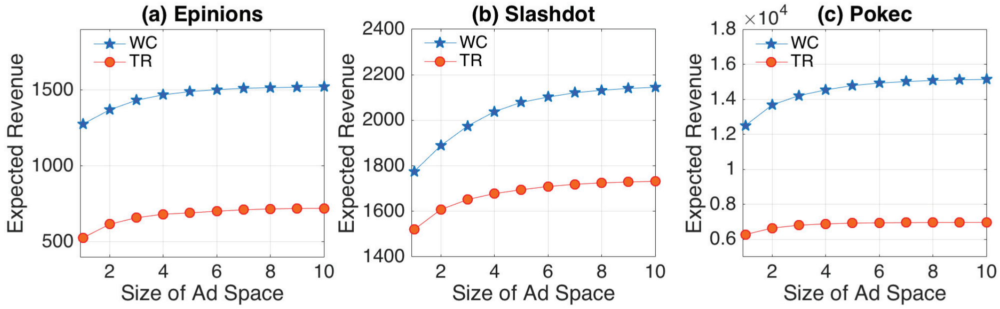

Figure 1: Expected revenue vs. size of ad space.

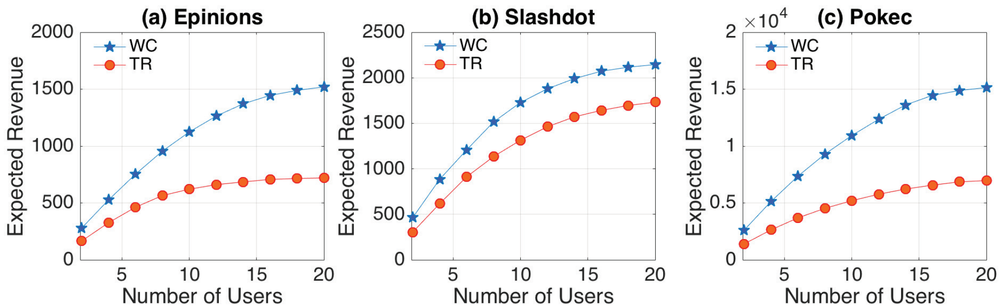

Figure 2: Expected revenue vs. number of users arriving online.

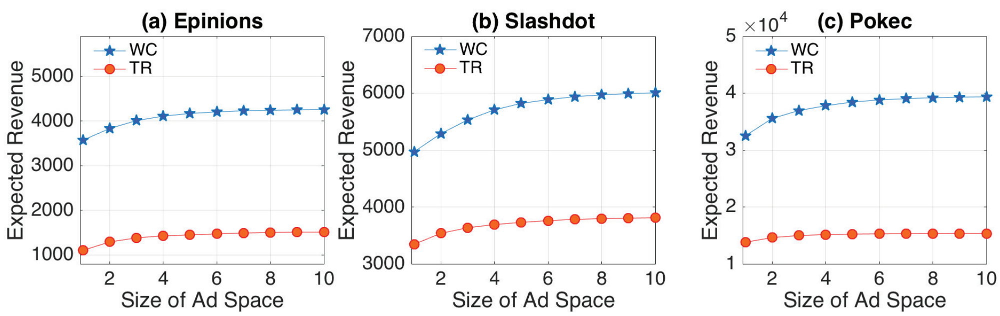

Figure 3: Expected revenue vs. size of ad space.

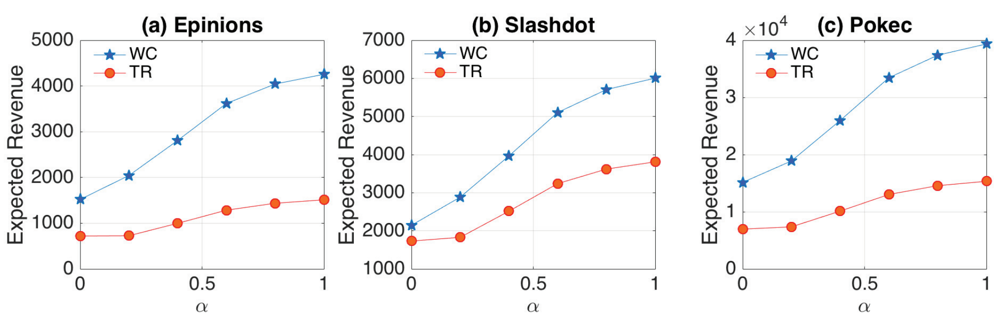

Figure 4: Expected revenue vs. α.

Table 1: Dataset characteristics

Dataset	***n***	***m***	**Type**	**Avg. degree**
Epinions	75,879	508,837	directed	13.4
Slashdot	82,168	948,464	directed	21.8
Pokec	1,632,803	30,622,564	directed	35.9

which is mostly widely used in the literature (Tang, Xiao, and Shi 2014; He and Kempe 2016). As for the *IC* model, we adopt two different methods to set the influence probability associated with the edges, and test how the quality of solutions changes under these two settings. In the first approach, we set the propagation probability of each directed edge as reciprocal of the in-degree of the node that the edge points to. Specifically, for each edge e we first identify the node v that e points to, and then set $p(e) = 1/d(v)$, where $d(v)$ denotes the in-degree of v. This setting of $p(e)$ is widely used in prior work (Tang, Xiao, and Shi 2014; Goyal, Lu, and Lakshmanan 2011; jun), it can be used to explain the influence spreading in the social networks where the recipiants recieve similar amount of influence regardless of her node degree. We call it `WC` setting. In the second approach, we set the influence probability of each directed edge randomly from $\{0.1, 0.01, 0.001\}$ as in (Chen, Wang, and Yang 2009; jun), representing the case when multiple types of personal relations (three types in this case) exist, we call it `TR` setting.

Parameters. In our experiments, unless otherwise specified, for each ad, we set the click probability and the continuation probability for all the users by sampling uniformly at random from $[0, 0.1]$. We set the number of candidate ads to 100. The pay-per-engagement revenue for each ad is sampled uniformly at random from $[1, 5]$. We vary the number of users arriving onine and the ad space size to evaluate the impact of these parameters on the quality of the solutions. For the adaptive greedy ad sequencing algorithm (Algorithm 2), we adjust the value of control parameter α to evaluate its impact on the performance of the solutions.

Experimental Results

Results for non-adaptive ad sequencing. Our first set of experiments evaluates our solutions produced by Algorithm 1 (abbreviated as *NG*) in terms of the expected revenue obtained from advertisers. Figure 1 shows the expected revenue yielded by *NG* on all tested datasets, with the size of ad space varying from 1 to 10. The x-axis holds the size of ad space, and the y-axis holds the expected revenue. We observe that as the size of ad space increases, the expected revenue obtained by *NG* increases at first, and then gradually converges as the size of ad space approaches 10. The expected revenue obtained under *WC* setting consistently achieves a higher value than that obtained under *TR* setting. The underlying reason is that under *WC* setting, the edges are associated with higher influence probabilities, which leads to a larger size of the expected influence spread, therefore the expected revenue increases accordingly. We also observe that the lines clearly illustrate the phenomenon of diminishing marginal returns, empirically illustrating submodularity.

Figure 2 shows the expected revenue yielded by *NG* on all tested datasets, with the number of users arriving online varying from 2 to 20. The x-axis holds the number of users that will arrive in an online manner, and the y-axis holds the expected revenue. In this set of experiments, we set the size of ad space to be 10. As the number of online users increases, we have more users to allocate advertisements, therefore, we will obtain more seed users to trigger a larger cascade of engagement (repost/like) for each ad throughout the entire social network, leading to a higher expected revenue earned from advertisers. We observe that again, *NG* yields a higher expected revenue under *WC* setting than it does under *TR* setting.

Results for adaptive ad sequencing. Figure 3 shows the expected revenue produced by Algorithm 2 (abbreviated as *AG*), with the size of ad space varying from 1 to 10. The x-axis holds the size of ad space, and the y-axis holds the expected revenue. In this set of experiments, we set the number of users arriving online to be 20 and $\alpha = 1$, indicating a fully adaptive ad sequencing strategy is computed. Again, we observe that the expected revenue increases as the size of ad space increases, eventually converges. We also observe that compared with *NG* under the same parameter setting, *AG* produces a must higher expected revenue than *NG*, because *AG* fully utilizes the observations on actual influence diffusion that can be made during the ad sequencing process. Thus *AG* provides a sophisticated yet effective ad sequencing strategy, leading to a higher expected revenue.

Figure 4 illustrates the expected revenue produced by *AG* with the value of α ranging from 0 to 1. The x-axis holds the value of α, and the y-axis holds the expected revenue. In this set of experiments, we set the number of users arriving online to be 20 and the size of ad space is set to be 10. Recall that when $\alpha = 0$, our model reduces to non-adaptive model; and when $\alpha = 1$, our model becomes fully adaptive model. As expected, the expected revenue produced by *AG* increases as α increases. The underlying reason is that when α approaches to 1, each decision made by our ad sequencing strategy is based on more observation of the actual influence diffusion triggered by the current seed set for each ad. In particular, when $\alpha = 1$, *AG* fully utilizes the observation of the diffusion realization, thus it produces a high quality ad sequence for each online user, leading to a higher expected revenue across all the sequences.

In summary, our experiments on benchmark datasets demonstrate that our proposed social ad sequencing algorithms are effective and efficient under various settings.

Conclusion

In this paper, we study social ad sequencing problem for influence maximization. We integrate viral marketing into existing ad sequencing problem, and study our problem under both non-adaptive and adaptive settings. We propose a serial of policies that achieve bounded competitive ratios. We also conduct extensive experiments to evaluate the performance of our algorithms, and the experiment results validate the effectiveness of our solutions.

References

Aggarwal, G.; Feldman, J.; Muthukrishnan, S.; and Pál, M. 2008. Sponsored search auctions with markovian users. In *International Workshop on Internet and Network Economics*, 621–628. Springer.

Chen, W.; Wang, C.; and Wang, Y. 2010. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *Proceedings of the 16th ACM SIGKDD*, 1029–1038. ACM.

Chen, W.; Wang, Y.; and Yang, S. 2009. Efficient influence maximization in social networks. In *Proceedings of the 15th ACM SIGKDD*, 199–208. ACM.

Craswell, N.; Zoeter, O.; Taylor, M.; and Ramsey, B. 2008. An experimental comparison of click position-bias models. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, 87–94. ACM.

Edelman, B.; Ostrovsky, M.; and Schwarz, M. 2005. Internet advertising and the generalized second price auction: Selling billions of dollars worth of keywords. Technical report, National Bureau of Economic Research.

Goldenberg, J.; Libai, B.; and Muller, E. 2001a. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing letters* 12(3):211–223.

Goldenberg, J.; Libai, B.; and Muller, E. 2001b. Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata. *Academy of Marketing Science Review* 2001:1.

Golovin, D., and Krause, A. 2011a. Adaptive submodular optimization under matroid constraints. *arXiv preprint arXiv:1101.4450*.

Golovin, D., and Krause, A. 2011b. Adaptive submodularity: Theory and applications in active learning and stochastic optimization. *Journal of Artificial Intelligence Research* 427–486.

Goyal, A.; Lu, W.; and Lakshmanan, L. V. 2011. Simpath: An efficient algorithm for influence maximization under the linear threshold model. In *Data Mining (ICDM), 2011 IEEE 11th International Conference on*, 211–220. IEEE.

He, X., and Kempe, D. 2016. Robust influence maximization. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, 885–894. New York, NY, USA: ACM.

Kempe, D., and Mahdian, M. 2008. A cascade model for externalities in sponsored search. In *International Workshop on Internet and Network Economics*, 585–596. Springer.

Kempe, D.; Kleinberg, J.; and Tardos, É. 2003. Maximizing the spread of influence through a social network. In *KDD*, 137–146. ACM.

Lahaie, S.; Pennock, D. M.; Saberi, A.; and Vohra, R. V. 2007. Sponsored search auctions. *Algorithmic game theory* 699–716.

Leskovec, J., and Krevl, A. 2014. SNAP Datasets: Stanford large network dataset collection. http://snap.stanford.edu/data.

Tang, S., and Yuan, J. 2016. Optimizing ad allocation in social advertising. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, 1383–1392. ACM.

Tang, S.; Yuan, J.; Mao, X.; Li, X.-Y.; Chen, W.; and Dai, G. 2011. Relationship classification in large scale online social networks and its impact on information propagation. In *INFOCOM, 2011 Proceedings IEEE*, 2291–2299. IEEE.

Tang, Y.; Xiao, X.; and Shi, Y. 2014. Influence maximization: Near-optimal time complexity meets practical efficiency. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*, 75–86. ACM.

Tang, S. 2017. Robust advertisement allocation. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, 4419–4425.

Tong, G.; Wu, W.; Tang, S.; and Du, D.-Z. 2017. Adaptive influence maximization in dynamic social networks. *IEEE/ACM Transactions on Networking (TON)* 25(1):112–125.

Varian, H. R. 2007. Position auctions. *international Journal of industrial Organization* 25(6):1163–1178.

Yang, Y.; Zeng, D.; Yang, Y.; and Zhang, J. 2015. Optimal budget allocation across search advertising markets. *INFORMS Journal on Computing* 27(2):285–300.

Yuan, J., and Tang, S.-J. 2017a. Adaptive discount allocation in social networks. In *Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, 22. ACM.

Yuan, J., and Tang, S. 2017b. No time to observe: Adaptive influence maximization with partial feedback. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, 3908–3914.