Online Influence Maximization in Non-Stationary Social Networks

Yixin Bao*, Xiaoke Wang*, Zhi Wang†, Chuan Wu*, Francis C.M. Lau*

*Department of Computer Science, The University of Hong Kong, Email: {yxbao,xkwang,cwu,fcmlau}@cs.hku.hk
†Graduate School at Shenzhen, Tsinghua University, Email: wangzhi@sz.tsinghua.edu.cn

Abstract—Social networks have been popular platforms for information propagation. An important use case is viral marketing: given a promotion budget, an advertiser can choose some influential users as the seed set and provide them free or discounted sample products; in this way, the advertiser hopes to increase the popularity of the product in the users’ friend circles by the world-of-mouth effect, and thus maximizes the number of users that information of the production can reach. There has been a body of literature studying the influence maximization problem. Nevertheless, the existing studies mostly investigate the problem on a one-off basis, assuming fixed known influence probabilities among users, or the knowledge of the exact social network topology. In practice, the social network topology and influence probabilities are typically unknown to the advertiser, which can be varying over time, i.e., in cases of newly established, strengthened or weakened social ties. In this paper, we focus on a dynamic non-stationary social network and design a randomized algorithm, RSB, based on multi-armed bandit optimization, to maximize influence propagation over time. The algorithm produces a sequence of online decisions and calibrates its explore-exploit strategy utilizing outcomes of previous decisions. It is rigorously proven to achieve an upper-bounded regret in reward and applicable to large-scale social networks. Practical effectiveness of the algorithm is evaluated using both synthetic and real-world datasets, which demonstrates that our algorithm outperforms previous stationary methods under non-stationary conditions.

I. INTRODUCTION

Influence maximization in social networks is an important problem that seeks the best seed users to maximize the spread of information [1]. Prominent use cases include advertising and viral marketing [1][2]. When a company is promoting a new product, it can engage some influential users as seeds in a social network, providing them samples for free or at discounted prices. These seed users may inform their friends of this product, and their friends will further influence other users, and so on. Through world-of-mouth distribution, the product will get to be known by more and more users in the social network. As it is common for a company to have a promotion budget, it is most beneficial to identify the best set of seeds so as to maximize the number of users that information can eventually reach.

The influence maximization problem has been studied on several probabilistic cascade models. In the independent cascade model [1], each node probabilistically activates (influences) its neighbors at each time stamp independently of the history thus far, and a node only attempts to activate a neighbor once. In the linear threshold model [1], a node will be activated only when the sum of influence probabilities from its neighbors exceeds a threshold. The influence probability in the above models, namely the probability for node $u$ to activate its neighbor $v$ after $u$ has been activated, is often decided empirically in studies designing influence maximization algorithms, e.g., according to inverse of the indegree of $v$.

Based on these information propagation models, existing studies mostly tackle the influence maximization problem on a one-off basis, assuming that both the social network topology and influence probabilities are fixed and available as input. Kempe et al. [1] prove that the influence maximization problem is NP hard but can be approximated to within a factor of $(1 - \frac{1}{e} - \epsilon)$ with a greedy hill-climbing method, where $\epsilon$ is any positive real number. A number of other approximation algorithms have also been proposed to achieve near-optimal time complexity, including CELF [3], CELF++ [4], TIM [5], and IMM [6]. In real-world social networks, exact network topology and influence probabilities are typically unknown to a third party advertiser, and are time-varying. For example, new social ties are set up when people make new friends, and the ties can be strengthened over time when they become more familiar; two people become connected when collaborating on a short-term project and the tie may weaken after the project has ended; a couple may break up and be no longer connected in the social network. It is therefore more realistic to describe the influence probabilities and social network topology as non-stationary. In addition, it is often hard to determine an accurate stochastic distribution assumption for the variance of influence probabilities, since no assumption may exist for human behavior.

To handle unknown underlying distributions in online optimization, multi-armed bandit optimization has been applied in related scenarios. The multi-armed bandit problem [7] is a problem in which an agent has multiple arms to choose from, and needs to decide a policy to select an arm at each time. When chosen, an arm provides a random reward from an unknown distribution specific to the arm, and the agent utilizes the outcome to update his strategy. The objective is to maximize the overall reward in the whole time span through selecting a sequence of arms, thus minimize regret, which is the gap between offline optimal overall reward and the actual overall reward the agent has obtained. The design of multi-
Multi-armed bandit optimization has also been applied to solve the influence maximization problem with unknown influence probabilities \[9\][13][14]. The existing algorithms have been relying on assumptions of the rewards to guarantee nice theoretical bounds on regret. For example, UCB \[15\] assumes that the reward distributions are stationary and obtains a regret bound of \(O(T)\) under the adversary settings, i.e., there is a rival assigning the rewards against the agent. There exist some studies dealing with non-stationary bandits \[16\], but none can be readily applied to the influence maximization problem. Detailed discussions of the existing literature in these aspects are given in Sec. \[II\].

This paper designs an online randomized algorithm, referred to as RSB, based on multi-armed bandit optimization, to maximize influence propagation in a dynamic non-stationary social network with unknown and non-stationary influence probabilities between pairs of users. Our algorithm design does not assume knowledge of the social graph and the influence probability distributions, nor requires any initialization stage. Regardless of the concrete influence probabilities or the topology of the social network, an \(O(\sqrt{T} N \ln N)\) regret bound is rigorously proven where \(T\) is the number of time stages in the entire system span and \(N\) is the number of nodes. To the best of our knowledge, this is the first influence maximization algorithm dealing with both unknown and non-stationary influence probabilities. We evaluate practical effectiveness of the algorithm using both synthetic and real-world datasets, which demonstrates that our algorithm outperforms previous stationary multi-armed bandit algorithms under non-stationary conditions.

The rest of the paper is organized as follows. We discuss related work in Sec. \[III\] and present the problem model in Sec. \[III\]. In Sec. \[IV\] and Sec. \[V\] we present the detailed online algorithm and provide theoretical analysis of its regret bound. Simulation results are presented in Sec. \[VI\]. We conclude the paper in Sec. \[VII\].
further investigate restless bandits with Markov rewards [11], where the states of an arm evolve dynamically over time no matter whether it has been played. The algorithm utilizes regenerative property of a Markov chain and achieves a regret near logarithmic on the total number of time stages. Both studies rely on an initialization stage, in which each arm is tried for at least once. This is impractical for influence propagation (e.g., market campaign) in a large-scale network, as the cost of trying all nodes is unaffordable. Granmo et al. [25] use Kalman filter to update estimation of the reward distribution, and evaluate their results by simulation without theoretical analysis. Kalman filter is only applicable to linear dynamic system and the states inherently form a Markov chain. It is not realistic to make the Markov chain assumption in influence propagation, since human behavior does not simply depend on one’s latest status.

### III. Problem Model

We model the social network as an influence graph \( G = (\mathcal{N}, \mathcal{E}) \). \( \mathcal{N} = \{1, 2, \ldots, N\} \) is the set of users (nodes), where \( N \) is the total number of nodes. \( \mathcal{E} \) is the set of social connections among the nodes. An unknown influence probability \( p^t_{n,m} \) is associated with each edge \((n,m)\) \( \in \mathcal{E} \), which is time varying following an unknown, non-stationary distribution: after user \( n \) is activated (e.g., obtained information of a product), he may activate his neighbor \( m \) (e.g., share information of the product) with different probabilities at different time stages \( t \). In this way, each edge \((n,m)\) is associated with a non-stationary Bernoulli distribution: in \( t \), user \( n \) may activate his neighbor \( m \) with probability \( p^t_{n,m} \), or not with probability \( 1 - p^t_{n,m} \). We do not assume any cascade model of the information propagation system (e.g., independent cascade model or linear threshold model), and our algorithm works with various cascade models as long as the information spread brought by an activate node can be modeled as a random variable.

Let \( T \) be the total number of time stages that the system spans. In each time stage, a set of \( K \) seeds are selected as information sources (e.g., the seed users that an advertiser directly promotes the product to, whose number is decided by the promotion budget), from which the information spreads to other nodes in the network. The seed set is repeatedly selected over different time stages. For example, a company may carry out a promotion campaign for a series of time stages, e.g., a number of consecutive days. After the promotion in each time stage via a potentially different set of seeds, the company collects statistics on the number of purchases of their promoted product(s) and utilizes this feedback to update its seed selection strategies in later time stages. The goal is to maximize the expected overall influence spread in the whole time span \( 1, 2, \ldots, T \), i.e., the expected total number of activated nodes. Let \( \mathcal{M} \) be the collection of all subsets of \( \mathcal{N} \). In our bandit optimization framework, we define \( a|S \), meaning node \( a \) under a given set \( S \in \mathcal{M} \), as an arm. The expected reward of selecting an arm \( a|S \) is the expected marginal gain by adding \( a \) into the existing seed set \( S \), i.e., the expected additional number of activated users after we add \( a \) into \( S \). Let \( f_t(S) \) be an influence spread function in time stage \( t \), indicating the total number of activated nodes in \( t \) based on seed set \( S \). The value of \( f_t(S) \) is a random variable. The expectation \( \mathbb{E}[f_t(S)] \) is non-negative, monotone and submodular, as proven in [26]. The submodularity of the spread function is useful such that we can utilize the benchmark based on greedy optimal value. The expected reward of selecting an arm \( a|S \) in \( t \) is hence \( \mathbb{E} [ f_t(S \cup \{a\}) ] - \mathbb{E} [ f_t(S) ] \). Note that the expectation \( \mathbb{E} [ \cdot ] \) is taken over both randomized rewards and randomized policies, where a policy refers to the agent’s strategy for seed selection, which is random given the random nature of our algorithm.

In each time stage \( t \), starting from an empty set \( S_t = \emptyset \), we obtain a seed set of size \( K \) by adding nodes to \( S_t \) one by one in some order. Let \( S_t = (a^t_1, \ldots, a^t_K) \) be the completed seed set, in which the \( k^{\text{th}} \) element is the \( k^{\text{th}} \) seed selected in this time stage. Let \( S^t_{(1:k-1)} \) represent the selected seed set with elements \( 1, \ldots, k-1 \), and \( a^t_k | S^t_{(1:k-1)} \) mean that node \( a^t_k \) is selected as the \( k^{\text{th}} \) seed in \( t \) given previous choices in \( S^t_{(1:k-1)} \). Let \( \tilde{r}_t(k^t | S^t_{(1:k-1)}) = \mathbb{E} [ f_t(S^t_{(1:k-1)} \cup \{a^t_k\}) ] - \mathbb{E} [ f_t(S^t_{(1:k-1)}) ] \) denote the expected marginal gain of choosing \( a^t_k \) as the \( k^{\text{th}} \) seed in \( t \). The expected total reward in time stage \( t \) is \( \tilde{r}_t(S_t) = \sum_{k=1}^K \tilde{r}_t(k^t | S^t_{(1:k-1)}) = \mathbb{E} [ f_t(S_t) ] \).

In this model, maximizing the expected total number of activated nodes in \( 1, \ldots, T \) is equivalent to maximizing the expected overall reward in the entire span. \( \sum_{t=1}^T \tilde{r}_t(S_t) = \sum_{S \in \mathcal{M}} \sum_{t=1}^T \mathbb{E} [ f_t(S) ] \). It is further equivalent to minimizing the regret, the gap between the expected overall reward that the agent can obtain by running our online algorithm and the offline optimal expected overall reward computed using full knowledge of the system. In our algorithm design, we aim to minimize the weak regret, i.e., the gap between the expected overall reward achieved by our algorithm and the offline expected overall reward achieved by using the same best seed set \( S^* \) in all time stages, namely \( S^* \in \arg \max_{S \in \mathcal{M}} \sum_{t=1}^T \mathbb{E} [ f_t(S) ] \), computed based on full knowledge of the entire system. Such a weak regret is the difference between the expected overall reward obtained by our algorithm and that achieved by the best single action, i.e., sticking with one seed set in all time stages. Weak regret is commonly used in the literature on analysing non-stationary bandit algorithms [24][11][12][27], and the key ingredient is to form accurate estimates on the average condition for each arm [28], so as to find the arm performing best in a long term. In particular, we analyze a greedy weak regret, with detailed definition given in Definition 2 in Sec. V, that compares the expected overall reward produced by our algorithm with the lower bound of an approximate offline overall reward achieved by a single best seed set derived by a greedy approach. Greedy weak regret is a concept narrowed down from weak regret, when the best single action is decided by a greedy algorithm. We apply this notion so as to compare with the lower bound of the greedy optimal value.
### TABLE I: Notation

| Symbol | Description |
|--------|-------------|
| $N$    | # of nodes |
| $\mathcal{N}$ | the set of nodes |
| $\mathcal{M}$ | the collection of subsets of $\mathcal{N}$ |
| $T$    | the total number of time stages |
| $C$    | input parameter to Alg. 1 |
| $K$    | the size of seed set |
| $\gamma$ | input parameter to Alg. 1 |
| $S_t^{1:k-1}$ | the set containing the first $k - 1$ selected seeds in $t$ |
| $a|S_t^{1:k-1}$ | an arm in $t$, selecting node $a$ given $S_t^{1:k-1}$ |
| $f_t(S)$ | the influence spread of seed set $S$ in $t$ |
| $\bar{r}_t^k(a|S_t^{1:k-1})$ | reward of choosing $a$ as $k^{th}$ seed based on $S_t^{1:k-1}$ in $t$ |
| $\bar{r}_t^k(a|S_t^{1:k-1})$ | expected reward of choosing $a$ as $k^{th}$ seed based on $S_t^{1:k-1}$ in $t$ |
| $Reg_G(T)$ | greedy weak regret in the whole system span |
| $Reg^k(t)$ | position regret for the $k^{th}$ seed in $t$ |
| $a^*_t$ | selected node as $k^{th}$ seed in $t$ |
| $\alpha^*_t$ | optimal node as $k^{th}$ seed in all time stages |
| $E[\cdot]$ | expectation taken over both random policies and random rewards |
| OPT    | the offline maximal value of the expected overall reward |

### IV. RSB: RANDOMIZED MULTI-ARMED BANDIT ALGORITHM FOR NON-STATIONARY SOCIAL NETWORKS

**Main Idea.** We next design an online multi-armed bandit algorithm to minimize the greedy weak regret. In each time stage, we select the best seed set by sequentially selecting the next best node given previous seed decisions. Given the set of already selected seeds, we associate weights with candidate arms, and deal with the varying environment (time-varying underlying distributions of influence probabilities) by adjusting the weights of arms based on rewards received due to previous seed selection (the exploitation component of our algorithm). Besides, we also include a constant $\frac{\gamma}{T}$ in the weight of each arm, where $\gamma \in (0, 1]$ is a gaugeable value, in order to enable exploration of arms never tried before. Different from deterministic stationary bandit algorithms, our algorithm is randomized in arm selection according to the weights, and hence even if the environment changes abruptly, the algorithm still has a chance to switch to the new best arm.

**Algorithm Steps.** Our multi-armed bandit algorithm for selecting the best seed set in each time stage $t$ is given in Alg. 1. Here $w_t^n|S_t^{1:k-1}$ is the weight for selecting node $n$ as the $k^{th}$ seed in time stage $t$, while the set of already selected seeds in $t$ is $S_t^{1:k-1}$, $w_t^n|S_t^{1:k-1}$ is an auxiliary quantity to compute the weights, updated based on the past reward information of arm $n|S_t^{1:k-1}$, as an exploration measure. $q_t^n|S_t^{1:k-1}$ is the probability of playing arm $n|S_t^{1:k-1}$ in $t$, derived from the weights of the arms. $\bar{r}_t^k(a|S_t^{1:k-1})$ denotes the realization of the reward (actual marginal influence spread) by choosing node $a$ as the $k^{th}$ seed in $t$. $C$ is an input parameter to the algorithm, which satisfies $C \geq \frac{\gamma}{\text{OPT} \times \text{OPT}}$, $\forall n \in \mathcal{N}$, $S \in \mathcal{M}$.

In Alg. 1 the $K$ seeds are selected sequentially (line 5). The weights $w$ associated with the nodes should be equal at the beginning of each time stage, and adjusted based on past random rewards. We will evaluate the impact of the input parameter $\gamma$ under practical settings in simulations. The input parameter $C$ is related to the largest spread brought by a seed, which is unknown before running the algorithm. In fact, requiring $C \geq \frac{\gamma}{\text{OPT} \times \text{OPT}}$, only needed for regret analysis. We can set the value of $C$ empirically when running the algorithm in practice, and will evaluate the performance of the algorithm under an empirical value of $C$ in simulations, which does not necessarily satisfies the above condition.

We note that our algorithm does not rely on any knowledge of the underlying social network topology and the influence probabilities, but only utilizes the outcomes that are decided by them. In addition, although the entire space of arms, $a|S, \forall a \in \mathcal{N}$, $S \in \mathcal{M}$, is exponential, the number of arms that need to be dealt with in each time stage in Alg. 1 (weights and probabilities computed and used in seed selection) is still polynomial, as given in the following theorem.

**Theorem 1.** The time complexity of Alg. 1 executed in each time stage $t$, is $O(KN)$.

**Proof:** In each time stage $t$, we select $K$ seeds. When selecting the $k^{th}$ seed based on already selected seeds in $S_t^{1:k-1}$, we compute/update weights, and compute selection probabilities for at most $N$ arms corresponding to $N$ nodes in the network. Therefore, the time complexity is $O(KN)$. □
Algorithm 1 RSB: Randomized Sequential Multi-armed Bandit Algorithm for Non-Stationary Networks

Input: $N$, $K$, $C$, $\gamma$
Output: the seed set $S_t^{(1:k)}$ for each time stage $t$

1: set $v_1^{n|S_t^{(1:k-1)}} = 1, \forall n \in N, k = 1, \ldots, K$
2: for $t = 1, 2, \ldots, T$ do
3:   for $k = 1, 2, \ldots, K$ do
4:     for each node $n \in N$ do
5:       set $w_t^{n|S_t^{(1:k-1)}} = (1 - \gamma)\frac{v_t^{n|S_t^{(1:k-1)}}}{\sum_{n' \in N} v_t^{n'|S_t^{(1:k-1)}}} + \frac{\gamma}{N}$
6:     end for
7:   for each node $n \in N \setminus S_t^{(1:k-1)}$ do
8:     $q_t^{n|S_t^{(1:k-1)}} = \frac{w_t^{n|S_t^{(1:k-1)}}}{\sum_{n' \in N \setminus S_t^{(1:k-1)}} w_t^{n'|S_t^{(1:k-1)}}}$
9:   end for
10:   draw an arm $a_t|S_t^{(1:k-1)}$ according to the distribution $\{q_t^{n|S_t^{(1:k-1)}}\}_{n \in N \setminus S_t^{(1:k-1)}}$
11:   receive a reward $r_t^{k}(a_t|S_t^{(1:k-1)})$
12:   set $S_t^{(1:k)} = S_t^{(1:k-1)} \cup \{a_t\}$
13:   set $\hat{r}_t^{k}(a_t|S_t^{(1:k-1)}) = \frac{r_t^{k}(a_t|S_t^{(1:k-1)})}{q_t^{a_t|S_t^{(1:k-1)}}}$
14:   for all $n \in N \setminus \{a_t\}$, set $\hat{r}_t^{k}(n|S_t^{(1:k-1)}) = 0$
15:   for each arm $n_t|S_t^{(1:k-1)}$, $\forall n \in N$ do
16:     update $v_{t+1}^{n_t|S_t^{(1:k-1)}} = v_t^{n_t|S_t^{(1:k-1)}} \exp\left\{\frac{\gamma r_t^{k}(n_t|S_t^{(1:k-1)})}{NC}\right\}$
17:   end for
18: end for
19: end for

V. REGRET ANALYSIS

We next analyze an upper bound of the greedy weak regret achieved by Alg. 1. Let $OPT$ denote the offline optimal value of the expected overall reward $\sum_{t=1}^{T} \bar{r}_t(S_t) = \sum_{t=1}^{T} E[f_t(S)]$ over all $S \in \mathcal{M}$, computed based on complete knowledge of the influence probability distributions and the social graph topologies in $1, \ldots, T$. Let $S^*$ be the offline optimal seed set, i.e., the single best seed set that maximizes $\sum_{t=1}^{T} \bar{r}_t(S)$.

A. REDUCTION FROM GREEDY WEAK REGRET TO POSITION WEAK REGRET

We define a position optimal reward $OPT^k$ as the sum of the expected marginal gains achieved by using the best $k^{th}$ seed in all time stages. The best $k^{th}$ seed maximizes $\sum_{t=1}^{T} \hat{r}_t^{k}(a_t|S_t^{(1:k-1)})$ based on full knowledge of the system, given the first $k - 1$ seeds in $S_t^{(1:k-1)}$ in each $t$ derived using RSB. The idea is to reduce the original problem of finding the best solution of the full set to a parallel bandit setting, finding the best $k^{th}$ element under the condition determined by our algorithm. Let $\tilde{a}^k$ denote this optimal $k^{th}$ seed, i.e., $\tilde{a}^k \in \arg\max_{a \in N} \sum_{t=1}^{T} \hat{r}_t^{k}(a|S_t^{(1:k-1)})$. Such a best $k^{th}$ seed may form different arms, $\tilde{a}^k|S_t^{(1:k-1)}$, under different seed sets $S_t^{(1:k-1)}$ in different time stages. We have $OPT^k = \max_{a \in N} \sum_{t=1}^{T} \bar{r}_t^{k}(a|S_t^{(1:k-1)}) = \sum_{t=1}^{T} \bar{r}_t^{k}(\tilde{a}^k|S_t^{(1:k-1)})$.

Definition 1. The position weak regret for the $k^{th}$ seed is

$$\text{Reg}^k(T) = \sum_{t=1}^{T} \bar{r}_t^{k}(\tilde{a}^k|S_t^{(1:k-1)}) - \sum_{t=1}^{T} \bar{r}_t^{k}(a_t|S_t^{(1:k-1)})$$

where $\tilde{a}^k \in \arg\max_{a_t \in N} \sum_{t=1}^{T} \bar{r}_t^{k}(a_t|S_t^{(1:k-1)})$ and $a_t^k|S_t^{(1:k-1)}$ is the arm selected by Alg. 1 in time stage $t$. The conditional set $S_t^{(1:k-1)}$ is also decided by Alg. 2.

The following theorem states the relationship between position weak regret and $OPT$, which will be used to bound the greedy weak regret in Theorem 3. Its proof can be found in Appendix A.

Theorem 2. For any position $k = 1, 2, \ldots, K$, we have

$$\sum_{t=1}^{T} \left( \bar{r}_t(S_t^{(1:k)}) - \bar{r}_t(S_t^{(1:k-1)}) \right) \leq \frac{1}{k} \left( OPT - \sum_{t=1}^{T} \bar{r}_t(S_t^{(1:k-1)}) \right) - \text{Reg}^k(T).$$

Let $F(S) = \sum_{t=1}^{T} E[f_t(S)]$, $\forall S \in \mathcal{M}$, which denotes the expected overall influence spread over the whole system span. $F(S)$ is a submodular function since it is the summation of submodular functions $E[f_t(S)]$, $\forall t = 1, \ldots, T$. Then we can design a greedy approach to compute a $S$ that approximately maximizes the expected overall reward $\sum_{t=1}^{T} \bar{r}_t(S) = \sum_{t=1}^{T} E[f_t(S)]$ based on full knowledge of the system: after deciding $S_t^{(1:k-1)}$, we select a local optimal node as the $k^{th}$ seed, that maximizes the expected marginal influence spread, i.e., node $u$ such that $u \in \arg\max_{v \in N \setminus S_t^{(1:k-1)}} \left\{ \frac{1}{v} \right\} - F(S_t^{(1:k-1)})$. We can easily prove that the approximate offline solution computed this way achieves an approximation ratio of $1 - \frac{1}{e}$, i.e., the expected overall reward it achieves is at least $(1 - \frac{1}{e})OPT$, following Theorem 3.5 in [26], based on submodularity of the spread function and local optimality when selecting each seed. The reason that we compute this approximate offline solution using the greedy approach (which runs in polynomial time) is that computing $S^*$ has been shown to be an NP hard problem.

Using the approximate offline overall reward computed as above, we define a greedy weak regret as follows, which we use to evaluate the performance of our algorithm RSB.

Definition 2. The greedy weak regret is defined as the gap between the lower bound of the approximate offline overall reward derived by the greedy approach and the expected overall reward produced by RSB in Alg 1, i.e.,

$$\text{Reg}_{GR}(T) = (1 - \frac{1}{e})OPT - \sum_{t=1}^{T} \bar{r}_t(S_t^{(1:k)})$$.
The following theorem shows that the overall position weak regret provides an upper bound of the greedy weak regret. The proof can be found in Appendix B.

**Theorem 3.** The greedy weak regret is upper bounded by the sum of position weak regrets over all positions \( k = 1, 2, \ldots, K \), i.e.,

\[
R_T^G \leq \sum_{k=1}^{K} R(T).
\]

Based on Theorem 3, we seek to bound the position weak regret for each \( k \), in order to derive an upper bound of \( R_T^G \).

**B. Bounding Greedy Weak Regret**

According to Definition 1, the position weak regret for the \( k \)-th seed is

\[
R_T^k = \sum_{t=1}^{T} \left( \max_{n \in S_T^k} r_T^k(n) - \sum_{n \in S_T^k} r_T^k(n) \right).
\]

Let \( D \) be the upper bound of the realization of reward, i.e., \( r_T^k(n) \leq D \). \( \forall n \in S_T^k \), \( S_T^k \subseteq S \subseteq M \). The following theorem states an upper bound of the position weak regret for each \( k \).

In particular, if \( \gamma \) is set to a special value, it can minimize the regret bound. The proof can be found in Appendix C.

**Theorem 4.** Let \( R_T^k = \max_{n \in S_T^k} \sum_{t=1}^{T} r_T^k(n) S_T^k(1:k-1) \) be the expected overall reward achieved by selecting the best \( k \)-th arm given \( S_T^k(1:k-1) \), \( \forall t = 1, \ldots, T \), derived by Alg. 2. Let \( R_k^{RSB} = \sum_{t=1}^{T} \mathbb{E}[r_T^k(a_T^k|S_T^k(1:k-1))] \) denote the expected overall marginal gain obtained by adding the \( k \)-th seeds into the given \( S_T^k(1:k-1) \), \( \forall t = 1, \ldots, T \). For any parameter \( \gamma \) \( \in (0, 1] \), we have

\[
R_T^k = R_T^k - R_k^{RSB} \leq (1 + (e - 2) \frac{D}{C}) \gamma R_T^k + \frac{NC \ln N}{\gamma}.
\]

If we set \( \gamma = \min \{ 1, \sqrt{\frac{NC \ln N}{1 + (e - 2) \frac{D}{C}}} \} \) where constant \( g \geq R_T^{max} \), \( \forall k = 1, \ldots, K \), we have the following minimum upper bound

\[
\sum_{k=1}^{K} R_T^k \leq 2K \sqrt{1 + (e - 2) \frac{D}{C} \sqrt{gCN \ln N}}.
\]

**Corollary 1.** The greedy weak regret achieved by Alg. 2 is upper bounded as follows:

\[
R_T^G \leq 2K \sqrt{1 + (e - 2) \frac{D}{C} \sqrt{DCTN \ln N}},
\]

i.e., the upper bound of the greedy weak regret of Alg. 2 is \( O(\sqrt{T \ln N}) \).

It shows that our greedy weak regret is sublinear with both \( N \) and \( T \). The proof can be found in Appendix D.
a confidence bound with each arm and chooses the arm with the highest upper confidence bound greedily.

We note that although a number of bandit algorithms have been proposed for influence maximization (as discussed in Sec. [1][I-A]), most are not directly comparable since they run with the complete knowledge of a social network. We compare with OG-UCB since it is the only existing bandit algorithm without requiring knowledge of the social graph topology. In addition, the bandit algorithms designed for non-stationary systems in Sec. [1][II-B] either deal with 1 arm or assume Markov rewards, and hence cannot be readily extended for comparison.

In computing greedy weak regret, we also compute the approximate offline optimal overall reward by the greedy offline algorithm discussed before Defn. 2 in Sec. [1][V].

D. Evaluation Results

To show greedy weak regret values in a unified range in our figures, we plot the ratio between greedy weak regret and the approximate offline optimal overall reward, i.e., approx. offline opt. overall reward/overall reward by RSB.

Especially, a data point at a specific T represents the average ratio computed using overall rewards in [1, T]. We set K = 5, γ = 0.2 (default), D = 120 and C = 1 in our experiments.

Fig. 5 show the results obtained using synthetic data or Tencent Weibo traces under different time-varying models of influence probabilities. We observe that the regret ratios (and hence information spread) achieved by RSB and the random algorithm are usually similar at the early stages of the system, when RSB has not cumulated much feedback. RSB gets better than the other algorithms (lower regret and hence better spread) after more time stages, validating that RSB can improve with more feedback received from the real system. Besides, OG-UCB performs the worst especially with the ongoing of time, showing that it is only suitable for fixed influence probability distributions and does not work well in cases of time-varying influence probabilities. The increase of cumulative regret by RSB with the increase of time stages, if any, is always slower than that of the other algorithms.

In Fig. 5 we compare the regret ratios of RSB achieved under different values of input parameter γ, using Tencent Weibo traces under the FR model. From line 5 of Alg. 1, we can see γ = 0 represents pure exploitation and γ = 1 indicates pure exploration. Although Theorem 4 requires γ ≥ 0 for the bound to be meaningful, we can still test the extreme case that γ = 0 when running the algorithm in practice. RSB performs worst in these extreme cases, γ max = 0.18 is computed following the formula in Theorem 4 which minimizes the theoretical upper bound. We observe that γ max achieves lowest regrets in actual execution of our algorithm under practical settings as well.

In Fig. 6 we evaluate the impact of different graph sizes N, by extracting subgraphs of different sizes using Tencent Weibo traces. We observe that the regret is larger in larger networks, but it always improve when the system runs for longer period of time.

VII. Conclusion

This paper investigates online influence maximization in dynamic social networks with non-stationary influence probability distributions among participants. We design a randomized algorithm based on multi-armed bandit optimization to guide source selection for information dissemination over multiple time stages, aiming to maximize the overall spread over the system span. The algorithm is simple and neat, relying on carefully designed, continuously updating preferences on seed selection, which exploit real-world feedback from previous decisions, as well as explore new choices. As the first in the literature, the algorithm does not require knowledge of the dynamic social graph topology, nor time-varying influence probabilities, but is able to achieve an upper-bounded weak regret, as compared to an approximate offline optimal reward. Simulations based on both synthetic and real-world datasets further validate that our algorithm is more adaptive to a changing environment than heuristic and stationary bandit algorithms. In addition, our algorithm is also applicable to many other real-world problems such as advertisement placement [9], as long as the reward functions are submodular or there exists an approximate offline algorithm that can achieve an approximation ratio of (1−1/e). In future work, we seek to apply similar algorithms to solve the other real-world problems.

REFERENCES

[1] D. Kempe, J. Kleinberg, and É. Tardos, “Maximizing the Spread of Influence through a Social Network,” in Proc. of the 9th ACM SIGKDD, 2003, pp. 137–146.
[2] M. Richardson and P. Domingos, “Mining Knowledge-Sharing Sites for Viral Marketing,” in Proc. of the 8th ACM SIGKDD, 2002, pp. 61–70.
[3] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, “Cost-effective Outbreak Detection in Networks,” in Proc. of the 13th ACM SIGKDD, 2007, pp. 420–429.
[4] A. Goyal, W. La, and L. V. Lakshmanan, “CELF++: Optimizing the Greedy Algorithm for Influence Maximization in Social Networks,” in Proc. of the 20th WWW. ACM, 2011, pp. 47–48.
[5] Y. Tang, X. Xiao, and Y. Shi, “Influence Maximization: Near-Optimal Time Complexity Meets Practical Efficiency,” in Proc. of ACM SIGMOD, 2014, pp. 75–86.
[6] Y. Tang, Y. Shi, and X. Xiao, “Influence Maximization in Near-Linear Time: A Martingale Approach,” in Proc. of ACM SIGMOD, 2015, pp. 1539–1554.
[7] P. Auer, N. Cesa-Bianchi, and P. Fischer, “Finite-time Analysis of the Multiarmed Bandit Problem,” Machine learning, vol. 47, no. 2-3, pp. 235–256, 2002.
[8] S. Bubeck, “Regret Analysis of Stochastic and Nonstochastic Multiarmed Bandit Problems,” Foundations and Trends® in Machine Learning, vol. 5, no. 1, pp. 1–122, 2012.
[9] W. Chen, Y. Wang, and Y. Yuan, “Combinatorial Multi-Armed Bandit: General Framework, Results and Applications,” in Proc. of the 30th ICML, 2013, pp. 151–159.
[10] L. Qin, S. Chen, and X. Zhu, “Contextual Combinatorial Bandit and its Application on Diversified Online Recommendation,” in Proc. of SIAM International Conference on Data Mining, 2014.
[11] Y. Gai, B. Krishnamachari, and M. Liu, “Online Learning for Combinatorial Network Optimization with Restless Markovian Rewards,” in Proc. of the 9th IEEE SECON, 2012, pp. 28–36.
[12] C. Tekin and M. Liu, “Online Learning in Opportunistic Spectrum Access: A Restless Bandit Approach,” in IEEE INFOCOM, 2011, pp. 2462–2470.
[13] S. Lei, S. Maniu, L. Mo, R. Cheng, and P. Senellart, “Online Influence Maximization,” in Proc. of the 21th ACM SIGKDD, 2015, pp. 645–654.
Fig. 1: Synthetic data and WC model.
Fig. 2: Synthetic data and TR model.
Fig. 3: Synthetic data and FR model.

Fig. 4: Tencent Weibo trace and FR model: $\gamma = 0.2$.
Fig. 5: Tencent Weibo trace and FR model: different values of $\gamma$.
Fig. 6: Tencent Weibo trace and TR model: different graph sizes $N$.

[14] S. Vaswani, L. Lakshmanan, and M. Schmidt, “Influence Maximization with Bandits,” in NIPS Workshop on Networks in the Social and Information Sciences, 2015.

[15] S. Bubeck and A. Slivkins, “The Best of Both Worlds: Stochastic and Adversarial Bandits,” in The 25th Annual Conference on Learning Theory (COLT), 2012.

[16] A. Garivier and E. Moulines, “On Upper-Confidence Bound Policies for Switching Bandit Problems,” in Algorithmic Learning Theory. Springer, 2011, pp. 174–188.

[17] L. Massoulié, M. I. Ohannessian, and A. Prouixière, “Greedy-Bayes for Targeted News Dissemination,” in Proc. of ACM SIGMETRICS, 2015, pp. 285–296.

[18] Y. Yue and C. Guestrin, “Linear Submodular Bandits and their Application to Diversified Retrieval,” in NIPS, 2011, pp. 2483–2491.

[19] T. Lin, J. Li, and W. Chen, “Stochastic Online Greedy Learning with Semi-bandit Feedbacks,” in NIPS, 2015, pp. 352–360.

[20] A. Garivier and E. Moulines, “On Upper-Confidence Bound Policies for Non-Stationary Bandit Problems,” arXiv preprint arXiv:0805.3415, 2008.

[21] J. Y. Yu and S. Mannor, “Piecewise-stationary Bandit Problems with Side Observations,” in Proc. of the 26th ICML, 2009, pp. 1177–1184.

[22] O. Besbes, Y. Gur, and A. Zeevi, “Stochastic Multi-Armed-Bandit Problem with Non-stationary Rewards,” in NIPS, 2014, pp. 199–207.

[23] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire, “The Non-stochastic Multi-armed Bandit Problem,” SIAM Journal on Computing, vol. 32, no. 1, pp. 48–77, 2003.

[24] Y. Gai, B. Krishnamachari, and M. Liu, “On the Combinatorial Multi-Armed Bandit Problem with Markovian Rewards,” in IEEE GLOBECOM, 2011, pp. 1–6.

[25] O.-C. Granmo and S. Berg, “Solving Non-Stationary Bandit Problems by Random Sampling from Sibling Kalman Filters,” in Trends in Applied Intelligent Systems. Springer, 2010, pp. 199–208.

[26] W. Chen, L. V. Lakshmanan, and C. Castillo, “Information and Influence Propagation in Social Networks,” Synthesis Lectures on Data Management, vol. 5, no. 4, pp. 1–177, 2013.

[27] H. Liu, K. Liu, and Q. Zhao, “Learning in a Changing World: Restless Multiarmed Bandit With Unknown Dynamics,” Information Theory, IEEE Transactions on, vol. 59, no. 3, pp. 1902–1916, 2013.

[28] Y. Liu and M. Liu, “An Online Approach to Dynamic Channel Access and Transmission Scheduling,” in Proc. of the 16th ACM MobiHoc, 2015, pp. 187–196.

[29] X. Yi, F. Liu, J. Liu, and H. Jin, “Building a Network Highway for Big Data: Architecture and Challenges,” Network, IEEE, vol. 28, no. 4, pp. 5–13, 2014.

[30] H. Zhuang, Y. Sun, J. Tang, J. Zhang, and X. Sun, “Influence Maximization in Dynamic Social Networks,” in Proc. of the 13th IEEE ICDM, 2013, pp. 1313–1318.

[31] W. Chen, C. Wang, and Y. Wang, “Scalable Influence Maximization for Prevalent Viral Marketing in Large-Scale Social Networks,” in Proc. of the 16th ACM SIGKDD, 2010, pp. 1029–1038.

[32] O. Besbes, Y. Gur, and A. Zeevi, “Optimal Exploration-Exploitation in a Multi-Armed-Bandit Problem with Non-stationary Rewards,” arXiv preprint arXiv:1405.3316, 2014.

**APPENDIX A**

**PROOF OF THEOREM 2**

**Proof:** At any time $t$, given fixed $S^*_t(1:k−1)$, there exists a node $a ∈ S^*$ so that $a ∈ \arg\max_{a ∈ S^*} \sum_{t=1}^{T} \bar{f}_t(a | S^*_t(1:k−1))$. Then $a$ can satisfy the following inequality.

$$\sum_{t=1}^{T} \bar{f}_t(a | S^*_t(1:k−1)) = \sum_{t=1}^{T} \left( \bar{f}_t(S^*_t(1:k−1) \cup \{a\}) - \bar{f}_t(S^*_t(1:k−1)) \right)$$

$$\geq \frac{1}{K} \left( \sum_{t=1}^{T} \bar{f}_t(S^* \cup S^*_t(1:k−1)) - \sum_{t=1}^{T} \bar{f}_t(S^*_t(1:k−1)) \right)$$

(1)
which is the expected value of marginal gain by adding a conditional set \( \tilde{a} \) of random policy’s actions, the optimal solution is deterministic over the whole time period, its marginal reward is equal or larger than the mean value of all nodes belonging to \( S^* \).

Note that although \( \tilde{r}_t^k(a)S_t(1:k-1) \) is the expectation taken of random policy’s actions, the optimal solution is deterministic thus \( \tilde{r}_t^k(a)S_t(1:k-1) \) reduces to the expectation of random reward only here.

Let \( \tilde{a}_k \) be the selected seed with full information fixing \( S_t(1:k-1) \), i.e., it maximizes the total marginal gain which is equal or larger than \( \sum_{t=1}^{T} f_k(a)S_t(1:k-1) \), \( \forall a \in N \) under the conditional set \( S_t(1:k-1) \). Thus we have

\[
\sum_{t=1}^{T} \tilde{r}_t^k(\tilde{a}_k|S_t(1:k-1)) \geq \frac{1}{K} \left( \sum_{t=1}^{T} \tilde{r}_t(S^*) - \sum_{t=1}^{T} \tilde{r}_t(S_t(1:k-1)) \right).
\]

Define \( \Delta(\tilde{r}_t^k) = \tilde{r}_t(S_t(1:k)) - \tilde{r}_t(S_t(1:k-1)) \). Noting that

\[
\sum_{t=1}^{T} \tilde{r}_t^k(\tilde{a}_k|S_t(1:k-1)) - Reg^k(T) = \sum_{t=1}^{T} \tilde{r}_t^k(a_k|S_t(1:k-1))
\]

which is the expected value of marginal gain by adding \( a_k \) to \( S_t(1:k-1) \). This implies that

\[
\sum_{t=1}^{T} \Delta(\tilde{r}_t^k) = \sum_{t=1}^{T} \tilde{r}_t^k(\tilde{a}_k|S_t(1:k-1)) - Reg^k(T).
\]

Then we have

\[
\sum_{t=1}^{T} \Delta(\tilde{r}_t^k) \geq \frac{1}{K} \left( OPT - \sum_{t=1}^{T} \tilde{r}_t(S_t(1:k-1)) \right) - Reg^k(T).
\]

**APPENDIX B**

**PROOF OF THEOREM 3**

**Proof:** We prove the following inequality for each position \( k \) by induction.

\[
OPT - \sum_{t=1}^{T} \tilde{r}_t(S_t(1:k)) \leq (1 - \frac{1}{K})^k OPT + \sum_{m=1}^{k} Reg^m(T)
\]

(2)

The base case \( k = 0 \) is trivial. In the induction, let

\[
Z_k = OPT - \sum_{t=1}^{T} \tilde{r}_t(S_t(1:k)) = OPT - \sum_{t=1}^{T} \sum_{m=1}^{k} \Delta(\tilde{r}_t^m).
\]

Thus \( Z_k = Z_{k-1} - \sum_{t=1}^{T} \Delta(\tilde{r}_t^k) \).

According to Theorem 2, we know that

\[
\sum_{t=1}^{T} \Delta(\tilde{r}_t^k) \geq \frac{1}{K} Z_{k-1} - Reg^k(T).
\]

Then we have \( Z_k \leq (1 - \frac{1}{K}) Z_{k-1} + Reg^k(T) \).

Combining with the induction hypothesis, we can obtain the inequality. By taking \( k = K \) and using \((1 - \frac{1}{K})K < \frac{1}{e}\), we have

\[
\sum_{t=1}^{T} \tilde{r}_t(S_t(1:K)) \geq (1 - \frac{1}{e}) OPT - \sum_{k=1}^{K} Reg^k(T).
\]

**APPENDIX C**

**PROOF OF THEOREM 4**

**Proof:** We show the following trivial facts derived from the definitions where \( a(S_t(1:k-1)) \) is the selected arm in Alg. [1]

\[
\sum_{n=1}^{N} w_t^{n(S_t(1:k-1))} \geq w_t^{(S_t(1:k-1))}, \forall n \in N \setminus S_t(1:k-1)
\]

(3)

\[
\sum_{n=1}^{N} w_t^{n(S_t(1:k-1))} \tilde{r}_t(n|S_t(1:k-1))^2
\]

\[
\leq w_t^{(S_t(1:k-1))} \tilde{r}_t(a|S_t(1:k-1)) \tilde{r}_t(a|S_t(1:k-1))
\]

(4)

We will prove the inequality under any position \( k \). In the end we will illustrate that it still holds for the whole \( K \) size seed set. Under the conditional set \( S_t(1:k-1) \), \( \forall t = 1, 2, \ldots, T \).

Let \( V_t = \sum_{n=1}^{N} v_t^{n(S_t(1:k-1))} \). Then for all actions in Algorithm 1 we have

\[
\frac{V_{t+1}}{V_t} = \frac{\sum_{n=1}^{N} v_t^{(S_t(1:k-1))}}{V_t} = \frac{\sum_{n=1}^{N} w_t^{n(S_t(1:k-1))} \exp \left( \frac{\gamma \tilde{r}_t(n|S_t(1:k-1))}{NC} \right)}{V_t}
\]

(5)

\[
\leq \frac{\sum_{n=1}^{N} w_t^{n(S_t(1:k-1))} \exp \left( \frac{\gamma \tilde{r}_t(n|S_t(1:k-1))}{NC} \right)}{V_t}
\]

(6)
The inequality \(3\) uses the fact that \(e^x \leq 1 + x + (e - 2)x^2\) for \(x \leq 1\) and the inequality \(6\) is derived by the facts \(3\) and \(4\). Using the inequality \(1 + x \leq e^x\) and taking logarithms, we have

\[
\ln \frac{V_{t+1}}{V_t} \leq \frac{\gamma}{\text{NC}} \bar{r}_k(a_i|S_t^{(1:k-1)}) + \frac{(e - 2)(\frac{\gamma}{\text{NC}})^2 D}{1 - \gamma} \sum_{n=1}^{N} \bar{r}_k(n|S_t^{(1:k-1)}).
\]

Let \(\bar{r}_k^{\text{RSB}} = \sum_{t=1}^{T} \bar{r}_k(a_i|S_t^{(1:k-1)})\) and thus \(R_k^{\text{RSB}} = \mathbb{E}[\bar{r}_k^{\text{RSB}}]\). Then summing over \(t\), we can get all reward obtained by the agent over period \(1, \ldots, T\) for the position \(k\) as follows.

\[
\ln \frac{V_{T+1}}{V_1} \leq \frac{\gamma}{\text{NC}} \sum_{t=1}^{T} \bar{r}_k(n|S_t^{(1:k-1)}) - \ln N.
\]

For any node \(n_j \in N\) whatever the agent selects it, we have

\[
\ln \frac{V_{T+1}}{V_1} \geq \frac{\gamma}{\text{NC}} \sum_{t=1}^{T} \bar{r}_k(n_j|S_t^{(1:k-1)}) - \ln N.
\]

Since \(v_{n_j}^{\gamma}(S_0^{(1:k-1)}) = v_{n_j}^{\gamma}(S_1^{(1:k-1)}) \prod_{t=1}^{T} \exp \left\{ \frac{\gamma r_k(n_j|S_t^{(1:k-1)})}{\text{NC}} \right\}\) and \(v_{n_j}^{\gamma}(S_1^{(1:k-1)}) = 1\), we can derive the following inequality.

\[
\ln \frac{V_{T+1}}{V_1} \geq \frac{\gamma}{\text{NC}} \sum_{t=1}^{T} \bar{r}_k(n_j|S_t^{(1:k-1)}) - \ln N.
\]

Based on the inequalities above, we can derive

\[
R_k^{\text{RSB}} \geq (1 - \gamma) \sum_{t=1}^{T} \bar{r}_k(n_j|S_t^{(1:k-1)}) - \frac{\text{NC} \ln N}{\gamma} - (e - 2)\frac{\gamma D}{\text{NC}} \sum_{t=1}^{T} \sum_{n=1}^{N} \bar{r}_k(n|S_t^{(1:k-1)}).
\]

Then we can get

\[
R_k^{\text{RSB}} \geq (1 - \gamma) \sum_{t=1}^{T} \bar{r}_k(n_j|S_t^{(1:k-1)}) - \frac{\text{NC} \ln N}{\gamma} - (e - 2)\frac{\gamma D}{\text{NC}} \sum_{t=1}^{T} \sum_{n=1}^{N} \bar{r}_k(n|S_t^{(1:k-1)}).
\]

Since over the whole period, the expected marginal gain of any node \(n_j \in N\) is no larger than that of the best seed, which is \(R_k^{\text{RSB}}\), it is apparent \(\sum_{t=1}^{T} \sum_{n=1}^{N} \bar{r}_k(n|S_t^{(1:k-1)}) \leq N R_k^{\text{RSB}}\). Since node \(n_j\) is chosen arbitrarily, we can choose it as the best seed \(\hat{a}_k\) under the conditional set \(S_t^{(1:k-1)}\), thus we have \(\sum_{t=1}^{T} \bar{r}_k(\hat{a}_k|S_t^{(1:k-1)}) = R_k\). Combined with these two results, we have

\[
R_k^{\text{RSB}} \leq (1 + (e - 2)\frac{D}{C})\gamma R_k^{\text{RSB}} + \frac{\text{NC} \ln N}{\gamma}.
\]

Note that the expectation above is under any conditional set \(S_t^{(1:k-1)}, \forall t = 1, 2, \ldots, T\), which is related to previous choices for position \(1, \ldots, k - 1\). This is consistent with Definition \(\ref{def:optimal-choice}\) that \(S_t^{(1:k-1)}\) is decided by Alg. \(\ref{alg:optimal-choice}\). The expectation in \(R_k^{\text{RSB}}\) is also reduced to randomizing on reward only.

Since \(\sum_{k=1}^{K} R_k^{\text{RSB}} \leq gK\), summing up for all positions \(1, 2, \ldots, K\), we can get

\[
\sum_{k=1}^{K} R_k^{\text{RSB}} \leq (1 + (e - 2)\frac{D}{C})gK + \frac{\text{NC} \ln N}{\gamma}.
\]

Taking the first derivative of the right part of inequality \(\ref{eq:optimal-choice}\), we can set \(\gamma = \min\{1, \sqrt{\frac{\text{NC} \ln N}{(1 + (e - 2)\frac{D}{C})g}}\}\) to minimize the bound, then in the period \(1, \ldots, T\)

\[
\sum_{k=1}^{K} \text{Reg}^k(T) = \sum_{k=1}^{K} R_k^{\text{RSB}} - \sum_{k=1}^{K} R_k^{\text{RSB}} \leq 2K \sqrt{1 + (e - 2)\frac{D}{C}}\sqrt{\text{gCN} \ln N}.
\]

Note that if \(\gamma = 1\), we have \(\sqrt{\text{NC} \ln N} \geq \sqrt{1 + (e - 2)\frac{D}{C}}\). The bound \(\ref{eq:optimal-choice}\) is larger than the maximal reward \(gK\), then it must holds.

\section*{Appendix D}

\section*{Proof of Corollary \(\ref{cor:optimal-choice}\)}

\textbf{Proof:} It is apparent that the expected number of activated nodes from a seed can not exceed \(C\), the size of the largest connected component in the social graph. Then the overall reward achieved over the entire system span can not exceed \(CT\). Thus we can set \(g = CT\). According to Theorems \(\ref{thm:optimal-choice}\) and \(\ref{thm:optimal-choice}\) we have

\[
\text{Reg}_G(T) \leq \sum_{k=1}^{K} \text{Reg}^k(T) \leq 2K \sqrt{1 + (e - 2)\frac{D}{C}}\sqrt{\text{DCTN} \ln N}.
\]

Then we have \(\text{Reg}_G(T)\) is bounded by \(O(\sqrt{T N \ln N})\).