Fast Thompson Sampling Algorithm with Cumulative Oversampling: Application to Budgeted Influence Maximization

Shatian Wang, Shuoguang Yang, Zhen Xu, Van-Anh Truong
Department of Industrial Engineering & Operations Research
Columbia University, New York, NY 10025
sw3219,sy2614,zx2235,vt2196@columbia.edu

Abstract

We propose a cumulative oversampling (CO) technique for Thompson Sampling (TS) that can be used to construct optimistic parameter estimates using significantly fewer samples from the posterior distributions compared to existing oversampling frameworks. We apply CO to a new budgeted variant of the Influence Maximization (IM) semi-bandits with linear generalization of edge weights. Combining CO with the oracle we designed for the offline problem, our online learning algorithm tackles the budget allocation, parameter learning, and reward maximization challenges simultaneously. We prove that our online learning algorithm achieves a scaled regret comparable to that of the UCB-based algorithms for IM semi-bandits. It is the first regret bound for TS-based algorithms for IM semi-bandits that does not depend linearly on the reciprocal of the minimum observation probability of an edge. In numerical experiments, our algorithm outperforms all UCB-based alternatives by a large margin.

1 Introduction

The stochastic multi-armed bandit (MAB) problem is a classical problem that models the exploration and exploitation trade-off. There is a slot machine with \( m \) arms, each having an unknown reward distribution. In each round of a finite-horizon game, an agent pulls one arm and observes its realized reward. The agent aims to maximize the cumulative expected reward, or equivalently to minimize the cumulative regret over all rounds. To do so, she needs to not only learn the arms’ reward distribution by playing each arm a sufficient number of times (explore), but also exploit her current estimates on the reward distributions to make good arm selections. Two widely used methods to address the exploration-exploitation trade-off are Upper Confidence Bound (UCB) [5], and Thompson Sampling (TS) [22,7]. UCB-based algorithms maintain estimates on upper confidence bounds of the mean arm rewards and treat the bounds as nominal means when making decisions. TS-based algorithms maintain a belief on the distributions of the parameters to be learned. In each round, they randomly sample the parameters from the distributions and treat the sampled parameters as the nominal ones when making decisions. After observing feedbacks, the two types of algorithms both update empirical beliefs accordingly.

TS was proposed by [22] more than 80 years ago and has since demonstrated superior empirical performance compared to other state-of-the-art methods including UCB for different variants of MAB [7,12]. The theoretical guarantees for TS-based algorithms are however much more limited compared to those of the UCB family, mainly due to the difficulty in controlling the deviations resulting from random sampling. Not until 2012 were some progress made on the theoretical analysis of TS applied to the linear contextual bandit variant, in which each arm has an associated known \( d \)-dimensional feature vector and the expected reward of each arm is given by the dot product of the...
feature vector and an unknown global vector $\theta^* \in \mathbb{R}^d$. [4] considers TS as a Bayesian algorithm with a Gaussian prior on $\theta^*$ that is updated and sampled from in each round, and proves a regret in the order of $O(d^{3/2} \sqrt{T})$. Following the intuition of [4], [2] shows that sampling from an actual Bayesian posterior is not necessary, and the same order of regret (frequentist) is achievable as long as the distribution that TS samples from obeys suitable concentration and anti-concentration properties, which can be achieved by oversampling the standard least-squares confidence ellipsoid by a factor of $\sqrt{d}$. [13] considers an online dynamic assortment selection problem with contextual information. The authors assume a multinomial logit choice model, in which the utility of each item is given by the dot product of a $d$-dimensional context vector and an unknown global vector $\theta^*$. [13] proposes a TS-based algorithm that is inspired by the oversampling idea. In each round of their TS-based algorithm, an optimistic sample set of size at least $11 \log(K)$ is drawn from a least-squares confidence ellipsoid to construct the optimistic utility estimations of the items in the choice set, where $K$ is the number of items to choose for the assortment. Once the optimistic utility estimations are constructed, they feed them into an efficient oracle to solve for the optimal assortment given the estimations. This oversampling idea can be applied to online learning problems whose corresponding offline problems are easy to solve to optimality. However, for bandits with NP-hard offline problems, the regret analysis of TS-based algorithms remains challenging (detailed in Section 5).

In this paper, we propose a new online learning problem: Budgeted Influence Maximization Semi-Bandits with linear generalization of edge weights (Lin-IMB-L). It is a budgeted extension of the Influence Maximization (IM) semi-bandits ([15, 23, 19, 24, 25]). In IM-L, a social network is given as a directed graph with nodes representing users and directed edges indicating relationships. An edge from $a$ to $b$ means that user $b$ is a follower of user $a$ and influence (for example, in the form of product adoption) can spread from $a$ to $b$. Given a finite horizon consisting of $T$ rounds and a cardinality constraint $K$, an agent selects a seed set of $K$ nodes in each round to start an influence diffusion process that usually follows the Independent Cascade (IC) diffusion model [13]. Initially, all nodes in the seed set are activated. Then in each subsequent time step, each node activated in the previous time step has a single chance to activate its downstream neighbors, independently, with success probabilities equal to the edge weights. Each round terminates as soon as no more nodes are activated in a diffusion step. IM-L assumes that the edge weights are initially unknown. The agent seeks to choose the seed sets in a way to simultaneously learn the edge weights and maximize the expected cumulative number of activated nodes. In these problems, edge semi-bandit feedback is usually assumed, namely, for each node activated during the IC process, the agent observes whether its attempts to activate its followers succeed or not. If an attempt is successful, we say that the observed realization of the corresponding edge is a success; otherwise it is a failure. The agent learns the edge weights using the edge semi-bandit feedback. IM-L can be cast as combinatorial semi-bandits with probabilistically triggered arms (CMAB-prob) [9], in which a set of arms (as opposed to a single arm) are pulled in each round, and the reward associated with each pulled arm is observed. Furthermore, pulled arms can probabilistically trigger other arms to be played and the rewards of which to be observed. In IM-L, the arms pulled by the agent in each round are the edges leaving the chosen seed set, and the probabilistically triggered ones are edges that are not played but whose realizations are observed.

IM-L is a very hard learning problem. Even when no learning is involved and the edge weights are known, the corresponding offline problem of finding an optimal seed set of cardinality $K$ is NP-hard [13]. Since the expected number of activated nodes as a function of seed sets is monotone and submodular, the greedy algorithm achieves an $1 - 1/e$ approximation guarantee if the function values can be computed exactly [17]. However, because this function is #P-hard to compute, it requires simulation-based estimations to be estimated [8]. Existing learning algorithms for IM-L thus all assume the existence of an $(\alpha, \beta)$-approximation oracle that returns a seed set whose expected reward is at least $\alpha$ times the optimal with probability at least $\beta$, with respect to the input edge weights and cardinality constraint. These learning algorithms use UCB- or TS-based algorithms (or sometimes epsilon greedy) to estimate the edge weights and feed the updated estimates to the oracle in each step to get a seed selection ([15, 23, 19, 24, 25]). [25] is the first to scale up the learning algorithms by assuming linear generalization of edge weights. That is, each edge has an associated $d$-dimensional feature vector that is known by the agent, and the weight on each edge is given by the dot product of the feature vector and an unknown global vector $\theta^* \in \mathbb{R}^d$. With this assumption, they propose a UCB-based learning algorithm for IM-L that achieves a scaled regret in the order of $\tilde{O}(dC, \sqrt{mT})$.

$1\tilde{O}$ is a variant of the big $O$ notation that ignores all the logarithmic dependencies.
where $C_s$ is a network topology-dependent parameter upper bounded by $n \sqrt{m}$, $n$ is the number of nodes in the graph, and $m$ is the number of edges. Their result improves upon the existing regret bound in [9] that have a linear dependency on $1/p^*$, where $p^*$ is the minimum observation probability of an edge. $1/p^*$ can be exponential in the number of edges.

Although IM-L has been extensively studied in the past, there are still gaps to be filled. The most noticeable being that despite the superior empirical performance of TS-based algorithms for IM-L [7, 11, 12], few regret analysis exists for TS-based algorithms. [11] proposes a TS-based algorithm for CMAB-prob. Without assuming linear generalization of edge weights, their regret still depends linearly on $1/p^*$. Even with linear generalization, extending the UCB analysis of [25] to TS-based algorithms is highly non-trivial, the reason of which will be detailed later in the paper. In general, TS for online learning problems with NP-hard offline problems is hard to analyze.

Our contribution We propose a novel TS-based cumulative oversampling technique (CO) that can be applied to IM-L and potentially to many other bandits with NP-hard offline problems. CO is inspired by the oversampling idea in [2] and [18], but requires significantly fewer samples compared to [18]. Exactly one sample needs to be drawn from the standard least-squares confidence ellipsoid in each round. Our key idea is to utilize all the samples up to the current round to construct optimistic parameter estimations that asymptotically concentrate closely around the true parameters as tighter upper confidence bounds compared to the ones constructed with UCB-based methods.

We apply CO to a new online learning problem which we call Budgeted Influence Maximization Semi-Bandits with linear generalization of edge weights (Lin-IMB-L). In it, each node charges a different commission to be included in a seed set. There is also a global budget $B$ that needs to be satisfied in expectation together with a finite time horizon $T$. The agent needs to allocate the budget to $T$ rounds as well as learning edge weights and maximize cumulative reward. For this problem, we analyze its corresponding offline problem and propose the first $(\alpha, \beta)$-approximation oracle for it. To develop this oracle, we extend the state-of-the-art Reverse Reachable Sets (RRS) simulation techniques for IM [6, 20, 21] to accurately estimate the reward of seed sets of any size. We combine our cumulative oversampling technique with our oracle into an online learning algorithm for Lin-IMB-L. We prove that the scaled regret of our algorithm is in the order of $O(dC_s \sqrt{mT})$, which matches the regret bound for the UCB-based algorithm for IM-L with linear generalization of edge weights proved in [25]. We further conduct numerical experiments on Twitter subnetworks and show that our algorithm outperforms all UCB-based algorithms by a large margin with or without perfect linear generalization of edge weights.

## 2 Budgeted IM Semi-Bandits

We mathematically formulate our new budgeted IM semi-bandits problem in this section. We model the topology of a social network using a directed graph $D = (V, A)$. Each node $v \in V$ represents a user, and an arc (directed edge) $(u, v) \in A$ indicates that user $v$ is a follower of user $u$ in the network and influence can spread from user $u$ to user $v$. For each arc $e = (u, v)$, we use $\bar{w}(e) \in [0, 1]$ to denote the edge weight on $e$. There are in total $n$ nodes and $m$ arcs in $D = (V, A)$. Throughout the text, we refer to the function $\bar{w} : A \to [0, 1]$ as edge weights.

Once a seed set $S \subseteq V$ is selected, influence spreads in the network from $S$ following the Independent Cascade Model (IC) [13]. IC specifies an influence spread process in discrete time steps. In the initial step, all influencers in $S$ are activated. In each subsequent step $s$, each user activated in step $s - 1$ has a single chance to activate its followers, or downstream neighbors, with success rates equal to the corresponding edge edge weights. This process terminates when no more users can be activated. We can equivalently think of the IC model as flipping a biased coin on each edge and observing connected downstream neighbors has a single chance to activate its followers, or

More specifically, after the influencers in the seed set $S$ are activated, the environment decides on the binary weight function $w$ by independently sampling $w(e) \sim \text{Bern}(\bar{w}(e))$ for each $e \in A$. A node $v_2 \in V \setminus S$ is activated by a node $v_1 \in S$ if there exists a directed path $e_1, e_2, \ldots, e_l$ from $v_1$ to $v_2$ such that $w(e_i) = 1$ for all $i = 1, \ldots, l$. Let $I(S, w) = \{v \in V | v \in S$ or $v$ is activated by a node $u \in S$ under $w \}$ be the set of nodes activated during the IC process given seed set $S$. We denote the expected number of activated nodes given seed set $S$ and edge weights $\bar{w}$ by $\hat{f}(S, \bar{w})$, i.e., $\hat{f}(S, \bar{w}) := E(I(S, \bar{w}))$. And refer to the realization of $w(e)$ as the realization of edge $e$. 

Below, we formally define our Budgeted Influence Maximization Semi-Bandits with linear generalization of edge weights (Lin-IMB-L). In it, an agent runs an influencer marketing campaign over \( T \) rounds to promote a product in a given social network \( \mathcal{D} \). The agent is aware of the structure of \( \mathcal{D} \) but initially does not know the edge weights \( \bar{w} \). In each round \( t \), it activates a seed set \( S_t \) of influencers in the network by paying each influencer \( u \in S_t \) a fixed commission \( c(u) \in \mathbb{R}^+ \) to promote the product. Influence of the product spreads from \( S_t \) to other users in the network in round \( t \) according to the IC model. For each round \( t \), we assume that the influence spread process in this round terminates before the next round \( t + 1 \) is initiated. The total cost of selecting seed set \( S_t \) is denoted by \( c(S_t) = \sum_{u \in S_t} c(u) \). A exogenous budget \( B \) is given at the very beginning. The campaign selects seed sets with the constraint that in expectation, the cumulative cost over \( T \) rounds cannot exceed \( B \), where the expectation is over possible randomness of \( S_t \), since it can be returned by a randomized algorithm. The goal of the agent is to maximize the expected total reward over \( T \) rounds.

As in \([25]\), we assume a linear generalization of \( \bar{w} \). That is, for each arc \( e \in \mathcal{A} \), we are given a feature vector \( x_e \in \mathbb{R}^d \) that characterizes the arc. Also, there exists a vector \( \theta^* \in \mathbb{R}^d \) such that the edge weight on arc \( e \), \( w(e) \), is closely approximated by \( x_e^\top \theta^* \). \( \theta^* \) is initially unknown. The agent needs to learn it over the finite horizon of \( T \) rounds through edge semi-bandit feedback \([9, 25]\).

That is, for each edge \( e = (u, v) \in \mathcal{A} \), it observes the realization of \( w(e) \) in round \( t \) if and only if \( u \in I(S_t, w) \), i.e., the head of the edge was activated during the IC process in round \( t \). We refer to the set of edges whose realizations are observed in round \( t \) as the set of observed edges, and denote it as \( \mathcal{A}^e_t \). Depending on whether or not the tail node of an observed edge is activated, the realization of the edge can be either a success \((w(e) = 1)\), or a failure \((w(e) = 0)\).

**Problem 1 Budgeted Influence Maximization Semi-Bandits with linear generalization of edge weights (Lin-IMB-L)**

Given a social network \( \mathcal{D} = (\mathcal{V}, \mathcal{A}) \), edge feature vectors \( x_e \in \mathbb{R}^d \forall e \in \mathcal{A} \), cost function \( c : \mathcal{V} \to \mathbb{R}^+ \), budget \( B \in \mathbb{R}^+ \), finite horizon \( T \in \mathbb{Z}^+ \); assume \( \bar{w}(e) = x_e^\top \theta^* \) for some unknown \( \theta^* \in \mathbb{R}^d \), and one observes edge semi-bandit feedback in each round \( t = 1, ..., T \). In each round \( t \), adaptively choose \( S_t \subseteq \mathcal{V} \) so that

\[
\{ S_t \}_{t=1}^T \in \text{arg max} \left\{ \mathbb{E} \left[ \sum_{t=1}^T f(S_t, \bar{w}) \right] : \mathbb{E} \left[ \sum_{t=1}^T c(S_t) \right] \leq B \right\} .
\] (1)

Lin-IMB-L presents three challenges. First, the agent needs to learn the edge weights through learning \( \theta^* \) over a finite time horizon. Second, the agent needs to allocate the budget to individual rounds. Third, the agent needs to make a good seeding decision in each round that balances exploration (gathering more information on \( \theta^* \)) and exploitation (maximize cumulative reward using gathered information). Our online learning algorithm uses a novel TS-based cumulative oversampling (CO) technique to construct an optimistic (thus exploratory) estimate \( w : \mathcal{A} \mapsto [0, 1] \) on the edge weights \( \bar{w} \) in each round using the edge semi-bandit feedback gathered so far (thus exploitative). It then feeds \( w \) together with the budget \( b \) allocated to the current round to an approximation oracle in order to decide on a seed set for the current round. In the next section, we propose the first such approximation oracle and prove its approximation guarantee. Then in Section 4, we detail our online learning algorithm and the CO idea behind it.

### 3 Approximation Oracle

Assume that we have an estimate \( w \) on the edge weights \( \bar{w} \) and an expected budget \( b \) for the current round, an important subproblem of Lin-IMB-L is that in each round, we want to choose a seed set that maximizes the expected reward with respect to \( w \) while respecting the budget constraint \( b \). We refer to this subproblem as IMB and formally define it below.

**Problem 2 IMB**

Given network \( \mathcal{D} = (\mathcal{V}, \mathcal{A}) \), budget \( b \), cost function \( c \), edge weights \( w \), find \( S \subseteq \mathcal{V} \) such that \( \mathbb{E}(c(S)) \leq b \), and \( \mathbb{E}(f(S, w)) \) is maximized. The expectations are over possible randomness of \( S \), since \( S \) can be returned by a randomized algorithm.

IMB is NP-hard (see Section C.1 for a reduction of the minimum set cover problem to it). Its NP-hardness does not follow trivially from the NP-hardness of IM. Given an IM instance with cardinality
constraint $K$, consider the “corresponding” IMB instance with the cost of each node being 1 and the expected budget $B$ being $K$. It is worth noting that the optimal solution to this IMB instance might be a probability distribution over several disjoint seed sets: if we let $f^*(k)$ be the optimal reward of the original IM instance with cardinality constraint $k$, then $f^*(\cdot)$ is not necessarily “concave”, i.e., for some $k < l$, we might have $f^*(l) - f^*(l - 1) > f^*(k) - f^*(k - 1)$. As a result, it is possible that the IMB solution does not give any information on the best deterministic solution to the original IM instance.

Below, we propose the first approximation oracle for IMB. We refer to it as ORACLE-IMB. Further note that $f(\cdot, w)$ is #P hard to compute \cite{13}, and thus we need efficient simulative-based methods to accurately estimate it with high probability. We defer the estimation of $f(\cdot, w)$ to Section \cite{13} where we detail how to modify our oracle to incorporate the estimation of $f(\cdot, w)$ and prove that the resulting algorithm’s $(\alpha, \beta)$-approximation ratio.

**Algorithm 1: ORACLE-IMB**

Data: $D = (V, A), b, c, w$

Result: $S \subseteq V$

initialization: $S_0 = \emptyset$

for $i = 1, 2, ..., n$ do

Compute $v_i = \arg \max_{v \in V \setminus S_{i-1}} (f(S_{i-1} \cup \{v\}, w) - f(S_{i-1}, w))/c(v)$;

Set $S_i = S_{i-1} \cup \{v_i\}$;

if $c(S_i) > B$ then

Set $S_- = S_{i-1}, S_+ = S_i$;

Break

end

Solve the following LP to get an optimal solution $(p^*, q^*)$:

$$\max p \cdot f(S_-, w) + q \cdot f(S_+, w) \text{ s.t } p \cdot c(S_-) + q \cdot c(S_+) \leq b; \ p + q = 1; \ p, q \geq 0;$$

Sample $S$ from $\{S_-, S_+\}$ with probability distribution $(p^*, q^*)$

We have the following approximation guarantee result for ORACLE-IMB (proof in Section \cite{22}). Note that the existing approximation algorithm for budgeted monotone submodular function maximization with a deterministic budget needs to evaluate all seed sets of size up to 3 to achieve an $1 - 1/e$-approximation \cite{14}. Our ORACLE-IMB does not have this computationally expensive partial enumeration step. With an expected budget, we have the same approximation guarantee.

**Theorem 1** For any IMB instance, $\mathbb{E}(f(S^{ora}, w)) \geq (1 - 1/e)\mathbb{E}(f(S^{opt}, w))$, where $S^{ora}$ is the seed set returned by ORACLE-IMB and $S^{opt}$ is the seed set selected by an optimal algorithm.

4 Online Learning Algorithm for Lin-IMcB-L

To utilize Thompson Sampling to solve Lin-IMB-L, we maintain a belief on the distribution of $\theta^*$. We sample a $\theta$ from the belief in each decision round, treating it as the nominal mean when making the seeding decision. Finally, we update the belief based on edge semi-bandit realizations. Thompson Sampling, while demonstrating superior performance in experiments, is hard to analyze, mainly due to the difficulty in controlling the deviations resulting from random sampling. \cite{2} shows that for linear contextual bandits, sampling from an actual Bayesian posterior is not necessary, and the same order of regret (frequentist) is achievable as long as the distribution TS samples from follows suitable concentration and anti-concentration properties, which can be achieved by oversampling the standard least-squares confidence ellipsoid by a factor of $\sqrt{d}$. The oversampling step is used to guarantee that the estimates have a constant probability of being optimistic. \cite{13} extends this idea to a dynamic assortment optimization problem with MNL choice models. Their oversampling inspired TS algorithm uses $11 \cdot \log(K)$ samples from the least-squares confidence ellipsoid in each round to construct the optimistic utility estimates of the items in the choice set, where $K$ is the number of items in the assortment. For both linear contextual bandits and the dynamic assortment optimization with MNL choice models, the optimal “arm” with respect to the estimates can be efficiently computed.

However, the technique of oversampling a constant number of samples in each round is insufficient to guarantee a small regret for bandits with NP-hard offline problems, for which there exist only...
(α, β)-approximation oracles returning an α-approximation “arm” with probability at least β. We postpone the explanations of the challenges to Section 5.

We propose an alternative cumulative oversampling (CO) technique that can be applied to Lin-IMB-L and potentially to other bandits with NP-hard offline problems to obtain bounded small regrets and superior empirical performance. Under CO, we sample exactly one θi from the Gaussian distribution $N(\theta_i, v^2M_{t-1})$ where $\theta_i$ is the regularized least squares estimator and $M_{t-1}$ is the corresponding design matrix. However, we use all the t samples we have collected so far, with some rescaling, to construct an optimistic estimate $\tilde{u}_t : A \rightarrow [0, 1]$ for the true edge weights $\bar{w}$. The details of the resulting TS-CO algorithm are summarized in Algorithm 2. Besides requiring much fewer samples compared to existing oversampling techniques that use a fixed number of samples in each round [18], CO has two major benefits. The first is the nice concentration properties of the resulting edge weight estimate $\tilde{u}_t$. We show that $|\tilde{u}_t(e) - \bar{w}(e)| \leq O(1/t^2)$ for all $e \in A$ with probability at least $1 - O(1/t^2)$ (see Lemma 11), which implies that $\tilde{u}_t$ concentrates around $\bar{w}$ and matches the $O(\log t)$ estimation error in the UCB analysis [25]. Since t samples, $\theta_1, ..., \theta_t$, are used to construct $\tilde{u}_t$, as $t$ increases, the probability that $\tilde{u}_t$ is an upper bound on the true parameter $\bar{w}$ approaches one. $\tilde{u}_t$ therefore asymptotically concentrates closely around $\bar{w}$ as a tighter upper confidence bound. Second, CO practically preserves the advantages of both TS- and UCB-based algorithms: CO is similar to TS with oversampling in the initial learning rounds, whose superior empirical performance over other state-of-art methods such as UCB has been shown. As the number of rounds increases, the weight estimate $\tilde{u}_t$ serves as a tighter upper confidence bound that achieves smaller regrets. This CO technique sheds light on designing algorithms with small regret guarantees and superior empirical performance for other NP-hard problems. The exact proof for the asymptotic concentration of the estimators constructed using the cumulative samples might differ from problem to problem, but the general regret analysis outline shall be fairly similar to the one presented in the next section.

5 Regret Analysis

We first explain why the existing oversampling technique does not alleviate the challenges in regret analysis of TS-based algorithms for bandits with NP-hard offline problems in general and show how CO can be employed to tackle the challenges. We then present the regret results of TS-CO and provide a proof sketch.

For any bandits whose offline problem can be solved to optimality efficiently, we can analyze the regret as follows. Use $S^*(\bar{w})$ to denote the optimal action given parameter $\bar{w}$. In each round, action $S_t = S^*(\tilde{w}_t)$ is taken given the parameter estimate $\tilde{w}_t$. Let $\bar{w}$ be the true parameter. The cumulative

\begin{algorithm}
\caption{TS-CO}
\begin{algorithmic}
\State \textbf{Data:} digraph $D = (\mathcal{V}, \mathcal{A})$, node costs $c : \mathcal{V} \rightarrow \mathbb{R}^+$, edge feature vector $x_e \in \mathbb{R}^d$, number of rounds $T$, hyperparameter $v^2 \in \mathbb{R}^+$. 
\State \textbf{Result:} $S_t \subseteq \mathcal{V}, t = 1, ..., T$. 
\State \textbf{Initialization:} $\mathbf{M}_0 = I \in \mathbb{R}^{d \times d}$, $\mathbf{B}_0 = 0 \in \mathbb{R}^d$, $\tilde{w}_0(e) = -\infty$ for all $e$. \textbf{for} $t = 1, 2, ..., T$ \textbf{do} \State Set $\theta_t = \mathbf{M}_t^{-1} \mathbf{B}_t^{-1}$; \State Sample $\tilde{\theta}_t$ from $N(\theta_t, v^2\mathbf{M}_t^{-1})$; \State Compute $\tilde{w}_t(e) = \left[ \max \left( (\tilde{w}_{t-1}(e) - x_e^\top \theta_{t-1})||x_e||_{\mathbf{M}_{t-1}^{-1}} / ||x_e||_{\mathbf{M}_{t-1}^{-2}} + x_e^\top \theta_t, x_e^\top \theta_t \right) \right]$, \State $\hat{u}_t(e) = \text{Proj}_{[0,1]} \tilde{w}_t(e)$ for all $e$; \State $S_t \leftarrow \text{ORACLE-IMB}(D, \mathcal{B}/T, c, \hat{u}_t)$; \State Select seed set $S_t$ and observe semi-bandit edge activation realizations; \State Update $\mathbf{M}_t = \mathbf{M}_{t-1} + \sum_{e \in \mathcal{A}^e : y_e^t = 1} x_e x_e^\top$, $\mathbf{B}_t = \mathbf{B}_{t-1} + \sum_{e \in \mathcal{A}^e : y_e^t = 0} x_e x_e^\top$, where $\mathcal{A}_e$ is the set of edges whose realizations are observed, and $y_e^t \in \{0,1\}$ is the realization of edge $e$ in round $t$. \end{algorithmic}
\end{algorithm}
regret $R(T)$ can be written as $R(T) = \sum_{t=1}^{T} R_t$ where
\[
R_t = f(S^*(\bar{w}), \bar{w}) - f(S_t, \bar{w}) = f(S^*(\bar{w}), \bar{w}) - f(S^*(\bar{w}_t), \bar{w})
\]
\[
= f(S^*(\bar{w}), \bar{w}) - f(S^*(\bar{w}_t), \bar{w}_t) + f(S^*(\bar{w}_t), \bar{w}_t) - f(S^*(\bar{w}_t), \bar{w}).
\]
(2)

While bounding $R^2_t$ is relatively straightforward using standard linear bandits techniques, bounding $R^1_t$ requires more careful analysis. Intuitively, however, when $\bar{w}_t$ and $\bar{w}$ are close enough, the difference between their optimal rewards is likely to be small as well. Indeed, this has been done for the stochastic linear bandits and the assortment optimization settings using the constant optimistic probability achieved with oversampling \cite{18}.

On the other hand, when the underlying problem is NP-hard, an $(\alpha, \beta)$-approximation oracle has to be used in the learning algorithm. It takes the parameter estimate $\bar{w}_t$ as input and returns an action $S_t$ such that $f(S_t, \bar{w}_t) \geq \alpha \cdot f(S^*(\bar{w}_t), \bar{w}_t)$ with probability at least $\beta$. As a result, a scaled regret analysis is performed instead, where one is interested in bounding $R^1(T) = \sum_{t=1}^{T} R^1_t$, where
\[
R^1_t = \mathbb{E}\left[ f(S^*(\bar{w}), \bar{w}) - \frac{f(S_t, \bar{w})}{\eta} \right], \quad \eta = \alpha \beta.
\]
\[
= \mathbb{E}\left[ f(S^*(\bar{w}), \bar{w}) - \frac{f(S_t, \bar{w}_t)}{\eta} \right] + \frac{1}{\eta} \mathbb{E} [f(S_t, \bar{w}_t) - f(S_t, \bar{w})].
\]
(3)

Again, the difficulty mainly arises in bounding $R^1_t$. Use $S^0(w)$ to denote any solution such that $\mathbb{E}[f(S^0(w), w)] \geq \eta \cdot \mathbb{E}[f(S^*(w), w)]$. By definition of $S^0(w)$ and the property of the $(\alpha, \beta)$-approximation oracle, we can establish the following two upper bounds for $R^1_t$:
\[
R^1_t \leq \mathbb{E}[f(S^0(\bar{w}), \bar{w}) - f(S_t, \bar{w}_t)]/\eta,
\]
\[
R^1_t \leq \mathbb{E}[f(S^*(\bar{w}), \bar{w}) - f(S^*(\bar{w}_t), \bar{w}_t)].
\]
(4)
(5)

In Eq. (4), even when $\bar{w}_t = \bar{w}$, $R^1_t$ does not necessarily diminish to 0. This is because $\mathbb{E}[f(S^0(\bar{w}), \bar{w})] \geq \eta \cdot \mathbb{E}[f(S^*(\bar{w}), \bar{w})]$ and $\mathbb{E}[f(S_t, \bar{w}_t)] \geq \eta \cdot \mathbb{E}[f(S^*(\bar{w}_t), \bar{w}_t)]$ do not guarantee $\mathbb{E}[f(S^0(\bar{w}), \bar{w})] = \mathbb{E}[f(S_t, \bar{w}_t)]$. To bound the right hand side (RHS) of Eq. (5) is also challenging, because all the observations gathered by the agent are under action $S_t$ and $\bar{w}$. Losing the dependency on $S_t$ means the observations under $S_t$ cannot be utilized to construct a good upper bound. On the other hand, with CO, we can prove that $\bar{w}_t$ asymptotically concentrates closely around $\bar{w}$ as a tighter upper confidence bound. As a result, the RHS of Eq. (5) can be upper bounded by 0 with a higher probability as $t$ increases. Below, we present the first regret analysis of TS-CO for IM with linear generalization of edge weights. Prior to this work, only regret bounds for UCB algorithms have been established \cite{25} to the best of our knowledge.

**Theorem 2** Assume that $\forall e \in \mathcal{E}, \bar{w}(e) = x_e^T \theta$ and $\|x_e\|_2 \leq 1$. The scaled regret of TS-CO is
\[
R^1(T) \leq \frac{(\alpha \beta + \beta \eta) C_\theta}{\eta} \sqrt{dT m \log(1 + \frac{T m}{\eta \sigma^2})} + n \cdot \left( 1 - \hat{p}^T \hat{p} + \frac{\pi^2}{3} \right),
\]
where $\eta = (1 - 1/e - \epsilon)$ for any $\epsilon > 0$, $\alpha_t = \frac{1}{\sigma^2} \sqrt{2d \log(1 + \frac{4m}{\eta \sigma^2})} + 2 \log t + \|	heta^*\|_2$, $\beta_t = v(\sqrt{2\log(t)} + \sqrt{2\log m + 4\log t})$, $\hat{p} = 1 - v \cdot e^{-\alpha^2/2v^2}/(4\alpha_1 \sqrt{\tau})$, $C^*$ is a network topology-dependent complexity metric that is upper bounded by $n \sqrt{m}$ (see definition in Section 4).

**Proof sketch:** For each round $t$, we define the favorable event $\xi_t$ (and its complement $\bar{\xi}_t$) as
\[
\xi_t := \left\{ x_e^T \theta_t - x_e^T \theta^* \leq \alpha_t \sqrt{x_e^T M_{-1}^{-1} x_e}, \forall e \in \mathcal{E} \right\} \cap \left\{ |\bar{u}_{t}(e) - x_e^T \theta_t| \leq \beta_t \sqrt{x_e^T M_{-1}^{-1} x_e}, \forall e \in \mathcal{E} \right\},
\]
namely the event that $x_e^T \theta_t$ and $\bar{u}_t$ are concentrated around their respective means. By decomposing $\mathbb{E}[R^1_t]$ and using the naive bound $R^1_t \leq n$, we have
\[
\mathbb{E}[R^1_t] = \mathbb{E}[f(S^*, \bar{w}) - \frac{1}{\eta} f(S_t, \bar{u}_t) | \xi_t] \cdot \mathbb{P}(\xi_t) + \frac{1}{\eta} \mathbb{E}[f(S_t, \bar{u}_t) - f(S_t, \bar{w}) | \bar{\xi}_t] \cdot \mathbb{P}(\bar{\xi}_t) + n \cdot \mathbb{P}(\bar{\xi}_t).
\]
We first show $\mathbb{P}(\xi_t) \geq 1 - 2/t^2$ using Lemma $10$ and $11$. Then, we observe that $Q_1 \leq n \cdot \mathbb{P}(f(S_t, \bar{u}_t)/\eta \geq f(S^*, \bar{w}) | \xi_t)$, which can be further bounded by $n\eta^t$. Finally, by Lemma 8, we have $Q_2 \leq \mathbb{E}\left[\sum_{e \in E_t} 1\{O_t(e, S_t, \bar{w})\}||\bar{u}_t(e) - \bar{w}(e)||\xi_t\right]$, where $O_t(e, S_t, \bar{w})$ is the event that in round $t$, edge $e$’s realization is observed given seed set $S_t$ and edge weights $\bar{w}$. Summing up the regret over all rounds, we can prove Theorem 2 using standard linear-bandit techniques. Please refer to Appendix D for details of the proof and the lemmas used.

6 Numerical Experiments

We conduct numerical experiments on two Twitter subnetworks. The first subnetwork has 25 nodes and 319 directed edges, and the second has 50 nodes and 249 directed edges. We obtain the network structures from [16], and construct node feature vectors using the node2vec algorithm proposed in [10]. We then use the element-wise product of two node features to get each edge feature vector. We adopt this setup from [25]. For the 25-node network, we hand-pick a $\theta$ vector so that the edge weight obtained by taking the dot product between each edge feature vector and this $\theta$ falls between $0.01$ and $0.15$. Thus we have a perfect linear generalization of edge weights. For the 50-node experiment, we randomly sample an edge weight from Unif$(0,0.1)$ for each edge. As a result, it is unlikely that there exists a vector $\theta$ that perfectly generalizes the edge weights. For each subnetwork, we compare the performance of TS-CO with three other learning algorithms, 1) TS assuming linear generalization 2) UCB assuming linear generalization and 3) CUCB assuming no linear generalization [9]. We set $T = 4,000$, $d = 10$, and $B = 8,000$ and use ORACLE-IMB-M as the seeding oracle. For UCB and CUCB, we perform 500 rounds of random seeding and belief updates for “pre-training” before starting the campaign. We average the cumulative regret over 5 realizations for each algorithm to produce the plots in Figure 1. As we can see, although UCB and CUCB are given advantage with pre-training to learn the parameters, TS-based algorithms still outperform the UCB-based algorithms by a large margin in both network instances. Also, with or without perfect linear generalization of edge weights, algorithms assuming linear generalization (i.e., TS, TS-CO, UCB) in general outperform the one that does not (CUCB). Our TS-CO strategy falls between TS and UCB because it uses cumulative samples of the updated beliefs on $\theta$ to produce optimistic estimates. We see that in practice, its performance is closer to TS and much better than UCB-based algorithms.
References

[1] Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In Proceedings of the 24th International Conference on Neural Information Processing Systems, NIPS’11, pages 2312–2320, USA, 2011. Curran Associates Inc. ISBN 978-1-61839-599-3. URL http://dl.acm.org/citation.cfm?id=2986459.2986717.

[2] Marc Abeille, Alessandro Lazaric, et al. Linear thompson sampling revisited. Electronic Journal of Statistics, 11(2):5165–5197, 2017.

[3] Milton Abramowitz and Irene A Stegun. Handbook of mathematical functions with formulas, graphs, and mathematical tables. national bureau of standards applied mathematics series 55. tenth printing. 1972.

[4] Shipra Agrawal and Navin Goyal. Thompson sampling for contextual bandits with linear payoffs. CoRR, abs/1209.3352, 2012. URL http://arxiv.org/abs/1209.3352.

[5] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. Mach. Learn., 47(2-3):235–256, May 2002. ISSN 0885-6125. doi: 10.1023/A:1013689704352. URL https://doi.org/10.1023/A:1013689704352.

[6] Christian Borgs, Michael Brautbar, Jennifer T. Chayes, and Brendan Lucier. Influence maximization in social networks: Towards an optimal algorithmic solution. CoRR, abs/1212.0884, 2012. URL http://arxiv.org/abs/1212.0884.

[7] Olivier Chapelle and Lihong Li. An empirical evaluation of thompson sampling. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger, editors, Advances in Neural Information Processing Systems 24, pages 2249–2257. Curran Associates, Inc., 2011. URL http://papers.nips.cc/paper/4321-an-empirical-evaluation-of-thompson-sampling.pdf.

[8] Wei Chen, Chi Wang, and Yajun Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’10, pages 1029–1038, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0055-1. doi: 10.1145/1835804.1835934. URL http://doi.acm.org/10.1145/1835804.1835934.

[9] Wei Chen, Yajun Wang, and Yang Yuan. Combinatorial multi-armed bandit and its extension to probabilistically triggered arms. CoRR, abs/1407.8339, 2014. URL http://arxiv.org/abs/1407.8339.

[10] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. CoRR, abs/1607.00653, 2016. URL http://arxiv.org/abs/1607.00653.

[11] Alihan Hüyük and Cem Tekin. Thompson sampling for combinatorial network optimization in unknown environments. ArXiv, abs/1907.04201, 2019.

[12] Emilie Kaufmann, Nathaniel Korda, and Rémi Munos. Thompson sampling: An asymptotically optimal finite-time analysis. In Nader H. Bshouty, Gilles Stoltz, Nicolas Vayatis, and Thomas Zeugmann, editors, Algorithmic Learning Theory, pages 199–213, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg. ISBN 978-3-642-34106-9.

[13] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’03, pages 137–146, New York, NY, USA, 2003. ACM. ISBN 1-58113-737-0. doi: 10.1145/956750.956769. URL http://doi.acm.org/10.1145/956750.956769.

[14] Andreas Krause and Carlos Guestrin. A note on the budgeted maximization of submodular functions. 01 2005.
We define a crucial network-dependent complexity metric $C_*$ which was originally proposed in [25]. Here we extend its definition to our budgeted settings. First recall that $A^*_{o,t}$ is the set of edges whose realizations are observed in round $t$. Let $P_{S,e} = \mathbb{P}\{e \in A^*_{o,t} | S\}$, i.e., the probability that edge $e$’s realization is observed in round $t$ given seed set $S$. In the IC model, $P_{S,e}$ purely depends on the network topology and the edge edge weights. We say an edge $e \in A$ is relevant to a node $v \in V \setminus S$ if there exists a path $P$ that starts from any influencer $s \in S$ and ends at $v$, such that $e \in P$ and $P$ contains only one influencer, namely the starting node $s$. Use $A_{S,v}$ to denote the set of edges relevant to node $v$ with respect to $S$. Note that $A_{S,v}$ depends only on the topology of the network. Similarly, from each edge’s perspective, define $N_{S,e} := \sum_{v \in V \setminus S} 1\{e \text{ is relevant to } v \text{ under } S\}$, i.e., number

\[ A^*_{o,t} = \{ e \in A^*_{o,t} \}, \quad P_{S,e} = \mathbb{P}\{e \in A^*_{o,t} | S\}, \quad A^*_{o,t} \quad \text{depends on } S. \]
of non-seed nodes that \( e \) is relevant to with respect to \( S \). \( N_{S,e} \) also depends only on the network topology. With this notation, we define

\[
C_* := \max_{S \subseteq V} \sqrt{\sum_{e \in A} N^2_{S,e} P_{S,e}}.
\]

Clearly, \( C_* \) depends only on the network topology and edge weights. Also, it is upper bounded by \( |V| \sqrt{|A|} \). \( C_* \) is referred to as the maximum observed relevance in [23].

### B Simulation of \( f(\cdot, w) \) Modified Approximation Oracle

In the main body of the paper, we give oracles for IMB with the assumption that \( f(\cdot, w) \) can be computed exactly. However, it is \#P-hard to compute this quantity [8], and thus we need to approximate it by simulation. In [13], the authors propose to simulate the random diffusion process and use the empirical mean of the number of activated users to approximate the expected influence spread. In their numerical experiments, they use 10,000 simulations to approximate \( f(S, w) \) for each seed set \( S \). Such a method greatly increases the computational burden of the greedy algorithm. [6] propose a very different method that samples a number of so-called Reverse Reachable (RR) sets and use them to estimate influence spread under the IC model.

Based on the theoretical breakthrough of [6], [21, 20] present Two-phase Influence Maximization (TIM) and Influence Maximization via Martingales (IMM) for IM with complexity \( O((m + n)K\epsilon^{-2}\log(n)) \), where \( m \) is the number of edges in the network, \( n \) the number of nodes, \( K \) the cardinality constraint of the seed sets, and \( \epsilon \in (0, 1) \) the size of the error. These two methods improve upon the algorithm in [6] that has a run time complexity of \( O((m + n)K\epsilon^{-3}\log(n)) \). All these three methods are designed solely for IM with simple cardinality constraints. Their analysis relies on the assumption that the optimal seed set is of size \( K \). As a result, the number of RR sets required in their methods does not guarantee estimation accuracy of \( f(S, w) \) for seed sets of bigger sizes. However, in our problems, the feasible seed sets can potentially be of any sizes. In particular, our ORACLE-IMB assigns a probability distribution to seed sets of cardinalities from 0 to \( n \). This means that our simulation method needs to guarantee accuracy for seed sets of all sizes.

In order to cater to this requirement of our budgeted problems, we extend the results in [6] and [20], and develop a Concave Error Interval (CEI) analysis which gives an upper bound on the number of RR sets required to secure a consistent influence spread estimate for seed sets of different sizes with high probability. We then detail how we modify our oracles using RR sets to estimate \( f(\cdot, w) \). We prove \( 1 - 1/\epsilon - \epsilon \)-approximation guarantees for the modified oracles. We also supply the run time complexity analysis. In the rest of the section, we suppress \( w \) as an argument of \( f(\cdot, w) \).

#### B.1 Reverse Reachable (RR) Set

To precisely explain our simulation method, we introduce the formal definition of RR sets.

**Definition 1 (Tang et al.: Reverse Reachable Set)** Let \( v \) be a node in \( V \), and \( H \) be a graph obtained by removing each directed edge \( e \) in \( A \) with probability \( 1 - p(e) \). The reverse reachable (RR) set for \( v \) in \( H \) is the set of nodes in \( H \) that can reach \( v \). That is, a node \( u \) is in the RR set if and only if there is a directed path from \( u \) to \( v \) in \( H \).

**Definition 2 (Tang et al.: Random RR Set)** Let \( W \) be the distribution on \( H \) induced by the randomness in edge removals from \( V \). A random RR set is an RR set generated on an instance of \( H \) randomly sampled from \( W \), for a node selected uniformly at random from \( V \).

[20] give an algorithm for generating a random RR set, which is given in Algorithm 3 below.

By [6], we have the following lemma:

**Lemma 1 (Borgs et al.)** For any seed set \( S \) and node \( v \), the probability that a diffusion process from \( S \) which follows the IC model can activate \( v \) equals the probability that \( S \) overlaps an RR set for \( v \) in a graph \( H \) generated by removing each directed edge \( e \) in \( A \) with probability \( 1 - p(e) \).
Algorithm 3: Random RR set

Data: digraph $D = (V, A)$, edge edge weights $w : A \rightarrow [0, 1]$
Result: Random RR set $R$

initialization: $R = \emptyset$, first-in-first-out queue $Q$

sample a node $v$ uniformly at random from $V$, add to $R$

for $u \in V$ s.t. $(u, v) \in A$ do
  flip a biased coin with probability $w(u, v)$ of turning head;
  if the coin turns head then
    Add $u$ to $Q$ and $R$

while $Q$ is not empty do
  extract the node $v'$ at the top of $Q$;
  for $u' \in V$ s.t. $(u', v') \in A$ do
    flip a biased coin with probability $w(u', v')$ of turning head;
    if the coin turns head then
      add $u'$ to $Q$ and $R$

Suppose we have generated a collection $R$ of random RR sets. For any node set $S$, let $F_R(S)$ be the fraction of RR sets in $R$ that overlap $S$. From Lemma 1, showed that the expected value of $nF_R(S) = \frac{f(S)}{OPT_B}$ equals the expected influence spread of $S$ in $V$, i.e., $E[nF_R(S)] = f(S)$. Thus, if the number of RR sets in $R$ is large enough, then we can use the realized value $nF_R(S)$ to approximate $f(S)$.

B.2 Concave Error Interval and Simulation Sample Size

In this section, we propose the Concave Error Interval (CEI) method of analysis which gives the number of random RR sets required to obtain a close estimate of $f(S)$ using $nF_R(S)$ for every seed set $S$. Our analysis uses the following Chernoff inequality.

Lemma 2 (Chernoff Bound) Let $X$ be the sum of $L$ i.i.d random variables sampled from a distribution on $[0, 1]$ with a mean $\mu$. For any $\eta > 0$,

$$\mathbb{P}(X/L - \mu \geq \eta \mu) \leq e^{-\frac{\eta^2 L\mu}{2e}},$$

$$\mathbb{P}(X/L - \mu \leq -\eta \mu) \leq e^{-\frac{\eta^2 L\mu}{2e}}.$$

B.2.1 Concave Error Interval

Let $OPT^B = \sum_S p_{opt}^S f(S)$ be the expected influence spread of OPTIMAL-IMB, the optimal stochastic strategy for IMB. Let $OPT^B$ be the expected influence spread of the optimal seed set for IMB, and $OPT^K$ be the expected influence spread of the optimal seed set for IM with cardinality constraint $K$.

We now introduce a concave error interval $I_S$ for each seed set $S$, and define an event $E_S$ as follows which limits the difference between $nF_R(S)$ and $f(S)$. We first focus on analyzing IMB. The analysis of IMB is similar, with modifications that we will specify. Suppose $\epsilon$ is given.

Definition 3 (Concave Error Interval $I_S$ and Event $E_S$)

$$I_S = \left[ -\frac{\epsilon}{1 + \sqrt{1 - 1/e}} \sqrt{OPT^B f(S)} , \frac{\epsilon}{1 + \sqrt{1 - 1/e}} \sqrt{OPT^B f(S)} \right],$$

$$E_S = \left\{ nF_R(S) - f(S) \in I_S \right\}.$$

The length of the error interval $I_S$ is $\frac{2\epsilon}{1 + \sqrt{1 - 1/e}} \sqrt{OPT^B f(S)}$, which is concave in $f(S)$. $E_S$ is the event that the difference between $nF_R(S)$ and its mean $f(S)$ is within $I_S$. With $L$ being the number
of randomly sampled RR sets, the likelihood of \( E_S \) can be bounded as follows.

\[
\mathbb{P}(E_S) = \mathbb{P}\left( |nF_R(S) - f(S)| \leq \frac{\epsilon}{1 + \sqrt{1 - 1/e} \sqrt{OPTB} f(S)} \right) = \mathbb{P}\left( |LF_R(S) - L \frac{f(S)}{n}| \leq \sqrt{\frac{OPTB}{f(S)}} \frac{\epsilon}{1 + \sqrt{1 - 1/e} L \frac{f(S)}{n}} \right).
\]

Let \( \eta = \sqrt{\frac{OPTB}{f(S)}} \frac{\epsilon}{1 + \sqrt{1 - 1/e}} \). By Lemma 2, we have that when \( \epsilon \leq \frac{3}{\sqrt{n}} \),

\[
\mathbb{P}(E_S) \geq 1 - 2e^{-\frac{\eta^2 n}{12} f(S)} \geq 1 - 2e^{-\frac{\epsilon OPTB L}{n(1+\sqrt{1-1/e})^2}}.
\]

Therefore, we have a uniform lower bound on \( \mathbb{P}(E_S) \) for every seed set \( S \), which implies the following lemma:

**Lemma 3** For any given \( l \), let

\[
L = \frac{7n(l \log n + n \log 2)}{OPTB \cdot \epsilon^2}
\]

If \( R \) contains \( L \) random RR sets and \( \epsilon \leq \frac{3}{\sqrt{n}} \), then for every seed set \( S \) in \( \mathcal{V} \), \( E_S \) happens with probability at least \( 1 - \frac{1}{n^2} \).

Since there are \( 2^n \) different seed sets, we have the following.

**Lemma 4** For any given \( l \), let

\[
L = \frac{7n(l \log n + n \log 2)}{OPTB \cdot \epsilon^2}
\]

If \( R \) contains \( L \) random RR sets and \( \epsilon \leq \frac{3}{\sqrt{n}} \), then

\[
\mathbb{P}(E_S \text{ holds for all } S) > 1 - \frac{1}{n^l}.
\]

So far, we have established the relationship between the number of random RR sets and the estimation accuracy through the CEI analysis. Later we will prove that ORACLE-IMB combined with the above RR sets simulation technique give at least \( (1 - 1/e - \epsilon) \)-approximation guarantee for IMB with high probability. Compared to the naive simulation proposed in [13], the RR sets simulation technique has the following advantages.

- While naive simulation constructs different samples to estimate the influence spread of different seed sets, the RR sets method generates the collection \( R \) of \( L \) random RR sets only once. The same \( R \) is used to estimate the expected influence spread of any seed set.
- Our Concave Error Interval analysis is able to deal with the budgeted variants and give a uniform accuracy bound for all seed sets.
- In [20], it is shown that TIM returns an \( (1 - 1/e - \epsilon) \)-approximation solution with an expected runtime of \( O\left(\frac{(k+l)(m+n) \log n}{\epsilon^2}\right) \), which is near-optimal under the IC diffusion model, as it is only a \( \log n \) factor away from the lower-bound established by [6]. As will be shown later, the expected runtime of ORACLE-IMB is \( O\left(\frac{n(l \log n + n)}{\epsilon^2}\right) \), which has an extra \( n \) compared to the lower-bound due to the flexible usage of the total budget. However, we give guaranteed influence spread estimates for all \( 2^n \) possible seed sets, while the TIM analysis only covers the \( \binom{n}{k} \) size-\( K \) seed sets. Under the \( \log \) operator, the difference in runtime is \( n \log 2 \) versus \( k \log n \).

**B.3 \((1 - 1/e - \epsilon)\)-approximation Ratio for Modified ORACLE-IMB**

We denote by ORACLE-IMB-\( M \) the modified version of ORACLE-IMB that includes the \( f(S) \) approximation. Assume \( l \) and \( \epsilon \leq \frac{3}{\sqrt{n}} \) are given.

To prove the approximation guarantee for ORACLE-IMB-\( M \), we need the following theorem.
We now prove the
Proof of Theorem 4:
By Lemma 4, we have that with probability
Proof of Theorem 3:
over seed sets computed in
ORACLE-IM
OPTIMAL-IM

By Jensen’s inequality,
Since

Algorithm 4: ORACLE-IM-M

Generate a collection \( R \) of \( \frac{7n(\log n + n \log 2)}{OPTB^{n+2}} \) random RR sets;
Run ORACLE-IM with the change that whenever a \( f(S) \) needs to be computed, use \( nF_{R}(S) \) instead

\[ \sum_{S} p(S)nF_{R}(S) \geq \sum_{S} p(S)f(S) - \frac{\epsilon}{1 + \sqrt{1 - 1/e}} \sqrt{OPTB \sum_{S} p(S)f(S)}, \quad (8) \]

and
\[ \sum_{S} p(S)f(S) \geq \sum_{S} p(S)nF_{R}(S) - \frac{\epsilon}{1 + \sqrt{1 - 1/e}} \sqrt{OPTB \sum_{S} p(S)nF_{R}(S)}. \quad (9) \]

Proof of Theorem 3: By Lemma 3 we have that with probability \( 1 - \frac{1}{n^t} \), for all \( S \)
\[ nF_{R}(S) \geq f(S) - \frac{\epsilon}{1 + \sqrt{1 - 1/e}} \sqrt{OPTBf(S)}. \]

Since \( \sqrt{OPTBf(S)} \) is concave in \( f(S) \), using Jensen’s inequality we have
\[ \sum_{S} p(S)nF_{R}(S) \geq \sum_{S} p(S)\left(f(S) - \frac{\epsilon}{1 + \sqrt{1 - 1/e}} \sqrt{OPTBf(S)}\right) \]
\[ \geq \sum_{S} p(S)f(S) - \frac{\epsilon}{1 + \sqrt{1 - 1/e}} \sqrt{OPTB \sum_{S} p(S)f(S)}. \]

Similarly, we have
\[ nF_{R}(S) \leq f(S) + \frac{\epsilon}{1 + \sqrt{1 - 1/e}} \sqrt{OPTBf(S)}. \]

By Jensen’s inequality,
\[ \sum_{S} p(S)nF_{R}(S) \leq \sum_{S} p(S)\left(f(S) + \frac{\epsilon}{1 + \sqrt{1 - 1/e}} \sqrt{OPTBf(S)}\right) \]
\[ \leq \sum_{S} p(S)f(S) + \frac{\epsilon}{1 + \sqrt{1 - 1/e}} \sqrt{OPTB \sum_{S} p(S)f(S)}. \]

We now prove the \( 1 - 1/e - \epsilon \)-approximation guarantee for ORACLE-IM-M.

Theorem 4: With probability at least \( 1 - \frac{1}{n^t} \), the expected influence spread of the seed set returned by ORACLE-IM-M is at least \( (1 - 1/e - \epsilon) \) that of the optimal spread.

Proof of Theorem 4: Let \( p^{opt}(S) \) be the probability of selecting seed set \( S \) for any \( S \subseteq V \) in OPTIMAL-IM when assuming \( f(S) \) can be computed exactly. Let \( p^* \) be the probability distribution over seed sets computed in ORACLE-IM-M where \( f(S) \) is approximated by RR sets. Since \( F_{R}(\cdot) \) is submodular, Theorem 3 implies that
\[ \sum_{S} p^*(S)nF_{R}(S) \geq (1 - 1/e) \sum_{S} p^{opt}(S)nF_{R}(S). \quad (10) \]
Now plugging $p^{opt}(S)$ into (8), we get
\[ \sum_S p^{opt}(S) n_{\mathcal{F}}(S) \geq \sum_S p^{opt}(S) f(S) - \frac{\epsilon}{1 + \sqrt{1 - 1/e}} \sqrt{OPT^B \sum_S p^{opt}(S) f(S)} = OPT^B - \frac{\epsilon}{1 + \sqrt{1 - 1/e}} OPT^B = (1 - \frac{\epsilon}{1 + \sqrt{1 - 1/e}})OPT^B. \] (11)

Furthermore, by plugging $p^*(S)$ into (9), we get
\[ \sum_S p^*(S) f(S) \geq \sum_S p^*(S) n_{\mathcal{F}}(S) - \frac{\epsilon}{1 + \sqrt{1 - 1/e}} \sqrt{OPT^B \sum_S p^*(S) n_{\mathcal{F}}(S)}. \] (12)

(10), (11) and (12) together give us that
\[ \sum_S p^*(S) f(S) \geq (1 - \frac{\epsilon}{1 + \sqrt{1 - 1/e}})OPT^B \]
with probability at least $1 - \frac{1}{n^l}$, which completes the proof. □

B.4 Runtime Complexity of ORACLE-IMB-M

The running time of ORACLE-IMB-M mostly falls on generating random RR sets. To analyze the corresponding time complexity, we first define expected coin tosses (EPT).

Definition 4 EPT is the expected number of coin tosses required to generate a random RR set following Algorithm 3.

With the definition above, the expected runtime complexity of ORACLE-IMB-M is $O(L \cdot EPT)$, where $L$ is the number of random RR sets required by the algorithm. [20] establishes a lower bound of $OPT^B$ based on EPT. We bound $OPT^B$ similarly in the following lemma.

Lemma 5
\[ \min(\frac{b}{\bar{c}}, 1)OPT^B \geq \frac{n}{m} EPT, \]
where $b$ is the budget and $\bar{c}$ is the maximum cost among all nodes.

Proof. Let $R'$ be a random RR set, and let $p_{R'}$ be the probability that a randomly selected edge from $\mathcal{D}$ points to a node in $R'$. Then, $EPT = \mathbb{E}[p_{R'} \cdot m]$, where the expectation is taken over the random choices of $R'$. Let $Y(v, R')$ be a boolean function that returns 1 if $v \in R'$, and 0 otherwise. Denote $\deg(v)$ as the in-degree of node $v$ in $\mathcal{D}$ and $deg = \sum_v \deg(v)$. Then
\[ \frac{EPT}{m} = \mathbb{E}[p_{R'}] = \sum_{R' \in \mathcal{R}} \mathbb{P}(R') \cdot p_{R'} = \sum_{R' \in \mathcal{R}} \mathbb{P}(R') \cdot \left( \sum_{v \in \mathcal{V}} \frac{\deg(v)}{\deg} Y(v, R') \right) = \sum_{v \in \mathcal{V}} \frac{\deg(v)}{\deg} \cdot \left( \sum_{R' \in \mathcal{R}} \mathbb{P}(R') Y(v, R') \right) = \sum_{v \in \mathcal{V}} \frac{\deg(v)}{\frac{deg}{\deg} \cdot p_v}, \]
where, by Lemma 1, \( p_v = \sum_{R' \in R} \mathbb{P}(R') Y(v, R') \) equals the probability that a randomly selected node is activated given \( v \) is in the seed set. Now consider a very simple policy \( \Pi^{\text{one}} \) that selects one node \( v \) as the seed set with probability \( \frac{\deg(v)}{\deg} \). Then \( \frac{n \cdot \text{EPT}}{m} \) is the average expected influence of \( \Pi^{\text{one}} \). It’s easy to show that \( \min \left( \frac{7}{n}, 1 \right) OPT^B \geq f(\Pi^{\text{one}}) = \frac{n \cdot \text{EPT}}{m} \), where \( f(\Pi^{\text{one}}) \) is the expected influence spread of the seed set returned by policy \( \Pi^{\text{one}} \). □

Moreover, \( \text{EPT} \) can be estimated by measuring the average width of RR sets, which is defined to be the average number of edges connecting to at least one node in a random RR set. If we choose \( L = 7n(\log n + n \log 2) \), then the complexity of \( \text{ORACLE-IMB} \) is \( O(L \cdot \text{EPT}) = O \left( \frac{m \log n + m}{n} \right) \), since by Lemma 5, \( L \leq \frac{7n(\log n + n \log 2) \min \left( \frac{7}{n}, 1 \right)}{n^{\frac{1}{2}}} \). Further note that the \( OPT \) in the denominator of \( L \) is not readily available. To closely approximate \( L \), we apply the idea by [20]. We compute \( L' = \frac{7n(\log n + n \log 2) \min \left( \frac{7}{n}, 1 \right)}{n^{\frac{1}{2}}} \), which is a lower bound of \( L \). We generate \( L' \) many RR sets and then estimate \( \text{EPT} \) using the average width of the generated RR sets, \( \text{EPT} \). If \( L' \leq \frac{7n(\log n + n \log 2) \min \left( \frac{7}{n}, 1 \right)}{n^{\frac{1}{2}}} \), we keep generating RR sets to reach the quantity specified by the right-hand side of the proceeding inequality, and update our estimate of \( \text{EPT} \) with all the available RR sets. We repeat this process until the current number of RR sets exceeds \( \frac{7n(\log n + n \log 2) \min \left( \frac{7}{n}, 1 \right)}{n^{\frac{1}{2}}} \).

## C Proofs of offline results

### C.1 NP-Hardness

**Theorem 5** IMB is NP-hard.

**Proof of Theorem 5** Given any instance of the minimum set cover problem in the following form:

\[ U = \{u_1, u_2, \ldots, u_n\} \] is a ground set with \( n \) elements, \( S = \{S_1, S_2, \ldots, S_m\} \) is a family of \( m \) subsets of \( U \). Find a minimal cardinality subset \( S' \) of \( S \) such that \( \cup_{S_i \in S'} S_i = U \).

Assume IMB can be solved efficiently. We show that the given minimum set cover instance can be solved efficiently.

First, construct a network \( D = (V, \mathcal{A}) \) as follows. For each \( S_i \in S \), there is a node in \( V \) that corresponds to it. For each \( u_j \in U \), there is a node that corresponds to it. \( (S_i, u_j) \in A \) if and only if \( u_j \in S_i \).

Use \( \text{IMB}(D, b, c, w) \)-OPT to denote the optimal solution for an IMB instance with budget \( b \), node costs \( c \) and edge weights \( w \). Note that this solution can be expressed as a probability distribution on seed sets that specifies the likelihood with which each seed set will be played.

Find the smallest integer \( k \) such that the expected number of activated nodes of \( \text{IMB}(D, k, 1, 1) \)-OPT is at least \( n + k \). Note such a \( k \) must exists and is smaller than \( m \) since by our assumption, \( S \) covers \( U \). Since \( \text{IMB}(D, b, 1, 1) \)-OPT can be obtained efficiently for each \( b \in \{0, 1, \ldots, m\} \), \( k \) can be found efficiently.

We claim that i) \( k \) is the smallest size of \( S' \), that is, the smallest number of subsets needed to cover \( U \); ii) any \( S \subseteq V \) with positive probability in \( \text{IMB}(D, k, 1, 1) \)-OPT must correspond to a set cover for \( U \) of size \( k \). To prove ii), note that with cardinality constraint \( k \) the maximum number of activated nodes in \( D \) is \( n + k \) due to the way we construct the network. Since the expected number of activated nodes of \( \text{IMB}(D, k, 1, 1) \)-OPT is at least \( n + k \), only \( S \in V \) such that \( f(S, w) = n + k \) can have positive probability. Without loss of generality, \( |S| \leq k \), since if \( |S| > k \), by cardinality constraint, there must exists a subset \( S' \) with positive probability such that \( |S'| < k \). Furthermore, \( f(S, w) \leq n + |S| \), and thus \( |S| \geq k \). As a result, \( |S| = k \). From ii), we know that we need at most \( k \) subsets in \( S \) to cover \( U \). If there exists a family \( S' \) of \( h \) subsets in \( S \) that covers \( U \), where \( h < k \), then \( p(S') = 1 \) is a feasible solution to IMB(\( D, h, 1, 1 \)) with objective value \( n + h \). Thus, the expected number of activated nodes of \( \text{IMB}(D, h, 1, 1) \)-OPT is at least \( n + h \), contradicting the assumption that \( k \) is the smallest integer such that the expected number of activated nodes of \( \text{IMB}(D, k, 1, 1) \)-OPT is at least \( n + k \). i) therefore follows. □
C.2 Proof of Theorem 1

We first study an alternative approximation oracle ORACLE-IMB-a detailed below. We show that the distribution obtained in ORACLE-IMB is an optimal solution to the LP in ORACLE-IMB-a using Lemma 6.

Algorithm 5: ORACLE-IMB-a

Data: D = (V, A), b, c, w
Result: S ⊆ V
Initialization: S₀ = ∅;
for i = 1, 2, ..., n do
  Compute vi = arg max:v∈V\Sᵢ₋₁ (f(Sᵢ₋₁ ∪ {v}, w) − f(Sᵢ₋₁, w)) / c(v);
  Set Sᵢ = Sᵢ₋₁ ∪ {vi};
Solve the following LP to get an optimal solution p* = (p₁*, p₂*, ..., pₙ*):
max ∑ₙ j=0 pⱼf(Sⱼ, w) s.t ∑ₙ j=0 c(Sⱼ)pⱼ ≤ b; ∑ₙ j=0 pⱼ = 1; pⱼ ≥ 0 ∀ j = 0, 1, ..., n;
Sample S from {S₀, S₁, ..., Sₙ} with probability distribution p*

Lemma 6 There exits an optimal solution p* to the LP in ORACLE-IMB-a that has the following properties: 1) at most two elements in p* = (p₁*, p₂*, ..., pₙ*) are non-zero; 2) if p₁*, p₂* are non-zero, then |i − j| ≤ 1.

Proof of Lemma 6 For conciseness, we suppress w as a argument of f(S, w). Also, without loss of generality, assume i ≤ j. Now suppose that there exists pᵢ*, pⱼ* such that |i − j| ≥ 2. Then the contribution of Sᵢ, Sⱼ to the objective value of the LP is pᵢ* f(Sᵢ) + pⱼ* f(Sⱼ) and the consumption of the budget is pᵢ* c(Sᵢ) + pⱼ* c(Sⱼ). Now let pᵢ′, pⱼ′ be the solution to the following system of equations:

\[ pᵢ′ + pⱼ′ = pᵢ* + pⱼ* \]
\[ pᵢ′c(Sᵢ₊₁) + pⱼ′c(Sⱼ₋₁) = pᵢ*c(Sᵢ) + pⱼ*c(Sⱼ) \] (13)

The system of equations has a unique solution since c(Sᵢ) ≤ c(Sᵢ₊₁) ≤ c(Sⱼ₋₁) ≤ c(Sⱼ). Note that pᵢ′f(Sᵢ₊₁) + pⱼ′f(Sⱼ₋₁) ≥ pᵢ*f(Sᵢ) + pⱼ*f(Sⱼ) by the observation that f(Sₖ, w) − f(Sₖ₋₁, w) ≥ f(Sₖ, w) − f(Sₖ₋₁, w) for 1 ≤ k ≤ l ≤ n. Therefore, allocating pᵢ′ to Sᵢ₊₁ and pⱼ′ to Sⱼ₋₁ costs the same as pᵢ* to Sᵢ and pⱼ* to Sⱼ, but it achieves at least the same expected influence spread.

By repeating the process, eventually at most two consecutive sets in {S₀, S₁, ..., Sₙ} will have positive probability.

Since the expected cost is at most b, by Lemma 6 we know that if two sets in {S₀, S₁, ..., Sₙ} have positive probability, then one of them is the largest set whose cost is below b, denoted by S−, and the other one is the smallest set whose cost is greater than b, denoted by S+. As a result, in the for loop of ORACLE-IMB-a, we can stop as soon as c(Sᵢ) ≥ b. When solving the LP, instead of having n + 1 variables, we solve for the optimal probability distribution on S− and S+ only. This simplified algorithm is what we presented as ORACLE-IMB.

Proof of Theorem 1 To prove Theorem 1, we first observe that IMB can be equivalently formulated as the following linear program (LP1):

\[ \max \sum_{S ∈ P(V)} p(S)f(S, w) \text{ s.t } \sum_{S ∈ P(V)} c(S)p(S) ≤ b; \sum_{S ∈ P(V)} p(S) = 1; p(S) ≥ 0 \forall S ∈ P(V), \]

where P(V) is the power set of V. We also need some definitions and lemmas.

Definition 5 For any budget b and edge weights w, let S(b, w) := arg maxₜ∈G(S),c(S)≤b f(S, w), and call it the best response set of budget b and influence w.
Lemma 7 If a family of seed sets \( S' = \{ S'_1, S'_2, \ldots, S'_L \} \) is a \( \beta \)-combo-approximation family with respect to \( w \), then for any budget \( b \), we can find two seed sets \( S'_i, S'_j \) in \( S' \) and probabilities \( q_i, q_j \) such that \( q_i c(S'_i) + q_j c(S'_j) \leq \beta c(S) \) and \( q_i f(S'_i, w) + q_j f(S'_j, w) \geq \beta f(S, w) \).

Proof of Lemma 7 Initialize \( p'_i = 0 \) \( \forall i = 1, \ldots, L \). For each \( S \in P(V) \), from the definition of \( \beta \)-combo-approximation family, we find two seed sets \( S'_i, S'_j \) in \( S' \) with different probabilities such that \( q_i c(S'_i) + q_j c(S'_j) \leq \beta c(S) \) and \( q_i f(S'_i, w) + q_j f(S'_j, w) \geq \beta f(S, w) \). Updating \( p'_i \leftarrow p'_i + p'_{opt}(S)q_i, p'_j \leftarrow p'_j + p'_{opt}(S)q_j \). After doing so for all \( S \in P(V) \), we have the probability distribution \( p' \) as desired. \( \square \)

To prove Theorem 1 we show that \( S = \{ S_0, S_1, \ldots, S_n \} \) as constructed in ORACLE-IMBA is a \( 1 - 1/e \)-combo-approximation family with respect to \( w \). As ORACLE-IMBA solves an LP to find the optimal distribution \( p' \) over \( S \), it indeed achieves an \( 1 - 1/e \) approximation ratio.

Consider the sequence of sets, \( S_0, S_1, \ldots, S_n \), constructed in the oracle. For any budget \( 0 < b \leq c(V) \), we can find a unique index \( i(b) \in \{ 1, 2, \ldots, n \} \) such that \( c(S_{i(b)} - 1) < b \leq c(S_{i(b)}) \), and a unique \( \alpha \in (0, 1) \) such that \( b = (1 - \alpha)c(S_{i(b) - 1}) + \alpha c(S_{i(b)}) \). We now only need to show that \( (1 - \alpha)f(S_{i(b)} - 1, w) + \alpha f(S_{i(b)}, w) \geq (1/e - 1)f(S(b, w), w) \).

Let \( r_j = \max_{x \in \mathbb{V} \setminus S_{i(j - 1)}} (f(S_{i(j - 1)} \cup \{ v \}, w) - f(S_{i(j - 1)}, w)) / c(v) \). Let \( x_0 = 0, x_j = c(S_j) \) for \( j = 1, 2, \ldots, i(b) - 1 \), and \( x_{i(b)} = b \). We define a density function \( p(x) \) on \([0, B]\) as \( p(x) := r_{j+1} \) if \( x \in [x_j, x_{j+1}] \). We denote \( h(x) := \int_0^x p(s)ds \).

Now as \( f(\cdot, w) \) is submodular and by the definition of \( r_{j+1} \), we have that \( f(S_j, w) = h(x_j) \) for \( j = 1, 2, \ldots, i(b) - 1 \), and

\[
\sum_{i=1}^L f(S'_i, w)p'_i \geq \beta \sum_{S \in P(V)} f(S, w)p'_{opt}(S) = \beta \cdot E(f(S'_{opt}, w)),
\]

where \( p'_{opt} \) is an optimal probability distribution over all \( |P(V)| \) possible seed sets used by an optimal oracle for IMB (i.e., \( p'_{opt} \) is an optimal solution to LP1), and \( S'_{opt} \) is the seed set returned by an optimal oracle.
D Regret analysis of TS-CO

D.1 Preliminaries & concentration results

Let \( f(S, w) \) be the expected reward function with seed set \( S \) under \( w \), let \( 1 \{ O_t(e, S_t, w) \} \) be the realization of edge \( e \) under \( S_t \) and \( w \). As \( S_t \) is chosen as the seed set and true diffusion probability is \( \bar{w} \), we are only able to observe \( 1 \{ O_t(e, S_t, \bar{w}) \} \). Define \( 1 \{ O_t(e) \} := 1 \{ O_t(e, S_t, \bar{w}) \} \). We first provide the following lemma:

Lemma 8 Let \( \bar{w} \) be the true diffusion probabilities. For any \( t \) and any diffusion probability \( u_t \), we have

\[
|f(S_t, u_t, v) - f(S_t, \bar{w}, v)| \leq \sum_{e \in E_t} \mathbb{E}[1 \{ O_t(e, S_t, \bar{w}) \}]|u_t(e) - \bar{w}(e)|.
\]

Proof. Let \( u^*_t(e) = \min(u_t(e), \bar{w}(e)) \), then we have

\[
|f(S_t, u_t, v) - f(S_t, \bar{w}, v)| = |f(S_t, u^*_t, v) + f(S_t, u^*_t, v) - f(S_t, \bar{w}, v)|
\]

\[
\leq \sum_{e \in E_t} \mathbb{E}[1 \{ O_t(e, S_t, u^*_t) \}]|u_t(e) - u^*_t(e)| + \sum_{e \in E_t} \mathbb{E}[1 \{ O_t(e, S_t, \bar{w}) \}]|u^*_t(e) - \bar{w}(e)|
\]

\[
\leq \sum_{e \in E_t} \mathbb{E}[1 \{ O_t(e, S_t, \bar{w}) \}]|u_t(e) - u^*_t(e)| + \sum_{e \in E_t} \mathbb{E}[1 \{ O_t(e, S_t, \bar{w}) \}]|u^*_t(e) - \bar{w}(e)|
\]

\[
= \sum_{e \in E_t} \mathbb{E}[1 \{ O_t(e, S_t, \bar{w}) \}]|u_t(e) - \bar{w}(e)|,
\]

where the first inequality follows from Theorem 3 in \([25]\), the second inequality is due to the fact that \( \mathbb{E}[1 \{ O_t(e, S_t, u^*_t) \}] \leq \mathbb{E}[1 \{ O_t(e, S_t, \bar{w}) \}] \) for \( u^*_t \leq \bar{w} \), and the last equality comes from the fact that \( u^*_t(e) = \min(\bar{w}(e), u_t(e)) \). \( \square \)

Lemma 9 In round \( t \), let \( Z_j \) for \( j = 1, \ldots, t \) be \( t \) independent standard normal random variables. For any \( s \), we have

\[
\mathbb{P}\left(\frac{\hat{w}_t(e) - x_e^\top \theta_t}{v\|x_e\|_{M_t^{-1}}} \leq s\right) = \mathbb{P}\left(\max_{j=1,\ldots,t} Z_j \leq s\right).
\]

That is, \( \hat{w}_t(e) \) follows the same distribution as \( \max_{i=1,\ldots,t} Z_i \).

Proof. We prove the lemma by induction. As \( \hat{w}_0(e) = -\infty \) for all \( e \in E \), it is easy to see \( \hat{w}_1(e) = x_e^\top \hat{\theta}_1 \), where \( x_e^\top \hat{\theta}_1 \) has mean \( x_e^\top \theta_1 \) and standard deviation \( v\|x_e\|_{M_1^{-1}} \). We have

\[
\mathbb{P}\left(\frac{\hat{w}_1(e) - x_e^\top \theta_1}{v\|x_e\|_{M_1^{-1}}} \leq s\right) = \mathbb{P}\left(\frac{x_e^\top \hat{\theta}_1 - x_e^\top \theta_1}{v\|x_e\|_{M_1^{-1}}} \leq s\right) = \mathbb{P}(Z_1 \leq s),
\]

which implies the correctness of this argument for \( t = 1 \).

Suppose the argument is true for round \( t - 1 \). Then in round \( t \), by definition, \( \hat{w}_t(e) = \max\left(\left(\hat{w}_{t-1}(e) - x_e^\top \theta_{t-1}\right)\|x_e\|_{M_{t-1}^{-1}}/\|x_e\|_{M_{t-2}^{-1}} + x_e^\top \theta_t, x_e^\top \theta_t\right) \). Then we have

\[
\mathbb{P}\left(\frac{\hat{w}_t(e) - x_e^\top \theta_t}{v\|x_e\|_{M_t^{-1}}} \leq s\right) = \mathbb{P}\left(\max\left(\frac{\hat{w}_{t-1}(e) - x_e^\top \theta_{t-1}}{v\|x_e\|_{M_{t-2}^{-1}}}, \frac{x_e^\top \theta_t - x_e^\top \theta_t}{v\|x_e\|_{M_{t-2}^{-1}}} \right) \leq s\right)
\]

\[
= \mathbb{P}\left(\max\left(\max_{i=1,\ldots,t-1} Z_i, Z_t\right) \leq s\right),
\]

which completes the proof. \( \square \)

We provide two concentration results for \( \theta_t \) and \( \hat{w}_t \).
Lemma 10 (Concentration of \( \theta_t \)) Let \( \alpha_t = \frac{1}{\sigma} \sqrt{2d \log(1 + \frac{t}{\sigma^2}) + 2 \log(t) + \|\theta^*\|_2} \). Then we have
\[
\|\theta_t - \theta^*\|_{M_t^{-1}} \leq \alpha_t
\]
for all \( t \) with probability at least \( 1 - 1/t^2 \). Under this event, we have \( \|x_e^t \theta_t - x_e^t \theta^*\| \leq \alpha_t \sqrt{x_e^t M_t^{-1} x_e} \) for all \( e \in E \).

Proof. This result comes from Theorem 1 \([1]\). □

Lemma 11 (Concentration of \( \tilde{u}_t \)) Let \( \beta_t = \sqrt{(2 \log(t) + \sqrt{2 \log |E| + 4 \log t})} \). Then for all \( e \in E \),
\[
|\tilde{u}_t(e) - x_e^T \theta_t| \leq \beta_t \sqrt{x_e^T M_t^{-1} x_e},
\]
with probability at least \( 1 - t^{-2} \).

Proof. By definition, for all \( e \in E \), \( \tilde{u}_t(e) = \max(\langle \tilde{w}_{t-1}(e) - x_e^T \theta_{t-1}, x_e \rangle_{M_t^{-1}} + x_e^T \theta_{t-1}, x_e^T \theta_t) \). By definition, each Gaussian random variable \( x_e^T \tilde{\theta}_t \) has mean \( x_e^T \theta_t \) and standard deviation \( v \|x_e\|_{M_t^{-1}} \). Then it is easy to see
\[
|\tilde{w}_t(e) - x_e^T \theta_t| = \max(\langle \tilde{w}_{t-1}(e) - x_e^T \theta_{t-1}, x_e \rangle_{M_t^{-1}} / \|x_e\|_{M_t^{-1}} + x_e^T \theta_{t-1}, x_e^T \theta_t) - x_e^T \theta_t)\]
\[
= \max(\tilde{w}_{t-1}(e) - x_e^T \theta_{t-1}) \|x_e\|_{M_t^{-1}} / \|x_e\|_{M_t^{-1}} + x_e^T \theta_{t-1}, x_e^T \theta_t)
\]
\[
= \max|Z_j| \leq v \|x_e\|_{M_t^{-1}} \cdot \max |Z_j|.
\]

By Lemma \([12]\) setting \( \delta = |E|t^2 \), we have \( \max_j |Z_j| \leq \sqrt{2 \log t + \sqrt{2 \log |E|} + 4 \log t} \) with probability at least \( 1 - 1/|E|^2 \). By union of probability, for all \( e \in E \),
\[
v \|x_e\|_{M_t^{-1}} \cdot \max_j |Z_j| \leq v \|x_e\|_{M_t^{-1}} \left( \sqrt{2 \log(t) + \sqrt{2 \log |E|} + 4 \log t} \right)
\]
with probability at least \( 1 - 1/t^2 \). As \( \tilde{u}_t(e) = \text{Proj}_{[0,1]} \tilde{w}_t(e) \), by setting \( \beta_t = \sqrt{(2 \log t + \sqrt{2 \log |E|} + 4 \log t}) \), we conclude that \( |\tilde{u}_t(e) - x_e^T \theta_t| \leq \beta_t \|x_e\|_{M_t^{-1}} \) with probability at least \( 1 - 1/t^2 \), which completes the proof. □

D.2 Proof of Theorem 2

Proof of Theorem 2. Let \( F_t \) be the history of past edge-level observations and actions by the end of round \( t \), and \( \mathcal{H}_t = F_{t-1} \cup \{\tilde{\theta}_1, \ldots, \tilde{\theta}_t\} \). Then \( \theta_t \) is \( \mathcal{H}_{t-1} \)-measurable, and \( \tilde{u}_t, \tilde{u}_t \) are \( \mathcal{H}_t \)-measurable. Define the event \( \xi_t \) that both \( x_e^T \theta_t \) and \( \tilde{u}_t(e) \) are concentrated around their respective means.
\[
\xi_t := \left\{ x_e^T \theta_t - x_e^T \theta^* \leq \alpha_t \sqrt{x_e^T M_t^{-1} x_e}, \forall e \in E \right\} \cap \left\{ |\tilde{u}_t(e) - x_e^T \theta_t| \leq \beta_t \sqrt{x_e^T M_t^{-1} x_e}, \forall e \in E \right\}
\]
By Lemma \([10]\), Lemma \([11]\) and union of probability, we have \( \mathbb{P}(\xi_t) \geq 1 - 2/t^2 \) and \( \mathbb{P}(\xi_t) \leq 2/t^2 \).
Let $S_t$ be the seed set selected in round $t$, and $R_t^\eta = f(S^*, \bar{w}) - \frac{1}{\eta} f(S_t, \bar{w})$ be the corresponding $\eta$-scaled regret. Given $\mathcal{F}_t$, it can be written as

$$
E[R_t^\eta] = E[f(S^*, \bar{w}) - \frac{1}{\eta} f(S_t, \bar{w})] \cdot P(\xi_t) + E[R_t^\eta | \xi_t] \cdot P(\xi_t)
$$

$$
= E[f(S^*, \bar{w}) - \frac{1}{\eta} f(S_t, \bar{w})] \cdot P(\xi_t) + \frac{1}{\eta} E[f(S_t, \bar{u}_t) - f(S_t, \bar{w})] \cdot P(\xi_t) + E[R_t^\eta | \xi_t] \cdot P(\xi_t)
$$

First, it is easy to see $f(S^*, \bar{w}) \leq |V|$ as the reward is bounded by number of nodes in $\mathcal{D}$ and thus we have $E[R_t^\eta | \xi_t] \leq |V|$. Further recall that $P(\xi_t) \leq 2/t^2$. As a result,

$$
E[R_t^\eta] \cdot P(\xi_t) \leq 2|V|/t^2.
$$

Then, consider $Q_2$. It is easy to see that

$$
Q_2 = E\left[ f(S_t, \bar{u}_t) - f(S_t, \bar{w}) \right] \leq E\left[ f(S_t, \bar{u}_t) - f(S_t, \bar{w}) \right] \xi_t \leq \sum_{e \in \mathcal{E}} \{O_t(e, S_t, \bar{w})\} \xi_t(e) \mid \xi_t.
$$

where the second inequality comes from Lemma 8. Under event $\xi_t$, we have $|\bar{u}_t(e) - \bar{w}(e)| \leq |\bar{u}_t(e) - x^e_\eta \theta_t| + |x^e_\eta \theta_t - \bar{w}| \leq (\alpha_t + \beta_t) \sqrt{x^e_\eta \mathbf{M}_t^{-1} x^e}$ for all $e \in \mathcal{E}$. Therefore, we have

$$
Q_2 \leq E\left[ \sum_{e \in \mathcal{E}} \{O_t(e, S_t, \bar{w})\} (\alpha_t + \beta_t) \sqrt{x^e_\eta \mathbf{M}_t^{-1} x^e} \xi_t \right].
$$

Finally, we consider $Q_1$. Let $S^*(\bar{u}_t)$ be the optimal seed set under diffusion probability $\bar{u}_t$. As $S_t$ is the $\eta$-approximation solution returned by ORACLE-IMB under $\bar{u}_t$, we have $f(S_t, \bar{u}_t)/\eta \geq f(S^*(\bar{u}_t), \bar{u}_t) \geq f(S^*, \bar{u}_t)$. Therefore, it is easy to see

$$
P\left( \frac{1}{\eta} f(S_t, \bar{u}_t) \geq f(S^*, \bar{w}) \mid \mathcal{F}_{t-1}, \xi_t \right) \geq P\left( f(S^*, \bar{u}_t) \geq f(S^*, \bar{w}) \mid \mathcal{F}_{t-1}, \xi_t \right)
$$

$$
\geq P\left( \bar{u}_t(e) \geq \bar{w}(e), \forall e \in \mathcal{E} \mid \mathcal{F}_{t-1}, \xi_t \right).
$$

For any $e \in \mathcal{E}$, as $\bar{u}_t = \text{Proj}_{[0,1]} \tilde{w}_t$ and $\bar{w}(e) \in [0,1]$, the events $\{\bar{u}_t(e) \geq \bar{w}(e)\}$ and $\{\bar{w}(e) \geq \bar{w}(e)\}$ are indeed equivalent to each other. Therefore, for any $e \in \mathcal{E}$, we have

$$
P\left( \bar{u}_t(e) \geq \bar{w}(e) \mid \mathcal{F}_{t-1}, \xi_t \right) = P\left( \bar{w}(e) \geq \bar{w}(e) \mid \mathcal{F}_{t-1}, \xi_t \right)
$$

$$
= P\left( \frac{\bar{w}(e) - x^e_\eta \theta_t}{\alpha_t \|x^e_\eta\| \mathbf{M}_t^{-1}} \geq \frac{\bar{w}(e) - x^e_\eta \theta_t}{\alpha_t \|x^e_\eta\| \mathbf{M}_t^{-1}} \mid \mathcal{F}_{t-1}, \xi_t \right)
$$

$$
= P\left( \frac{v}{\alpha_t} \cdot \max_{j=1,\ldots,t} Z_j \geq \frac{\bar{w}(e) - x^e_\eta \theta_t}{\alpha_t \|x^e_\eta\| \mathbf{M}_t^{-1}} \mid \mathcal{F}_{t-1}, \xi_t \right),
$$

where the last equality comes from Lemma 9. Under event $\xi_t$, for any edge $e \in \mathcal{E}$, we have $|\bar{w}(e) - x^e_\eta \theta_t| \leq \alpha_t \|x^e_\eta\| \mathbf{M}_t^{-1}$. Therefore, the preceding inequality becomes

$$
P\left( \bar{u}_t(e) \geq \bar{w}(e) \mid \mathcal{F}_{t-1}, \xi_t \right) \geq P\left( \max_{j=1,\ldots,t} Z_j \geq \frac{\alpha_t}{v} \cdot \frac{\bar{w}(e) - x^e_\eta \theta_t}{\alpha_t \|x^e_\eta\| \mathbf{M}_t^{-1}} \mid \mathcal{F}_{t-1}, \xi_t \right)
$$

$$
\geq P\left( \max_{j=1,\ldots,t} Z_j \geq \alpha_t/v \mid \mathcal{F}_{t-1}, \xi_t \right)
$$

$$
= 1 - P(Z_j \leq \alpha_t/v)^t.
$$

By union of probability, we have

$$
P\left( \frac{1}{\eta} f(S_t, \bar{u}_t) \geq f(S^*, \bar{w}) \mid \mathcal{F}_{t-1}, \xi_t \right) \geq 1 - |\mathcal{E}| P(Z_j \leq \alpha_t/v)^t
$$

$$
\geq 1 - \left(1 - \frac{v}{4\alpha_t \sqrt{\pi} e^{-\alpha_t^2/2v^2}}\right)^t,
$$

21
where the last inequality comes from Lemma 13 with $z = \alpha_t/v$. This further implies $\mathbb{P}\left(\frac{1}{\eta} f(S_t, \tilde{u}_t) \leq f(S^*, \bar{w}) \bigg| F_{t-1}, \xi_t \right) \leq 1 - \frac{v}{4\alpha_t \sqrt{\pi}} e^{-\alpha_t^2/2v^2}$. Denote $p_t := 1 - \frac{v}{4\alpha_t \sqrt{\pi}} e^{-\alpha_t^2/2v^2}$, we obtain

$$E[Q_1|F_{t-1}] = E[f(S^*, \bar{w}) - \frac{1}{\eta} f(S_t, \tilde{u}_t)|F_{t-1}, \xi_t] \leq |\mathcal{V}| \cdot \mathbb{P}\left(\frac{1}{\eta} f(S_t, \tilde{u}_t) \leq f(S^*, \bar{w}) \bigg| F_{t-1}, \xi_t \right) \leq |\mathcal{V}| \cdot p_t^\epsilon.$$  

Combining Eqs. (16), (17) and (18) together, we have

$$E[R_t^{\eta}] \leq \frac{\alpha_T + \beta_T}{\eta} \mathbb{E} \left[ \sum_{S_t \in \mathcal{S}} \sum_{e \in E_{S_t,v}} 1\{O_t(e)\} \frac{x^\top M_{t-1}^{-1} x}{\eta} \right] + |\mathcal{V}| \cdot \sum_{t=1}^{T} (p_t^\epsilon + 2/t^2),$$

where $N_{S_t,v} = \sum_{v \in \mathcal{V}\setminus S_t} 1\{v \in \mathcal{V} \setminus S_t\}1\{e \in E_{S_t,v}\}$ is defined in Section 5. Therefore, we have

$$R^n(T) \leq \frac{\alpha_T + \beta_T}{\eta} \mathbb{E} \left[ \sum_{t=1}^{T} \sum_{e \in E} 1\{O_t(e)\} N_{S_t,e} \sqrt{x^\top M_{t-1}^{-1} x} \right] + |\mathcal{V}| \cdot \sum_{t=1}^{T} (p_t^\epsilon + 2/t^2).$$

Moreover, by definition, $\alpha_t \geq \alpha_{t-1}$ for all $t$, thus $p_t \leq p_{t-1}$ holds for all $t$. Let $\bar{p} := p_1$, we have

$$R^n(T) \leq \frac{\alpha_T + \beta_T}{\eta} \mathbb{E} \left[ \sum_{t=1}^{T} \sum_{e \in E} 1\{O_t(e)\} N_{S_t,e} \sqrt{x^\top M_{t-1}^{-1} x} \right] + |\mathcal{V}| \cdot \sum_{t=1}^{T} (\bar{p}_t^\epsilon + 2/t^2),$$

whereas the last inequality is due to the fact that $\sum_{k=1}^{\infty} 1/k^2 = \pi^2/6$. By Lemma 1 in [25], the term $\sum_{t=1}^{T} \sum_{e \in E} 1\{O_t(e)\} N_{S_t,e} \sqrt{x^\top M_{t-1}^{-1} x}$ can be bounded as

$$\sum_{t=1}^{T} \sum_{v \in \mathcal{V}} 1\{O_t(e)\} N_{S_t,v} \sqrt{x^\top M_{t-1}^{-1} x} \leq \left( \sum_{t=1}^{T} \sum_{e \in E} 1\{O_t(e)\} N_{S_t,e}^2 \right) \frac{\log(1 + T\mathbb{E}[\xi])}{\log(1 + 1/\sigma^2)}.$$  

Moreover, for any $t$, we have

$$\mathbb{E}\left[ \sum_{e \in E} 1\{O_t(e)\} N_{S_t,e}^2 \right] = \sum_{e \in E} \mathbb{E}[1\{O_t(e)\} N_{S_t,e}^2] \leq C^2.$$  

Taking the expectation over the random oracle, we have

$$\mathbb{E}\left[ \sum_{t=1}^{T} \sum_{e \in E} 1\{O_t(e)\} N_{S_t,e}^2 \right] \leq \sum_{t=1}^{T} \sum_{e \in E} \mathbb{E}[1\{O_t(e)\} N_{S_t,e}^2] \leq nC^2.$$  

Combining the above inequality with Eqs. (19) and (20), we obtain

$$R^n(T) \leq \frac{\alpha_T + \beta_T}{\eta} C \sqrt{T |\mathcal{E}| \log(1 + T\mathbb{E}[\xi]) \log(1 + 1/\sigma^2)} + |\mathcal{V}| \cdot (1 - \bar{p}_T + \pi^2/6),$$

which completes the proof. □
D.3 Auxiliary Lemmas

Lemma 12 (Oh and Iyengar [18]) Let $Z_i \sim N(0, 1), i = 1, \cdots, n$ be $n$ standard Gaussian random variables. Then we have

$$\mathbb{P} \left( \max_i |Z_i| \leq \sqrt{2 \log(2n) + \sqrt{2 \log \frac{1}{\delta}}} \right) \geq 1 - \delta.$$

Lemma 13 (Abramowitz and Stegun [3]) For a Gaussian random variable $Z$ with mean $\mu$ and variance $\sigma^2$, for any $z \geq 1$,

$$\frac{1}{2\sqrt{\pi}z} e^{-z^2/2} \leq \mathbb{P}(|Z - \mu| \geq \sigma z) \leq \frac{1}{\sqrt{\pi}z} e^{-z^2/2}.$$