Maximizing Social Welfare in a Competitive Diffusion Model

Prithu Banerjee† Wei Chen‡ Laks V.S. Lakshmanan†

†University of British Columbia, {prithu,laks}@cs.ubc.ca ‡Microsoft Research, weic@microsoft.com

ABSTRACT
Influence maximization (IM) has garnered a lot of attention in the literature owing to applications such as viral marketing and information containment. It aims to select a small number of seed users to adopt an item such that adoption propagates to a large number of users in the network. Competitive IM focuses on the propagation of competing items in the network. Existing works on competitive IM have several limitations. (1) They fail to incorporate economic incentives in users’ decision making in item adoptions. (2) Majority of the works aim to maximize the adoption of a particular item, and ignore the collective role that different items play. (3) They focus mostly on one aspect of competition – pure competition. To address these concerns we study competitive IM under a utility-driven propagation model called UIC, and study social welfare maximization. The problem in general is not only NP-hard but also NP-hard to approximate within any constant factor. We, therefore, devise instant dependent efficient approximation algorithms for the general case as well as a $(1-1/e-\epsilon)$-approximation algorithm for a restricted setting. Our algorithms outperform different baselines on competitive IM, both in terms of solution quality and running time on large real networks under both synthetic and real utility configurations.

PVldb Reference Format:
http://dx.doi.org/10.14778/xxxx-xxxx

1. INTRODUCTION
Influence maximization (IM) on social and information networks is a well-studied problem that has gained a lot of traction since it was introduced by Kempe et al. [28]. Given a network, modeled as a probabilistic graph where users are represented by nodes and their connections by edges, the problem is to identify a small set of $k$ seed nodes, such that by starting a campaign from those nodes, the expected number of users who will be influenced by the campaign, termed influence spread, is maximized. Here, the expectation is w.r.t. an underlying stochastic diffusion model that governs how the influence propagates from one node to another. The “item” being promoted by the campaign may be a product, a digital good, an innovative idea, or an opinion.

Existing works on IM typically focus on two types of diffusion models – single item diffusion and diffusion of multiple items under pure competition. The two classic diffusion models, Independent Cascade (IC) and Linear Threshold (LT), were proposed in [28]. Advances on these lines of research have led to better scalable approximation algorithms and heuristics [18, 43, 44]. Most studies on multiple-item diffusion focus on two items in pure competition [20, 36, 41, 45], that is, every node would only adopt at most one item, never both. The typical objective is to select seeds for the second item (the follower item) to maximize its number of adoptions, or minimize the spread of the first item [20].

There are a number of key issues on multiple item diffusion that are not satisfactorily addressed in most prior studies. First, most propagation models are purely stochastic, in which if a node $v$ is influenced by a neighboring node $u$ on certain item, it will either deterministically or probabilistically adopt the item, without any consideration of the utility of that item for the node. This fails to incorporate economic incentives into the user adoption behavior. Second, most studies focus on pure competition, where each node adopts at most one item, and ignore the possibility of nodes adopting multiple items. For instance, when items are involved in a partial competition, their combined utility may still be more than the individual utility, although it may be less than the sum of their utilities. Third, most studies on competition focus on the objective of maximizing the influence of one item given other items, or minimizing the influence of existing items, and do not consider maximizing the overall welfare caused by all item adoptions.

The study by Banerjee et al. [6] is unique in addressing the above issues. It proposes the utility-based independent cascade model UIC, in which: (a) each item has a utility determined by its value, price and a noise term, and each node selects the best item or item-set that offers the highest utility among all items that the node becomes aware of thanks to its neighbors’ influence; and (b) the utility-based adoption naturally models the adoption of multiple items, in a framework that allows arbitrary interactions between items, based on chosen value functions. Banerjee et al. [6] study the maximization of expected social welfare, defined as the total sum of the utilities of items adopted by all network nodes, in expectation. However, their study is confined to the complementary item scenario, where item utilities increase when bundled together.

In this paper, we complement the study in [6] by considering the social welfare maximization problem in the UIC model when items are purely or partially competitive. Partial (pure) competition means adopting an item makes a user less likely (resp., impossible) to adopt another item. To motivate the problem, we note that for a social network platform owner (also called the host), one natural objective might be to optimize the advertising revenue, as studied by Chalermsook et al. [15], or a proxy thereof, such as expected
number of item adoptions. On the other hand, one of the key assets of a network host is the loyalty and engagement of its user base, on which the host relies for its revenue from advertising and other means. Thus, while launching campaigns, it is equally natural for the host to take into account users’ satisfaction by making users aware of itemsets that increase their utility. Social Welfare, being the sum of utilities of itemsets adopted by users, is directly in line with this objective.

As a real application, consider a music streaming platform such as the Last.fm. Benson et al. [7] using their discrete choice model showed existence of competition across different genres of songs in the Last.fm dataset. In a platform such as Last.fm, the platform owner (i.e., host) completely controls the promotion of songs and the host would like to keep making engaging recommendations to the users. Even when there are multiple competing songs from different genres, the host should recommend based on users’ preferences, i.e., the users’ utility. A similar idea extends to different competing products that an e-retailer like Amazon sells directly. Those products are already procured by the e-retailer and it has full control over how it wants to sell them. Once again, in this setting, keeping users’ satisfaction from adopting these products high helps maintain a loyal and engaged user base. Thus maximizing the overall social welfare is in line with the goal of the platform. While [6] studies this problem for complementary items, social welfare maximization under competing products is open. Moreover, under pure competition, the bundling algorithm of [6] would lead to nodes adopting at most one of several competing items, leading to poor social welfare.

Compared to [6], we also consider a more flexible setting where the allocation of some items has been fixed (e.g., the items had the seeds selected by the host earlier) and the host is only allocating seeds for the remaining items. Once again, the objective is still to maximize the total social welfare of all users in the network. We call this the CWelMax problem (for Competitive Welfare Maximization).

As it turns out, CWelMax under UIC is significantly more difficult than the welfare maximization problem in the complementary setting studied in [6]. We show that when treating the allocation as a set of item-node pairs, the welfare objective function is neither monotone nor submodular. Moreover, with a non-trivial reduction, we prove that CWelMax is in general NP-hard to approximate to within any constant factor. In contrast a constant approximation was possible in the setting considered in [6].

Despite all these difficulties, we design several algorithms that either provide an instance-dependent approximation guarantee in the general case, or better (constant) approximation guarantee in the general setting with a utility-dependent approximation guarantee, or have better (constant) approximation guarantees in special cases [5].

We conducted an extensive experimental evaluation over several real social networks comparing our algorithms with existing algorithms. Our results show that our algorithms significantly dominate existing algorithms and validate that our algorithms both deliver good quality and scale to large networks (§6).

Background and related work are discussed in §2. We conclude the paper and discuss future work in §7. Proofs compressed or omitted for lack of space can be found in [1].

2. BACKGROUND & RELATED WORK

One Item IM: A directed graph $G = (V, E, p)$ represents a social network with users $V$ and a set of connections (edges) $E$. The function $p : E \to [0, 1]$ specifies influence probabilities between users. Independent cascade (IC) model is a commonly used discrete time diffusion model [20, 28]. Given a seed set $S \subseteq V$, at time $t = 0$, only the seed nodes in $S$ are active. For $t > 0$, if a node $u$ becomes active at $t - 1$, then it makes one attempt to activate its every inactive out-neighbor $v$, with success probability $p_{uv} := p(u, v)$. The diffusion stops when no more nodes can become active.

For a seed set $S \subseteq V$, we use $\sigma(S)$ to denote the influence spread of $S$, i.e., the expected number of active nodes at the end of diffusion from $S$. For a seed budget $k$ and a diffusion model, influence maximization (IM) problem is to find a seed set $S \subseteq V$ with $|S| \leq k$ such that the influence spread $\sigma(S)$ under the model is maximized [28].

A set function $f : 2^V \to \mathbb{R}$ is monotone if $f(S) \leq f(T)$ whenever $S \subseteq T \subseteq V$; submodular if for any $S \subseteq T \subseteq V$ and any $x \in V \setminus T$, $f(S \cup \{x\}) - f(S) \geq f(T \cup \{x\}) - f(T)$; $f$ is supermodular if $f$ is submodular and $f$ is modular if it is both submodular and supermodular.

Under the IC model, IM is intractable [18, 19, 28]. However $\sigma(\cdot)$ is monotone and submodular. Hence, using Monte Carlo simulation for estimating the spread, a simple greedy algorithm delivers a $(1 - 1/e)$-approximation to the optimal solution, for any $\epsilon > 0$ [26–28]. The concept of reverse reachable (RR) sets proposed by Borgs et al. [11], has led to a family of scalable state-of-the-art approximation algorithms such as IMM and SSA for IM [16, 25, 31, 39, 44].

Multiple item competitive IM: More recently, IM has been studied involving independent items [21], and competing items [8, 12, 26, 28]. We extensively test our algorithms against state-of-the-art IM algorithms under seven different utility configurations including both real and synthetic ones, which capture different aspects of competition. Our results on real networks show that our algorithms produce social welfare up to five times higher than the baselines. Furthermore, they easily scale to large networks with millions of nodes and billions of edges. We also empirically test the effect of social welfare maximization on adoption count and show that whereas the overall adoption count remains the same, social welfare is maximized by reducing adoption of just the inferior items. To summarize, our major contributions are as follows:

- We are the first to study the competitive social welfare maximization problem CWelMax under the utility-based UIC model (§3).
- We show that social welfare is neither monotone, submodular, nor supermodular; furthermore, it is NP-hard to approximate CWelMax within any constant factor, in general (§4).
- We provide several algorithms that either solve the CWelMax in the general setting with a utility-dependent approximation guarantee, or have better (constant) approximation guarantees in special cases (§5).

We conducted an extensive experimental evaluation over several real social networks comparing our algorithms with existing algorithms. Our results show that our algorithms significantly dominate existing algorithms and validate that our algorithms both deliver good quality and scale to large networks (§6).
In this section, we first briefly review the utility driven independent cascade model (UIC for short) proposed in [6]. Then we describe the competitive setting of UIC studied in this paper and formally state the new problem we address.

Review of UIC Model: UIC integrates utility driven adoption decision of nodes, with item propagation. Every node has two sets of items – desire set and adoption set. Desire set is the set of items that the node has been informed about (and thus potentially desires), via propagation or seeding. Adoption set, is the subset of the desire set that has the highest utility, and is adopted by the user. The utility of an itemset $I \subseteq \mathcal{I}$ is defined as $U(I) = V(I) - P(I) + N(I)$, where $V(\cdot)$ denotes users’ latent valuation for an itemset, $P(\cdot)$ denotes the price that user needs to pay, and $N(\cdot)$ is a random noise term that denotes our uncertainty in users’ valuation.

Budget vector $\vec{b} = (b_1, ..., b_N)$ represents the budgets associated with the items, i.e., the number of seed nodes that can be allocated with that item. An allocation is a relation $\mathcal{S} \subseteq \mathcal{V} \times \mathcal{I}$ such that $\forall i \in \mathcal{I}$ : $|\{(v, i) \mid v \in \mathcal{V}\}| \leq b_i$. $S^S_i := \{v \mid (v, i) \in \mathcal{S}\}$ denotes the seed nodes of $\mathcal{S}$ for item i and $S^D_i := \bigcup_{s \in S} S^S_i$. When the allocation $\mathcal{S}$ is clear from the context, we write $S$ (resp., $S_i$) to denote $S$ (resp., $S_i^S$).

Before a diffusion begins, the noise terms of all items are sampled, and they are used until the end of that diffusion. The diffusion proceeds in discrete time steps, starting from $t = 1$. $\mathcal{R}^S(v, t)$ and $\mathcal{A}^S(v, t)$ denote the desire and adoption sets of node $v$ at time $t$. At $t = 1$, the seed nodes have their desire sets initialized according to the allocation $\mathcal{S}$ as, $\mathcal{R}^S(v, 1) = \{i \mid (v, i) \in \mathcal{S}\}$, $\forall v \in \mathcal{S}^D$. These seed nodes then adopt the subset of items from the desire set that maximizes the utility. The propagation then unfolds recursively for $t \geq 2$ in the following way. Once a node $u$ adopts an item $i$ at time $t - 1$, it influences its out-neighbor $u$ with probability $p_{iu}^u$, and if it succeeds, then $i$ is added to the desire set of $u$ at time $t$. Subsequently $u$ adopts the subset of items from the desire set of $u$ that maximizes the utility. Adoption is progressive, i.e., once a node adopts an item, it cannot unadopt it later. Thus $\mathcal{A}^S(u, t) = \arg \max_{T \subseteq S^D(u)} |\{\mathcal{U}(T) \mid T \supseteq \mathcal{A}^S(u, t - 1) \land \mathcal{U}(T) \geq 0\}|$. The propagation converges when there is no new adoption in the network. For more details readers are referred to [6].

Social welfare maximization relative to a fixed seed set: Let $G = (V, E, p)$ be a social network, $\mathcal{I}$ the universe of items under consideration. We consider a utility-based objective called social welfare, which is the sum of all users’ utilities of itemsets adopted by them after the propagation converges. Formally, $\mathcal{E}(\mathcal{U}(\mathcal{A}^S(u)))$ is the expected utility that a user $u$ attains for a seed allocation $\mathcal{S}$ after the propagation ends. The expected social welfare for $\mathcal{S}$, is $\rho(\mathcal{S}) = \sum_{u \in \mathcal{N}} \mathcal{E}(\mathcal{U}(\mathcal{A}^S(u)))$, where the expectation is over both the randomness of propagation and randomness of noise terms $N(\cdot)$.

In a social network, a campaign may often be launched on top of other existing campaigns, where the seeds for some items $I_1 \subseteq \mathcal{I}$ may already be fixed. Let $\mathcal{S}'$ be this fixed allocation for items in $I_1$. Then $I_2 = \mathcal{I} \setminus I_1$ is the set of items for which the seeds are to be selected. We define the problem of maximizing expected social welfare, on top of a fixed seed allocation as follows.

Welfare maximization under competition: To model competition, we assume that $\mathcal{V}$ is submodular [13], i.e., the marginal value of an item with respect to an itemset $I \subseteq \mathcal{I}$ decreases as $I$ grows. We assume $\mathcal{V}$ is monotone, since it is a natural property for valuations. We set $\mathcal{V}(\emptyset) = 0$. For $i \in \mathcal{I}$, $N(i) \sim \mathcal{D}_i$ denotes the noise term associated with item $i$, where the noise may be drawn from any distribution $\mathcal{D}_i$, having a zero mean. Every item has an independent noise distribution. For a set of items $I \subseteq \mathcal{I}$, we assume the noise and price to be additive. Since noise is drawn from a zero
mean distribution, $E[U(I)] = V(I) - P(I)$. Below, we refer to $V, P, \{D_i\}_{i \in I}$ as the model parameters and denote them collectively as Param.

**Problem 1 (CWelMax)** Given $G = (V, E, p)$, the set of model parameters Param, an existing fixed allocation $S^P$, and budget vector $b$, find a seed allocation $S^*$ for items $I_2$, such that $\forall i \in I_2$, $|S^P| \leq b$, and $S^*$ maximizes the expected social welfare, i.e., $S^* = \arg \max S^P \rho(S \cup S^P)$.

Note that this problem subsumes the typical "fresh campaigns" setting as a special case where $I_1 = \emptyset$ and hence $S^P = \emptyset$.

**An equivalent possible world model:** In [6], the authors proposed an equivalent possible world interpretation of the diffusion under UIC, which we will find useful. We briefly review this below. Let $(G, Param)$ be an instance of CWelMax, where $G = (V, E, p)$. A possible world $w = (w_1, w_2)$ consists of an edge possible world (edge world) $w_1$ and a noise possible world (noise world) $w_2$: $w_1$ is a deterministic graph sampled from the distribution associated with $G$, where each edge $(u, v) \in E$ is sampled in an independent probability of $p_{uv}$; and $w_2$ is a sample of noise terms for items in $I$, drawn from noise distributions in Param. Note that propagation and adoption in $w$ is fully deterministic. In a possible world $w$, $N_u(i)$ is the noise for item $i$ and $U_u(I)$ is the (deterministic) utility of itemset $I$. The social welfare of an allocation $S$ in $w$ is $\rho_w(S) := \sum_{v \in V} U_v(A_v^s(v))$, where $A_v^s(v)$ is the adoption set of $v$ at the end of the propagation in world $w$. The expected social welfare of an allocation $S$ is $\rho(S) := E_w[\rho_w(S)] = E_{w_1}[E_{w_2}[\rho_{w_1}(S)]]$.

**4. PROPERTIES OF UIC**

It is easy to see that CWelMax is NP-hard.

**Proposition 1 CWelMax in the UIC model is NP-hard.**

**Proof Sketch.** Classic IM is a special case of CWelMax.

**Item blocking.** Under the complementary setting in [6] leveraged the reachability property: if a node $v$ adopts an item $i$ in any possible world $w$, then all the other nodes that are reachable from $v$ in $w$ will also adopt it. This property does not hold under the competitive setting. In fact, adoption of one particular item can block the propagation of another item, making social welfare non-monotone and non-submodular.

**Theorem 1** Expected social welfare is not monotone, and neither submodular nor supermodular, with respect to sets of node-item allocation pairs.

**Proof.** We show a counterexample for each of the three properties. Consider a simple network with two nodes $u$ and $v$, and a directed edge $(u, v)$ with probability 1. Assume that there is no noise, i.e., noise is 0. There are three items in propagation whose utility configuration is shown in Fig. 1 (a).

**Monotonicity.** Consider two allocations $S^1 = \{(u, i_1)\}$ and $S^2 = \{(u, i_1), (v, i_2)\}$. Clearly $S^1$ is a subset of $S^2$. Under $S^1$, both $u$ and $v$ adopt $i_1$, thus $\rho(S^1) = 8$. However under $S^2$, $u$ adopts $i_1$ but $v$ adopts $i_2$. Thus $\rho(S^2) = 7 < \rho(S^1)$.

**Submodularity.** Consider $S^1 = \{(v, i_2), (v, i_3)\}$, $S^2 = \{(v, i_2), (v, i_3), (u, i_1)\}$ and $(u, i_1)$. Clearly $S^1$ is a subset of $S^2$. Under $S^1$, only $v$ adopts $i_2$. Under $S^2$, $u$ adopts $i_1$ and $v$ adopts $i_2$. So $\rho(S^1) = 4$. Under $S^2$, $v$ adopts $i_2$. So $\rho(S^2) = 5 > \rho(S^1)$. Hence $\rho(S^2) + \rho(S^1) = 5 > \rho(S^1)$. Hence $\rho(S^2) = 5 > \rho(S^1)$.

**Supermodularity.** Consider $S^1 = \emptyset$, $S^2 = \{(u, i_2)\}$ and $(u, i_1)$. Clearly $S^1$ is a subset of $S^2$. Under $S^1$, there is no adoption by any node. Under $S^1$ and $(u, i_1)$, $u$ and $v$ both adopt $i_1$. So $\rho(S^1) = 8$. Under $S^2$, $v$ adopts $i_2$. So $\rho(S^2) = 5 > \rho(S^1)$.

The absence of these properties makes CWelMax really hard to approximate, as shown next.

**Theorem 2** CWelMax in the UIC model is NP-hard. Further there is no PTIME algorithm that can approximate CWelMax within any constant factor $c$, $0 < c < 1$, unless $P = NP$.

**Proof.** We show that Influence maximization under the IC model is an NP-hard problem. As a special case of CWelMax.

The result follows from the fact that the IM problem under the IC model is a special case of CWelMax: let $I = \{i\}, \forall i \in I$, and let $\rho(i) = 1$, and set the noise term for item $i$ to 0. This makes $U(i) = 1$ so any influenced node will adopt it. Thus, the expected social welfare is simply the expected spread. We know maximizing expected spread under the IC model is NP-hard [28].

**Hardness of approximation**

We prove the theorem by a gap introducing reduction from SET COVER. Suppose there is a PTIME $c$-approximation algorithm $A$ for CWelMax, for some $0 < c \leq 1$. Given an instance $\mathcal{I} = (\mathcal{F}, X)$ of SET COVER, where $\mathcal{F} = \{S_1, \ldots, S_r\}$ is a collection of subsets over a set of ground elements $X = \{g_1, \ldots, g_n\}$, and a number $k (k < r < n)$, the question is whether there exist $k$ subsets from $\mathcal{F}$ that cover all the ground elements, i.e., whether $\exists C \subseteq \mathcal{F} : |C| = k$ and $\cup_{S \in C} S = X$. We can transform $\mathcal{I}$ in polynomial time into an instance $\mathcal{J}$ of CWelMax.

As an overview, our reduction will show that for a YES-instance of SET COVER, the optimal expected welfare in the corresponding CWelMax instance is high and for a NO-instance, it is low. More precisely, let $x_n^+$ (resp., $x_n^-$) be the optimal welfare on the transformed instance $\mathcal{J}$ whenever the given instance $\mathcal{I}$ is a YES-instance (resp., NO-instance). Our reduction ensures that $x_n^+ < cx_n^-$. In this case, running $A$ on $\mathcal{J}$ will clearly allow us to decide if $\mathcal{I}$ is a YES-instance or not, which is impossible unless $P = NP$.

For the rest of the discussion we assume no noise, i.e., noise distribution has 0 mean and variance. Also all the edge probabilities of the graph are set to 1. The details of the reduction follow.

**Value, price and utility.** We consider four items $i_1, i_2, i_3$ and $i_4$, with the following utility configuration: $i_1$ competes with $i_2$ and $i_3$, and $i_4$ has a higher individual utility than both. However $\{i_2, i_3\}$ as a bundle has higher utility than $i_4$. $i_4$ has a very high utility, much higher than that of any other individual item. A node adopting $i_1$ adopts $i_4$ if it arrives later. However if a node adopts the bundle $\{i_2, i_3\}$, then it will not adopt $i_4$ later. We use this configuration in the following way. For a YES-instance, $i_1$ blocks $i_2$ and $i_3$, consequently allowing a large number of nodes to adopt $i_4$. For a NO-instance, however, most nodes adopt $\{i_2, i_3\}$, blocking $i_4$ adoption. Thus by setting $c(\mathcal{U}(i_4) > \mathcal{U}(\{i_2, i_3\}$), the desired gap in the optimal welfare is achieved. Lastly, as we will see later in the proof that we need $\mathcal{U}(\{i_2, i_3\}) < c/4 \mathcal{U}(\{i_1, i_4\})$.

Assuming no noise terms, Fig. 1 (b) provides an abstract summary of this utility configuration, focusing on the items $i_1, i_2, i_3, i_4$. Note that in addition to the aforementioned constraints, the value function is monotone and submodular as required. Further, we give one such complete configuration (over all four items $i_1, i_2, i_3, i_4$) in Table 1, for $c = 0.4$.

**The network.** The graph instance constructed from the given instance of SET COVER is illustrated in Fig. 2 (a). We first create a
bipartite graph having two partitions of \( r \) nodes \( \{s_1, \ldots, s_r\} \) corresponding to the sets \( S_i \) and \( n \) nodes \( \{g_1, \ldots, g_n\} \) corresponding to the ground elements \( g_j \) respectively. There is a directed edge from \( s_i \) node to \( g_j \) node iff \( g_j \in S_i \) in the SET COVER instance. There are also \( m \) number of “\( a \)”, “\( b \)”, “\( c \)”, “\( f \)” nodes. Node \( a_i \) is connected with a directed edge to the corresponding \( g_j \) node. For each \( g_j \), there is an incoming (directed) edge from \( a_i \), and an outgoing edge to \( f_j \). Each node \( b_i \) is connected to \( f_j \) with a path of length 2, i.e., \( b_i \rightarrow c_i \rightarrow f_j \), where \( c_i \) is the intermediate node between \( b_i \) and \( f_j \). This construction creates the following behavior. If all the \( g \) nodes adopt \( a_i \) then all the \( f \) nodes adopt \( i \). However if any one of the “\( g \)” nodes adopts \( j \) and all the “\( e \)” nodes adopt \( i \), then all the “\( f \)” nodes adopt \( \{i, j\} \). The significance of this behavior will be clear in the remaining part of the proof.

For a large \( N \gg n \) that is a multiple of \( n \), we create \( d_1, \ldots, d_{2n} \). For \( 1 \leq i \leq n \), we add the edges \((f_i, d_{i(N−n)/2}) \), \((f_i, d_{i+n}) \). This gadget helps create the gap in the welfare that we are aiming for.

We create \( n \) copies of “\( j \)” nodes. Each \( j_i \) is connected to \( c_i \) by a directed path from \( j_i \) to \( c_i \) of length 3, where \( i \) and \( m_i \) are the intermediate nodes. Similar to \( f_i \), each \( c_i \) is connected to \( N/n \) “\( d \)” nodes – \( d_{i+1} \). As a preview, the “\( a \)” nodes (resp., “\( b \)” nodes, “\( c \)” nodes) will serve as seeds for item \( i \) (resp., item \( i_3 \) and item \( i_4 \)). Note that the length of the paths from “\( j \)” nodes (seeds of \( i_2 \)) to “\( d \)” nodes is 4, while the paths from the seeds of \( i_2 \) and \( i_3 \) to “\( d \)” nodes are of length 3. Thus, if \( \{i_2, i_3\} \) are not blocked by \( i_1 \), all the “\( d \)” nodes will adopt \( \{i_2, i_3\} \) and cannot adopt \( i_4 \) when it arrives later. This completes one copy of the graph, shown in Fig. 2(a). All edge probabilities are set to 1. We will explain the significance of node color and the surrounding box soon.

**Budgets and seed allocation**: We set the budgets for \( i_2, i_3, i_4, \ldots, n \) each. The “\( a \)” nodes are seeded with \( i_2 \), “\( b \)” nodes are seeded with \( i_3 \) and “\( c \)” nodes are seeded with \( i_4 \). These seeds are fixed (see Fig. 2). The budget for \( i_1 \) is set to \( k \) and these seeds are to be selected so as to maximize the expected social welfare. We complete the construction of the instance \( J \) of CWelMax by making \( N \) copies of the graph described above.

Notice for YES-instance of the set cover, the \( N \) number of “\( d \)” nodes adopt \( \{i_1, i_4\} \). Hence we have,

**Claim 1**: Suppose \( \mathcal{I} = (\mathcal{F}, X) \) is a YES-instance and \( \mathcal{J} \) the transformed instance of CWelMax corresponding to Fig. 2(a) and the seed allocation described above. The optimal welfare on \( \mathcal{J} \) is \( x^* > N \times U(\{i_1, i_4\}) \).

**Proof**: For a YES-instance, choosing the corresponding \( s \) nodes of the SET COVER solution maximizes the welfare. In this case, every “\( g \)” node has at least one in-neighbor that adopted \( i_1 \) at time \( t = 1 \). Thus, all “\( g \)” nodes adopt \( i_1 \) at time \( t = 2 \). Consequently all the “\( f \)” and “\( d \)” nodes adopt \( i \) at time \( t = 3 \) and \( t = 4 \) respectively. Later at \( t = 5 \) when \( i_4 \) arrives, those “\( d \)” nodes adopt \( \{i_1, i_4\} \). The optimal welfare in this case \( x^* > N \times U(\{i_1, i_4\}) \).

**Claim 2**: Suppose \( \mathcal{I} = (\mathcal{F}, X) \) is a YES-instance and \( \mathcal{J} \) the transformed instance of CWelMax corresponding to Fig. 2(b) and the

| Item  | \( \mathcal{V} \) | \( \mathcal{P} \) | \( \mathcal{U} \) | Constraints |
|-------|-----------------|-----------------|-----------------|-------------|
| \( i_1 \) | \( v - d_1 \) | \( p - d_1 - \delta \) | \( u_1 \) | \( \delta > 0 \) |
| \( i_2 \) | \( v \) | \( p \) | \( u_2 \) | \( u_1 = u_2 + \delta \) |
| \( i_3 \) | \( v \) | \( p \) | \( u_3 \) | \( u_1 = u_3 + \delta \) |
| \( i_4 \) | \( 8v/c \) | \( 4p/c \) | \( u_4 \) | \( u_4 > \delta \) |
| \( i_1, i_2 \) | \( 2v - d_1 - \epsilon \) | \( 2p - d_1 - \epsilon \) | \( u_2 \) | \( \epsilon > \delta \) |
| \( i_1, i_3 \) | \( 2v - d_1 - \epsilon \) | \( 2p - d_1 - \epsilon \) | \( u_2 \) | \( \epsilon > \delta \) |
| \( i_1, i_4 \) | \( v - d_1 - \epsilon \) | \( 3p - d_1 - \epsilon \) | \( u_1, i_4 \) | \( u_4 > \delta \) |
| \( i_2, i_3 \) | \( 2v \) | \( 2p \) | \( u_2 \) | \( u_2 > \delta \) |
| \( i_1, i_2, i_3 \) | \( 3v - d_1 - \epsilon \) | \( 3p - d_1 - \epsilon \) | \( u_1, i_4 \) | \( u_4 > \delta \) |
| \( i_1, i_3, i_4 \) | \( 2v + 8e/c - d_1 - e \) | \( 2p - d_1 - 4e/c \) | \( u_2 \) | \( u_2 > \delta \) |
| \( i_2, i_3, i_4 \) | \( 2v + 8e/c - d_1 - e \) | \( 2p - d_1 - 4e/c \) | \( u_2 \) | \( u_2 > \delta \) |
| \( i_1, i_2, i_3, i_4 \) | \( \max \{v_1, v_2, v_3, v_4\} \) | \( 3p - d_1 - 4e/c \) | \( u_2 \) | \( u_2 > \delta \) |

**Figure 1**: Utility configurations: (a) Used in Theorem 1; (b) Used in Theorem 2
seed allocation described above. The optimal welfare on \( J \) is 
\[
x^* > N^2 \times U(\{i_1, i_4\}).
\]

Proof. There are \( N^2 \) “d” nodes in all. For a YES instance, the optimal seeds for \( i_1 \) are exactly the solution of SET COVER. It follows from Claim 1 that the optimal welfare in that case is
\[
x^*_y > N^2 \times U(\{i_1, i_4\}).
\]

For a NO-instance, maximum number of “d” nodes adopt \( i_4 \). The only candidate seeds which could achieve that are “s”, “g”, “f”, “c”, and “o” nodes. We show that regardless of which \( k \) seeds are chosen for item \( i_1 \), the welfare achieved is \( x^*_y < cN^2U(i_4) \). First, observe that choosing \( k \) “g” nodes as seeds of item \( i_1 \) achieves a welfare no less than that of any other choice of \( k \) seeds for \( i_1 \). In the \( k \) copies where \( g \) nodes are seeded with \( i_1 \) we have the following adoptions. \( k \) number of “g” and \( n \) number of “f” nodes adopt \( i_1 \); \( n \) number of “o” adopt \( i_2 \); \( n \) number of “m” and “o” nodes adopt \( i_3 \); and \( N \) number of “d” nodes adopt \( \{i_1, i_4\} \). For the remaining \( N - k \) copies we have: \( n \) number of “g” nodes adopt \( i_2 \); \( n \) number of “e” nodes adopt \( i_3 \); \( 3n \) number of “m”, “n” and “e” nodes adopt \( i_4 \); \( N \) number of “d” nodes and \( n \) number of “f” nodes adopt \( \{i_2, i_3\} \). Lastly from the seeds, \( n \) number of “n” nodes adopt \( i_2 \); \( n \) number of “b” nodes adopt \( i_3 \) and \( n \) number of “j” nodes adopt \( i_4 \). So the total welfare for a NO-instance is:
\[
k((n + 1)U(i_1) + (2n - 1)U(i_2) + 2nU(i_2) + 4nU(i_4) + NU(\{i_1, i_4\}) + (N - k)(2nU(i_2) + 2nU(i_4) + 4nU(i_4) + (N + n)U(\{i_2, i_3\})) + nU(i_2) + U(i_3) + U(i_4).
\]

Since \( U(\{i_2, i_3\}) > U(i_1) > U(i_2) = U(i_4),\)

\[
(*) < (k + 2n + 5Nn + N^2 - kN)U(\{i_2, i_3\}) + (3Nn + n)U(i_4) + NkU(\{i_1, i_4\}) = N^2U(\{i_2, i_3\}) + (k + 2n + 5Nn - kN)U(\{i_2, i_3\}) + (3Nn + n)U(i_4) + NkU(\{i_1, i_4\}) = N^2U(\{i_2, i_3\}) + (8Nn - kN)U(\{i_2, i_3\}) + 4NnU(i_4) + NkU(\{i_1, i_4\}) \quad (**)
\]

Hence combining this with Claim 2 we get,

**Claim 3** Given a SET COVER instance \( I \), transform it into an instance \( J \) of CWelMax and run the algorithm \( A \) on \( J \). \( I \) is a YES-instance iff the welfare returned by \( A \) is \( > cN^2U(\{i_1, i_4\}) \). \( I \) is a NO-instance iff the welfare returned by \( A \) is \( < cN^2U(\{i_1, i_4\}) \).

Proof. Suppose \( I \) is a YES-instance. Then, by Claim 2, the optimal welfare of \( J \) is \( > N^2U(\{i_1, i_4\}) \), so the welfare returned by Algorithm \( A \) on \( J \) is \( > cN^2U(\{i_1, i_4\}) \). Suppose \( I \) is a NO-instance. Then the optimal welfare of \( J \) is \( x^*_n < cN^2U(\{i_1, i_4\}) \).

Thus, even if Algorithm \( A \) returned the optimal welfare on the NO-instance \( J \), it would be strictly less than the welfare returned on the corresponding YES-instance.

For a NO-instance the optimal welfare is upper bounded by Eq. 2. Hence, the claim follows.

The theorem follows, \( A \) cannot exist unless \( P = NP \).

5. APPROXIMATION ALGORITHMS

Since the CWelMax problem cannot be approximated within any constant factor in general, in this section we propose several approximation algorithms that either produce a non-constant approximation guarantee dependent on the problem instance or a constant approximation guarantee for a special case of CWelMax. We first define some important notions.

**Truncated utility.** For accounting the social welfare of an allocation, we develop the notion of **truncated utility** of an item. Recall that when the noise of an item makes its utility negative, no node adopts the item. Hence what contributes to the final expected social welfare is the set of non-negative contributions to utility. We call this the **truncated utility**, denoted \( \mathcal{U}'(I) := \max(0, \mathcal{U}(I)) \). Thus
for a (node, item) allocation pair \((v, i)\), its expected social welfare (when there are no other allocations) is \(p(v, i) = E[4^+ (i)]\), where \(\sigma([v])\) is the influence spread of \(v\).

**Minimum and maximum utility bundle.** We define \(u_{\text{min}} = \min_{i \in I} E[4^+ (i)]\) as the minimum expected truncated utility of any item in \(I\), and \(u_{\text{max}} = E[\max_{i \in I} 4^+ (I)]\) as the expected maximum truncated utility of any item bundle in \(I\). Note that the definitions of \(u_{\text{min}}\) and \(u_{\text{max}}\) are not symmetric: (a) \(u_{\text{min}}\) takes the minimum of an expectation, while \(u_{\text{max}}\) takes the expectation of a maximum; and (b) \(u_{\text{min}}\) takes minimum on single items while \(u_{\text{max}}\) takes maximum among all bundles. The reason of this asymmetry will be clear in our analysis.

**Superior and inferior item.** A given itemset \(I\) is said to have a superior item \(i_u\), if the least possible utility of \(i_u\) is strictly higher than the highest possible utility of any item in \(I \setminus \{i_u\}\). Notice the definition of superior item entails that the noise distribution should be bounded in some way. We discuss a practical way to bound the noise in our experiments (§6). Given a superior item, all the other items of the itemset are called inferior items.

In what follows, we present three different algorithms with progressively better theoretical guarantees, under progressively stronger assumptions. As a preview, our first algorithm SeqGRD provides a \(u_{\text{min}} / m(1 - 1/\epsilon)\)-approximation in the most general case. Our second algorithm, MaxGRD, assumes no prior allocations, i.e., \(S^p = \emptyset\). Under this assumption, it provides a \(1 / m(1 - 1/\epsilon)\)-approximation, where \(m\) is the number of items. By simply returning the better of the two allocations produced by SeqGRD and MaxGRD, the bound is improved to \(\max (u_{\text{min}} / m, \frac{1}{m} (1 - 1/\epsilon))\), when \(S^p = \emptyset\). Our final algorithm SupGRD assumes that there exists a superior item in the itemset, the allocations for all inferior items are fixed, and that items exhibit pure competition. Under these assumptions, it provides a \(1 / (1 - 1/\epsilon)\)-approximation.

### 5.1 SeqGRD Algorithm

The pseudocode of our first algorithm SeqGRD is shown in Algorithm 1.

**Algorithm 1 SeqGRD**

\[\begin{align*}
1: \quad & S^p \leftarrow \text{Seed nodes of the allocation } S^p \\
2: \quad & S_{\text{Seq}} \leftarrow \text{PRIMA }^+: (G, \epsilon, \ell, S^p, \bar{b}, \sum_{i \in I} b_i) \\
3: \quad & S_{\text{Seq}} \leftarrow \emptyset \\
4: \quad & \text{Sort } I_2 \text{ in decreasing order of the expected truncated utility} \\
5: \quad & \text{Added} \leftarrow \emptyset \\
6: \quad & \text{for } i \in I_2 \text{ do} \\
7: \quad & \quad \text{top } b_i \text{ nodes from } S_{\text{Seq}} \\
8: \quad & \quad \text{if } \rho(S_{\text{Seq}} | S^p) < \rho((S_{\text{Seq}} \cup \{i\}) | S^p) \text{ then} \\
9: \quad & \quad S_{\text{Seq}} \leftarrow S_{\text{Seq}} \cup \{i\} \\
10: \quad & \quad \text{Remove those } b_i \text{ nodes from } S_{\text{Seq}} \\
11: \quad & \quad \text{Added} \leftarrow \text{Added} \cup \{i\} \\
12: \quad \text{end if} \\
13: \quad \text{end for} \\
14: \quad \text{for } i \in I_2 \setminus \text{Added} \text{ do} \\
15: \quad \quad S_{\text{Seq}} \leftarrow \text{top } b_i \text{ nodes from } S_{\text{Seq}} \\
16: \quad \quad S_{\text{Seq}} \leftarrow S_{\text{Seq}} \cup \{i\} \\
17: \quad \quad \text{Remove those } b_i \text{ nodes from } S_{\text{Seq}} \\
18: \quad \text{end for} \\
19: \quad \text{Return } S_{\text{Seq}}
\end{align*}\]

Algorithm SeqGRD considers the general setting where a set of items have already been seeded and \(S^p\) corresponds to this partial allocation. Let \(S^p := \{v \mid (v, i) \in S^p\}\) be the seed set allocated in \(S^p\) and let \(I_2\) denote the remaining items which have yet to be allocated. The algorithm takes a graph \(G\), to-be-allocated itemset \(I_2\), item budget vector \(\bar{b}\) for the items in \(I_2\), accuracy parameter \(\epsilon\), tolerance parameter \(\ell\), the partial allocation \(S^p\) as input. It first selects a seedset \(S_{\text{Seq}}\) of size \(\bar{b}\), where \(\bar{b} := \sum_{i \in I} b_i\) (line 2). To select the seeds it uses an algorithm, called PRIMA\(^+\), which delivers a set of seeds that are approximately optimal w.r.t. the marginal gain \(\sigma(S | S^p)\). We present the PRIMA\(^+\) algorithm in §5.2.1 and establish its properties.

SeqGRD then sorts the items based on their truncated utility (line 4). Starting from the item \(i\) having the highest truncated utility, it tries to allocate the item to the top \(b_i\) nodes of \(S_{\text{Seq}}\), \(S^p\). If the allocation \(S_{\text{Seq}} \times \{i\}\) yields a positive marginal welfare, it is added to the existing allocation and nodes of \(S_{\text{Seq}}\) are removed for future considerations (Lines 8-12). The items that are not allocated in this iteration are appended following an arbitrary order (lines 14-18) and allocated at the end.

Let \(\Gamma_w(S)\) be the set of nodes reachable from a seed set \(S\) in the possible world \(w\). We first establish the following lemma.

**Lemma 1** Let \(S\) be an allocation, \(S\) be its seedset, let \(w\) be a random possible world. Then for any node \(v \in V\), we have

\[\frac{u_{\text{min}}}{\epsilon} \leq E_u \left[ U_w(A_w(v)) \mid v \in \Gamma_w(S) \right] \leq u_{\text{max}}.\]

**Proof.** Let \(w = (w_1, w_2)\) where \(w_1\) is the edge possible world and \(w_2\) is the noise possible world. Note that, (a) reachable set \(\Gamma_w(S)\) is only determined by the edge possible world \(w_1\), so we can use \(\Gamma_{w_1}(S)\) to represent it; (b) utility function \(U_w(\cdot)\) is only determined by the noise possible world \(w_2\), so we can use \(U_{w_2}(\cdot)\) to represent it; and (c) adoption set \(A_w(v)\) is determined by both \(w_1\) and \(w_2\), so we use \(A_{w_1, w_2}(v)\) to represent it.

\[\frac{u_{\text{min}}}{\epsilon} \leq \frac{\sum_{v \in \Gamma_w(S)} U_w(A_w(v))}{\epsilon} \leq u_{\text{max}}.\]

We first prove the lower bound \(u_{\text{min}}\). To do so, we prove that for any fixed edge possible world \(w_1\) and conditioned on \(v \in \Gamma_{w_1}(S)\), we have \(E_{w_2} [U_{w_2}(A_{w_1, w_2}(v))] \geq u_{\text{min}} = \min_{i \in I} E[4^+ (i)] = \min_{i \in I} E_{w_2}[4^+ (i)]\). Once this is proved, Eq. (3), we immediately have \(E_u \left[ U_w(A_w(v)) \mid v \in \Gamma_w(S) \right] \geq E_{w_1} [u_{\text{min}}] \geq u_{\text{min}}\).

Consider first a seed \(u \in S\). Let \(A_{w_1, w_2}(u, 1)\) be the set of items adopted by \(u\) initially at time 1 before the propagation starts. Let \(I^u\) be the set of items allocated to \(u\) in \(S\). Note that \(I^u\) is determined purely by the fixed allocation \(S\) and is not affected by the noise or edge possible world. By our model, node \(u\) will select the best item bundle in \(I^u\) and adopt them as \(A_{w_1, w_2}(u, 1)\). Then we know that \(U_{w_2}(A_{w_1, w_2}(u, 1)) \geq u_{\text{min}} = \min_{i \in I} E[4^+ (i)]\). Therefore, we have

\[\frac{u_{\text{min}}}{\epsilon} \leq \frac{\sum_{v \in \Gamma_w(S)} U_{w_2}(A_{w_1, w_2}(v, 1))}{\epsilon} \leq \max_{i \in I^u} U_{w_2}(i) \leq u_{\text{min}}.\]

This means that for the initial seed adoption, we have that their expected utility is at least \(u_{\text{min}}\). Now for any \(v \in \Gamma_w(S)\), \(v\) is reachable from some seed node \(u \in S\) via some shortest path in the edge possible world \(w_1\). By the propagation model, then the utility of \(v\)’s final adoption \(U_{w_2}(A_{w_1, w_2}(v))\) should be at least the utility of \(u\)’s initial adoption, \(U_{w_2}(A_{w_1, w_1}(u, 1))\). Then we have \(E_u [U_{w_2}(A_{w_1, w_2}(v))] \geq u_{\text{min}} \frac{E_{w_2}[U_{w_2}(A_{w_1, w_2}(u, 1))]}{\epsilon} \geq u_{\text{min}}\).

This concludes the proof.

\(^1\text{We assume } I_1 \cap I_2 = \emptyset.\)
The proof of the upper bound $u_{\text{max}}$ is straightforward: 
\[ E_w\left[ \mathcal{U}_w(\mathcal{A}^S(v)) \mid v \in \Gamma_w(S) \right] \leq E_w\left[ \max_{I \subseteq I_2} \mathcal{U}^+(I) \mid v \in \Gamma_w(S) \right] = u_{\text{max}}. \]

**Lemma 2** Let $\mathcal{S}$ be an allocation and $\mathcal{S}$ its corresponding seed nodes. Then $u_{\text{min}} \cdot \sigma(S) \leq \rho(\mathcal{S}) \leq u_{\text{max}} \cdot \sigma(S)$.

**Proof.** The lower bound is derived below:

\[ \rho(\mathcal{S}) = E_w\left[ \sum_v \mathcal{U}_w(A^S(v)) \right] = \sum_v E_w\left[ \mathcal{U}_w(A^S(v)) \mid v \in \Gamma_w(S) \right] \geq \sum_v P_r[v \in \Gamma_w(S)] \cdot E_w\left[ \mathcal{U}_w(A^S(v)) \mid v \in \Gamma_w(S) \right] \]

where the inequality is by Lemma 1. The upper bound can be shown in a similar way.

We are now ready to prove the following bound for SeqGRD.

**Theorem 3** Let $\mathcal{S}^{\text{Seq}}$ be the allocation returned by the Algorithm SeqGRD. Given $\epsilon, \ell > 0$, we have $\rho(\mathcal{S}^{\text{Seq}} \cup \mathcal{S}^{P}) \geq \frac{u_{\text{min}}}{\max} (1 - \frac{1}{\ell} - \epsilon) \rho(\mathcal{S}^A \cup \mathcal{S}^P)$ w.p. at least $1 - \frac{1}{\sqrt{\ell}}$, where $\mathcal{S}^A$ is any arbitrary allocation of items in $I_2$ respecting the budget constraint.

**Proof.** Let $\mathcal{S}^{\text{Seq}}, \mathcal{S}^A$ and $\mathcal{S}^P$ be the seed sets of the allocations $\mathcal{S}^{\text{Seq}}, \mathcal{S}^A$ and $\mathcal{S}^P$ respectively. Then $|\mathcal{S}^A| \leq \sum_{i \in I_2} b_i$. By SeqGRD, $\mathcal{S}^{\text{Seq}}$ exhausts all budgets for items in $I_2$, so $|\mathcal{S}^{\text{Seq}}| = \sum_{i \in I_2} b_i$. Since $\mathcal{S}^{\text{Seq}}$ computes the expected marginal social welfare of the allocation $\rho(S_{\text{seq}}^{\text{Seq}} \times \{i\}) \mid \mathcal{S}^{P})$, we have that w.p. at least $1 - \frac{1}{\sqrt{\ell}},$

\[ \sigma(S_{\text{seq}}^{\text{Seq}} \mid \mathcal{S}^{P}) \geq (1 - \frac{1}{\ell} - \epsilon) \sigma(S^A \mid \mathcal{S}^{P}). \]

From this, it follows that

\[ \sigma(S^{\text{Seq}} \cup \mathcal{S}^{P}) \geq (1 - \frac{1}{\ell} - \epsilon) \sigma(S^A \cup \mathcal{S}^{P}). \]

Therefore, we have

\[ \rho(\mathcal{S}^{\text{Seq}} \cup \mathcal{S}^{P}) \geq \frac{u_{\text{min}}}{\max} \cdot \sigma(S^{\text{Seq}} \cup \mathcal{S}^{P}) \]

\[ \geq \frac{u_{\text{min}}}{\max}(1 - \frac{1}{\ell} - \epsilon) \cdot \sigma(S^A \cup \mathcal{S}^{P}) \]

\[ \geq \frac{u_{\text{min}}}{\max}(1 - \frac{1}{\ell} - \epsilon) \rho(S^A \cup \mathcal{S}^{P}), \]

Computing marginals involves sampling, which takes significant time in large networks. On the other hand, the marginal check avoids the phenomenon of items with lower (truncated) utility blocking those with higher utility, to some extent. Thus even though SeqGRD-NM is faster than SeqGRD and has the same approximation guarantee, under certain utility configurations, the welfare produced by SeqGRD-NM can be worse than that of SeqGRD. We explore this in our experiments in §6. On the other hand, we still append all items in the end to exhaust the budget in SeqGRD (lines 14–18). To really discard a certain itemset, we need to exhaustively search through all itemset combinations, which is time-consuming. So we only do a simple marginal check in SeqGRD, and append all items at the end to ensure the theoretical guarantee.

### 5.2 MaxGRD Algorithm

Our next algorithm MaxGRD provides $\frac{1}{m}(1 - \frac{1}{\ell})$-approximation, when $\mathcal{S}^{P} = \emptyset$, i.e., no prior allocation. The pseudocode is shown in Algorithm 2. Like SeqGRD, MaxGRD also selects its seedset $S^{\text{Max}}$ using PRIMA $^{+}$, but the size of the seedset is different: $b := \max_{i \in I_2} b_i$, i.e., the maximum budget of any unallocated item (line 1). Then for every item $i \in I_2$, it computes the expected marginal social welfare of the allocation $\rho((S_{\text{max}}^{\text{Max}} \times \{i\}) \mid \mathcal{S}^{P})$, where $S_{\text{max}}^{\text{Max}}$ is the set of first $b_i$ nodes of $S^{\text{Max}}$. It returns the allocation with the maximum welfare (line 3).

**Algorithm 2** MaxGRD($G, \epsilon, \ell, \mathcal{S}^{P}, I_2, \bar{b})$

1: $S_{\text{Max}}^{\text{Max}} \leftarrow \text{PRIMA}^{+}(G, \epsilon, \ell, \mathcal{S}^{P}, \bar{b}; \max_{i \in I_2} b_i)$

2: $S_{\text{Max}} \leftarrow \text{top} b_i$ nodes of $S_{\text{Max}}^{\text{Max}}, \forall i \in I_2$

3: $\mathcal{S}^{\text{max}} \leftarrow \arg \max_{i \in I_2} \{\rho(S_{\text{max}}^{\text{Max}} \times \{i\}) \mid \mathcal{S}^{P}\}$

4: Return $\mathcal{S}^{\text{Max}} \times \{\mathcal{S}^{\text{max}}\}$

Notice that MaxGRD is applicable even when $\mathcal{S}^{P} \neq \emptyset$, so we have provided the algorithm for this general case. However, it enjoys an approximation bound only for the special case, when $\mathcal{S}^{P} = \emptyset$. We prove the following lemma under this constraint, which is instrumental in the proof of the approximation bound. A key observation is that given a possible world $w$, the utility function $U_w(\cdot)$ in that possible world is submodular. This follows from the fact that valuation is submodular and price and noise, being additive are both modular.

**Lemma 3** Let $\mathcal{S} := \bigcup_{i=1}^{m}(S_i \times \{i\})$ be an arbitrary allocation, where $S_i$ is the set of seed nodes of item $i$. Then $\rho(\bigcup_{i=1}^{m}(S_i \times \{i\})) \leq \sum_{i=1}^{m} \rho((S_i \times \{i\}))$.

**Proof Sketch.** Consider an arbitrary but fixed possible world $w$ and an arbitrary item $i \in I_2$. Let $v$ be any node that adopts $i$ in $w$ under the allocation $\bigcup_{i=1}^{m}(S_i \times \{i\})$. We can show that $v$ must also adopt $i$ in $w$ when the allocation is only $S_i \times \{i\}$. The lemma follows from this.

**Theorem 4** Suppose that $\mathcal{S}^{P} = \emptyset$. Let $S^{\text{Max}}$ be the allocation produced by MaxGRD. Given $\epsilon, \ell > 0$, we have $\rho(S^{\text{Max}}) \geq \frac{1}{m}(1 - \frac{1}{\ell} - \epsilon) \rho(S^A)$ w.p. at least $1 - \frac{1}{\sqrt{\ell}}$, where $\mathcal{S}^A$ is any arbitrary allocation.

**Proof.** Recall that item $i$ has a budget $b_i$ and expected utility $u_i$. Since in an arbitrary allocation $|S_{\text{max}}^A| \leq b_i$, from the prefix preserving property of PRIMA $^{+}$ we have,

\[ \sigma(S_{\text{max}}^{\text{Max}}) \geq \left(1 - \frac{1}{\ell} - \epsilon\right) \sigma(S^A). \]

Let $E[t^+(i)]$ be the expected positive utility of item $i$. We have $\rho(S_{\text{max}}^{\text{Max}} \times \{i\}) = E[t^+(i)] \cdot \sigma(S_{\text{max}}^{\text{Max}})$ and $\rho(S^A \times \{i\}) = \sigma(S^A)$.
Sampling Reachable Sets (RR-sets), as proposed in IMM [44]. Here, we construct a marginal RR-set as follows. Choose a root node $b \in V$ uniformly at random, add it to $R$, and start a BFS from $v$. Whenever $u \in R_v$, sample each incoming edge $(u', u)$ w.p. $p_{u'u}$ and add it to $R_u$. Stop when no new nodes are added to $R_v$. If any stage $R_v \cap S_p^b \neq \emptyset$, then set $R_v := \emptyset$. That is, whenever a generated RR-set “hits” $S_p^b$, just set it to $\emptyset$. Algorithm 3 shows the pseudo code of this marginal RR-set sampling process.

Given graph $G$, a number $\theta$ denoting how many RR-sets needs to be sampled and a fixed seed nodes $S_p^b$. Marginal_Sampling generates $\theta$ number of RR-sets to $\mathcal{R}$ from $G$, based on the marginal on $S_p^b$.

**Algorithm 3 Marginal_Sampling($G, \mathcal{R}, \theta, S_p^b$)**

1. while $|\mathcal{R}| \leq \theta$ do
   2. Select $v$ from $G$ uniformly at random
   3. $R \leftarrow$ BFS($v$)
   4. if $R \cap S_p^b \neq \emptyset$ then
      5. $R \leftarrow \emptyset$
   6. end if
   7. $\mathcal{R} \leftarrow \mathcal{R} \cup R$
   8. end while
   9. Return $\mathcal{R}$

PRIMA + using Marginal_Sampling, achieves the property of prefix preservation on marginals. Its pseudo code is shown in Algorithm 4.

**Algorithm 4 PRIMA ($G, \epsilon, \ell, S_p^b, \bar{b}, \bar{\bar{b}}$)**

Initialize $\mathcal{R} = \emptyset, s = 1, n = |V|, i = 1, \epsilon' = \sqrt{2} \cdot \epsilon$, budgetSwitch = false, $\bar{b} = \bar{b} \cup \bar{\bar{b}}$
1. $\ell = \ell + \log 2/\log n, \ell' = \log_{\epsilon'}(\bar{\bar{b}})$
while $i \leq \log_2(n) - 1$ and $s \leq \bar{b}$ do
   2. $k = b_s, LB = 1$
   3. $x = \frac{\theta_s}{\ell'}; \theta_s = \lambda_s^b / LB$, where $\lambda_s^b$ is defined in Eq. (8)
   4. Marginal_Sampling($G, \mathcal{R}, \theta_s, S_p^b$)
   5. if budgetSwitch then
      6. $S_k = \{k \} \text{ in the ordered set } S_{s-1}$ returned from the previous call to NodeSelection
   7. else $S_k = \text{NodeSelection}(R, k)$
   8. end if
   9. if $n \cdot F_{\mathcal{R}}(S_k) \geq (1 + \epsilon') \cdot x$ then
      10. $LB = n \cdot F_{\mathcal{R}}(S_k)/(1 + \epsilon')$
      11. $\theta_k = \lambda_s^b / LB$
   12. end if
      13. $s = s + 1; \text{budgetSwitch} = \text{false}$
   14. end if
   15. end while
   16. if $s \leq \bar{b}$ then
      17. $\theta_s = \lambda_s^\mathcal{B} / LB$
   18. end if
   19. end if
20. if $\mathcal{R} = \emptyset$
   21. Marginal_Sampling($G, \mathcal{R}, \theta_s, S_p^b$)
   22. $S_\mathcal{B} = \text{NodeSelection}(\mathcal{R}, \bar{b})$ return $S_{\mathcal{B}}$ as the final seed set;

It runs in time $O((\bar{b} + \ell + \log_{\epsilon'}(\bar{\bar{b}}))(n + m) \log n \cdot \epsilon^{-2})$, where $\bar{b} := \sum_{i \in \ell_2} b_i$ for SeqGRD and $\bar{b} := \max_{i \in \ell_2} b_i$, for MaxGRD.

**5.3 SupGrd Algorithm**

Our third algorithm SupGRD provides a constant $(1 - \frac{1}{\ell})$-approximation. The bound holds under more restrictive conditions as given below.

**Conditions required for SupGRD approximation bound.** (i) There exists a superior item (defined in §5) $i_w$ in the item set: i.e., under any noise possible world $w_2, U_{w_2}(i_w) > U_{w_2}(i), \forall i \in \mathcal{I}$
Lemma 4 Given $S^P$ and $I_2$, let $S_1$ and $S_2$ be two allocations over $I_2$ such that $S_1 \subseteq S_2$. Then $\rho(S_1 \cup S'^P) \leq \rho(S_2 \cup S'^P)$.

Proof. In an arbitrary but fixed world $w = (w_1, w_2)$, $U_{w_2}(A_{w_1,w_2}^S(v)) \leq U_{w_2}(A_{w_1,w_2}^{S_2}(v))$. This is because if $v$ changes its adoption between the two allocations, then it must change it to $i_m$ since all other inferior items are fixed. Since $i_m$ is the superior item, the claim holds. Since this holds for every $w$, the lemma follows.

Lemma 5 Given $S^P$ and $I_2$, let $S_1$ and $S_2$ be two allocations over $I_2$ such that $S_1 \subseteq S_2$. Let $s = (u, i_m) \notin S_2$ be an allocation pair. Then $\rho(s \mid S_1 \cup S'^P) \geq \rho(s \mid S_2 \cup S'^P)$.

Proof. Let $w = (w_1, w_2)$ be an arbitrary but fixed possible world. Let $C_1$ denote the set of all nodes that adopt $i_m$ under allocation $s$ but not under $S_1 \cup S'^P$, $i = 1, 2$. Then $C_2 \subseteq C_1$. Thus,

$$\rho_{w_1,w_2}(s \mid S_2 \cup S'^P) = \sum_{v \in C_2} (U_{w_2}(\{i_m\}) - U_{w_2}(A_{w_1,w_2}^{S_2}(v))) \leq \sum_{v \in C_1} (U_{w_2}(\{i_m\}) - U_{w_2}(A_{w_1,w_2}^{S_1}(v))) = \rho_{w_1,w_2}(s \mid S_1 \cup S'^P).$$

Since this holds for every $w$, the lemma follows.

Since social welfare is monotone and submodular, a standard greedy selection based on the marginal welfare will have $(1 - \frac{1}{e})$-approximation. However since computing spread itself is #P-hard, computing the exact marginal is not feasible. In IMM, sampling using RR-sets has been used to achieve some of the art performance. In what follows, by extending IMM [44], we adopt a martingale approach for seed selection in SupGRD. Given $\epsilon$ and $\ell$, SupGRD returns a seed set that has a $(1 - \frac{1}{e}) - \epsilon$-approximation w.p. at least $1 - \frac{1}{\ell^2}$.

In the classical setting, RR-set samples are used to compute an unbiased estimation of the spread. In our case we need to estimate the marginal welfare using the RR-sets. Towards that we define a notion of weight for every RR-set. The weight of an RR-set $R_v$ denotes the marginal gain in the expected social welfare achieved by activating the root $v$ of the RR-set $R_v$. Thus it is the difference between the expected truncated utility of the item that the root $v$ adopts under the existing partial allocation $S^P$ and that of $i_m$. To ensure that the root $v$ indeed adopts $i_m$, the path from some seed of $i_m$ to $v$ should be no longer than that from any seed of $S^P$ to $v$. Thus a weighted RR-set is constructed as follows.

Definition 2 (Weighted Reverse Reachable Set). For a given fixed allocation $S^P$ and a node $v \in G$, a weighted RR-set of $v$, $R_v$ is obtained by starting $R_v = \{v\}$ and starting a BFS from $v$ such that: for $u \in R_v$, sample each incoming edge $(u', u)$ w.p. $p_{u',u}$ and add it to $R_v$; stop when either no new nodes are added or $R_v$ overlaps $S^P$ (the distance from any node in $R_v$ to $v$ along the reversely generated edges is at most the distance from $S^P$ to $v$). Then, the weight of $R_v$ is $w(R_v) = U^v(\{i_m\}) - \max_{i \in I^v \mid i \in S^P \cap R_v} U^v(i)$, where $I^v$ denotes the items allocated to node $s$ in the allocation $S^P$.

Algorithm 5 NodeSelection($R, b'$)

1: Initialize $S_{b'} = \emptyset, i = 1$
2: while $i \leq b'$ do
3: Select $v \in V \setminus S_{b'}$ which has the highest $M_R(S_{b'} \cup v) - M_R(S_{b'})$
4: $S_{b'} \leftarrow S_{b'} \cup \{v\}$
5: Remove $R$ from $R$ if $v \in R$
6: end while
7: return $S_{b'}$ as the final seed set

SupGRD samples RR-sets using an early termination as described in Definition 2. This construction ensures that if any member of a weighted RR-set is seeded with $i_m$, the root of the RR-set, $v$, will adopt $i_m$. In what follows, we first establish the connection between marginal social welfare and weighted RR-sets and then present efficient seed selection and RR-set sampling algorithms to maximize the marginal social welfare.

For a node set $S$, let $\mathbb{E}[\cdot]$ be an indicator function denoting whether $S$ covers the (weighted) RR set $R$, i.e., $\mathbb{E}[S \cap R \neq \emptyset] = 1$, if $S \cap R \neq \emptyset$, 0 otherwise. Also let $\mathcal{L}(G)$ denote the distribution of all the live edge graphs, then extending the result of Borg et al., we now prove the following lemma for weighted RR-sets.

Lemma 6 For given seed sets $S$ and $S^P$, we have $\mathbb{E}_{\omega \sim \mathcal{L}(G)}[\rho_{w_1}(S \mid S^P)] = n \cdot \mathbb{E}_{v \sim V \setminus S, w \sim \mathcal{L}(G)}[\mathbb{I}(S \cap R_v \neq \emptyset) \cdot w(R_v)]$ where $n = |V|$ is the number of nodes in $G$.

Proof.

$$\mathbb{E}_{w_1 \sim \mathcal{L}(G)}[\rho_{w_1}(S \mid S^P)] = \mathbb{E}_{v \sim V \setminus S, w \sim \mathcal{L}(G)} \mathbb{I}(S \cap R_v \neq \emptyset) \cdot w(R_v) = n \cdot \mathbb{E}_{v \sim V \setminus S, w \sim \mathcal{L}(G)} \mathbb{I}(S \cap R_v = \emptyset)$$

We now extend the RR-set based efficient approximation IM algorithm, IMM, for maximizing welfare. Similar to IMM, our algorithm SupGRD has two key phases, namely, NodeSelection and Sampling. The NodeSelection phase is similar to that of IMM, except we consider the weight of RR-sets while selecting seed nodes. For a node set $S$ and a collection of weighted RR-sets $R$, define $M_R(S) := \sum_{R \in R} \mathbb{I}[S \cap R \neq \emptyset] \cdot w(R)$. Let $b = b_m$ be the budget of the superior item $i_m$. Given a set $R$, NodeSelection selects $b'$ seeds that maximizes $M_R$ (Algorithm 5).

Next, the goal of the Sampling phase (Algorithm 6) is to generate $R$ such that $|R| \geq \lambda \cdot OPT$, where $OPT$ is the optimal welfare, and $\lambda$ is defined as follows,

$$\lambda = 2 \alpha \cdot (1 - 1/e) \cdot \alpha + \beta^2 \cdot e^{-2},$$

where, $\alpha = \sqrt{\ell \log n + \log 2}$ and $\beta = \sqrt{(1 - 1/e) \cdot \log \left(\frac{\ell}{n} \right) + \ell (\log n + \log 2)}$.

Since OPT is unknown, the Sampling first ensures that it finds a lower bound to OPT w.h.p. For that it deploys a statistical test using a binary search on the range of OPT. The maximum possible value of the welfare OPT is $UB = n \times u_{max}$, where every node in the network adopts the superior item $i_m$, $u_{max}$ is the utility of $i_m$. Thus the binary search ranges from 1 to $UB$ (Line 2).

A good lower bound is found when the condition of Line 8 is satisfied. Different from [44], this condition directly operates on welfare. This is a key step in the correctness of the algorithm, hence we prove it explicitly in Lemma 7.
Lemma 7 Let $x \in [1, UB]$, $\epsilon'$ and $\delta \in (0, 1)$, then if we invoke NodeSelection with $|R| = 0$, where

$$\theta \geq \frac{(2 + \frac{1}{2} \epsilon') \cdot (\log(n) + \log(1/\delta))}{\epsilon'^2} \cdot \frac{n}{x} \quad (7)$$

and $S$ is the output NodeSelection returns, then if $OPT < x$, $\frac{n}{\theta} \cdot M_R(S) < (1 + \epsilon') \cdot x$, w.p. at least $(1 - \delta)$. \square

Proof. Let $x_i$ be a random variable for each $R_i \in R$ defined as, $x_i = \min\{w(R_i) \mid S \neq R_i\}$, where $w_{max}$ is the maximum weight possible for any RR set. Thus, $0 \leq x_i \leq 1$, which ensures the martingale property. Now let $F_R(S) = \frac{M_R(S)}{w_{max}}$, $P = E[F_R(S)]$ and $\alpha = \frac{(1 + \epsilon') \cdot x}{np \cdot w_{max}} - 1$. Using Lemma 6 and linearity of expectation,

$$p = E[F_R(S)] = E\left[\frac{M_R(S)}{w_{max}}\right] = E[p(S \mid S^p)]/(w_{max})$$

$$\leq OPT/(w_{max} \cdot n) \leq x/(w_{max} \cdot n)$$

Consequently $\alpha > \epsilon' \cdot x/(np)$ and from Lemma 6 of [44],

$$P \geq \frac{n}{\theta} \cdot M_R(S) \geq (1 + \epsilon') \cdot x \leq \delta/(n^p)$$

Finally by applying union bound we get $\frac{n}{\theta} \cdot M_R(S) < (1 + \epsilon') \cdot x$, w.p. at least $(1 - \delta)$. \square

Thus by setting $\lambda'$ using Eq. (8), we get Theorem 2 of [44]

$$\lambda' = \frac{(2 + \frac{1}{2} \epsilon') \cdot (\log(n) + \epsilon' \cdot \log n + \log \log_2 n) \cdot n}{\epsilon'^2} \cdot \frac{n}{\theta}$$

The rest of the proof is similar to that of [44], which gives us the following result.

Theorem 5 Let $S^P$ be a partial allocation on the inferior items. Let $S^G$ be the allocation of the superior item produced by SupGrd. Given $\epsilon, \ell > 0$, we have $P(S^G \cup S^P) \geq (1 - \frac{1}{2} - \epsilon)P(S^A \cup S^P)$ w.p. at least $1 - \frac{1}{|V|^7}$, where $S^A$ is an arbitrary allocation.

### Table 2: Network Statistics

| Network        | # edges | avg. degree | type   |
|----------------|---------|-------------|--------|
| NetHEPT        | 15\,2K  | 4.13        | undirected |
| Douban-Book    | 31.4K   | 6.5         | directed |
| Douban-Movie   | 141K    | 7.8         | directed |
| Orkut          | 27.4K   | 7.9         | undirected |
| Twitter        | 3.07M   | 77.5        | directed |
| # nodes        |         |             |        |
| NetHEPT        | 47\,547 |             |        |
| Douban-Book    | 70.5\,M |             |        |
| Douban-Movie   | 41.7\,M |             |        |
| Orkut          | 1.47\,G |             |        |
| Twitter        | 1.47\,G |             |        |

Running time: Let $w_{min}$ be the minimum weight of an RR set. Then using Lemma 9 of [44], the expected total time to generate $R$ is determined by,

$$E\left[\sum_{R \in R} \text{wid}(R)\right] = E[|R|] \cdot EPT$$

$$\leq O((b' + \ell)n \log n \cdot \epsilon^{-2})/OPT \cdot \frac{m}{n} OPT/w_{min}$$

$$= O((b' + \ell)(n + m) \log n \cdot \epsilon^{-2}/w_{min})$$

Notice that generating an RR-set from scratch for the final node selection (line 13), following the fix of [17], only adds a multiplicative factor of 2. Hence the overall asymptotic running time to generate $R$ remains unaffected.

6. EXPERIMENTS

6.1 Experiment Setup

All our experiments are run on a Linux machine with Intel Xeon 2.6 GHz CPU and 128 GB RAM.

6.1.1 Networks. Our experiments were conducted on five real social networks: NetHEPT, Douban-Book, Douban-Movie, Twitter, and Orkut, whose characteristics are summarized in Table 2. Of these, NetHEPT, Douban-Book, and Douban-Movie are benchmarks in IM literature [37], while Twitter and Orkut are two of the largest public networks available at [2].

6.1.2 Algorithms compared. In the experiments our four algorithms – SeqGRD, SeqGRD-NM, MaxGRD, and SupGRD are compared against three baselines – TCIM, Balance-C and greedyWM. There is no previous work that can deal with both arbitrary degree of competition and multiple items in propagation. Our first two baselines each covers one aspect. TCIM [34] in particular assumes a propagation model which is an extension of the IC model under pure competition. It can, however, handle more than two items. Given fixed seed sets of other competing items, TCIM selects seeds of an item under a budget constraint, such that the number of adoptions of that item is maximized. When we run TCIM for multiple items, we select seeds for each of the items one by one, while keeping the seeds of other items fixed and then report the allocation that produces the maximum welfare.

In contrast, Balance-C [23] does not assume pure competition, but it works only when number of items in propagation is two. Given an initial seed placement of the two items, Balance-C chooses the remaining seeds such that at the end of the propagation, the number of nodes seeing either both the items or none, is maximized. Thus for competing ideas, Balance-C ensures that there is a balanced exposure of the two ideas to the most number of nodes. It is non-trivial to extend Balance-C for more than two items hence we compare against it only in two item set up.

Both TCIM and Balance-C aim to maximize adoption count, not social welfare. Our third baseline greedyWM maximizes the social welfare directly. It greedily selects iteratively the (node, item) pair that maximizes the marginal social welfare, till the budgets are exhausted. Below, by deterministic utility of an itemset $I$, we mean $\mathcal{V}(I) - \mathcal{P}(I)$, i.e., its utility with the noise term ignored.
6.1.3 Default parameters. Following previous works [25, 39] we set probability of edge $e = (u, v)$ to $1/d_{in}(v)$, where $d_{in}(v)$ is the in-degree of node $v$. Unless otherwise specified, we use $\epsilon = 0.5$ and $\ell = 1$ as our default in all the algorithms that use these parameters. We test the algorithms across a wide variety of utility configurations to cover different aspects of competition. We will describe the configurations as we present the corresponding experiments. Whenever marginal gains are required, we run 5000 simulations and take the average result.

6.2 Experiments with two items

For our first set of experiments we restrict the number of items to two so that we can compare against all of the mentioned baselines. We also consider four different configurations to capture different kinds of competition. The details of the configurations are given in Table 3. In configurations C1 and C2, the items exhibit pure competition. In C1, items have comparable individual utility. In C2, the difference between individual utility is high: $i$’s deterministic utility is 1, 10 times higher than that of $j$. C3 and C4 exhibit soft competition. Except for C4, in all configurations we consider the same budget for both items (uniform); budget is varied from 10 to 50 in steps of 10. In C4, we fix the budget of $i$ to 50 and vary $j$’s budget (non-uniform) from 30 to 100 in steps of 20. We assume $S^0 = \emptyset$ in these configurations. Since it does not meet constraints required by SupGRD, we defer the comparison until §6.2.3.

6.2.1 Running time. First we compare the running time of the algorithms using C1 as a representative case. Fig. 3 shows the results on four networks. SeqGRD-NM is orders of magnitude faster than other algorithms in every network. The reason is that SeqGRD-NM does not compute any marginal. Each marginal computation requires iterating over 5000 samples, which significantly increases the running time. For the same reason greedyWM and Balance-C exhibit exorbitantly high running time: they do not in fact complete in 6 hours on a large network such as Orkut. Hence they are not included in Fig. 3(d). Except for SeqGRD-NM, none of the other algorithms scale to the largest network Twitter. We will compare SeqGRD-NM and SupGRD on Twitter later. Performance on other configurations show similar trends, and hence omitted for brevity.

6.2.2 Social welfare. We now compare the expected social welfare achieved by the algorithms on the four configurations (Fig. 4). We show the results only for Douban-Movie, since the trend of the results is similar on other networks. In all configuration SeqGRD, SeqGRD-NM and greedyWM outperform all other algorithms. The difference in welfare is up to 3× higher. MaxGRD in particular allocates just one of the two items. Thus when items exhibit soft competition (C3 and C4), MaxGRD performs significantly worse. Balance-C performs comparatively better under soft competition (C3), however for a non-uniform budget again its performance drops. TCIM on the other hand aims to maximize the adoption count of the item being allocated. Thus it also ends up allocating both the items in same seed nodes. This reduces the overall social welfare for configuration such C1, where both Balance-C and TCIM perform comparatively worse. Social welfare produced by greedyWM is consistently good, but its running time is exorbitantly high, which prohibits its applicability on any decently sized network. SeqGRD-NM on the other hand is the fastest algorithm, which produces similar welfare across all these four configurations.

6.2.3 Comparison against SupGRD. In this experiment we compare SupGRD and SeqGRD-NM on the two largest networks, Orkut and Twitter. We use utility configurations of C1 and C2, but adopt the seed placements needed to meet the constraints required for SupGRD. Recall that for SupGRD the seeds for the inferior items need to be fixed. Hence we select the top 50 nodes using IMM and set them as seeds of $j$. Then, the seeds of $i$ are selected using SupGRD and SeqGRD-NM with the budget being varied from 10 to 50 in steps of 10. We call these new configurations C5 and C6 respectively.

Since the top nodes in terms of the spread are given to $j$, these two cases pose a unique challenge of dealing with arbitrary degree of competition when maximizing welfare. When items’ utilities are
similar, in C5, new seeds of $i$ should be chosen in a way that minimizes $i$’s overlap with $j$ propagation. Instead in C6, when $i$ has much higher utility, it should be allocated to the top seed nodes. That way, the number of nodes that can be reached by $i$ is much higher and that helps boost the overall social welfare. As can be seen from our results next, that SupGRD can navigate through these varied “strategies”, while SeqGRD-NM cannot.

Fig. 5 (a) and (b) shows the result on the expected social welfare on Orkut and Twitter respectively. “SeqGRD-NM-C5” (resp. “SupGRD-C5”) refers to SeqGRD-NM (resp. SupGRD) on C5 and “SeqGRD-NM-C6” (resp. “SupGRD-C6”) on C6. Notice that in C5 the welfare produced by the two algorithms are comparable. However in C6, where the gap between the individual utilities of the two items is higher, difference between the welfare of SupGRD and SeqGRD-NM is also larger. The reason for that is as follows. SupGRD-NM uses PRIMA’ to select the seeds of $i$. Consequently to maximize the marginal gain in spread, it minimizes the overlap in the spread of $i$ and $j$ and hence allocates $i$ to lower ranked nodes in terms of spread. However $i$ is the superior item, so allocating lower ranked nodes to $i$ decreases the overall welfare.

Fig. 5(c) and (d) compares the running time of the two algorithms on Orkut and Twitter. Both the algorithms scale on these large networks. Unlike SeqGRD-NM, running time SupGRD depends on the utility configurations as well. As our running time analysis ($\S$5.3) suggests, when the minimum utility of an item is lower, the running time of SupGRD is higher. However as can be seen, even on large networks, the difference in the running times is not very high: e.g., in configuration C6, the running time of SupGRD is only a 2× that of SeqGRD-NM, whereas in C5 the running times are similar. To summarize, SupGRD addresses this unique challenge of dealing with an arbitrary degree of competition, with a slightly higher running time.

6.3 More than two items

Except for Balance-C, all the algorithms can deal with multiple items. In this section, we study their performances when the number of items is more than two. First, we show the impact of increasing the number of items on the running time and social welfare produced. Then we study how the algorithms behave under some challenging configurations designed using multiple items.

6.3.1 Impact of number of items. For this experiment, the configuration we test is as follows. Each individual item has expected utility of 1 and the items exhibit pure competition. Every item has budget 50 and $S^i = \emptyset$.

Fig. 6(a) and (b) show respectively, the running time and social welfare produced by the five algorithms. Since Balance-C cannot run on more than two items, it is omitted. Running time of algorithms greedyWM, TCIM, MaxGRD, and SeqGRD increases significantly w.r.t the number of items. As the number of items increases, the number of times marginal check is needed for these algorithms, also increases. The marginal check is the most time consuming portion in their running time. SeqGRD-NM on the other hand relies solely on RR-sets and does not do any marginal checks. Hence the growth in running time is not high. With higher number items, the difference between the running time of SeqGRD-NM and other algorithms, increases.

In terms of social welfare, TCIM and MaxGRD perform worse than the other algorithms. MaxGRD selects only one item in the final allocation, hence it misses out on the additional welfare that could come from allocating the remaining items. Similarly TCIM tries to maximize the spread of the last allocated item, at cost of propagation of other items. Thus their welfare does not increase with more items, unlike the other algorithms.

6.3.2 Effect of marginal check. In our experiments so far, social welfare of SeqGRD-NM has been similar to other algorithms that perform marginal checks. One exception being SupGRD ($\S$6.2.3), but SupGRD assumes specific constraints that are not general. By not performing the marginal check, SeqGRD-NM runs much faster compared to other algorithms. This begs the question if there is any
advantage of using the marginal check altogether. In this experiment we show how marginal check helps avoid item-blocking that SeqGRD-NM fails to circumvent.

For this experiment, we consider three items in the propagation. Their expected utilities are specified in Table 4. $i$ has the highest expected utility, followed by $j$ and $k$ has the least. $i$ and $k$ exhibit soft competition hence bundle $\{i,k\}$ has a positive utility, but all other item bundles have negative utilities, exhibiting pure competition. We set the budget of $i$ to 500, and increase the budget of $j$ and $k$ from 100 to 500 each in steps of 100 and study the effect on the welfare produced by SeqGRD-NM and SeqGRD.

Fig. 6(c) shows the result on the NetHept network. Both algorithms first allocate $i$ as it has the highest individual utility. Then SeqGRD-NM allocates $j$ next, however this allocation is “adjacent” to $i$ since NetHept is small, and blocks propagation of $i$ more. Since the utility of $i$ is significantly higher than $j$, allocating $j$ this way in fact causes a negative marginal. SeqGRD, using marginal check, postpones allocation of $i$. After $i$, it instead allocates $k$. Although $k$ also has a low individual utility, because of soft competition, it does not block propagation of $i$ and the marginal is non-negative. It later allocates $j$, which is now further apart from $i$, hence cannot block $i$’s propagation. Thus SeqGRD produces a social welfare which is higher than that of SeqGRD-NM. Further, as the budget of $j$ increases, the amount of blocking also increases, hence the welfare difference between the two algorithms also goes up.

6.3.3 Scalability of SeqGRD-NM. Our next experiment shows the impact of network size on SeqGRD-NM using Orkut with two types of edge probabilities: (1) $1/d_{in}(v)$ and (2) fixed 0.01. We use a uniform budget of 50 for all three items. Instead of using the full network, we use breadth-first-search to progressively increase the network size so that it includes a certain percentage of the total nodes in the network. At 100%, the full network is used. Fig. 6(d) shows the results. “SeqGRD-NM, time 1” and “SeqGRD-NM, time 2” depict the running time of SeqGRD-NM on the two types of edge probabilities respectively. As the network size increases, the running time in both cases roughly has a linear increase.

6.4 Real item experiments

In this section, we learn the utilities of items from real dataset instead of the synthetic utilities used in earlier experiments. The dataset used is the LastfmGenres generated from the listening behavior of users of the music streaming service Last.fm [14, 30]. This dataset was used in [7] to learn the adoption probabilities of different items, where each genre is treated as an item. This dataset also echoes our first motivating example in the introduction. We next establish the connection between the adoption probabilities and the utilities, which enables us to learn the parameters using [7].

6.4.1 Learning the utilities. In [7], every item $i$ is associated with an adoption probability $p_i$. Adoption probability of an itemset $I = \{i, \ldots, k\}$ is $p_I = \gamma_I \prod_{j \in I} p_j + q_I$, where $q_I$ is a correction received depending on the way items in $I$ interact with each other: if the items are complementary, then the correction is positive, if competing then it is negative, and 0 if the items are independent. These probabilities and corrections are learnt in [7] from the dataset of how frequently items are selected together by the users.

According to Observation 2.2 of [7], $p_i = e^{v_i}/\sum_j e^{v_j}$, where $v_i$ is the expected utility of item $i$ as per our utility model. Given a set of learnt $p_i$, we first set $\sum_j e^{v_j} = 10000$. Then for every $i$, we set $v_i = \ln(10000 \cdot p_i)$. We choose the number 10000 to ensure that the corresponding utilities are positive. Finally we set the expected utility of item $i$, $U(i) = v_i$. Next for an itemset $I = \{i, \ldots, k\}$, [7] learns two parameters $\gamma_I$ and $q_I$. By using $q_I = \gamma \cdot q_I$ the probability of adopting the bundle, $p_I$ is derived. The expected utility of the bundle is similarly set to be $U(I) = \ln(10000 \cdot p_I)$. Notice that the exact values of utilities are not as important as the relative order of utilities of different itemsets. The way utilities are learnt is in correspondence with the adoption probabilities learned in [7].

Table 5 shows the utilities of four different items (i.e., genres) in the dataset: rock, indie, industrial and progressive, learned using the above described method. Larger bundles are either not present in the dataset or have smaller learned utilities compared to the individual items in the bundle, suggesting that items are in pure competition in our utility model.

6.4.2 Results using real parameters.

We use the learned utility configuration to compare the social welfare produced by the algorithms on two networks, namely NetHept and Orkut. For the experiment we set uniform budget for all the four items, which varies from 10 to 40 in steps of 10. The algorithms compared are TCIM, MaxGRD, SeqGRD and SeqGRD-NM. The results are shown in Fig 7.

In terms of running time the results are similar to our previous experiments (Fig 7(a)-(b)). SeqGRD-NM outperforms the other algorithms by orders of magnitude, since it does not require the time consuming marginal gain computation. For social welfare, notice that the real utility configuration exhibits pure competition. As noted earlier, under pure competition, social welfare produced by SeqGRD and SeqGRD-NM coincide. MaxGRD and TCIM on the other hand typically encourage the adoption of one single item. Hence the difference in social welfare produced by these algorithms compared to SeqGRD-NM is higher since the number of items are also more than the previous configurations we used.

6.4.3 Social welfare vs adoption.

Our final set of experiments compare the relationship between the social welfare and item adoptions. In particular, we want to investigate whether maximizing welfare for competing items could result in a significant drop in the number of item adoptions. For this experiment, we focus on two utility configurations – (i) Real utility of Table 5, which exhibits pure competition and (ii) Synthetic utility of Table 4, which exhibits a mix of partial and pure competition. NetHept and Orkut are the two networks used and each item’s budget is set to two different values, 10 and 40.

We compare our algorithm SeqGRD-NM against two baselines. After selecting the seed nodes, the first baseline allocates items to
the nodes in a round robin manner, hence it is called Round – robin. The second baseline, called Snake, is similar to Round-robin, but it flips the order for every successive sequence of allocations. To illustrate, if there are 4 seed nodes $s_1, s_2, s_3, s_4$, in order, and two items $i, j$, SeqGRD-NM allocates as $s_1: i, s_2: j, s_3: i, s_4: j$. Round-robin allocates as $s_1: i, s_2: j, s_3: i, s_4: j$ and Snake allocates as $s_1: i, s_2: j, s_3: j, s_4: i$. Table 6 shows the adoption count of each item and the social welfare produced by these algorithms under different configurations.

In terms of the social welfare objective, SeqGRD-NM dominates across all different configurations. Round-robin produces the next highest welfare. Hence we report the fractional change (+ denotes increase and − denotes decrease), in comparison to Round-robin, next to each entry of the table. The entries that deserve more attention are highlighted in green.

As can be seen, the total number of adoptions of all the items remains the same across all three algorithms. However, SeqGRD-NM generally increases the adoption of the superior product to increase the welfare, while reducing the adoption of the inferior item. On NetHEPT, for budget 10, the maximum drop in adoptions happens for the most inferior item (progressive meta), by 15.8%. For a higher budget, the drop increases (to 18.9%), because when budget increases for the superior item, SeqGRD-NM allocates lower ranked seeds for the inferior item.

The highest drop in adoption i.e., 50.1%, also happens for the item progressive meta for budget 40 on Orkut. This is because when items are purely competing, number of items is high and each item has a large budget, the inferior items’ seeds are in fact much lower ranked. However, if it exhibits partial competition with a superior item, then leveraging it the adoption does not decrease that much. That is why in Orkut even when the budget is 40, for the synthetic utility configuration, the highest drop in adoptions for the inferior items is only 23.7%. Also notice SeqGRD-NM produces significantly higher social welfare compared to the baselines, the increase being up to 37.8%. In summary, we see that our welfare maximization algorithm provide more adoptions to the superior items and fewer adoptions to the inferior items, but the amount of change is not too drastic. We argue that this is the “price” of enhancing the overall user satisfaction; also the drop in the adoptions of the inferior items is exactly because they are not as competitive.

To conclude this section, we generally observe that: (a) when the conditions required by SupGRD are met, it is the best option providing the best social welfare and competitive running time; (b) in the general case, SeqGRD-NM performs well in most cases and has the best running time, but when item blocking is significant, its marginal-checking version SeqGRD could provide better social welfare, at the cost of higher running time; (c) MaxGRD could be used to enhance the theoretical guarantee when the utility difference is high, but its superiority is not typically observed in large networks; (d) our algorithms outperform all baselines on social welfare and running time and scale to large networks. Our algorithms achieve superior welfare at the expense of a reasonable drop in the adoption count of inferior items, keeping the total adoption count unchanged.

7. CONCLUSIONS AND FUTURE WORK

In this paper, we study the problem of maximizing social welfare over competing items under the UIC model. The problem is not only NP-hard but is also NP-hard to approximate within a constant factor. Further we find that due to conflicting requirements, it is challenging to design a single algorithm that can work effectively for all different utility configurations. Yet we propose a cohort of efficient algorithms that not only provide approximation guarantees but also scale well to real large networks, and their performance is validated through extensive experiments on real-world networks.

Although welfare maximization under competition ensures that users’ total utility from adoptions is maximized, it does not directly ensure fairness. For a campaigner who often pays for advertising, ensuring that her item is seen at least by a certain number of users is critical. While fairness in IM has been studied recently, incorporating fairness in social welfare maximization will be an interesting challenge. Further, this paper and [6] studied competition and complementarity in isolation. Designing algorithms for an arbitrary mix of competing and complementary items is an intriguing problem.

| Network | Budget | Algorithm | Real Utility Configuration (as shown in Table 3) | Synthetic Utility Configuration (as shown in Table 4) |
|---------|--------|-----------|--------------------------------------------------|--------------------------------------------------|
| NetHEPT 10 | RR | 201 | 217 | 191 | 196 | 4473.08 |
| | Snake | 208 | 190 | 182 | 204 | 4686.06 |
| | SGRD-NM | 595 | 592 | 195 | 197 | 4686.06 |
| NetHEPT 40 | RR | 396 | 495 | 491 | 475 | 10953.4 |
| | Snake | 486 | 486 | 486 | 486 | 10758.2 |
| | SGRD-NM | 976 | 976 | 976 | 976 | 10758.2 |
| Orkut 10 | RR | 35791 | 35791 | 35791 | 35791 | 217 |
| | Snake | 38241 | 38241 | 38241 | 38241 | 217 |
| | SGRD-NM | 32900 | 32900 | 32900 | 32900 | 217 |
| Orkut 40 | RR | 58142 | 58142 | 58142 | 58142 | 1227.3 |
| | Snake | 58142 | 58142 | 58142 | 58142 | 1227.3 |
| | SGRD-NM | 10587 | 10587 | 10587 | 10587 | 1227.3 |

Table 6: Adoption count of different items and the overall social welfare.

![Image of Table 6](image-url)

...
8. REFERENCES

[1] Full report: Maximizing social welfare in a competitive diffusion model. https://www.dropbox.com/s/ewfj2mqc10g60ue/cepic.pdf?dl=0. Accessed: 2020-05-30.
[2] Twitter and Orkut dataset. https://snap.stanford.edu/data/. Accessed: 2020-05-30.
[3] B. Abramowitz and E. Anshelevich. Utilitarians without utilities: Maximizing social welfare for graph problems using only ordinal preferences. In AAI, pages 894–901, 2018.
[4] C. Aslay, W. L. F. Bonchi, A. Goyal, and L. V. Lakshmanan. Viral marketing meets social advertising: Ad allocation with minimum regret. Proceedings of the VLDB Endowment, 8(7), 2015.
[5] C. Aslay, F. B. L. V. Lakshmanan, and W. Lu. Revenue maximization in incentivized social advertising. Proceedings of the VLDB Endowment, 10(11), 2017.
[6] P. Banerjee et al. Maximizing welfare in social networks under a utility driven influence diffusion model, 2019.
[7] A. R. Benson, R. Kumar, and A. Tomkins. A discrete choice model for subset selection. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, pages 37–45, 2018.
[8] S. Bharathi et al. Competitive influence maximization in social networks. In IWWIE, 2007.
[9] S. Bhattacharya et al. Welfare maximization with friends-of-friends network externalities. Theory of Computing Systems, 61(4):948–986, 2017.
[10] R. W. Boadway and N. Bruce. Welfare economics. B. Blackwell New York, 1984.
[11] C. Borgs et al. Maximizing social influence in nearly optimal time. In SODA, 2014.
[12] C. Budak et al. Limiting the spread of misinformation in social networks. In WWW, 2011.
[13] R. Carbaugh. Contemporary Economics: An Applications Approach. Routledge, 8th edition, 2016.
[14] O. Celma. Last.fm Dataset – 1K users. http://www.dtic.upf.edu/~ocelma/MusicRecommendationDataset/lastfm-1K.html, 2010.
[15] P. Chalermsook et al. Social network monetization via sponsored viral marketing. In SIGMETRICS, 2015.
[16] S. Chen et al. Online topic-aware influence maximization. In VLDB, pages 666–677, 2015.
[17] W. Chen. An issue in the martingale analysis of the influence maximization algorithm imm. arXiv preprint arXiv:1808.09363, 2018.
[18] W. Chen et al. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In KDD, 2010.
[19] W. Chen et al. Scalable influence maximization in social networks under the linear threshold model. In ICDM, 2010.
[20] W. Chen et al. Information and influence propagation in social networks. Morgan & Claypool Publishers, 2013.
[21] S. Datta et al. Viral marketing for multiple products. In ICDM, 2010.
[22] U. Feige and J. Vondræk. The submodular welfare problem with demand queries. TOC, 2010.
[23] K. Garimella, A. Gionis, N. Parotsidis, and N. Tatti. Balancing information exposure in social networks. In Advances in Neural Information Processing Systems, pages 4663–4671, 2017.
[24] X. He et al. Influence blocking maximization in social networks under the competitive linear threshold model. In ICDM, 2012.
[25] K. Huang et al. Revisiting the stop-and-stare algorithms for influence maximization. VLDB, 2017.
[26] K. Jung et al. Irie: Scalable and robust influence maximization in social networks. In ICDM, 2012.
[27] M. Kapralov et al. Online submodular welfare maximization: Greedy is optimal. In SODA, 2013.
[28] D. Kempe et al. Maximizing the spread of influence through a social network. In KDD, 2003.
[29] N. Korula et al. Online submodular welfare maximization: Greedy beats 1/2 in random order. In TOC, 2015.
[30] P. Lamere. LastFM-ArtisTags2007 dataset. http://musicmachinery.com/2010/11/10/lastfm-artissttags2007/, 2008.
[31] H. Li et al. Conformity-aware influence maximization in online social networks. In VLDB, pages 117–141, 2015.
[32] Y. Li et al. Influence maximization on social graphs: A survey. TKDE, 2018.
[33] Y. Li, J. Fan, G. Ovchinnikov, and P. Karras. Maximizing multifaceted network influence. In 2019 IEEE 35th International Conference on Data Engineering (ICDE), pages 446–457. IEEE, 2019.
[34] Y. Lin and J. C. Lui. Analyzing competitive influence maximization problems with partial information: An approximation algorithmic framework. Performance Evaluation, 91:187–204, 2015.
[35] W. Lu, F. Bonchi, A. Goyal, and L. V. Lakshmanan. The bang for the buck: fair competitive viral marketing from the host perspective. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 928–936, 2013.
[36] W. Lu et al. The bang for the buck: fair competitive viral marketing from the host perspective. In KDD, 2013.
[37] W. Lu et al. From competition to complementarity: Comparative influence diffusion and maximization. In VLDB, 2016.
[38] R. B. Myerson. Optimal auction design. Mathematics of operations research, 1981.
[39] H. T. Nguyen et al. Stop-and-stare: Optimal sampling algorithms for viral marketing in billion-scale networks. In SIGMOD, 2016.
[40] N. Nisan et al. Algorithmic game theory. Cambridge university press, 2007.
[41] N. Pathak et al. A generalized linear threshold model for multiple cascades. In ICDM, 2010.
[42] T. Sun et al. Participation maximization based on social influence in online discussion forums. In ICMWS, 2011.
[43] J. Tang et al. Online processing algorithms for influence maximization. In SIGMOD, 2018.
[44] Y. Tang et al. Influence maximization in near-linear time: A martingale approach. In SIGMOD, 2015.
[45] Y. Zhu et al. Minimum cost seed set for competitive social influence. In INFOCOM, 2016.