Mitigating Filter Bubbles Under a Competitive Diffusion Model

PRITHU BANERJEE, University of British Columbia, Canada
WEI CHEN, Microsoft Research, China
LAKS V.S. LAKSHMANAN, University of British Columbia, Canada

While social networks greatly facilitate information dissemination, they are well known to have contributed to the phenomena of filter bubbles and echo chambers. This in turn can lead to societal polarization and erosion of trust in public institutions. Mitigating filter bubbles is an urgent open problem. Recently, approaches based on the influence maximization paradigm have been proposed in our community for mitigating filter bubbles by balancing exposure to opposing viewpoints. However, existing works ignore the inherent competition between the adoption of opposing viewpoints by users.

In this paper, we propose a realistic model for the filter bubble problem, which unlike previous work, captures the *competition* between opposing opinions *propagating* in a network as well as the *complementary* nature of the reward for *exposing* users to both those opinions. We formulate an optimization problem for mitigating filter bubbles under our model. We establish several evidences of the intrinsic difficulty in developing constant approximation to the problem and develop a heuristic and two instance-dependent approximation algorithms. Our experiments over 4 real datasets show that our heuristic far outperforms two state-of-the-art baselines as well as other algorithms in both efficiency and mitigating filter bubbles. We also empirically demonstrate that our best heuristic performs close to the optimal objective, which is obtained by utilizing the theoretical bounds of our approximation algorithms.

CCS Concepts: • **Information systems** → **Social networks**.

Additional Key Words and Phrases: social networks, filter bubble, influence propagation

ACM Reference Format:
Prithu Banerjee, Wei Chen, and Laks V.S. Lakshmanan. 2023. Mitigating Filter Bubbles Under a Competitive Diffusion Model. *Proc. ACM Manag. Data* 1, 2, Article 175 (June 2023), 26 pages. https://doi.org/10.1145/3589320

1 INTRODUCTION

With the proliferation of social networks, new ways have emerged to provide users with an abundance of information [16] and engage them in the sharing of information [3, 14]. Although access to information has never been easier, social media have also led to increased societal polarization [6, 22, 25]. Users are often found to be confined in filter bubbles, where, in an attempt to improve users' engagement, algorithms present to users only those types of information that align with the users' viewpoint [45, 46]. Existence of such filter bubbles impedes natural and fair opinion formation [42]. This can in turn inhibit free and open discourse among people with different viewpoints and can lead to one-sided policy decisions [47], and potentially lead to reduced social trust [43]. Attempts to address these issues have resulted in research across several dimensions.

Authors' addresses: Prithu Banerjee, University of British Columbia, Vancouver, Canada, prithu@cs.ubc.ca; Wei Chen, Microsoft Research, Beijing, China, weic@microsoft.com; Laks V.S. Lakshmanan, University of British Columbia, Vancouver, Canada, laks@cs.ubc.ca.

2836-6573/2023/6-ART175 $15.00
https://doi.org/10.1145/3589320

Earlier works focused on measuring the extent of polarization [1, 4]. Later, by identifying the groups of users to whom "counter-information" could be propagated [21, 27, 57], and by creating new links between users of opposing viewpoints, researchers have attempted to mitigate filter bubbles [5, 42, 55, 60]. To address the filter bubble problem, a recent body of research has explored the use of the influence propagation paradigm in social networks [23, 41, 48, 58]. Please see [2, 36] for detailed surveys on this topic.

Influence propagation is extensively studied in the context of the influence maximization (IM) problem. Starting from a small set of users, called seeds, in a given social network, influence cascades unfold following a stochastic diffusion model which specifies how influence propagates from one user to another in the network. IM based approaches have been found to be effective in solving a myriad of real-world problems such as gang violence [49], promotion of new prescription drugs using genuine social contagion [32], and IM for social good, i.e., public health and welfare of marginal communities [59]. This trend has motivated researchers to apply IM in the context of solving the filter bubble problem [23, 41, 58]. While the classical IM objective is to select k seeds to maximize the expected number of influenced users, some works on the filter bubble problem aimed at developing methods for balancing [23, 58] and diversifying [41] information exposure. These works assume that information spreads through the network following the same stochastic information propagation model used in the classical IM literature. As a result, they suffer from several limitations specific to the filter bubble problem.

Many studies on the filter bubble problem have reported that *items containing opinion-challenging information spread less readily than other items* [25, 56]. That is, from the propagation point of view, *items contributing to different filter bubbles are competing by nature.* We illustrate this using an example next.

Example 1. Suppose that a user has been exposed to article A arguing for relaxing gun control. Assume that the user is swayed by the article and has adopted the viewpoint that it promotes. Consider an article B arguing that gun ownership should be more restricted, citing studies revealing a strong correlation between gun ownership and violent crime. If the user is exposed to article B later, she may not readily "adopt" the viewpoint of B at the same time and may not propagate information about article B to her social peers. □

Earlier works completely ignore this competition aspect. Instead, they tacitly assume that a user, once influenced by her peers, and exposed to two opposing viewpoints, will happily adopt both of them! Thus, earlier works focus on maximizing the objective of balancing users' exposure to opposing viewpoints, assuming no competition between their adoption. We argue that in order to truly capture the goal of countering filter bubbles, it is necessary to model two seemingly conflicting requirements: (i) in terms of propagation, the items (e.g., opposing viewpoints on an issue) need to be *competing*, and (ii) the objective function measuring the effectiveness of a strategy for countering the filter bubble needs to treat the two items as *complementary* in that the reward for a user adopting both items should be significantly greater than the sum of rewards for two users adopting each of the items alone. *To our knowledge, none of the existing works is able to capture this.*

In addition to ignoring competition, the problem settings studied in prior works such as [41, 58] do not consider the task of *mitigating* the filter bubble, instead aiming to minimize the *formation* of bubbles altogether. Consequently, they require that seeds for all the items are selected and the corresponding campaigns launched synchronously, to minimize the formation of bubbles. However, for various reasons, filter bubbles may already exist in social networks. To illustrate, consider Example 1 again. Suppose article A was published before article B. A campaign based on A could well be started early by choosing its seeds, which could result in filter bubbles involving A. Later when B is published in response to article A, the network host may want to select its seeds such that the existing bubbles of A can be mitigated. Therefore, in the mitigation task, seeds of one item are fixed, while the seeds of the other item need to be selected so as to minimize the effect of filter bubbles. Our paper considers such a mitigation objective. Further, in [58] it was noted that the objective of [23] is not natural as it rewards a strategy for every user who does not adopt any item! Similarly, [41] uses an objective where the score improves even when users adopt more items that have the same or similar (political) leaning, which is contrary to the intuition behind balancing exposure.

Recent works [7, 40] on IM have used propagation models where influence is decoupled from item adoption. Users become *aware* of the items via influence, and then *adopt* a subset of the items they are aware of, following a separate adoption decision logic. We leverage the expressive power of the newly proposed models to address the filter bubble problem. In particular, we introduce a model where reward is based on a combination of complementary and competitive aspects. The first component of the reward is a deterministic complementary function that awards a reward for a user adopting two items from opposite viewpoints, which is often significantly higher than the sum of rewards for two users, each adopting any one item. However, the second component is a stochastic competition parameter which controls the probability of a user adopting the second item when she has already adopted the first item. That is, the user's adoption decision captures the inherent competition between opposing viewpoints. We formulate an optimization problem FBRewMax involving two items: given seed placement for a first item a, find k seed users for a second, opposing item b, such that the *net reward*, i.e., the sum of expected rewards of all the users at the end of the propagation, is maximized. In many real-world scenarios, such as in two-party elections, information is already limited to two topics (pro party a or b). Further, even when there are more than two candidate topics (such as in gun control: pro, anti, assault rifles, background check, etc.), we can pick any one topic (e.g., assault rifles ban) and label content as leaning for or against it. Prior works on filter bubble have similarly studied two item propagation [23, 58], but ignored the competition aspect.

While competition based adoption helps to realistically model the requirements of the filter bubble problem, it poses some unique technical challenges in terms of effectively solving FBRewMax. In the context of the IM problem, maximization of welfare, i.e., sum of expected user utilities, has been studied in [7, 8]. However, these works do not study the filter bubble problem and are restricted either to only complementary [7] or only competing items [8]. Even in the case of only competing items, welfare maximization is NP-hard to approximate within any constant factor [8]. Hence it is no surprise that in our case, for a reward which is a combination of competing and complementary functions, the optimization problem is difficult. Specifically, the objective function is neither monotone nor submodular, hence an approximation algorithm cannot be obtained by leveraging these properties. In fact, we show that it remains non-monotone and non-submodular even under several natural simplifying assumptions.

We therefore develop two instance-dependent approximation algorithms for the general problem. Our first algorithm, SpreadGRD, has an approximation bound that is tight. As a result, using this algorithm we can upper bound the optimal value of our objective which helps to empirically compare the performance of all the algorithms (which include a heuristic) with the optimal. Our second algorithm, SandwichGRD, leverages sandwich approximation [40] after bounding the non-submodular objective function with submodular functions. However, none of the above algorithms explicitly optimizes for the reward maximization objective. To that end, we design a non-trivial heuristic – RewGRD using the reverse influence sampling (RIS) approach, which is extensively used by the state-of-the-art IM algorithms [44, 51]. RewGRD owes its efficiency to the fact that it directly extends the RIS approach for the net reward maximization problem.

In sum, we make the following contribution: **(1)** We introduce a model supporting both competitive aspect of item adoption and complementary aspect of item exposure. Then we formally develop the FBRewMax problem to mitigate the filter bubble problem (§ 3). **(2)** We show that the objective of FBRewMax is neither monotone nor submodular for the general case and even under several natural simplifying assumptions. Thus, it is difficult to design a constant approximation algorithm for our problem (§ 4). **(3)** Since it is difficult to get a constant approximation, we devise two instance-dependent approximation algorithms for FBRewMax – SpreadGRD and SandwichGRD (§ 5). SpreadGRD helps calibrate an upper bound of our objective using which the effectiveness of other algorithms w.r.t. the optimal is measured empirically in § 6. **(4)** We also develop an effective heuristic called RewGRD. RewGRD makes non-trivial extensions to RIS samples such that RIS samples can be used for net reward maximization per FBRewMax (§ 5). **(5)** We conduct extensive experiments on four real-world networks. Our experiments reveal that RewGRD outperforms four state-of-the-art baselines on various experiment configurations (§ 6). We discuss related work in § 2 and summary and future work in § 7.

2 BACKGROUND & RELATED WORK

In this section, we introduce the basic concepts of the classical influence maximization (IM) problem and review related work.

Influence Maximization. A social network is represented as a directed graph $G = (V, E, p)$ with users (nodes) V and connections (edges) E, where function $p : E \to [0, 1]$ specifies the influence probabilities between users. Influence propagates following a diffusion model stating from a seed set $S \subset V$. Under a budget constraint k, the *influence maximization* (IM) problem is to find a seed set $S \subset V$ with $|S| \leq k$ such that the expected influence spread under the specified diffusion model is maximized [35].

A set function $f : 2^V \to \mathbb{R}$ is *monotone* if $f(S) \leq f(T)$ whenever $S \subseteq T \subseteq V$; it is *submodular* if for any $S \subseteq T \subseteq V$ and any $x \in V \setminus T$, $f(S \cup \{x\}) - f(S) \geq f(T \cup \{x\}) - f(T)$. Under the classical diffusion models, IM is NP-hard and also it is #P-hard to compute the spread for a given S [18, 19, 35], but the spread function is *monotone* and *submodular*. Hence, using Monte Carlo (MC) simulations for estimating the spread, a simple greedy algorithm delivers a $(1 - 1/e - \epsilon)$-approximation to the optimal solution, for any $\epsilon > 0$ [33–35]. The MC simulation, however is slow and shown to be unscalable for real world large social networks [18]. Borgs et al. [15] proposed the concept of *Reverse Influence Sampling* (RIS), that has contributed to a family of scalable state-of-the-art approximation algorithms for IM [31, 37, 44, 50, 51].

Competitive IM has been studied extensively [12, 39, 61] mostly focusing on pure competition between two or more items, from a follower perspective [39] or a host perspective [8]. In the former, given seeds of some items, the goal is to choose seeds for the “follower" item so as to either block the influence of the previous item(s) or maximize the spread of the follower item. In the host perspective, the network host chooses seeds for multiple items synchronously. Recently, complementary item propagation has been studied [7, 40]. [53] studies IM under dynamic personal affinity for different items. None of the above works study or are directly applicable to filter bubbles as they do not aim to maximize the co-exposure of opposing viewpoints.

Filter bubbles. When search and recommendation algorithms present personalized content to users for improved accuracy or relevance, users tend to get exposed to a narrow world view, as a result of which filter bubbles are formed [6, 46, 54]. They are exacerbated by echo chambers wherein users only interact with like-minded individuals and get exposed to information only from them [6, 25]. This leads to polarization in discourse [1, 4, 20, 24].

Mitigating filter bubbles and more generally polarization is important. The mitigation task raises multifaceted research questions such as which users to target, what viewpoints to promote, or how best to present opposing viewpoints to users [38]. Some works [5, 42, 55, 60] attempted to solve the problem by building connections between users from groups of opposing viewpoints. These works use the opinion-formation model and assume that the underlying graph can be manipulated by adding or removing edges.

In contrast, our work leverages the power of influence cascade using a propagation model, to solve the filter bubble problem. A recent body of work [10, 23, 41, 58], which is most similar to our work, aims to tackle the problem using an influence propagation setting; they aim for a balanced exposure of conflicting opinions on an issue, for the maximum number of users in the network. However, these works do not differentiate between exposure and adoption: in real life, a user being exposed to opposing viewpoints is not guaranteed to adopt both viewpoints and certainly not to share *both* of them on with her social peers. *One significant aspect lacking in the above body of works is the competition among the different opinions that are spreading.* These works assume that when exposed to conflicting viewpoints, a user will adopt and share all of them with their peers! There are several studies on filter bubbles that have established that items containing opinion-challenging information spread less than other items [25, 26, 56]. Our reward based propagation model adequately captures the competition effect, by separating awareness from adoption. Further, as pointed out in § 1, the objectives studied in the previous papers do not align well with mitigating the filter bubbles problem. In our work, we address these concerns by using reward parameters that provide a higher reward only for adopting items of the opposite polarity.

Fairness in IM has recently received significant attention [21, 27, 57]. These studies, where there are no competing items, are orthogonal to filter bubbles.

In summary, to our knowledge, *our study is the first to use the power of an influence propagation model where both competition and complementarity are incorporated to accurately model the unique requirements of the filter bubble mitigation problem.*

3 RIC-FB MODEL FOR FILTER BUBBLE

In this section, we first present our propagation model, called *reward driven independent cascade* model for the *filter bubble* problem, RIC-FB for short. Similar to prior works on filter bubbles [23, 58], we consider two opposing opinions. Let a and b denote the two opposite opinions (items) on a polarizing issue, being propagated through a network. We first describe the propagation model and formally state the new problem that we study under the RIC-FB model; then we discuss our design decisions in detail in § 3.3.

3.1 The RIC-FB Model

Independent Cascade (IC) is used extensively in the literature to model information adoption and propagation [17, 51, 52]. IC bears a close resemblance with propagation models used in marketing research [28, 29] where models similar to IC have been found to be useful to capture the advertisement spread. Our model, RIC-FB, builds on IC by distinguishing between awareness and adoption, where item adoption decision by users recognizes the competition between the two propagating items. Specifically, we have a social network $G = (V, E, p)$, where for each edge $(u, v) \in E$, $p_{uv} := p(u, v)$ denotes the influence probability of u over v. Each node maintains two sets of items – an *awareness set* and an *adoption set*. The awareness set of a node is the set of items it has been made aware of. Propagation or seeding populates the awareness set of a node. The node then adopts a subset of its awareness set which constitutes its adoption set. A node's awareness set could be any subset of $\{a, b\}$. Only nodes adopting an item can propagate it to its peers.

Competition. Recall that items a and b are polarizing opinions and hence users adopting any one of them may typically experience some resistance or hesitancy in adopting the second item. To capture this, we use a *competition parameter* $0 \le \kappa \le 1$ associated with adopting a second item. Specifically, a node certainly adopts the first item it is made aware of. Once it adopts a first item, which may be a or b, the probability that it adopts the second item is κ, i.e., $Pr[a|\emptyset] = Pr[b|\emptyset] = 1$ and $Pr[a|b] = Pr[b|a] = \kappa$. In § 3.3 we discuss our design decisions and consequences of relaxing them.

Reward. While there is competition between the two items a, b for adoption by a node, from the perspective of solving the filter bubble problem, as noted in § 1, a solution that makes a node adopt both items is preferable to one that makes the node adopt only one item. Our reward parameters capture this *complementary* aspect – a node that adopts both items contributes a higher reward toward the solution, than one adopting only one item. Further, the network owner, i.e., the *host*, would like to maximize its revenue, which is driven by nodes adopting (one or more) items. Hence a node adopting (any) one item contributes a positive reward, compared to a node adopting no item, which contributes zero reward. Notice that this is *orthogonal* to the two items being competitive. We stress that *this is a necessary requirement for faithfully capturing the effect of polarization and filter bubble.* Specifically, the reward for adopting no items is $\mathcal{R}(\{\}) = 0$, for adopting one item is $\mathcal{R}(\{a\}) = \mathcal{R}(\{b\}) = \delta$, and for adopting both items is $\mathcal{R}(\{a, b\}) = \delta + \Delta$, where $\Delta > \delta > 0$. We will explain our rationale for choosing these reward parameter values in discussion on our design choices in § 3.3.

Propagation. For an item $i \in \{a, b\}$, let S_i denote the corresponding seed nodes of i. In a social network, filter bubbles may be formed because of existing campaigns, whereby groups of nodes may adopt only one item. To make this concrete, we assume w.l.o.g. that the seed set of item a, S_a is fixed. Given a budget k, up to k seeds of item b, S_b, are to be selected.

The diffusion proceeds in discrete time steps. Let $S = (S_a, S_b)$ be a given seed allocation where S_b may possibly be empty. Let $\mathcal{I}^S(v, t)$ and $\mathcal{A}^S(v, t)$ denote the awareness and adoption sets of node v at time t for the given seed allocation S. At $t = 1$, the seed nodes have their awareness sets initialized according to S as: $\mathcal{I}^S(v, 1) = \{i \mid v \in S_i, i \in \{a, b\}\}, \forall v \in V$. Notice that awareness sets of non-seed nodes are initially empty.

These seed nodes then adopt the items from the awareness set following the competition parameter. If a seed node becomes aware of the two items simultaneously, then it breaks the tie arbitrarily to decide which one of the two it adopts first. 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 makes one attempt to influence its out-neighbor v, which succeeds w.p. p_{uv}. If it fails then the edge is permanently blocked. If it succeeds, then i is added to the awareness set of v at time t, if it's not already present in that set. If v has not adopted any item yet, then v adopts i, (w.p. 1); if v has already adopted an item $j \neq i$, it adopts item i w.p. κ.

If a node u has not adopted any item and is influenced by multiple in-neighbors at the same time t, then a random permutation π_u of u's in-neighbors is generated. Then u adopts the item (first) adopted by its first active in-neighbor in the permutation. There is no time delay between becoming aware and adopting an item. That is, after becoming aware, a node makes its adoption decision instantly. Adoption is progressive, i.e., once a node adopts an item, it cannot unadopt it later. The propagation converges when there is no new adoption in the network.

Stochastic reward from adoption. Owing to the competition between items a and b, the expected reward after a node becomes aware of both items is $\kappa(\delta + \Delta) + (1 - \kappa)\delta = \delta + \kappa\Delta$. This follows from the fact that $Pr[b|a] = Pr[a|b] = \kappa$.

3.2 Reward maximization to counter filter bubble

Given a social network $G = (V, E, p)$ and a seed allocation S, we consider an objective based on *net reward*, which is the sum of expected rewards of itemsets adopted by all network nodes after the propagation converges. Formally, $\mathbb{E}[\mathcal{R}(\mathcal{A}^S(u))]$ is the reward that the network (host) attains from a user u in expectation when the propagation ends, under a seed allocation S. If the propagation and adoptions are deterministic, then the reward is simply $\mathcal{R}(\mathcal{A}^S(u))$. The *expected net reward* for S, is $\rho(S) = \sum_{u \in V} \mathbb{E}[\mathcal{R}(\mathcal{A}^S(u))]$, where the expectation is over both the randomness of propagation and adoption.

In this paper, we take the *follower's* perspective on countering the filter bubble problem. Thus, we assume the seeds of one item, say a, S_a, are fixed. The campaign manager for item b approaches the network host for selecting seeds for b so as to maximize the net reward. We next formally state the problem we study.

Problem 1 (FBRewMax). *Given $G = (V, E, p)$, competition parameter κ, reward values δ and Δ, a fixed seedset S_a of item a, and a budget k for item b, find the optimal seedset for item b:*
$S_b^* = \arg\max_{S_b : |S_b| \le k} \rho(S)$, *where* $S = (S_a, S_b)$.

3.3 Design choices and Novelty

A key goal of a campaign for countering filter bubbles is to encourage more users to be exposed to both items (opposite opinions). Our choice of $\Delta > \delta$ reflects this, as the reward of adopting both items is greater than the sum of rewards of adopting each item. Choosing zero reward for adopting no items is natural; besides, it avoids the unnatural behavior of favoring strategies that can make users adopt no items, unlike previous work on filter bubbles such as [23]. Also, users adopting no items is incompatible with the fundamental objective of a network host which may want to maximize user engagement via item adoptions. Previous work has motivated this by assuming either social conscience or law enforcement may force a network owner to sacrifice revenue in order to reduce polarization. By contrast, we propose a more realistic model whereby the host can control the reward parameters δ, Δ to balance their trade-off between revenue and reducing polarization. In § 6, we report on experiments with varying values of the reward parameters δ, Δ, including the special case $\delta = 0$, which gives no reward for single item (i.e., solo) adoption.

In our problem, we consider the follower perspective, analogously to competitive IM literature [13, 39, 61]: given a fixed seedset for one item, the seedset of the second item is to be selected. However, *the goals are considerably different.* Unlike in competitive IM, the goal of our counter campaign is not to neutralize (or block) the first (i.e., "leader") campaign or even maximize adoption of the second item. This follows from our choice of the reward function, which is more aligned with exposing users to both items.

Lastly, our competition parameter captures the reluctance of users to adopt opposing viewpoints. Such reluctance could stem from prior bias as well, possibly causing $Pr[a \mid \emptyset] < 1$ (or $Pr[b \mid \emptyset] < 1$). If $Pr[a \mid \emptyset] = Pr[b \mid \emptyset] < 1$, we can create an equivalent model by adding shadow nodes and turning $Pr[a \mid \emptyset] = Pr[b \mid \emptyset]$ into edge probabilities; our algorithms and approximation results continue to hold. Under the more general case when $Pr[a \mid \emptyset] \neq Pr[b \mid \emptyset]$ or $Pr[a \mid b] \neq Pr[b \mid a]$, the approximation guarantees of our algorithms becomes worse. Proof for this general case can be found in our full report *here*.

Under an even more general case, where the parameters are local, i.e., are user-dependent, if local parameters can be upper and lower bounded by some global parameters, we can use these global bounds and the sandwich approximation to achieve an approximation guarantee. However, directly using the local parameters it is difficult to achieve an approximation guarantee. A similar difficulty is also noted in earlier works such as [7, 8]. Our algorithms, however, *do* apply even to this general case, albeit without the guarantees.

3.4 Practicality of RIC-FB

As discussed before, RIC-FB extends the classic IC model used extensively in IM literature to model influence propagation [17, 51]. IC uses edge probability p_{uv} between node pair u and v to model how influence propagates from u to v. There is a line of research on learning these probabilities from real propagation cascades [9, 30], which is orthogonal to the IM problem. We experimentally study RIC-FB using learned probabilities in §6.5.

One simplifying assumption that the classic IC model makes is that a node influenced with an item necessarily adopts the item, while influence and adoption are distinguished in RIC-FB. In practice, influence and adoption can be measured as follows. Notice that between end users, campaign advertisers, and the network host, the host is the party most interested in filter bubble mitigation. Whenever a user u clicks (or likes) a post related to an item, the host counts that as u being influenced by the item. Whenever u sends a post about item a or replies to such a post, then the host treats that as u adopting item a (and starting to try to influence her friends in the network). Similar to earlier works [8], the parameter κ can be learned using works on utility choice theory [11]. The two additional parameters δ and Δ are *not* learned but are set by the network host as needed to control the trade-off between revenue and filter bubble mitigation.

IC-like models have been found to be useful in several real world problems ranging from gang violence [49], to drug promotions [32], and to social good [59], which in turn has recently motivated researchers to apply similar models to solve the filter bubble problem [23, 41, 58]. By combining the competition and the reward aspects with the propagation, RIC-FB makes IC more realistic in the application context of the filter bubble mitigation problem.

3.5 An equivalent possible world model

For our analysis later, it will be useful to have a possible world interpretation of the RIC-FB propagation model.

An instance of FBRewMax is a pair $\langle G, M\rangle$, where $G = (V, E, p)$, and $M = (\kappa, \delta, \Delta)$ denotes the set of model parameters.

A *possible world* $w = (w_1, w_2)$, consists of an *edge possible world* (edge world) w_1, and an *adoption possible world* (adoption world) w_2. Here, w_1 is obtained by sampling a deterministic graph from the distribution associated with G, where each edge $(u, v) \in E$ is sampled with an independent probability of p_{uv}; w_2 is obtained by (i) choosing, for each node $v \in V$, a random permutation o_v over its in-neighbors and (ii) sampling a value $t_v \in [0, 1]$. If v is influenced by more than one in-neighbor at the same timestep t, it selects an in-neighbor u following the order o_v and adopts the item (first) adopted by u as the first item. Node v adopts the second item after becoming aware of it, only if $t_v \le \kappa$.

Note that propagation and adoption in a given possible world w is fully deterministic. The *net reward* of a given seed allocation S in w is $\rho_w(S) := \sum_{v\in V} \mathcal{R}(\mathcal{A}_w^S(v))$, where $\mathcal{A}_w^S(v)$ is the adoption set of v at the end of the propagation in world w. The *expected net reward* of an allocation S is $\rho(S) := \mathbb{E}_w[\rho_w(S)] = \mathbb{E}_{w_1}[\mathbb{E}_{w_2}[\rho_w(S)]] = \mathbb{E}_{w_2}[\mathbb{E}_{w_1}[\rho_w(S)]]$.

Further, given a seed set S' (for any item) and an edge world w, we use $\phi_w(S')$ to denote the set of nodes reachable from S' in w. Note that $\phi_w(S')$ does not depend on the adoption world sampled.

4 PROPERTIES OF RIC-FB

In this section, we show not only that FBRewMax is intractable but several natural attempts at designing approximation algorithms are beset with difficulties.

PROPOSITION 1. *FBRewMax in the RIC-FB model is NP-hard.*

PROOF. IM in the IC model is a special case of FBRewMax. □

Given this, we examine whether net reward satisfies monotonicity or submodularity.

Theorem 1. *Given a fixed seedset S_a for item a, the net reward is neither monotone nor submodular, with respect to seedset S_b for item b under the RIC-FB model.*

Proof. We show a counterexample for both properties. Consider the network shown in Figure 1a. The edge probabilities are all 1.
Monotonicity: Consider the fixed a seedset $S_a = \{v_1\}$. Let $S_b^1 = \{v_4, v_5\}$ and $S_b^2 = \{v_3, v_4, v_5\}$. Clearly $S_b^1 \subset S_b^2$. Let $S^i = (S_a, S_b^i)$, for $i \in \{a, b\}$. Under S^1, v_1, v_2 and v_3 adopt a w.p. 1; both v_4 and v_5 adopt b w.p. 1, and later when v_3 adopts a, v_4 and v_5 adopt a w.p. κ. Thus $\rho(S^1) = 3\delta + 2(\delta + \kappa\Delta)$. However under S^2, v_1 and v_2 adopt a w.p. 1; v_3 adopts b w.p. 1 and then a w.p. κ; v_4 and v_5 adopt b w.p. 1, and a w.p. κ^2 (because they can adopt a only after v_3 also adopts a). Thus $\rho(S^2) = 2\delta + (\delta + \kappa\Delta) + 2(\delta + \kappa^2\Delta)$; $\rho(S^1) - \rho(S^2) = \kappa\Delta(1 - 2\kappa) > 0$, for any $\kappa < \frac{1}{2}$ which violates monotonicity.
Submodularity: Again consider the fixed a seedset $S_a = \{v_1\}$. Let $S_b^1 = \{v_4, v_5\}$ and $S_b^2 = \{v_2, v_4, v_5\}$. Clearly $S_b^1 \subset S_b^2$. Let $x = v_3 \notin S_b^2$ be an additional b seed node. Note that adding x to S^2 does not change the adoption of any node. Therefore, $\rho(x \mid S^2) = 0$. Whereas, from the counterexample to monotonicity above, we have $\rho(x \mid S^1) < 0 = \rho(x \mid S^2)$, (when $\kappa < \frac{1}{2}$). This breaks submodularity. □

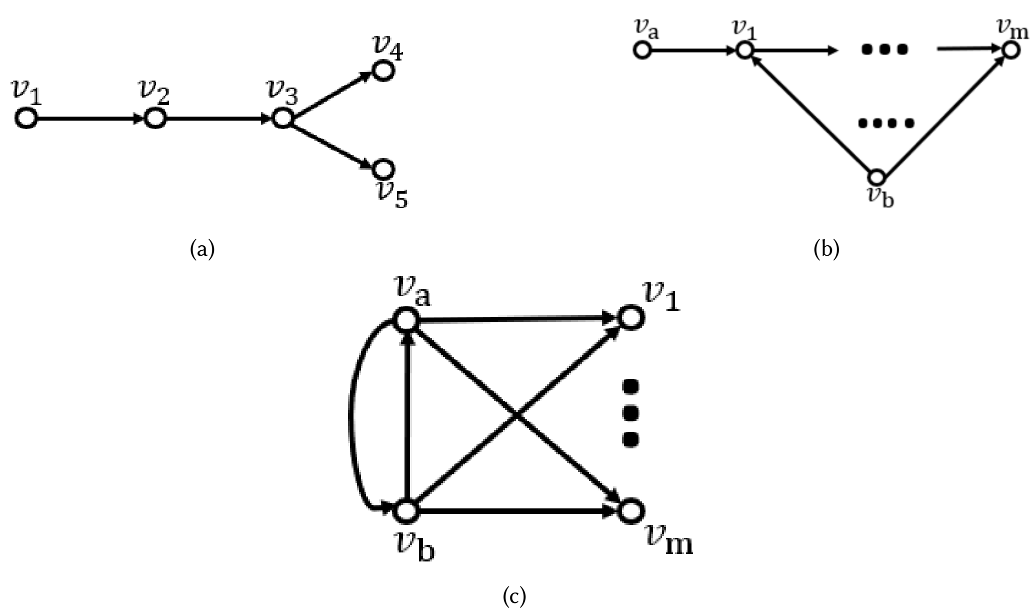


Fig. 1. Example networks showing (a) Net reward is non-monotone and non-submodular (b) RIC-FB is worse than RIC-FB-seq and (c) RIC-FB-seq is worse than RIC-FB.

The lack of these properties makes FBRewMax difficult to approximate. In our attempts to alleviate this, we explore a restricted propagation model and a surrogate objective next.

A restricted model. We examine the following question. Can we identify a restriction to the original RIC-FB model which is well behaved and admits a net reward which upper or lower bounds that under RIC-FB? If so, it would help design an approximation algorithm. One such restriction is obtained by assuming that the propagation of one item, say b, is not started until propagation of the other item, a, completes. We call this restricted model RIC-FB-seq. We next show that monotonicity and submodularity hold for RIC-FB-seq.

THEOREM 2. *Given a fixed a seedset S_a, the net reward is monotone and submodular, with respect to b seedset S_b under the propagation model RIC-FB-seq.*

PROOF. We show that monotonicity and submodularity hold for any arbitrary but fixed possible world w. Recall in w both edges and adoption are deterministic.
Monotonicity: Consider any fixed a seedset S_a. Let S_b^1 and S_b^2 be two b seedsets where $S_b^1 \subset S_b^2$. We show that for any node v, $\mathcal{A}_w^{S^1}(v) \subseteq \mathcal{A}_w^{S^2}(v)$, from which it follows that $\rho_w(S^1) \leq \rho_w(S^2)$.

If $a \in \mathcal{A}_w^{S^1}(v)$, then $v \in \phi_w(S_a)$, hence $a \in \mathcal{A}_w^{S^2}(v)$ because S_a is fixed and RIC-FB-seq lets a propagate first, so propagation of a in RIC-FB-seq is monotone. If $b \in \mathcal{A}_w^{S^1}(v)$, then $v \in \phi_w(S_b^1)$. By monotonicity of spread, we have $v \in \phi_w(S_b^2)$. Also if the same node v adopted a under allocation S^1, that implies $t_v \leq \kappa$ in w. Since w remains fixed, $b \in \mathcal{A}_w^{S^2}(v)$ must be true as well. Monotonicity of $\rho_w(.)$ follows from this.
Submodularity: Let the fixed a seedset be S_a and let S_b^1 and S_b^2 be two b seedsets where $S_b^1 \subset S_b^2$. Let $x = u \notin S_b^2$ be an additional b seed. Consider any node $v \in \phi_w(x \mid S_b^2)$. By submodularity of spread, we have $v \in \phi_w(x \mid S_b^1)$. Therefore, if v adopts only b under $\{x\} \cup S^2$, v will adopt b under $\{x\} \cup S^1$ as well. If v adopted b as a second item after adopting a under S^2, then since S_a is fixed, v would adopt a under S^1, and also since $t_v \leq \kappa$, v will adopt b as well under S^2. Submodularity of $\rho_w(.)$ follows from this. □

The net reward is monotone and submodular under RIC-FB-seq, hence a simple greedy algorithm achieves $(1 - \frac{1}{e})$-approximation under this model. While this is good news, as we show below, surprisingly, net reward under RIC-FB-seq is neither an upper nor a lower bound for that under RIC-FB. Thus, unfortunately this does not translate to an approximation guarantee for net reward under RIC-FB.

LEMMA 1. *Given a seed allocation S, $\rho(S)$ under RIC-FB can be arbitrarily worse than $\rho(S)$ under RIC-FB-seq and vice versa.*

PROOF.
RIC-FB arbitrarily worse than RIC-FB-seq: Consider the network shown in Figure 1b, where all edge probabilities are 1. Let $S_a = \{v_a\}$ and $S_b = \{v_b\}$.

Under RIC-FB-seq, a propagates first. Therefore, all $m + 1$ nodes $v_a, v_1, ..., v_m$ adopt a. Later, when propagation of b starts from v_b, $v_1, ..., v_m$ adopt it with probability κ. Therefore the net reward $2\delta + m(\delta + \kappa\Delta)$.

Under RIC-FB, a and b start propagating at the same time. They arrive only at v_1 at the same time, but for all the other nodes $v_2, ..., v_m$, b arrives first. The net reward in this case $2\delta + \sum_{i=1}^{m}(\delta + \kappa^i\Delta)$. As m increases the net reward under RIC-FB becomes arbitrarily worse than that of RIC-FB-seq.
RIC-FB-seq arbitrarily worse than RIC-FB: Consider the network in Figure 1c where all edge probabilities are 1. Again, $S_a = \{v_a\}$ and $S_b = \{v_b\}$.

Under RIC-FB-seq when a propagates first, all the nodes in the network, including v_b, adopt a. Later when propagation of b begins, at node v_b, the expected reward is $\delta + \kappa\Delta$. Every other node earns an expected reward of $\delta + \kappa^2\Delta$. Therefore the net reward is $(\delta + \kappa\Delta) + (m + 1)(\delta + \kappa^2\Delta)$.

Under RIC-FB at time $t = 1$ v_a adopts a and v_b adopts b, at $t = 1$ all nodes experience first level competition. Therefore the expected reward is $(m + 2)(\delta + \kappa\Delta)$. Thus for a large m and small κ the net reward is arbitrarily better than that under RIC-FB-seq. □

A Surrogate objective. Next, we explore whether we can bound the net reward under the RIC-FB model with a well-behaved surrogate objective. Consider an arbitrary but fixed edge possible world,

w_1. The maximum expected reward any node can achieve in w_1 is $\delta + \kappa\Delta$. Intuitively, a node v can achieve this reward provided v adopts an item (say a) first and there is a path from a seed node of the other item (say b) to v on which every node except v adopts the other item (i.e., b) first. Thus, at node v, there is a *first level competition* between a and b. The surrogate objective aims to maximize the number of such first level competition (FLC) nodes. Notice that given a seed allocation S, the FLC nodes are reached by both the items, and also every FLC node adopts the second item with a probability of at least κ. Let $\psi_{w_1}(S)$ denote the set of FLC nodes in w_1. Our surrogate objective aims to maximize the size of the set $E_w[|\psi_w(S)|]$. Since the expected reward at a FLC node is the maximum possible under the RIC-FB model, maximizing $E_w[|\psi_w(S)|]$ indeed helps maximize the net reward. In fact, it can be shown that for any allocation S, the objective $E_w[|\psi_w(S)|] \cdot (\delta + \kappa\Delta)$ lower bounds our original objective $\rho(S)$. However, as it turns out, $E_w[|\psi(S)|]$ is not monotone nor submodular.

Theorem 3. *Given a fixed a seedset S_a, the expected number of FLC nodes, $E_w[|\psi_w(S)|]$, is neither monotone nor submodular with respect to the b seedset S_b under the propagation model RIC-FB.*

Proof.
Monotonicity: Revisit the graph in Figure 1a, where all edge probabilities are 1. Thus, the only edge possible world w_1 is the entire graph. Let $S_a = \{v_1\}$, $S_b^1 = \{v_4, v_5\}$ and $S_b^2 = \{v_3, v_4, v_5\}$; $S_b^1 \subset S_b^2$. Under allocation S^1, both v_4 and v_5 are FLC nodes. Therefore $\psi_{w_1}(S^1) = \{v_4, v_5\}$. Under S^2, $\psi_{w_1}(S^2) = \{v_3\}$. Since $|\psi_{w_1}(S^2)| < |\psi_{w_1}(S^1)|$ and w_1 is the only edge possible world, it follows that monotonicity does not hold for $E_w[|\psi_w(S)|]$.
Submodularity: Let $S_a = \{v_1\}$, $S_b^1 = \{v_4, v_5\}$, $S_b^2 = \{v_2, v_4, v_5\}$ and $x = v_3 \notin S_b^2$ be the additional b seed. $\psi_{w_1}(S^2) = \psi_{w_1}(\{x\} \cup S^2) = \{v_2\}$. $\psi_{w_1}(S^1) = \{v_4, v_5\}$, but $\psi_{w_1}(\{x\} \cup S^1) = \{v_3\}$. Hence $|\psi_{w_1}(\{x\} \cup S^2)| - |\psi_{w_1}(S^2)| = 0 > |\psi_{w_1}(\{x\} \cup S^1)| - |\psi_{w_1}(S^1)| = -1$. Since w_1 is the only edge possible world, it follows that submodularity does not hold for $E_w[|\psi_w(S)|]$. □

In this section, we explored ways of bounding the original objective with well-behaved objectives in two ways: (i) using a restricted propagation model and (ii) using a simpler objective. They both fail, but for different reasons. Approach (i) fails since the resulting objective funtion, while monotone and submodular, does not bound our desired objective; approach (ii) while being a lower bound for our desired objective, is neither monotone nor submodular. This attests to the inherent difficulty of solving FBRewMax.

5 ALGORITHMS

In this section, we develop several algorithms having non-constant approximation guarantees as well as an efficient heuristic that is later shown to have good empirical performance (§ 6). Our first algorithm SpreadGRD ignores the reward objective and chooses seeds based on spread, and yet satisfies an approximation guarantee. The second algorithm SandwichGRD leverages the sandwich approximation proposed in [40]. It also produces a larger net reward than SpreadGRD. Our last algorithm RewGRD extends the state-of-the-art Reverse Influence Set based method for our reward maximization objective. It is not only efficient but it is also found to produce the highest net reward in all our experiments (§ 6).

5.1 SpreadGRD Algorithm

For a given seed set S', let $\sigma(S')$ denote the expected spread, i.e., $\sigma(S') = \mathbb{E}_w[|\phi_w(S')|]$, where the expectation is taken over the set of all possible worlds. Given graph G, fixed a seeds S_a, budget of item b as k, accuracy parameter ϵ, tolerance parameter ℓ as inputs, SpreadGRD selects S_b such that $\sigma(S_b|S_a) \geq (1 - \frac{1}{e} - \epsilon)\sigma(S_b^*|S_a)$ w.p. at least $1 - \frac{1}{|V|^\ell}$, where S_b^* is the optimal b seedset of size k,

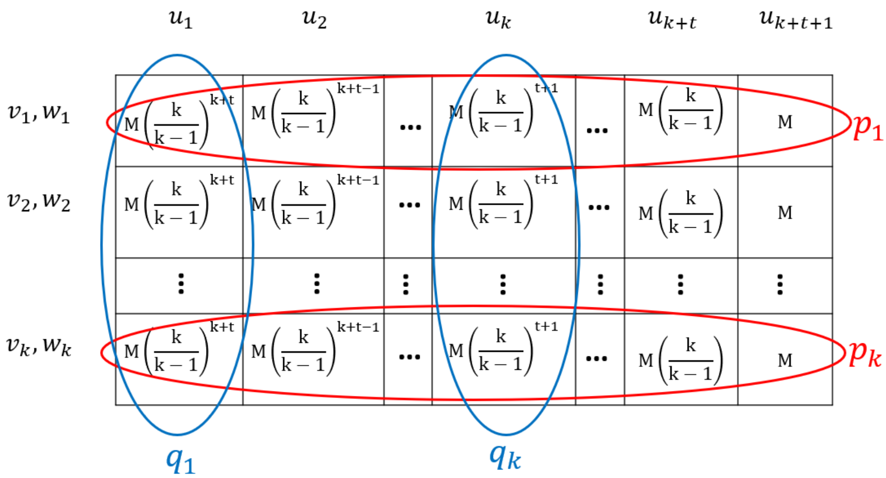


Fig. 2. Node clusters

and $\sigma(S_b|S_a) = \sigma(S_a \cup S_b) - \sigma(S_a)$. To achieve this, SpreadGRD uses a marginal sampling proposed in [8], which runs in time $O(k_a + k + \ell)(n + m) \log\ n \cdot \epsilon^{-2})$, where $|S_a| = k_a$.

Even though SpreadGRD ignores reward objective and focuses only on maximizing marginal spread, it still has a non-constant approximation guarantee. We show that the bound is tight, and this bound enables us to compute the upper bound of the optimal objective. Therefore we can compare the performance of our other algorithms and the baselines in regards to the optimal value that can be achieved.

Before proving the guarantee of SpreadGRD, we first show a bound on the expected net reward for any arbitrary seed allocation S.

Lemma 2. *Given seed allocation* $S = S_a, S_b$*, its expected net reward has the following bound,* $\delta\sigma(S_a \cup S_b) \leq \rho(S) \leq (\delta + \kappa\Delta)\sigma(S_a \cup S_b)$

Proof. Consider an arbitrary but fixed edge possible world w. The minimum expected reward for any node $v \in \phi_w(S_a \cup S_b)$ is δ, because v must adopt at least one item. Thus we have,

$$\rho_w(S) = \sum_{v \in V} \mathcal{R}(\mathcal{A}_w^S(v)) = \sum_{v \in \phi_w(S_a \cup S_b)} \mathcal{R}(\mathcal{A}_w^S(v)) \geq \delta\sigma_w(S_a \cup S_b)$$

The upper bound can be argued in the same way. Further, since it holds for every possible w, the lemma follows. □

We now show the following guarantee for SpreadGRD.

Theorem 4. *Let* S_b *be the seedset returned by SpreadGRD. Given fixed* S_a *and* $\epsilon, \ell > 0$*, we have,* $\rho(S_a, S_b) \geq \frac{\delta}{\delta + \kappa\Delta}(1 - \frac{1}{e} - \epsilon)\rho(S_a, S_b^*) + (\frac{1}{e} + \epsilon)\rho(S_a, \emptyset)$*, w.p. at least* $1 - \frac{1}{|V|^\ell}$*, where* S_b^* *is the optimal* b *seedset of size* k*.*

Proof. Using the upper bound of Lemma 2, for the optimal allocation S_b^*, we know that,

$$\rho(S_a, S_b^*) \leq (\delta + \kappa\Delta)\sigma(S_a \cup S_b^*)$$
$$\rho(S_b^* \mid S_a) + \rho(S_a, , \emptyset) \leq (\delta + \kappa\Delta)(\sigma(S_b^* \mid S_a) + \sigma(S_a))$$

where $\rho(S_b^* \mid S_a) = \rho(S_a, S_b^*) - \rho(S_a, \emptyset)$. By rearranging the above we get,

$$\sigma(S_b^* \mid S_a) \geq \frac{\rho(S_b^* \mid S_a) + \rho(S_a, \emptyset)}{\delta + \kappa\Delta} - \sigma(S_a) \quad (1)$$

Now for allocation S_b we have,

$$\begin{aligned}\rho(S_a, S_b) &= \rho(S_b \mid S_a) + \rho(S_a, \emptyset) \\ &\geq \frac{\delta}{\delta + \kappa\Delta}(1 - \frac{1}{e} - \epsilon)\rho(S_a, S_b^*) + (\frac{1}{e} + \epsilon)\rho(S_a, \emptyset)\end{aligned}$$

The entire derivation can be found in our full report *here*. The theorem follows. □

LEMMA 3. *Bound of Theorem 4 is a tight bound.*

PROOF. Consider a network of nodes shown in Figure 2. Each v_i and w_i node is directly connected to all the nodes of row i and there are p_i number of nodes in row i. Let $p = \sum_{i \in [k]} p_i$. Likewise, each u_j node is directly connected to all the nodes of column j and there are q_j number of nodes in row j (these nodes are different from row nodes). Let $q = \sum_{j \in [k]} q_j$. Clearly, $\frac{q}{p} = 1 - \frac{1}{e}$.

Now let $k_a = k$, $S_a = \{v_1, ..., v_k\}$. Therefore $\rho(S_a) = p\delta$. Also let $k_b = k$. Therefore $S_b^* = \{w_1, ..., w_k\}$, then $\rho(S_a \cup S_b^*) = p(\delta + \kappa\Delta)$.

SpreadGRD selects $\{u_1, .., u_k\}$ as the b seeds. The welfare in this case,

$$\begin{aligned}\rho(S_a, S_b) &= \delta(p + q) = p(1 - \frac{1}{e})\delta + p\delta \\ &= \frac{\delta}{\delta + \kappa\Delta}(1 - \frac{1}{e})p(\delta + \kappa\Delta) + p\delta \\ &= \frac{\delta}{\delta + \kappa\Delta}(1 - \frac{1}{e})\rho(S_a, S_b^*) + o(1)(\frac{1}{e})\rho(S_a)\end{aligned}$$

Hence the bound is tight. □

5.2 Sandwich Approximation Algorithm

The sandwich approximation (SA) was proposed in [40] to provide a data-dependent approximation guarantee for non-submodular maximization under a cardinality constraint. Given a non-submodular objective function $f : 2^V \longrightarrow \mathcal{R}_{\geq 0}$, let f_l and f_u be monotone submodular functions on 2^V such that, $f_l(I) \leq f(I) \leq f_u(I), \forall I \subseteq V$. Then run the greedy algorithm on f_l, f, and f_u and let S_l, S, and S_u be the corresponding greedy solutions. Set $S_{sand} = \arg\max_{T \in \{S_l, S, S_u\}} f(T)$, then we have:
$f(S_{sand}) \geq max\{\frac{f(S_u)}{f_u(S_u)}, \frac{f_l(S^*)}{f(S^*)}\}(1 - \frac{1}{e})f(S^*)$, where S^* is the optimal solution that maximizes f. The above approximation ratio improves for tighter f_l and f_u. Even if we have just an upper bound, SA can still be leveraged. We thus seek a tight upper bound for our objective ρ.

5.2.1 Establishing upper bound. We slightly tweak our propagation model, solely for deriving an upper bound. Recall that, according to our original model, RIC-FB, if a node u has adopted item, say a first, and if it is influenced by b later, then it does not adopt b w.p. $(1 - \kappa)$. When u does not adopt an item, in RIC-FB, it blocks the propagation by not propagating the item any further. In the tweaked model, called RIC-Tattler, a node u propagates all the items it is influenced by, even if u does not adopt all. We keep the reward function unchanged. Thus if both a and b are in the awareness set of a node u and u adopts one item, the reward at u is δ, but u propagates both the items.

We use $\rho_T(\cdot)$ to denote the objective under RIC-Tattler. We now show that $\rho_T(\cdot)$ is monotone and submodular under RIC-Tattler. Towards that, we first prove the following lemma.

Lemma 4. *Given a possible world w, a seed nodes S_a and b seed nodes S_b, every node that is reachable from a seed node, is influenced by the corresponding item.*

Proof We show it by induction on the hop count t from seed nodes. As a base case, when $t = 0$, the claim holds. Let this be true for a node u which is t hops away from a seed node. An out-neighbor of u, node v, will also be influenced via u irrespective of whether v adopts the item. Hence for the node v that is $t + 1$ hops away, the claim holds. □

Lemma 5. *[Monotonicity] Given a fixed a seedset S_a, $\rho(S)$ is monotone in S_b, i.e., for S_b^1 and S_b^2, where $S_b^1 \subseteq S_b^2$, $\rho(S^1) \leq \rho(S^2)$, where $S^i = (S_a, S_b^i)$ for $i \in \{1, 2\}$.*

Proof We show that the claim holds for any arbitrary but fixed edge possible world w. Note that if a node is influenced by one item only, then the expected reward of that node is δ, whereas if it is influenced by both the items then the expected reward is $(\kappa\Delta + \delta) > \delta$.

Let $\phi_w(S_a)$ be the set of nodes reached by a when the seed nodeset is S_a. From Lemma 4, $\phi_w(S_a)$ remains the same, when the allocation is changed to S^2. Also, $\phi_w(S_b^1) \subseteq \phi_w(S_b^2)$, from monotonicity of spread. Consequently, $\phi_w(S_a) \cap \phi_w(S_b^1) \subseteq \phi_w(S_a) \cap \phi_w(S_b^2)$.

Therefore, if a node is influenced by both the items under S^1, the node will be influenced by two items under S^2 as well. Further, as a direct consequence of Lemma 4, if a node is influenced by one item under S^1, it will be influenced by at least one item under S^2. Hence $\rho(S)$ is monotone. □

Lemma 6. *[Submodularity] Given a fixed a seedset S_a, $\rho(S)$ is submodular in S_b, i.e., for S_b^1 and S_b^2, where $S_b^1 \subseteq S_b^2$, and $x = v \notin S_b^2$, $\rho(x \mid S^2) \leq \rho(x \mid S^1)$, where $S^i = (S_a, S_b^i)$, for $i \in \{1, 2\}$.*

Proof. We show that the claim holds for any arbitrary but fixed edge possible world w. Consider a node $v \in \phi_w(x \mid S_b^2)$. Using monotonicity of reachability, $v \in \phi_w(x \mid S_b^1)$ must be true. Further since S_a is fixed, if $v \in \phi_w(x \mid S_a)$, it will be so under both S^1 and S^2. Therefore for any v which is in the reachable node set of $x \mid S_b^2$, its expected reward under $x \mid S^1$ is at least that of under $x \mid S^2$. Hence submodularity holds. □

Therefore using a greedy algorithm we get $1 - \frac{1}{e}$ approximation under RIC-Tattler.

We now show that the net reward under the RIC-Tattler model is an upper bound on the net reward produced by our model, RIC-FB.

Lemma 7. *Given seed allocation S, $\rho_T(S) \geq \rho(S)$.*

Proof. We show this holds in any arbitrary edge possible world w. In RIC-FB model, if a node v in w has expected reward greater than δ, then v definitely is reachable from S_a and S_b. Therefore under RIC-Tattler model, v will have a expected reward of $\delta + \kappa\Delta$.

Alternatively if v has expected reward δ in RIC-FB, then v is reachable from at least one node of $S_a \cup S_b$. Therefore under the RIC-Tattler, v will have a expected reward of at least δ.

Since this is true for every node v in w, the claim follows. □

Let S_T and S_G be the greedy solutions on ρ_T and ρ respectively and $S = \arg\max_{T \in \{S_T, S_G\}} \rho(T)$. Then from SA we get,

$$\rho(S) \geq \frac{\rho(S_T)}{\rho_T(S_T)}(1 - \frac{1}{e})\rho(S^*)$$

where S^* is the optimal set of b seeds for $\rho(\cdot)$.

5.3 RewGRD

Our previous algorithms, although they have approximation guarantees w.r.t. their own objectives, do not attempt to maximize our reward objective directly. Since maximizing expected reward is difficult to approximate, in this section, we propose a non-trivial heuristic, called RewGRD (Net Reward Greedy) to that effect.

RewGRD uses RR-sets, which are used in all state-of-the-art IM algorithms. However, our reward based objective poses some unique challenges as compared to traditional spread based objectives. Further, the stochastic competition parameter between items requires some additional steps in the sampling process. To address this, we modify the standard RR set construction process and develop an efficient node weighing technique that helps us estimate the marginal reward of a potential new seed node.

In what follows, we first formally describe the new RR set construction process. We then present an efficient recursive method to compute node weights of a RR set. Then we present RewGRD, which uses node weights of RR sets to greedily select seed nodes.

RR set construction In classical IM, RR sets are sampled to compute an unbiased estimate of the spread. Since propagation is deterministic in a sampled RR set, any node in the RR set, when selected as seed, certainly influences the root node. Thus, after sampling enough number of RR sets, in classical IM, nodes are greedily selected so as to influence the most number of root nodes.

In contrast, we need to estimate the *marginal reward* using RR sets. Towards that, we define a notion of weight for each node in an RR set. The node weight denotes the node's marginal contribution to the reward of the root of the RR set, if the node is selected as a seed node. These weights are non-uniform and cannot be computed efficiently if the RR set is an arbitrary graph. Hence we restrict an RR set to be a tree. In addition, we fix the randomness associated with the competition parameter while constructing RR tree. The process is described below.

Given a graph $G = (V, E, p)$ and an a seedset S_a, we randomly select a node $v \in V$ as the root of the RR tree RT_v. Set $RT_v = \{v\}$, then start a BFS such that: for $u \in RT_v$, sample each incoming edge (u', u) w.p. $p(u', u)$ and add it to RT_v if $u' \notin RT_v$. Stop when no new nodes can be added. For each $u \in RT_v$, set: (i) $t_u = 2$ w.p. κ, or $t_u = 1$ w.p. $(1 - \kappa)$; (ii) o_u as a random permutation of u's in-neighbors; (iii) w_u as the weight of u, denoting the reward contribution of node u to the root v when u is selected as a b seed. We next elaborate on how to compute the weight w_u.

Node weight assignment Since propagation is fully deterministic in a RR tree, each node can contribute one of the three possible *marginal* reward values – $0, \delta$, or Δ, which is set as w_u. The following lemmas identify the conditions that determine the w_u values.

Lemma 8. *If $RT_v \cap S_a = \emptyset$, then for any node $u \in RT_v$, $w_u = \delta$*

Proof. Since $RT_v \cap S_a = \emptyset$, root v cannot adopt a. Thus when any node u is selected as a b seed, v definitely adopts b and the marginal reward gain is δ. □

Lemma 9. *If $RT_v \cap S_a \neq \emptyset$ and $t_v = 1$, then for any node $u \in RT_v$, $w_u = 0$*

Proof. Since $t_v = 1$, v can never adopt two items. Also $RT_v \cap S_a \neq \emptyset$, v adopts a already and obtains a reward of δ which is the maximum reward v can have. Therefore in this case, for every $u \in RT_v$, $w_u = 0$. □

Lemma 10. *If $RT_v \cap S_a \neq \emptyset$ and $t_v = 2$, then for a node $u \in RT_v$, w_u can be either* 0 *or* Δ.

Proof. Without any b seed, v definitely adopts a as $RT_v \cap S_a \neq \emptyset$. Given a b seed node u, root v adopts one item if u entirely blocks propagation of a, or the existing a seed blocks b to reach v from u. In that case $w_u = 0$. Otherwise selecting u as a b seed causes v to adopt both a and b then for that u. Hence, $w_u = \Delta$. □

Checking the conditions of Lemmas 8 and 9 is straightforward. However, for Lemma 10, we need to know the existing a propagation paths. Since RT_v is a tree, there is a unique path from each $u_a \in S_a$ to any u; let $a(u)$ be the count of total number of such paths to u. Further, let r be the total number a paths to root v, i.e., $a(v) = r$. We then begin computing the weights for the case of Lemma 10 using recursion starting from the root v. See Algorithm 1.

Algorithm 1: $RewWeight(u, r, S_a)$

1 Set $w_u = \Delta$
2 **for** $u' \in v.children$ **do**
3   **if** $t_{u'} = 1$ *and* $a(u') = r$ **then**
4         $setWeight(u', 0)$
5   **if** $t_{u'} = 1$ *and* $a(u') < r$ **then**
6         $setABasedWeight(RT_{u'}, S_a)$
7   **if** $t_{u'} = 2$ **then**
8         $RewWeight(u', r, S_a)$

Each child is categorized into one of the following three subcases.
(i) Consider a child u such that $t_u = 1$ and $a(u) = r$. This means that all the a paths to the root pass via u, and u can adopt only one item. Then for all the nodes u', of the subtree rooted at u (including u), $w_{u'} = 0$ (line 4). This holds because if for any b seed selected in this subtree, u adopts b, then v cannot adopt a as there is no other path via which v can adopt a. Thus, v only adopts b in this case. Otherwise, if u does not adopt b, then v only adopts a.

Example 2. *Consider a sample RT_v shown in Figure 3. Let $S_a = \{a_1, a_2\}$ and thus $a(v) = a(u_1) = 2$, and $a(u_2) = 1$. Let $t_v = 2$, and assume the tie-breaking rule : (a) if a seed node is assigned both items, then it adopts and propagates item a first; and (b) at node u_1 its random permutation of in-neighbors orders a_2 first before u_2. Following Algorithm 1, $RewWeight(v, r, S_a)$ will set $w_v = \Delta$. Since $t_{u_1} = 1$ and $a(u_1) = a(v) = 2$, all the a-paths go through u_1 but u_1 can adopt at most one item, so assigning any node in RT_{u_1} will not make v to adopt more than one items. In this case we will call $setWeight(u_1, 0)$ (line 4) to set all nodes rooted at u_1 with weight 0, namely $w_u = 0$ for $u \in RT_{u_1} = \{u_1, u_2, a_1, a_2\}$.*

(ii) Now suppose $t_u = 1$ and $a(u) < r$ (note $a(u) > r$ is impossible). In this case there is at least one a path that reaches v but not via u. Hence any node in RT_u (i.e., the subtree rooted at u), that as a seed makes u adopt b, has weight Δ. All other nodes of the subtree RT_u have weight 0. We set these weights by doing a level-order traversal on RT_u. The pseudocode of the traversal in shown in Algorithm 3. Clearly, when there is no a seed in RT_u (therefore $a(u) = 0$), any node when selected as a b seed leads to b adoption at u, hence it has weight Δ (Line 3). If there is an a seed, then any node that is at a level below the level of the first a seed, has a weight 0, because a would reach u before b can reach from any such node (Line 15). For a node u' that is at the same level of the first a seed s, we compute the path from u' and s to the root of RT_u. If there is a node where these paths intersect, then at that node both a and b arrive together. If there is a node in that intersecting path that adopts only one item, and the first node of the intersecting point, say x, has a first in its order o_x, it will block propagation of b, resulting in a weight of 0 for u' (Line 24). If no such blocking exists then u' has weight Δ (Line 17).
(iii) Lastly, when $t_u = 2$ (note this is the case for root v as well), $w_u = \Delta$ (Line 1). In this case, u adopts both items, hence v will do so. Further, for the subtree rooted at this node we can recursively encounter one of the three subcases again as described above.

Example 3. *Consider Figure 3 again. S_a, t_v and the tie breaking rules remain the same as previous example; but now consider $t_{u_1} = 2$. We will recursively call $RewWeight(u_1, r, S_a)$ (line 8), which corresponds to the subcase (iii) above. In the recursive call, we set $w_{u_1} = \Delta$, because assigning it as one b seed will cause v to adopt both items, increasing v's utility by Δ. Next, the algorithm investigates the in-neighbors of u_1. We consider u_2 to continue our example. If $t_{u_2} = 1$, since $a(u_2) = 1 < a(v)$, we call $setABasedWeight(RT_{u_2}, S_a)$ (line 6) to set the weights of nodes in the subtree RT_{u_2} by determining whether a candidate b-seed can be blocked by an a-path. This corresponds to the subcase (ii) above. In*

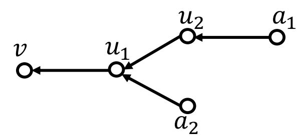


Fig. 3. Example RR tree

our example, u_2 as a b-seed cannot be blocked by any a-path in the subtree RT_{u_2}, so $w_{u_2} = \Delta$, but a_1 as one b-seed candidate will be blocked by a due to tie-breaking, so $w_{a_1} = 0$. If $t_{u_2} = 2$, we will further recursively call $RewWeight(u_2, r, S_a)$ to complete the weight assignment.

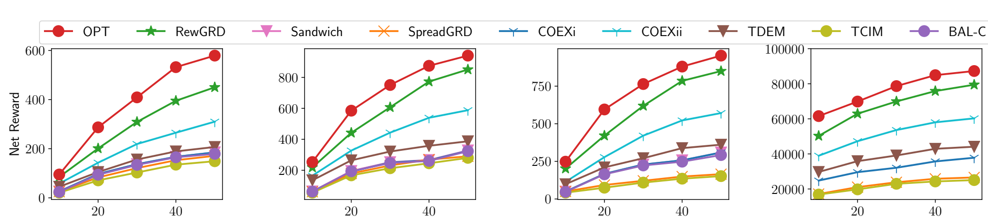
We now present the seed selection algorithm, RewGRD.

Algorithm 2: $setWeight(u, w)$

```
Set w_u = w
for u′ ∈ v.children do
    setWeight(u′, w)
```

Seed selection Pseudocode of RewGRD is shown in Algorithm 4. Similar to the classical IM algorithm, IMM [51], RewGRD first samples a set of RR trees $\mathcal{R}$, where $|\mathcal{R}| = \theta$, and θ is provided as an input parameter. Then for every node $v \in \mathcal{R}$, it computes the sum of its weight contributions in each RR tree (Line 7). After that, it greedily finds the node that has the highest weight contribution (Line 9). If the weight contribution of the found node is greater than 0, then it is selected as a b seed. Every time a new node is selected as a seed, the RR trees, where the node is present, are discarded from future consideration (Line 16). This process is repeated k times, where k is the budget.

6 EXPERIMENTS



(a) NetHept (b) Douban-Book (c) Douban-Movie (d) Orkut

Fig. 4. Expected net reward of the algorithms

6.1 Experiment Setup

Our experiments are performed on a Linux machine with Intel Xeon 2.6 GHz CPU and 128 GB RAM. Code is available *here*.

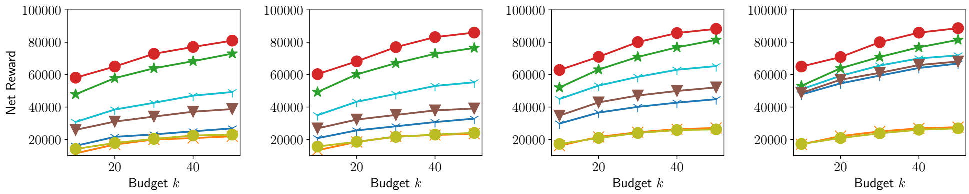
Networks We conduct our experiments of this section on four real social networks: NetHEPT, Douban-Book, Douban-Movie, and Orkut; properties of these networks are summarized in Table 1. Among these networks, NetHEPT, Douban-Book, and Douban-Movie are benchmark datasets in the IM literature [40]. Orkut is a large publicly available network that is made available *here*.

Algorithm 3: $setABasedWeight(RT_u, S_a)$

```
Set u = root(RT_u)
if a(u) = 0 then
    setweight(u, Δ)
    Return
Set Q = ∅
enqueue(u) while Q.notEmpty() do
    u′ = Q.dequeue
    S = u′.sibblings ∩ S_a
    if S = ∅ then
        w_{u′} = Δ
        for u″ ∈ u′.children do
            enqueue(u″)
    else
        for u″ ∈ u′.children do
            setWeight(u″, 0)
        else
            w_{u′} = Δ
            P_b = path(u′, u)
            for s ∈ S do
                P_a = path(s, u)
                Nodes = P_b ∩ P_a
                for node ∈ Nodes do
                    if t_{node} = 1 and o_{node} = {a, b} then
                        w_{u′} = 0
                        break
                if w_{u′} = 0 then
                    break
```

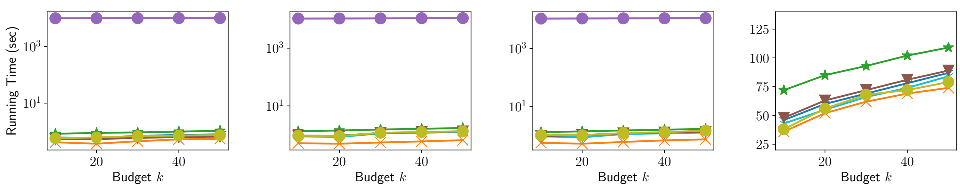
Algorithm 4: $RewGRD(G = (V, E, p), \theta, S_a, k)$

```
Sample θ number of RR trees from G in ℛ
for v ∈ V do
    Set w(v) = 0
Set S_b = ∅
for RT ∈ ℛ do
    for v ∈ RT do
        add to w_v the weight of v in RT
for i = 1 to k do
    v_max = arg max_{v∈V} w(v)
    if w(v_max) ≤ 0 then
        break
    S_b = S_b ∪ v_max
    for RT ∈ ℛ do
        if v_max ∈ RT then
            for v ∈ RT do
                subtract from w_v the weight of v in RT
Return S_b
```



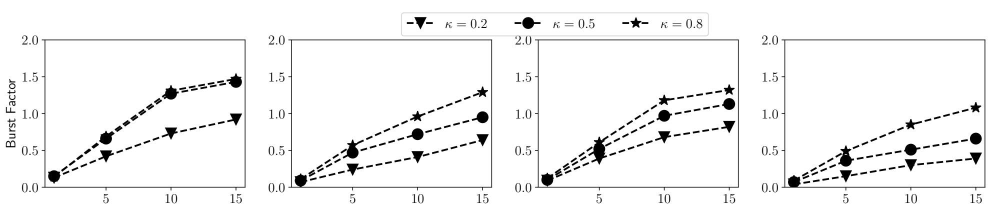
(a) $\kappa = 0.2$ (b) $\kappa = 0.4$ (c) $\kappa = 0.6$ (d) $\kappa = 0.8$

Fig. 5. Effect of κ on the algorithms using Orkut network



(a) NetHept (b) Douban-Book (c) Douban-Movie (d) Orkut

Fig. 6. Running time of the algorithms



(a) RewGRD (b) COEXii (c) RewGRD (d) COEXii

Fig. 7. BF of RewGRD and COEXii on (a-b) NetHept and (c-d) Orkut

	NetHEPT	Douban-Book	Douban-Movie	Orkut
# nodes	15.2K	23.3K	34.9K	3.07M
# edges	31.4K	141K	274K	117M
avg. deg.	4.13	6.5	7.9	77.5
type	undirected	directed	directed	undirected

Table 1. Network Statistics

Algorithms compared We compare the three algorithms we developed , namely SpreadGRD, SandwichGRD, and RewGRD against four baselines – TCIM [39], Balance-C [23], COEX [58], TDEM [41].

Our first baseline TCIM maximizes influence under pure competition, where each node can adopt at-most one item. Given a fixed seed set of the first item, TCIM selects seeds of the second item under a budget constraint, such that the number of adoptions of the second item is maximized.

Balance-C does not assume pure competition, but it also includes nodes adopting no item in its maximization objective. 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.

COEX is an improvement over Balance-C, and it considers propagation of two items where the goal is to select seeds of both items, such that at the end of the propagation, the number of nodes influenced by both items (i.e., co-exposure) is maximized. Unlike the other algorithms, COEX has an additional constraint that seed sets of the two items be disjoint. To ensure a fair comparison we use two versions - (i) $COEXi$, which does not enforce the constraint, and (ii) $COEXii$, which does.

TDEM, by contrast, relies on the "leaning scores" of nodes (users) and items, provided as input. An exposure quality function $f(\cdot)$ is used to compute a score (i.e., reward) based on the leaning scores of the items and of the node which is made aware of the items via propagation. The goal is to maximize the sum of exposure qualities over all the nodes in the network. We will establish a connection between this f function and our reward parameters next for experimental comparison. It is worth noting that no existing work on filter bubbles, including COEX and TDEM, distinguishes awareness from adoption, and consequently *does not model competition* between items. In other words, they assume $\kappa = 1$.

Default configuration Unless explicitly stated, all the parameters are set to the default values as mentioned in this section. Following previous works [31, 44] we set probability of edge $e = (u, v)$ to $1/d_{in}(v)$, where $d_{in}(v)$ is the in-degree of node v. This setting captures the intuition that as the number of connections into a node increases, the probability of that node getting influenced by any one specific in-neighbor decreases. Note that whether edge probabilities are learned from available data or are being set and how they are set is orthogonal to our optimization problem (as well as to IM in general). Nevertheless, in §6.5, we study the performance of the algorithms using Flixster where real edge probabilities are learned from actual action traces [9]. We use $\epsilon = 0.1$ and $\ell = 1$ as our default in all the algorithms that use these parameters.

The default value of the competition parameter is set as $\kappa = 0.5$. Our default choice of reward parameters is restricted by the leaning scores of TDEM so that a comparison with TDEM is possible. TDEM requires the leaning scores to be in the range $[-1, 1]$, where -1 and 1 denote the two extremes and 0 denotes neutral. Therefore, for each node $v \in V$, we set $l(v) = 0$, considering each node as a neutral node. Further, since δ needs to be the same for each of the two items, the leaning scores can only be symmetrically positioned in $[-1, 1]$, where the score 0 is already covered by the node. This restricts the choice of δ and Δ that our model permits. It is also known that the exposure quality is the highest for TDEM, when the items' leaning scores are evenly spaced out in $[-1, 1]$ ([41]). Hence, we select the leaning scores of items a and b as -0.5 and 0.5 respectively, to provide TDEM with the best possible configuration. Lastly, TDEM also assumes that each node adopts the two extreme (fictitious) items, with leaning scores -1 and 1 by default. To offset that we add the corresponding reward score $f(\{-1, 0, 1\}) = 0.5$ to every adoption. Consequently, $\delta = f(\{-1, -0.5, 0, 1\}) - f(\{-1, 0, 1\}) = f(\{-1, 0, 0.5, 1\}) - f(\{-1, 0, 1\}) = 0.125$ and $\Delta = f(\{-1, -0.5, 0, 0.5, 1\}) - (\delta + f(\{-1, 0, 1\})) = 0.1875$. Note that the difference between δ and Δ is very small, which puts our reward driven approach for filter bubble mitigation at a disadvantage. Yet to enable a comparison with TDEM, we limit our model in this way. In § 6.3, where we test the effect of different values of model parameters, we vary the values of δ and Δ more widely.

Unless specified otherwise, the budget of an item i is $k_i = 50$, where $i \in \{a, b\}$. Whenever marginal gains are required, we run 5000 MC simulations and take the average result. If an algorithm does not complete in six hours, it is omitted from comparison.

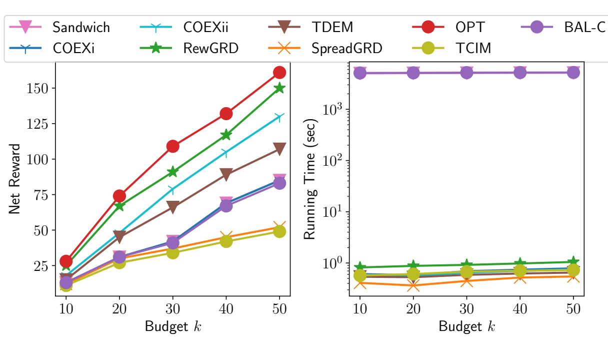
6.2 Seed quality experiment

We first compare the quality of the seeds selected by the algorithms in terms of the net reward produced. For this experiment, the seeds of item a are kept fixed. Given the default budget $k_a = 50$,

S_a is set to be the k_a seeds returned by IMM [51]. Budget of item b ranges from 10 to 50 in steps of size 10.

As noted in §5.1, we can calibrate the upper bound on the optimal net reward that can be achieved, based on the reward achieved by SpreadGRD and the approximation guarantee given in Theorem 4. For the sake of comparison we include this *upper bound of the optimal reward*, denoted as OPT, in the results that are shown in Figures 4 and 5. Since RewGRD is competition aware and explicitly maximizes the net reward, it dominates other algorithms across all the networks, producing up to 50% more reward than the closest baseline *COEXii*. In fact, RewGRD achieves net reward upto 90% of OPT, indicating that RewGRD performs quite close to the best possible solution. In contrast, SpreadGRD and TCIM ignore reward objective altogether. While TCIM aims to maximize the spread of b, SpreadGRD aims to maximize the marginal spread of b, both of which result in low number of co-adoptions. Consequently SpreadGRD and TCIM consistently attain the lowest net reward. SandwichGRD indirectly focuses on the objective but does not take competition into account while selecting the seeds. Hence it ends up selecting the same seeds for a and b to maximize co-adoption which gives the highest reward at a specific node.

None of the baselines is designed with competition factored in. They focus on maximizing co-adoptions pretending there is no competition. As a result, similar to SandwichGRD, *COEXi* selects the same seeds for a and b as that results in the most number of co-adoptions *if* there is no competition. Since *COEXii* is explicitly constrained to choose disjoint seeds for a and b, so it partially bypasses the effect of competition and hence produces the best net reward amongst the baselines.



(a) Net Reward (b) Running time

Fig. 8. Effect of using real edge probabilities on Flixster.

6.3 Impact of the model parameters

For the first subset of experiments, we hold all the parameters, except κ, at their default values. Value of κ is varied over − 0.2, 0.4, 0.6 and 0.8. The results corresponding to Orkut are shown in Figure 5(a-d). The other networks showed a similar trend.

As κ decreases, competition increases, and consequently net reward produced by any algorithm decreases. However, the drop for SpreadGRD is the minimum because SpreadGRD optimizes marginal spread, hence increasing competition has very little impact on its performance. In other words, RewGRD already factors in the effect of competition in selecting seeds. Note that, since the baselines are not designed for competition, their performance deteriorates fast for a low value of κ. In fact for a value of $\kappa = 0.2$, *COEXi* performs almost similar to SpreadGRD. RewGRD also experiences a drop in its net reward. However, unlike some of the baselines, RewGRD does not blindly co-allocate the seeds. Hence even for a low value of $\kappa = 0.2$, its performance is still much better than others.

For a high value of κ, the performance of the baselines improves drastically. As expected, the difference between the reward produced by RewGRD and the baselines reduces as κ increases. When $\kappa = 0.8$, $COEXi$ performs similar to $COEXii$ and TDEM, because the effect of competition is low for such a high value of κ.

6.4 Scalability

Using the same setup of §6.2, we next compare the running times of the algorithms on the four networks. The results are presented in Figure 6. When the network size increases or when the budget of b increases, the running times of all the algorithms increase. However, it is seen that COEX, SpreadGRD, TDEM, and RewGRD are orders of magnitude faster than SandwichGRD and Balance-C. This is because SandwichGRD and Balance-C both use costly MC simulations to select seeds. In comparison, COEX, SpreadGRD, TDEM, and RewGRD use a much faster alternative of RIS sampling [31, 37, 44, 50, 51]. In fact, SandwichGRD and Balance-C do not complete after six hours on Orkut, hence they are excluded from the last plot.

6.5 Study using real edge probabilities

Using real-world and publicly available datasets, Barbieri et al. [9] constructed the Flixster network where the edge probabilities are learned from actual action traces. The network has $6,353$ nodes and $84,606$ directed edges. We use this network to test the performance of the algorithms; the results are shown in Figure 8.

We keep the parameters at their default values, whereas the budget of item b is varied from 10 to 50. The overall trend is similar to our earlier findings: RewGRD dominates the other algorithms in terms of net reward produced, while being efficient w.r.t. running time.

6.6 Effectiveness in bursting bubbles

After establishing the effectiveness of RewGRD in terms of net reward, we next test its behavior in terms of reducing filter bubbles using adoption counts under various reward settings. For comparison purposes and to minimize clutter, we choose COEXii, the strategy that produces the highest net reward among the baselines (see §6.2), and include a comparison with COEXii. A direct way to quantify the efficacy of filter bubble mitigation is to measure the ratio of the number of co-adoptions over solo (i.e., single item) adoptions in the network. We call this ratio the *Burst Factor* (BF); a higher BF (i.e., #co-adoptions/#solo-adoptions) intuitively indicates less prominent bubbles. Values of BF closer to 0 indicate the presence of extreme filter bubbles. Figures 7 (a-b), and 7 (c-d) show the effect of varying the reward parameters (Δ and δ) on the BF achieved by RewGRD and COEXii over two networks, NetHEPT and Orkut respectively. Parameter κ is varied among three values – 0.2, 0.5 and 0.8, while other parameters are kept at their default values. The ratio $\frac{\Delta}{\delta}$ is increased from 1 to 15 in steps of 5. The y axis shows the BF for different $\frac{\Delta}{\delta}$ ratios.

When $\frac{\Delta}{\delta} = 1$, there is no incentive for co-adoption, hence the burst factor is low. As $\frac{\Delta}{\delta}$ approaches a higher value, the reward incentive for co-adoption and thus for mitigating filter bubbles increases. Hence both RewGRD and COEXii choose seeds that encourage more co-adoptions. Indeed, the BF increases dramatically, by more than 10× when Δ/δ is varied from 1 to 15, demonstrating increased effectiveness in mitigating filter bubbles. In our experiment, we find that the BF significantly increases as the $\frac{\Delta}{\delta}$ ratio goes up. E.g., on orkut, for $\kappa = 0.2$, BF jumps from 0.09 to 0.82 for *RewGRD*, and from 0.039 to 0.4 for $COEXii$. The host can thus set this ratio to control the desired level of mitigation.

Additionally, as κ increases, the effect of competition decreases, hence co-adoption increases as a whole for both algorithms. Note that since RewGRD is competition aware, its performance is significantly better than COEXii for lower values of κ. On the other hand, the marginal increase in the BF of RewGRD over COEXii decreases as we increase κ, which is understandable. Since in most real-world settings where filter bubbles form around competing opinions, κ is expected to have a low value, applications will benefit more from using RewGRD over any of the baselines.

7 SUMMARY AND FUTURE WORK

Mitigating filter bubbles is an important and urgent open problem for which one of the existing approaches is to use the influence propagation paradigm to balance exposure. While prior work ignores competition between opposing opinions for adoption, we take the first step toward realistically modeling filter bubbles formed by opposing items (opinions) which are inherently competitive, while encouraging co-adoption of the items via a reward function that treats the items as complementary from a "bubble breaking" perspective. Breaking away from the co-allocation of same seeds to both items that existing baselines generally end up making, we propose more effective algorithms and demonstrate their superiority via experiments on real data. Further research is needed for extending our framework to address filter bubbles formed by more than two competing items and for learning the model parameters from available data.

REFERENCES

[1] L. Akoglu. Quantifying political polarity based on bipartite opinion networks. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 8, 2014.

[2] F. Alatawi, L. Cheng, A. Tahir, M. Karami, B. Jiang, T. Black, and H. Liu. A survey on echo chambers on social media: Description, detection and mitigation. *arXiv preprint arXiv:2112.05084*, 2021.

[3] H. Allcott and M. Gentzkow. Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2):211–36, 2017.

[4] V. Amelkin, P. Bogdanov, and A. K. Singh. A distance measure for the analysis of polar opinion dynamics in social networks. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 13(4):1–34, 2019.

[5] V. Amelkin and A. K. Singh. Fighting opinion control in social networks via link recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 677–685, 2019.

[6] E. Bakshy, S. Messing, and L. A. Adamic. Exposure to ideologically diverse news and opinion on facebook. *Science*, 348(6239):1130–1132, 2015.

[7] P. Banerjee, W. Chen, and L. V. Lakshmanan. Maximizing welfare in social networks under a utility driven influence diffusion model. In *Proceedings of the 2019 International Conference on Management of Data*, pages 1078–1095, 2019.

[8] P. Banerjee, W. Chen, and L. V. Lakshmanan. Maximizing social welfare in a competitive diffusion model. *Proceedings of the VLDB Endowment*, 14(4):613–625, 2020.

[9] N. Barbieri, F. Bonchi, and G. Manco. Topic-aware social influence propagation models. *Knowledge and information systems*, 37(3):555–584, 2013.

[10] R. Becker, F. Corò, G. D'Angelo, and H. Gilbert. Balancing spreads of influence in a social network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 3–10, 2020.

[11] 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.

[12] S. Bhagat, A. Goyal, and L. V. Lakshmanan. Maximizing product adoption in social networks. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 603–612, 2012.

[13] S. Bharathi, D. Kempe, and M. Salek. Competitive influence maximization in social networks. In *International Workshop on Web and Internet Economics*, pages 306–311. Springer, 2007.

[14] D. Bhattacharya and S. Ram. Sharing news articles using 140 characters: A diffusion analysis on twitter. In *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 966–971. IEEE, 2012.

[15] C. Borgs, M. Brautbar, J. Chayes, and B. Lucier. Maximizing social influence in nearly optimal time. In *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*, pages 946–957. SIAM, 2014.

[16] M. Castells. The information age. *Media Studies: A Reader*, 2(7):152, 2010.

[17] W. Chen, L. V. Lakshmanan, and C. Castillo. Information and influence propagation in social networks. *Synthesis Lectures on Data Management*, 5(4):1–177, 2013.

[18] W. Chen, C. Wang, and Y. 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*, pages 1029–1038. ACM, 2010.

[19] W. Chen, Y. Yuan, and L. Zhang. Scalable influence maximization in social networks under the linear threshold model. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, pages 88–97. IEEE, 2010.

[20] X. Chen, J. Lijffijt, and T. De Bie. Quantifying and minimizing risk of conflict in social networks. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1197–1205, 2018.

[21] G. Farnad, B. Babaki, and M. Gendreau. A unifying framework for fairness-aware influence maximization. In *Companion Proceedings of the Web Conference 2020*, pages 714–722, 2020.

[22] S. Flaxman, S. Goel, and J. M. Rao. Filter bubbles, echo chambers, and online news consumption. *Public opinion quarterly*, 80(S1):298–320, 2016.

[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] K. Garimella, G. D. F. Morales, A. Gionis, and M. Mathioudakis. Quantifying controversy on social media. *ACM Transactions on Social Computing*, 1(1):1–27, 2018.

[25] R. K. Garrett. Echo chambers online?: Politically motivated selective exposure among internet news users. *Journal of computer-mediated communication*, 14(2):265–285, 2009.

[26] M. Gentzkow and J. M. Shapiro. Ideological segregation online and offline. *The Quarterly Journal of Economics*, 126(4):1799–1839, 2011.

[27] S. Gershtein, T. Milo, B. Youngmann, and G. Zeevi. Im balanced: influence maximization under balance constraints. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 1919–1922, 2018.

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

[29] J. Goldenberg, B. Libai, and E. Muller. Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata. *Academy of Marketing Science Review*, 9(3):1–18, 2001.

[30] A. Goyal, F. Bonchi, and L. V. Lakshmanan. Learning influence probabilities in social networks. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 241–250, 2010.

[31] K. Huang, S. Wang, G. Bevilacqua, X. Xiao, and L. V. S. Lakshmanan. Revisiting the stop-and-stare algorithms for influence maximization. *Proc. VLDB Endow.*, 10(9):913–924, 2017.

[32] R. Iyengar, C. Van den Bulte, and T. W. Valente. Opinion leadership and social contagion in new product diffusion. *Marketing science*, 30(2):195–212, 2011.

[33] K. Jung, W. Heo, and W. Chen. Irie: Scalable and robust influence maximization in social networks. In *Data Mining (ICDM), 2012 IEEE 12th International Conference on*, pages 918–923. IEEE, 2012.

[34] M. Kapralov, I. Post, and J. Vondrák. Online submodular welfare maximization: Greedy is optimal. In *Proceedings of the Twenty-fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '13, pages 1216–1225, 2013.

[35] D. Kempe, J. Kleinberg, and É. 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*, pages 137–146. ACM, 2003.

[36] L. V. S. Lakshmanan. On a quest for combating filter bubbles and misinformation. In Z. Ives, A. Bonifati, and A. E. Abbadi, editors, *SIGMOD '22: International Conference on Management of Data, Philadelphia, PA, USA, June 12 - 17, 2022*, page 2. ACM, 2022.

[37] H. Li, S. S. Bhowmick, A. Sun, and J. Cui. Conformity-aware influence maximization in online social networks. In *VLDB*, volume 24, pages 117–141. Springer, 2015.

[38] Q. V. Liao and W.-T. Fu. Can you hear me now? mitigating the echo chamber effect by source position indicators. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 184–196, 2014.

[39] Y. Lin and J. C. Lui. Analyzing competitive influence maximization problems with partial information: An approximation algorithmic framework. *Performance Evaluation*, 91:187–204, 2015.

[40] W. Lu, W. Chen, and L. V. Lakshmanan. From competition to complementarity: comparative influence diffusion and maximization. *Proceedings of the VLDB Endowment*, 9(2):60–71, 2015.

[41] A. Matakos, C. Aslay, E. Galbrun, and A. Gionis. Maximizing the diversity of exposure in a social network. *IEEE Transactions on Knowledge and Data Engineering*, 2020.

[42] C. Musco, C. Musco, and C. E. Tsourakakis. Minimizing polarization and disagreement in social networks. In *Proceedings of the 2018 World Wide Web Conference*, pages 369–378, 2018.

[43] C. T. Nguyen. Echo chambers and epistemic bubbles. *Episteme*, 17(2):141–161, 2020.

[44] H. T. Nguyen, M. T. Thai, and T. N. Dinh. Stop-and-stare: Optimal sampling algorithms for viral marketing in billion-scale networks. In *Proceedings of the 2016 International Conference on Management of Data*, pages 695–710. ACM, 2016.

[45] D. Nikolov, D. F. Oliveira, A. Flammini, and F. Menczer. Measuring online social bubbles. *PeerJ computer science*, 1:e38, 2015.

[46] E. Pariser. *The filter bubble: What the Internet is hiding from you.* Penguin UK, 2011.

[47] M. Perrone and B. Wieder. Pro-painkiller echo chamber shaped policy amid drug epidemic. *The Center for Public Integrity*, 2016.

[48] F. Pierri and S. Ceri. False news on social media: a data-driven survey. *ACM Sigmod Record*, 48(2):18–27, 2019.

[49] P. Shakarian, J. Salmento, W. Pulleyblank, and J. Bertetto. Reducing gang violence through network influence based targeting of social programs. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1829–1836, 2014.

[50] J. Tang, X. Tang, X. Xiao, and J. Yuan. Online processing algorithms for influence maximization. In *SIGMOD*, 2018.

[51] Y. Tang, Y. Shi, and X. Xiao. Influence maximization in near-linear time: A martingale approach. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, pages 1539–1554. ACM, 2015.

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

[53] Y.-W. Teng, Y. Shi, C.-H. Tai, D.-N. Yang, W.-C. Lee, and M.-S. Chen. Influence maximization based on dynamic personal perception in knowledge graph. In *2021 IEEE 37th International Conference on Data Engineering (ICDE)*, pages 1488–1499. IEEE, 2021.

[54] M. Tomlein, B. Pecher, J. Simko, I. Srba, R. Moro, E. Stefancova, M. Kompan, A. Hrckova, J. Podrouzek, and M. Bielikova. An audit of misinformation filter bubbles on youtube: Bubble bursting and recent behavior changes. In *Fifteenth ACM Conference on Recommender Systems*, pages 1–11, 2021.

[55] A. Tommasel, J. M. Rodriguez, and D. Godoy. I want to break free! recommending friends from outside the echo chamber. In *Fifteenth ACM Conference on Recommender Systems*, pages 23–33, 2021.
[56] P. Törnberg. Echo chambers and viral misinformation: Modeling fake news as complex contagion. *PLoS one*, 13(9):e0203958, 2018.
[57] A. Tsang, B. Wilder, E. Rice, M. Tambe, and Y. Zick. Group-fairness in influence maximization. In *IJCAI*, 2019.
[58] S. Tu, C. Aslay, and A. Gionis. Co-exposure maximization in online social networks. *Advances in Neural Information Processing Systems*, 33, 2020.
[59] A. Yadav, H. Chan, A. X. Jiang, H. Xu, E. Rice, and M. Tambe. Using social networks to aid homeless shelters: Dynamic influence maximization under uncertainty. In *AAMAS*, volume 16, pages 740–748, 2016.
[60] L. Zhu, Q. Bao, and Z. Zhang. Minimizing polarization and disagreement in social networks via link recommendation. *Advances in Neural Information Processing Systems*, 34, 2021.
[61] Y. Zhu, D. Li, and Z. Zhang. Minimum cost seed set for competitive social influence. In *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*, pages 1–9. IEEE, 2016.

Received October 2022; revised January 2023; accepted February 2023