Disincentivizing Polarization in Social Networks

Christian Borgs[1], Jennifer Chayes[1], Christian Ikeokwu[1,*] and Ellen Vitercik[2,*]

[1]UC Berkeley, Berkeley, California, USA
[2]Stanford University, Stanford, California, USA

Abstract

On social networks, algorithmic personalization drives users into filter bubbles where they rarely see content that deviates from their interests. We present a model for content curation and personalization that avoids filter bubbles, along with algorithmic guarantees and nearly matching lower bounds. In our model, the platform interacts with n users over T timesteps, choosing content for each user from k categories. The platform receives stochastic rewards as in a multi-arm bandit. To avoid filter bubbles, we draw on the intuition that if some users are shown some category of content, then all users should see at least a small amount of that content. We first analyze a naive formalization of this intuition and show it has unintended consequences: it leads to "tyranny of the majority" with the burden of diversification borne disproportionately by those with minority interests. This leads us to our model which distributes this burden more equitably. We require that the probability any user is shown a particular type of content is at least γ times the average probability all users are shown that type of content. Full personalization corresponds to $\gamma = 0$ and complete homogenization corresponds to $\gamma = 1$; hence, γ encodes a hard cap on the level of personalization. We also analyze additional formulations where the platform can exceed its cap but pays a penalty proportional to its constraint violation. We provide algorithmic guarantees for optimizing recommendations subject to these constraints. These include nearly matching upper and lower bounds for the entire range of $\gamma \in [0, 1]$ showing that the cumulative reward of a multi-agent variant of the Upper-Confidence-Bound algorithm is nearly optimal. Using real-world preference data, we empirically verify that under our model, users share the burden of diversification and experience only minor utility loss when recommended more diversified content.

Keywords

polarization, filter bubbles, bandit algorithms, social networks, recommender systems, personalization, diversification, polarization tax, exposure diversity, echo chambers

1. Introduction

Over the past decade, large internet platforms have amassed an unprecedented level of social and political power. Research has shown that the feedback loops generated by algorithmic recommendations increase polarization [1, 2, 3]. Echo chambers created by algorithmic recommendations on these platforms can have a wide range of adverse effects, such as amplifying and creating glass ceilings for minorities [4], as well as limiting exposure and job recommendations [5]. They also lead to disinformation and propaganda being disproportionately spread to minoritized groups [6].

In this paper, we propose an approach to content recommendation that simultaneously preserves the positive aspects of personalization while avoiding the pitfalls of filter bubbles. We do so by introducing a model that ensures that if some users are served a particular category of content, then all users will see at least a small amount of that content. For example, if a network includes individuals across a political spectrum, then every user will be exposed to at least a small amount of news from opposing perspectives. This allows a platform to present diverse content without forcing content on its users that no one is interested in. This approach builds upon seminal work by Celis et al. [7] who initiated the study of algorithmic approaches to reducing polarization. However, our approach to avoiding filter bubbles is different and our analysis techniques diverge significantly, as detailed in Section 1.2.

We model a platform recommending content to users with a standard multi-armed bandit formulation. There are k categories of content—such as fashion, sports, left-learning political content, right-leaning political content, and so on—and n users. For each user and content category, the platform receives a stochastic reward from an unknown distribution for showing the user content from that category, measured, for example, in terms of engagement or ad revenue. The platform interacts with the n users over T timesteps, at each timestep choosing a distribution over content categories for each user. The platform's goal is to maximize its cumulative reward. Standard bandit algorithms would eventually learn for each user the category with maximal expected reward and only show them content from that category, at which point the user's content recommendations would be caught in a filter bubble.

1.1. Our contributions

We propose a flexible approach to disincentivizing filter bubbles that adapts to the interests of the individuals on the network. We summarize our contributions along the following two axes.

1.1.1. Modeling contributions

We first analyze an approach that requires that the distribution of content shown to any one user is not far from the distribution shown to the population, so users cannot be siloed into disjoint filter bubbles. However, we show that the optimal recommendations exacerbate *tyranny of the majority*: the burden of diversification is borne by groups with minority interests (as often happens with naive approaches to diversification). A majority group will exclusively see content that they most enjoy while recommendations for minority users become far less relevant.

An equitable approach to preventing filter bubbles. The intuition behind our revised approach is that in order to avoid filter bubbles *and* tyranny of the majority, (1) users should primarily see content that they are most interested in (thus avoiding tyranny of the majority), and (2) if some users

AiOfAi '23: 3rd Workshop on Adverse Impacts and Collateral Effects of Artificial Intelligence Technologies, Macao, China

*Corresponding authors
borgs@berkeley.edu (C. Borgs); jchayes@berkeley.edu (J. Chayes); christian_ikeokwu@berkeley.edu (C. Ikeokwu); vitercik@stanford.edu (E. Vitercik)

CEUR Workshop Proceedings (CEUR-WS.org)

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

are shown a particular type of content, then all users should see at least a small amount of that content (thus avoiding filter bubbles). When both requirements are satisfied, users with majority interests will be exposed to content that interests minority groups and vice versa.

Formally, for each user, we impose the following constraint: we require that the probability she is shown content from a particular category must be at least γ times the average probability the entire population is shown content from that category, where $\gamma \in [0, 1]$ is a tunable parameter. We refer to this model as **Formulation 1.** Setting $\gamma = 0$ corresponds to complete personalization and setting $\gamma = 1$ requires that everyone see the same distribution of content. Moreover, if no one on the network is interested in some type of content, there is no requirement that users be shown that content. When $\gamma \leq \frac{1}{2}$, we show that conditions (1) and (2) are met and thus the burden of diversification is borne more equally among all users. We also provide a second formulation, called **Formulation 2**, where instead of imposing hard constraints, the platform is penalized based on the the extent to which it violates the γ constraint.

Taxation without knowledge of the true content distribution. The penalization described above depends on the true, underlying probabilities that the platform assigns to different types of content at each timestep. To augment the flexibility of our approach, we also analyze a model where an auditor only has access to a dataset describing the types of content that users were actually shown, as opposed to a description of the true distributions. In this model, the platform is penalized at the end of the T timesteps based on the extent to which the *empirical* distribution over content shown to each user violates the γ constraint described above. We refer to this model as **Formulation 3.**

1.1.2. Technical contributions

Since the platform does not know the reward distributions (corresponding to the users' preferences for the different types of content), it must learn a high-reward policy over the course of the T rounds. We analyze the *regret* of the Upper-Confidence-Bound (UCB) algorithm. The key challenges we face are providing nearly-matching lower bounds—which depend on structure exhibited by the specific constraints that we impose—and bounding the regret under Formulation 3, under which the optimal policy may be *history-dependent.*

Regret upper bounds. Under Formulation 1, we measure regret as the difference between (1) the cumulative reward of the optimal distribution over content that satisfies our γ constraint and (2) the cumulative reward of the platform's learning algorithm. Crucially, the optimal distribution (1) is defined by the users' reward distributions, but these are unknown to the learning algorithm. When $\gamma = 1$, a variant of the UCB algorithm achieves a regret of $\tilde{O}(\sqrt{nkT})$ and for $\gamma < 1$, another variant achieves a regret of $\tilde{O}(n\sqrt{kT})$. Under Formulations 2 and 3, we measure regret with respect to the optimal policy that maximizes the cumulative reward minus the penalty. Our regret bounds are $\tilde{O}(n\sqrt{kT})$.

Key challenge. Under Formulation 3, the optimal policy may be *history-dependent*: it may dynamically adjust its recommendations based on the empirical distribution over content thus far, and thus the magnitude of the final penalty. This is in contrast to Formulations 1 and 2, where the optimal policy is a fixed distribution over content.

Regret lower bounds. We provide a nearly-matching lower bound on regret under Formulation 1. As in the upper bound, our lower bound transitions from an $\Omega(n)$ dependence for small γ to an $\Omega(\sqrt{n})$ dependence for large γ. For $k = 2$ arms, we prove a lower bound of $\Omega(n\sqrt{T})$ for $\gamma < \frac{1}{2}$. Meanwhile, for all $k \geq 2$ and all $\gamma \in [0, 1]$, we prove a lower bound of $\Omega(\sqrt{nkT})$. This means that no algorithm has regret better than $\Omega(n\sqrt{T})$ for $\gamma < \frac{1}{2}$ or $\Omega(\sqrt{nkT})$ for any $\gamma \in [0, 1]$.

This transition from a $\Theta(n)$ to $\Theta(\sqrt{n})$ dependence elucidates a tension between the reward of the optimal policy and the ability of the learning algorithm to compete with the optimal policy. As γ grows, the set of distributions that the platform can show the user while still satisfying the γ constraint shrinks. Thus, the optimal policy comes from an increasingly restricted set so the regret benchmark is smaller. Likewise, as γ grows, the learner has to use an increasingly restricted set of policies to compete with the optimal policy. Since regret shrinks as γ grows, we show that the optimal policy's reward diminishes at a faster rate than the learner's handicap in competing with the optimal policy.

Key challenge. Lower bounds for bandit problems typically follow by identifying two worst-case problem instances that are similar enough that any algorithm would not be able to statistically distinguish between them, but are distinct enough to ensure that even if an algorithm has low regret on one instance, it will have high regret on the other. Simply creating n copies (one for each user) of the worst-case problem instances used in standard bandit lower bounds would lead to a large statistical difference between problem instances, thus precluding an $\Omega(n)$ dependence. Our lower bound construction therefore takes advantage of structure specific to our model.

Experiments. We analyze the optimal policies under the formulations from Section 1.1.1 using real user preference data [8]. We empirically verify that when users' preferences are heterogeneous, subgroups share the burden of diversification. We also show that users experience only a minor loss in utility when recommended diversified content.

1.2. Related work

There has been significant interest in understanding the mechanics of how recommender systems affect large-scale opinion dynamics, and if and when they lead to polarization [e.g., 9, 10, 11]. Most of the analysis has focused on how recommender systems impact network structure [12] and how this affects the spread of information and the opinions of members on the network. Recently there have been growing calls to algorithmically increase "exposure diversity" and combat filter bubbles [13, 14, 15]. Castells et al. [16] discuss methodologies and metrics to assess recommendation diversity, and Halpern et al. [17] analyze the trade-off between diversity and engagement in recommendation algorithms.

The most related research to ours is seminal work by Celis et al. [7], who initiated the study of algorithmic approaches to reducing polarization. There are a variety of differences between our work and theirs, highlighted below.

- *Modeling approach.* Celis et al. [7] suggest that a regulator should place pre-determined, fixed upper and lower bounds on the probability that each arm is played so that no user can exclusively see one type of content. Choosing bounds for each type of content, however, may be challenging. (For example, how should bounds on fashion content and major world events compare?) Moreover, if no user is interested in a type of content, it may not make sense to force all users to see it. The regulator would have to make these differential decisions, which would be a divisive and controversial task. These concerns are largely ameliorated under our model.

- *Stronger assumption on the regulator's knowledge.* Celis et al. [7] assume the regulator can control the exact probabilities that the platform shows different types of content to users. In contrast, in our Formulation 3, we propose a tax based on the content that the platform actually showed the user. As we describe in Section 1.1.2, this introduces technical challenges in providing a no-regret algorithm for the platform.
- *Lower bounds.* Our nearly-matching lower bounds help develop a complete understanding of this problem.

Since the multi-armed bandit problem was proposed [18], many variants have been studied, such as bandits with budgets [19, 20, 21], bandits with constraints [22, 23, 24, 25], and bandits with floors on content [26, 27]. Only a few variants [e.g., 28] study multi-agent settings. However, they usually still involve a common reward like in the classical multi-armed bandit problem. There has also been recent work on fairness in multi-armed bandits [e.g., 28, 29] but none of these focus on the issues of filter bubbles and polarization in social networks.

2. Notation and model

We use $\mathcal{P}^{d-1} = \{\boldsymbol{x} \in [0,1]^d : \|\boldsymbol{x}\|_1 = 1\}$ to denote the probability simplex and $[k]$ to denote the set $[k] = \{1, 2, \ldots, k\}$.

Problem definition. There are n users and k categories of content—for example, fashion, sports, right-leaning news, left-leaning news, and so on—each modeled as an *arm* of a *k-armed bandit.* An instance of our problem, denoted $\nu = \{\mathcal{D}_{i,j} : i \in [n], j \in [k]\}$, is defined by reward distributions $\mathcal{D}_{i,j}$ over $[0,1]$ with density function $f_{i,j} : [0,1] \to \mathbb{R}_{\geq 0}$. This distribution models the platform's reward for showing user i content from category j, measured in terms of engagement or ad revenue, for example. The set of all instances ν is denoted $\mathcal{E}^{n,k}$. The mean of user i's reward distribution for arm j is denoted $\mu_{i,j} \in [0,1]$, with $\boldsymbol{\mu}_i = (\mu_{i,1}, \ldots, \mu_{i,k})$. The instance ν is unknown to the platform.

Interaction between platform and users. This interaction takes place over T timesteps. At each timestep $t \in [T]$:

1. The platform selects an *action*, which is a distribution over arms for each user. This distribution corresponds to a random variable $\boldsymbol{A}_t \in [k]^n$ over arm choices for each of the n users. We use the notation $\boldsymbol{a}_t \in [k]^n$ to denote the specific set of arms the platform plays on round t, so it is a realization of the random variable $\boldsymbol{A}_t$.
2. Given the set of arms $\boldsymbol{a}_t = (a_{t,1}, \ldots, a_{t,n}) \in [k]^n$, the platform receives a reward for each user. The reward for user i is drawn from the distribution $\mathcal{D}_{i,a_{t,i}}$. We use the random variable $\boldsymbol{X}_t = (X_{t,1}, \ldots, X_{t,n}) \in [0,1]^n$ to denote the platform's reward on round t. We also use $\boldsymbol{x}_t \in [0,1]^n$ to denote a realization of this random variable.

Platform's learning algorithm. The platform uses a learning algorithm, or *policy*, π to decide the distribution over arms at each timestep. On timestep $t \in [T]$, the (randomized) policy π takes as input the history $\boldsymbol{h}_{t-1} = (\boldsymbol{a}_1, \boldsymbol{x}_1, \ldots, \boldsymbol{a}_{t-1}, \boldsymbol{x}_{t-1}) \in ([k]^n \times [0,1]^n)^{t-1}$ and returns the set of arms $\boldsymbol{a}_t \in [k]^n$ that will be played on round t. The conditional probability that $\boldsymbol{A}_t = \boldsymbol{a}_t$ given the history $\boldsymbol{A}_1 = \boldsymbol{a}_1, \boldsymbol{X}_1 = \boldsymbol{x}_1, \ldots, \boldsymbol{A}_{t-1} = \boldsymbol{a}_{t-1}, \boldsymbol{X}_{t-1} = \boldsymbol{x}_{t-1}$ is denoted $\pi(\boldsymbol{a}_t \mid \boldsymbol{a}_1, \boldsymbol{x}_1, \ldots, \boldsymbol{a}_{t-1}, \boldsymbol{x}_{t-1})$, or more compactly as $\pi(\boldsymbol{a}_t \mid \boldsymbol{h}_{t-1})$. The notation $\Pi^{n,k}$ denotes the set of all policies π.

Distribution over outcomes. Since the reward distributions are independent, the conditional distribution of the reward $\boldsymbol{X}_t \in [0,1]^n$ given $\boldsymbol{A}_t = \boldsymbol{a}_t = (a_{t,1}, \ldots, a_{t,n}) \in [k]^n$ has density function

$$f_{\boldsymbol{a}_t}(\boldsymbol{x}_t) = \prod_{i=1}^{n} f_{i,a_{t,i}}(x_{t,i}).$$

The interaction between the policy π and the instance ν induces a distribution $\mathbb{P}_{\pi\nu}$ over outcomes with density function

$$\begin{aligned} & f_{\pi\nu}(\boldsymbol{a}_1, \boldsymbol{x}_1, \ldots, \boldsymbol{a}_T, \boldsymbol{x}_T) \\ &= \prod_{t=1}^{T} \pi(\boldsymbol{a}_t \mid \boldsymbol{a}_1, \boldsymbol{x}_1, \ldots, \boldsymbol{a}_{t-1}, \boldsymbol{x}_{t-1}) f_{\boldsymbol{a}_t}(\boldsymbol{x}_t). \end{aligned} \tag{1}$$

Platform's goal. The platform's overall goal is to choose a policy π that optimizes its total reward

$$\mathop{\mathbb{E}}_{\pi\nu}\left[\sum_{i=1}^{n}\sum_{t=1}^{T} X_{i,t}\right]. \tag{2}$$

For each user $i \in [n]$, the optimal policy would choose the arm j_i that maximizes expected reward: $j_i = \operatorname{argmax}_{j \in [k]} \{\mu_{i,j}\}$. Classic bandit algorithms will eventually converge to this policy. However, repeatedly showing user i content from category j_i traps the user in a filter bubble. In the next sections, we limit the platform's ability to form filter bubbles.

3. A first attempt to disincentivize filter bubbles

We begin with a naive first attempt at disincentivizing filter bubbles and show that it has the harsh unintended consequence of exacerbating "tyranny of the majority": the burden of diversification is borne by those with minority interests. Interestingly, this issue mirrors real-world attempts at diversification where the labor associated with diversification is put disproportionately on members of the underrepresented groups.

To motivate this first attempt, we observe that in a network with severe filter bubbles, members are partitioned into groups which are exposed to disparate types of content. Thus, our first attempt at avoiding filter bubbles ensures that the content recommendations are not too "spread out." We formalize this intuition by requiring that each user's distribution over content is not too far from the average distribution over content shown to the entire population.

More formally, building on the notation from Section 2, let $\pi_i(j \mid \boldsymbol{h}_{t-1})$ denote the marginal probability that the platform shows user i arm j on timestep t given the history $\boldsymbol{h}_{t-1}$, with $\boldsymbol{\pi}_i(\boldsymbol{h}_{t-1}) = (\pi_i(1 \mid \boldsymbol{h}_{t-1}), \ldots, \pi_i(k \mid \boldsymbol{h}_{t-1}))$. Next, let $\bar{\boldsymbol{\pi}}(\boldsymbol{h}_{t-1}) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\pi}_i(\boldsymbol{h}_{t-1})$ denote the average of these marginal distributions. The j^{th} component of $\bar{\boldsymbol{\pi}}(\boldsymbol{h}_{t-1})$, denoted $\bar{\pi}(j \mid \boldsymbol{h}_{t-1})$, measures the average probability that arm j is shown to any user. Under our naive first approach, we require that the distance between the vectors $\boldsymbol{\pi}_i(\boldsymbol{h}_{t-1})$ and $\bar{\boldsymbol{\pi}}(\boldsymbol{h}_{t-1})$ is small under the ℓ_∞-norm:

$$\begin{aligned} &\left\| \boldsymbol{\pi}_i(\boldsymbol{h}_{t-1}) - \bar{\boldsymbol{\pi}}(\boldsymbol{h}_{t-1}) \right\|_\infty \\ &= \max_{j \in [k]} \left| \pi_i(j \mid \boldsymbol{h}_{t-1}) - \bar{\pi}(j \mid \boldsymbol{h}_{t-1}) \right| \leq \Delta \end{aligned} \tag{3}$$

for some $\Delta > 0$. (The ℓ_∞-norm could be replaced by any norm, but we use the ℓ_∞-norm for this exposition.)

We now show that the optimal policy $\boldsymbol{p}_1^*, \dots, \boldsymbol{p}_n^* \in \mathcal{P}^{k-1}$ leads to tyranny of the majority, where

$$\begin{aligned} &\boldsymbol{p}_1^*, \dots, \boldsymbol{p}_n^* \\ &= \underset{\boldsymbol{p}_1, \dots, \boldsymbol{p}_n}{\operatorname{argmax}} \left\{ \sum_{i=1}^{n} \boldsymbol{\mu}_i \cdot \boldsymbol{p}_i : \left\| \boldsymbol{p}_i - \frac{1}{n} \sum_{i'=1}^{n} \boldsymbol{p}_{i'} \right\|_\infty \leq \Delta, \forall i \right\}. \end{aligned}$$

To illustrate the pitfalls of this approach, we analyze a setting where there are two types of content (e.g., left- and right-leaning political content) and the users can be partitioned into disjoint sets where one set only likes content from the first category (i.e., $\boldsymbol{\mu}_i = (1, 0)$). Meanwhile, the other set only likes content from the second category (i.e., $\boldsymbol{\mu}_i = (0, 1)$). Without loss of generality, we assume that the former set—which we denote as N—is the majority.

When $\Delta \geq \frac{|N|}{n}$, the constraints are meaningless and allow for full personalization: $\boldsymbol{p}_i^* = (1, 0)$ if $i \in [N]$ and $\boldsymbol{p}_i^* = (0, 1)$ if $i \notin [N]$. Therefore, we analyze the case where $\Delta < \frac{|N|}{n}$. We show that under the optimal policy, the majority group will be able to exclusively see the content that they enjoy: $\boldsymbol{p}_i^* = (1, 0)$ if $i \in N$. Meanwhile, the minority group's recommendations take a hit in order to ensure that the constraints are satisfied. In particular, for all $i \notin N$, $\boldsymbol{p}_i^* = \left(1 - \frac{n\Delta}{|N|}, \frac{n\Delta}{|N|}\right)$. The proof of the following lemma is in the full version of the paper linked here

Lemma 3.1. *Suppose that there are $k = 2$ arms and for some set $N \subseteq [n]$ with $|N| \geq \frac{n}{2}$, $\boldsymbol{\mu}_i = (1, 0)$ for all $i \in N$ and $\boldsymbol{\mu}_i = (0, 1)$ for all $i \notin N$. If $\Delta < \frac{|N|}{n}$, then $\boldsymbol{p}_i^* = (1, 0)$ if $i \in N$ and $\boldsymbol{p}_i^* = \left(1 - \frac{n\Delta}{|N|}, \frac{n\Delta}{|N|}\right)$ otherwise.*

Lemma 3.1 illustrates that under this approach, tyranny of the majority prevails at the expense of minority interests.

4. Equitable approaches to disincentivizing filter bubbles

Motivated by Section 3, we propose three different formulations for disincentivizing filter bubbles that avoid tyranny of the majority. The intuition behind these approaches is built upon the following two pillars:

1. To avoid tyranny of the majority, users should primarily be recommended content they are most interested in,
2. But to avoid filter bubbles, that content must contain a flavor of the content shown to the entire population.

We show that it is possible to achieve both of these ends. If both conditions are satisfied, then a policy like that of Lemma 3.1 where the majority group sees no minority content is not possible. By the first requirement, groups with minority interests will be recommended content that they are interested in, which means that by the second requirement, the majority group's content recommendations will contain a small amount of that minority content, and vice versa.

4.1. Formulation 1: Personalization constraint

In our first formulation, we require that for each user $i \in [n]$, $\boldsymbol{\pi}_i(\boldsymbol{h}_{t-1})$ is at least $\gamma\bar{\boldsymbol{\pi}}(\boldsymbol{h}_{t-1})$ for some $\gamma \in [0, 1]$:

$$\boldsymbol{\pi}_i(\boldsymbol{h}_{t-1}) \geq \gamma\bar{\boldsymbol{\pi}}(\boldsymbol{h}_{t-1}). \tag{4}$$

Each user's recommendations become less personalized as γ grows.

To illustrate the benefit of this approach over that of Section 3, we analyze the same polarized example where there is a majority group N with $\boldsymbol{\mu}_i = (1, 0)$ for all $i \in N$. For the minority group, $\boldsymbol{\mu}_i = (0, 1)$ for all $i \notin N$. For all $\gamma \leq \frac{1}{2}$, we show that under the optimal policy, the majority of each group's content recommendations match their interests, but both groups see some content that appeals to the opposing group. In this case the optimal policy is defined as

$$\begin{aligned} &\boldsymbol{p}_1^*, \dots, \boldsymbol{p}_n^* \\ &= \underset{\boldsymbol{p}_1, \dots, \boldsymbol{p}_n}{\operatorname{argmax}} \left\{ \sum_{i=1}^{n} \boldsymbol{\mu}_i \cdot \boldsymbol{p}_i : \boldsymbol{p}_i \geq \frac{\gamma}{n} \sum_{i'=1}^{n} \boldsymbol{p}_{i'}, \forall i \in [n] \right\}. \end{aligned} \tag{5}$$

The proof of the following lemma is in the full version of the paper linked here

Lemma 4.1. *Suppose that there are $k = 2$ arms and for some set $N \subseteq [n]$, $\boldsymbol{\mu}_i = (1, 0)$ for all $i \in N$ and $\boldsymbol{\mu}_i = (0, 1)$ for all $i \notin N$. For $\gamma \leq \frac{1}{2}$, the optimal policy has the following form:*

$$\boldsymbol{p}_i^* = \begin{cases} \left(1 - \frac{\gamma(n - |N|)}{n}, \frac{\gamma(n - |N|)}{n}\right) & \textit{if } i \in N \\ \left(\frac{\gamma|N|}{n}, 1 - \frac{\gamma|N|}{n}\right) & \textit{if } i \notin N. \end{cases}$$

Since $\gamma \leq \frac{1}{2}$, this policy ensures that users are mostly recommended content that they are interested in: $\boldsymbol{\mu}_i \cdot \boldsymbol{p}_i^* \geq 1 - \gamma \geq \frac{1}{2}$ for all $i \in [n]$. However, they are still shown a small fraction of content that the other set of the population is interested in. We note that when N is the majority group $\left(|N| \geq \frac{n}{2}\right)$, the minority group $[n] \setminus N$ still sees more content that they are not interested in than the majority group because $\frac{\gamma|N|}{n} \geq \frac{\gamma(n - |N|)}{n}$. However, the burden of diversification is split far more equally among the two groups than in Lemma 3.1. The policy mirrors a typical mode of community forum discussions where members split time between listening to the opinions of each person in the entire group (for a γ-fraction of the time) and breaking into focus groups about specific topics (for a $(1 - \gamma)$-fraction of the time).

In Section 5, we provide upper and lower bounds on the platform's *regret* with respect to the optimal policies $\boldsymbol{p}_1^*, \dots, \boldsymbol{p}_n^*$ defined in Equation (5). Regret measures the difference between the total reward of the optimal policy and that of the platform's policy π. In other words, for any instance ν and policy π, the expected regret is defined as

$$R_{T,1}(\pi, \nu) = T \sum_{i=1}^{n} \boldsymbol{p}_i^* \cdot \boldsymbol{\mu}_i - \mathop{\mathbb{E}}_{\pi\nu}\left[\sum_{i=1}^{n}\sum_{t=1}^{T} X_{i,t}\right]. \tag{6}$$

4.2. Formulation 2: Personalization penalty

We analyze a second formulation where there are no constraints on the platform's policy, but the platform is penalized based on the extent to which Equation (4) is violated. Given a parameter $\eta \geq 0$, this penalty is defined as

$$\eta \sum_{i=1}^{n} \sum_{j=1}^{k} \max\left\{\gamma \bar{\boldsymbol{\pi}}(j \mid \boldsymbol{h}_{t-1}) - \pi_i(j \mid \boldsymbol{h}_{t-1}), 0\right\}.$$

In other words, the platform's goal is to maximize its cumulative reward

$$\begin{aligned}
&\mathrm{rew}_2(\pi, \nu; \eta, \gamma) \\
&= \mathop{\mathbb{E}}_{\pi\nu}\left[\sum_{i=1}^{n}\left(\sum_{t=1}^{T} X_{i,t}\right.\right. \\
&\left.\left. - \eta \sum_{j=1}^{k} \max\left\{\gamma \bar{\boldsymbol{\pi}}(j \mid \boldsymbol{h}_{t-1}) - \pi_i(j \mid \boldsymbol{h}_{t-1}), 0\right\}\right)\right] \\
&= \sum_{t=1}^{T} \mathop{\mathbb{E}}_{\pi\nu}\left[\sum_{i=1}^{n}\left(\boldsymbol{\mu}_i \cdot \boldsymbol{\pi}_i(\boldsymbol{h}_{t-1})\right.\right. \\
&\left.\left. - \eta \sum_{j=1}^{k} \max\left\{\frac{\gamma}{n}\sum_{i'=1}^{n} \pi_{i'}(j \mid \boldsymbol{h}_{t-1}) - \pi_i(j \mid \boldsymbol{h}_{t-1}), 0\right\}\right)\right]
\end{aligned} \tag{7}$$

The policy that maximizes Equation (7) is history independent and can be written as $\boldsymbol{p}^* = (\boldsymbol{p}_1^*, \dots, \boldsymbol{p}_n^*)$ with $\boldsymbol{p}_i^* \in \mathcal{P}^{k-1}$. The expected regret of a policy π under this formulation is $R_{T,2}(\pi, \nu) = \mathrm{rew}_2(\boldsymbol{p}^*, \nu; \eta, \gamma) - \mathrm{rew}_2(\pi, \nu; \eta, \gamma)$.

4.3. Formulation 3: Personalization penalty on the empirical distribution

Sections 4.1 and 4.2 describe models in which the platform is subject to constraints or penalties based on the *true* distribution over content that it shows users. However, an auditor may only have access to the *realizations* of those distributions—that is, the set of arms $a_{t,i} \in [k]$ shown to each user i at timestep t. Formulation 3 covers a setting in which a regulator penalizes the platform at the end of the T timesteps based on the empirical distribution over content. Specifically, let $\hat{p}_{i,j} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}_{\{A_{t,i}=j\}}$ be the average number of times that the platform pulls arm j for user i. At the end of the T timesteps, the platform is penalized based on how small $\hat{p}_{i,j}$ is compared to $\frac{\gamma}{n}\sum_{i'=1}^{n} \hat{p}_{i',j}$. In particular, given a normalizing factor η, we define a penalty that is the analogue of Equation (7):

$$\eta \sum_{i=1}^{n} \sum_{j=1}^{k} \max\left\{\frac{\gamma}{n}\sum_{i'=1}^{n} \hat{p}_{i',j} - \hat{p}_{i,j}, 0\right\}.$$

The platform's goal is therefore to maximize their expected total payoff minus this penalty, which is equal to

$$\begin{aligned}
\mathrm{rew}_3(\pi, \nu; \eta, \gamma) &= \mathop{\mathbb{E}}_{\pi\nu}\left[\sum_{i=1}^{n}\left(\sum_{t=1}^{T} X_{i,t}\right.\right. \\
&\left.\left. - \eta \sum_{j=1}^{k} \max\left\{\frac{\gamma}{n}\sum_{i'=1}^{n} \hat{p}_{i',j} - \hat{p}_{i,j}, 0\right\}\right)\right] \\
&= \sum_{i=1}^{n}\left(\sum_{t=1}^{T} \mathop{\mathbb{E}}_{\pi\nu}\left[\boldsymbol{\mu}_i \cdot \boldsymbol{\pi}_i(\boldsymbol{h}_{t-1})\right]\right. \\
&\left. - \eta \sum_{j=1}^{k} \mathop{\mathbb{E}}_{\pi\nu}\left[\max\left\{\frac{\gamma}{n}\sum_{i'=1}^{n} \hat{p}_{i',j} - \hat{p}_{i,j}, 0\right\}\right]\right).
\end{aligned} \tag{8}$$

Let π^* be the policy that maximizes Equation (8). The regret of π is $R_{T,3}(\pi, \nu) = \mathrm{rew}_3(\pi^*, \nu; \eta, \gamma) - \mathrm{rew}_3(\pi, \nu; \eta, \gamma)$.

A key difference between Equation (7) and Equation (8) is that in Equation (7), the platform is penalized at every timestep whereas in Equation (8), the platform is penalized at the end of the T timesteps. We make this distinction because the empirical distribution over content at a single timestep would be extremely noisy.

5. Regret analysis

In this section, we discuss algorithms that the platform can use to minimize regret in the three formulations from Section 4. We also provide a nearly-matching lower bound on regret for Formulation 1 in Section 5.1.3.

5.1. Regret analysis for Formulation 1

We begin with lower bounds on regret under Formulation 1. In Section 5.1.1, we show that a variant of the UCB algorithm has regret $O(n\sqrt{Tk})$ for $\gamma < 1$ and in Section 5.1.2, we show that a different variant of UCB has regret $O(\sqrt{nkT})$ for $\gamma = 1$. We then prove in Section 5.1.3 that these bounds are nearly optimal: for $\gamma \leq \frac{1}{2}$ and $k = 2$, no algorithm can achieve regret better than $\Omega(n\sqrt{T})$, and for all $k \geq 2$ and $\gamma \in [0, 1]$ (including $\gamma > \frac{1}{2}$) our bound is $\Omega(\sqrt{nkT})$.

The transition from a $\Theta(n)$ to a $\Theta(\sqrt{n})$ dependence illustrates that as γ grows, the platform is better able to compete with the optimal policy subject to the γ constraints. As γ grows, the platform has a smaller set of distributions that it can use to compete with the optimal policy while obeying the γ constraints. However, for the same reason, the cumulative reward of the optimal policy shrinks as γ grows. Intuitively, the transition from a $\Theta(n)$ to a $\Theta(\sqrt{n})$ dependence as γ grows illustrates that the optimal policy's reward degrades faster than the platform's ability to compete with that policy.

5.1.1. Regret upper bound when $\gamma < 1$

We analyze a multi-agent variant of the UCB algorithm, which we call n-UCB, and show that it has regret $O(n\sqrt{Tk})$ when $\gamma < 1$. The n-UCB algorithm essentially runs a copy of classic UCB for each user, but coordinates amongst these n UCB copies to ensure that they satisfy the global constraints. This requires n-UCB to play distributions over arms from the set of distributions $(\boldsymbol{p}_1, \dots, \boldsymbol{p}_n)$ that satisfy the constraints: $\boldsymbol{p}_i \geq \frac{\gamma}{n}\sum_{i'=1}^{n} \boldsymbol{p}_{i'}$ for all $i \in [n]$. This is in contrast to the classic case where UCB plays a single arm at each timestep. For completeness, we include a full description of n-UCB and the proof of the following theorem in the full version of the paper linked here

Theorem 5.1. *Let π be the policy of n-UCB. Then $R_{T,1}(\pi, \nu) = \tilde{O}(n\sqrt{kT})$.*

5.1.2. Regret upper bound when $\gamma = 1$

When $\gamma = 1$, all users must be shown the same distribution of content. We can therefore reduce our problem to a single-agent bandit problem with the reward distributions $\mathcal{D}_j = \sum_{i=1}^{n} \mathcal{D}_{i,j}$ for all arms $j \in [k]$. We adapt the robust-UCB framework by Bubeck et al. [30] with the median-of-means estimator [31], as summarized by Algorithm 1. The full proof of the following theorem is in the full version of the paper linked here

Algorithm 1 Robust-UCB (defined by parameter δ)

```
Require: Failure probability $\delta \in (0,1)$, median-of-means estimator $\bar{\boldsymbol{\mu}}(t,\delta)$
Set $N_j(0) = 0$, $\hat{\mu}_j^{(0)} = 0 \ \forall j \in [k]$
for $t \in \{1, \ldots, T\}$ do
    if $t \in \{1, \ldots, k\}$ then
        Set $\boldsymbol{p}^{(t)} = \boldsymbol{e}_t$
    else
        Set $\boldsymbol{p}^{(t)} = \operatorname{argmax}_{\boldsymbol{p} \in \mathcal{P}^{k-1}} \boldsymbol{p} \cdot \hat{\boldsymbol{\mu}}^{(t-1)}$
    end if
    Draw an arm $j^{(t)} \sim \boldsymbol{p}^{(t)}$
    Receive reward $r^{(t)} \sim \mathcal{D}_{j^{(t)}}$
    Set $N_{j^{(t)}}(t) = N_{j^{(t)}}(t-1) + 1$    ▷ Increment the counter for arm $j^{(t)}$
    Set $N_j(t) = N_j(t-1), \ \forall j \neq j^{(t)}$    ▷ Do not increment the other counters
    Set $\beta_j^{(t)} = \sqrt{\frac{24n}{N_j(t)} \log \frac{Tk}{\delta}}, \ \forall j \in [k]$    ▷ Define confidence intervals
    $\hat{\mu}_j^{(t)} = \bar{\boldsymbol{\mu}}_j(N_j(t), \delta) + \beta_j^{(t)}, \ \forall j \in [k]$    ▷ Get mean rewards estimates
end for
```

Theorem 5.2. *Let π be the policy of Robust-UCB. Then $R_{T,1}(\pi, \nu) = \tilde{O}(\sqrt{nkT})$.*

5.1.3. Regret lower bound

In this section, we provide nearly-matching regret lower bounds. Our first bound holds when there are $k = 2$ arms, $\gamma \leq \frac{1}{2}$, and n is sufficiently large ($n > 100$). In this case, we prove a regret lower bound of $\Omega(n\sqrt{T})$. Meanwhile, for all $k \geq 2$ and $\gamma \in [0,1]$ (including $\gamma > \frac{1}{2}$), we provide a bound of $\Omega(\sqrt{nkT})$. We begin with our main result (Theorem 5.3) and show in Corollary 5.4 that it implies a regret bound of $\Omega(n\sqrt{T})$ for $\gamma \leq \frac{1}{2}$ and $n > 100$.

Theorem 5.3. *For all $T \geq 4$, the regret is lower bounded as follows:*

$$\inf_{\pi \in \Pi^{n,2}} \sup_{\nu \in \mathcal{E}^{n,2}} R_{T,1}(\pi, \nu) \geq \max\left\{ \sqrt{\frac{T}{8}} \left(\frac{n}{8e} - \gamma \left(\frac{n}{8e} + \sqrt{\frac{n}{2\pi}} \right) \right), \frac{\sqrt{nT}}{16e} \right\}.$$

Proof. The proof of this theorem can be found in the full version of the paper linked here □

Corollary 5.4. *For all $n > 100$, $\gamma \leq \frac{1}{2}$, and $T \geq 4$, the regret is lower bounded as*

$$\inf_{\pi \in \Pi^{n,2}} \sup_{\nu \in \mathcal{E}^{n,2}} R_{T,1}(\pi, \nu) \geq \frac{n\sqrt{T}}{900}.$$

5.2. Regret analysis for Formulation 2

Under Formulation 2, a variation on UCB we call Penalty-UCB (Algorithm 2) achieves regret $\tilde{O}(n\sqrt{kT})$. Penalty-UCB maintains estimates $\hat{\boldsymbol{\mu}}_i^{(t)}$ of each $\boldsymbol{\mu}_i$ and selects the distribution maximizing the estimated reward minus the penalty:

$$\left(\boldsymbol{p}_i^{(t)}\right)_{i \in [n]} = \operatorname{argmax} \left\{ \sum_{i=1}^{n} \boldsymbol{p}_i \cdot \hat{\boldsymbol{\mu}}_i^{(t)} - \eta \sum_{j=1}^{k} \max\left\{ \frac{\gamma}{n} \sum_{i'=1}^{n} p_{i',j} - p_{i,j}, 0 \right\} \right\}.$$

For completeness, we include the proof of the following theorem in the full version of the paper linked here

Algorithm 2 Penalty-UCB (defined by parameter δ)

```
Require: Failure probability $\delta > 0$
Set $N_{i,j}(0) = 0, \ \forall i \in [n], j \in [k]$; $\hat{\boldsymbol{\mu}}_i^{(0)} = \mathbf{0}, \ \forall i \in [n]$
for $t \in \{1, \ldots, T\}$ do
    if $t \in \{1, \ldots, k\}$ then
        Set $\boldsymbol{p}_i^{(t)} = \boldsymbol{e}_t$
    else
        Set $\left(\boldsymbol{p}_i^{(t)}\right)_{i \in [n]} = \operatorname{argmax} \left\{ \sum_{i=1}^{n} \boldsymbol{p}_i \cdot \hat{\boldsymbol{\mu}}_i^{(t-1)} - \eta \sum_{j=1}^{k} \max\left\{ \frac{\gamma}{n} \sum_{i'=1}^{n} p_{i',j} - p_{i,j}, 0 \right\} \right\}$
    end if
    Draw $j_i^{(t)} \sim \boldsymbol{p}_i^{(t)} \ \forall i \in [n]$
    Receive reward $r_i^{(t)} \sim X_{i, j_i^{(t)}}$
    $N_{i,j_i^t}(t) = N_{i,j_i^t}(t-1) + 1, \ \forall i \in [n]$
    $N_{i,j}(t) = N_{i,j}(t-1), \ \forall i \in [n]$ and $j \neq j_i^t$
    $\beta_{i,j}^{(t)} = \sqrt{\frac{\log(2Tnk/\delta)}{N_{i,j}(t)}}, \ \forall i \in [n], j \in [k]$
    $\hat{\boldsymbol{\mu}}_{i,j}^t = \frac{1}{N_{i,j}(t)} \sum_{\tau=1}^{t} r_i^{(\tau)} \mathbb{1}\left\{ j_i^{(\tau)} = j \right\} + \beta_{i,j}^{(t)}, \ \forall i \in [n], j \in [k]$
end for
```

Theorem 5.5. *Let π be the policy of Penalty-UCB. Then $R_{T,2}(\pi, \nu) = \tilde{O}(n\sqrt{kT})$.*

5.3. Regret analysis for Formulation 3

A key challenge under Formulation 3 is that the platform's optimal strategy, given perfect information about the reward distributions $\mathcal{D}_{i,j}$, may be *history dependent.* For example, the platform may choose to increase or decrease personalization dynamically based on the empirical distribution of content chosen thus far. Nonetheless, we show that Penalty-UCB (Algorithm 2) competes with the optimal history-dependent policy by reducing our analysis to that of Section 5.2. We use the notation π^* to denote the optimal policy that maximizes Equation (8).

First, we show that under Formulation 2, the optimal policy obtains a larger reward (measured in terms of rew$_2$) than π^* under Formulation 3 (measured in terms of rew$_3$).

Lemma 5.6. *Let $\boldsymbol{p}^* = (\boldsymbol{p}_1^*, \ldots, \boldsymbol{p}_n^*)$ with $\boldsymbol{p}_i^* \in \mathcal{P}^{k-1}$ be the policy that maximizes rew$_2\left(\boldsymbol{p}, \nu; \frac{\eta}{T}, \gamma\right)$. Then*

$$\mathit{rew}_2\left(\boldsymbol{p}^*, \nu; \frac{\eta}{T}, \gamma\right) \geq \mathit{rew}_3(\pi^*, \nu; \eta, \gamma).$$

Proof. The proof of this lemma can be found in the full version of the paper linked here □

Next, we show that for any policy π that deterministically plays each of the k arms once in the first k rounds, the difference between its rewards under Formulations 2 and 3 is bounded. This condition holds for Penalty-UCB (Algorithm 2) and could be removed with a slightly more involved analysis.

Lemma 5.7. *Let π be any policy such that $\pi_i(t \mid \boldsymbol{h}_{t-1}) = 1$ for all $t \leq k$ and $i \in [n]$. For any instance ν,*

$$\begin{aligned} &\mathit{rew}_2\left(\pi, \nu; \frac{\eta}{T}, \gamma\right) \\ &\leq \mathit{rew}_3(\pi, \nu; \eta, \gamma) + \eta n k(\gamma + 1)\sqrt{\frac{10 \log T}{T}}. \end{aligned}$$

Proof. The proof of this lemma can be found in the full version of the paper linked here □

Our regret bound follows from Lemmas 5.6 and 5.7 as well as Theorem 5.5. The proof is in the full version of the paper

Theorem 5.8. *Let π be the policy played by Algorithm 2. Then the regret is bounded as*

$$\begin{aligned} &\mathit{rew}_3(\pi^*, \nu; \eta, \gamma) - \mathit{rew}_3(\pi, \nu; \eta, \gamma) \\ &= \tilde{O}\left(n\sqrt{kT} + \frac{\eta n k(1+\gamma)}{\sqrt{T}}\right). \end{aligned}$$

Even if η grows linearly in T, the regret bound in Theorem 5.8 will only grow with $\sqrt{T}$.

6. Empirical Results

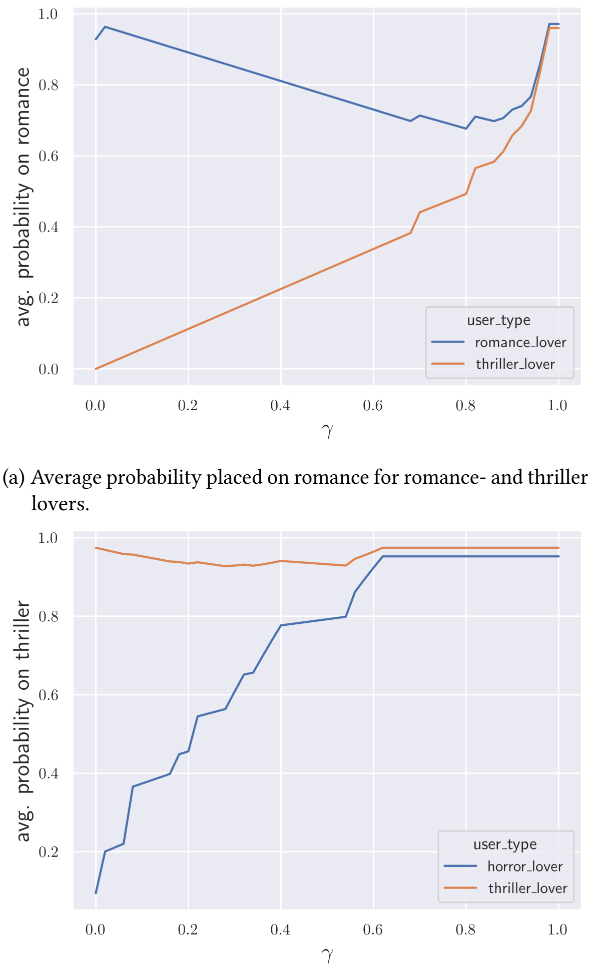
To explore how our framework impacts exposure diversity in practice, we test it out on real world data: the MovieLens dataset [8] which describes people's expressed preferences for movies[1]. These preferences take the form of <user, item, rating, timestamp> tuples, each the result of a user giving a 0–5 star rating for a movie at a particular time.

6.1. Experimental setup

There are $n = 58$ users, randomly selected from the database, and a set $\mathcal{K}$ of $k = 18$ movie genres: $\mathcal{K} = \{$Action, Adventure, Animation, Children, Comedy, Crime, Documentary, Drama, Fantasy, Film-Noir, Horror, Musical, Mystery, Romance, Sci-Fi, Thriller, War, Western$\}$. Each genre is paired with an associated index in $[k]$ determined by alphabetically ordering $\mathcal{K}$. For each movie $m \in \mathcal{M}$, where $\mathcal{M}$ is the set of all movies, there is an associated genre set $m_K \subseteq [k]$ with $|m_K| \geq 1$ (a movie could belong to multiple genres). We use the ratings data to generate preferences for the users. For each movie $m \in \mathcal{M}_i$, where $\mathcal{M}_i$ is the set of movies watched by user $i \in [n]$, the user gives a numeric rating $r_{i,m}$ on a 5-star scale with half-star increments: $r_{i,m} \in \{0.5, 1, 1.5, \ldots, 5.0\}$. We sum these ratings by genre and divide by the number of movies that user i watched from that genre. This results in an average rating $\mu_{i,j} \in [0, 5]$ per genre $j \in [k]$. Finally, we divide $\mu_{i,j}$ by 5 so that $\mu_{i,j} \in [0, 1]$. In the end,

$$\mu_{i,j} = \frac{\sum_{m \in \mathcal{M}_i} r_{i,m} \cdot \mathbb{1}\{j \in m_K\}}{\sum_{m \in \mathcal{M}_i} \mathbb{1}\{j \in m_K\}} \cdot \frac{1}{5}.$$

Using the $\boldsymbol{\mu}_i$s as the mean reward vectors, we use linear programs (LPs) to solve for the optimal policy under no constraints and under both our polarization cap and polarization tax frameworks.



(a) Average probability placed on romance for romance- and thriller-lovers.

(b) Average probability placed on thriller for horror- and thriller-lovers.

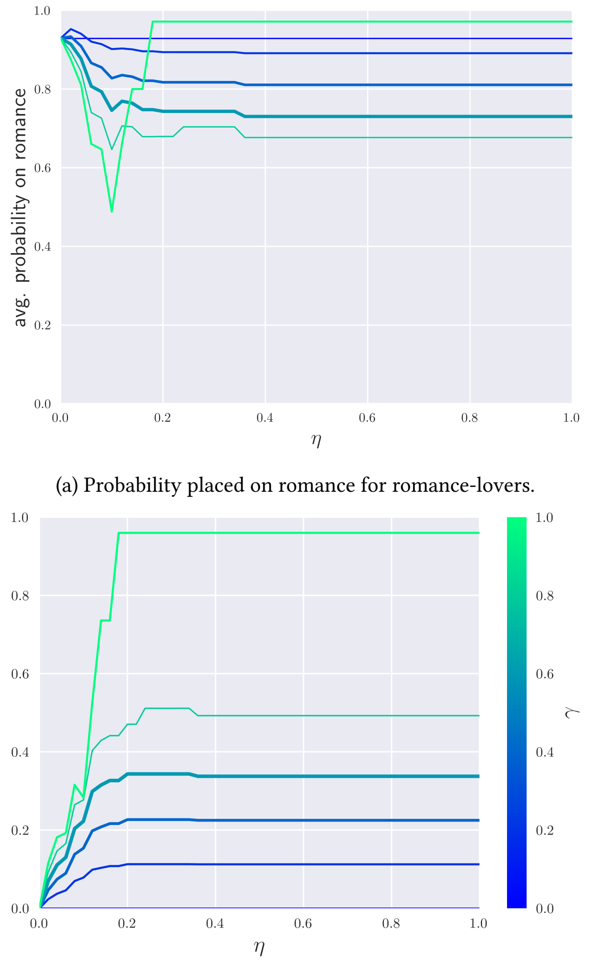
Figure 1: Polarization cap: Content changes as a function of γ for 2 user groups. We compute the optimal policy for 50 values of γ equally spaced between $[0, 1]$.

6.2. Effect of the polarization cap and tax on content recommendations

We begin by investigating the effects that our constraints from Formulation 1 (Section 4.1) have on the optimal content distribution. These experiments provide a parallel to Lemma 4.1, which shows that in a polarized population, users share the burden of diversification. To model a polarized society, we restrict our attention to two dissimilar genres: thriller and romance. In this restricted space, $\boldsymbol{\mu}_i \in [0, 1]^2$. We call the users who prefer the thriller genre *thriller-lovers* and those who prefer the romance genre *romance-lovers.* In Figure 1a, we plot the probability placed on romance by the optimal policy (which maximizes $\sum_{i=1}^{n} \boldsymbol{\mu}_i \cdot \boldsymbol{p}_i$ such that $\boldsymbol{p}_i \geq \frac{\gamma}{n} \sum_{i'=1}^{n} \boldsymbol{p}_{i'}$) as a function of γ. For comparison, we run the same experiments for two similar genres: thriller and horror. In Figure 1b, we plot the probability placed on thriller.

In both Figures 1a and 1b, as γ increases, the content recommendations become more homogeneous. However, the rates at which the recommendations become homogeneous are significantly different. In Figure 1a where the users are polarized, the content recommendations converge slowly. It is not until $\gamma = 0.9$ that the content is completely homogeneous.

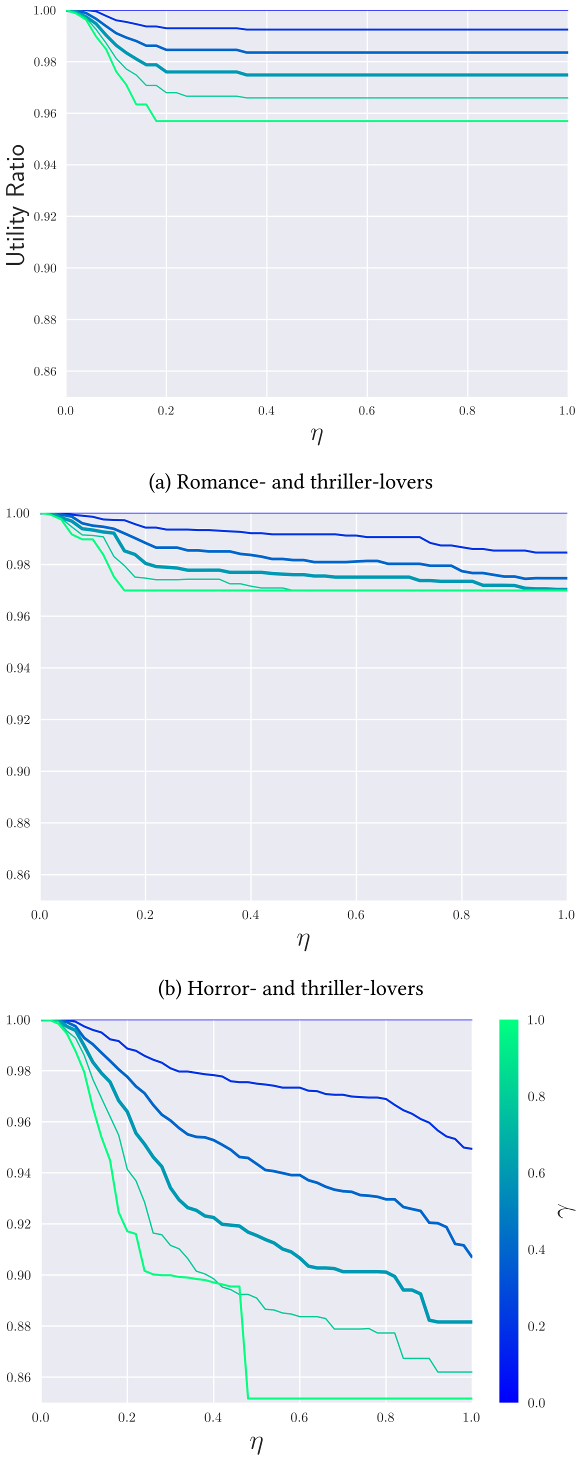
[1] We use this dataset in order to analyze our methods on real-world user preferences, recognizing that movie recommendation filter bubbles would likely not be as pernicious as political news filter bubbles, for example.



(a) Probability placed on romance for romance-lovers.

(b) Probability placed on romance for thriller-lovers.

Figure 2: Polarization tax: Content changes as function of γ and η for romance- and thriller-lovers. We compute the optimal policy for 6 values of γ equally spaced between $[0, 1]$ and 50 values of η equally spaced between $[0, 1]$.



(a) Romance- and thriller-lovers

(b) Horror- and thriller-lovers

(c) All user types

Figure 3: Multiplicative utility loss as a function of γ and η.

Meanwhile, in Figure 1b where the groups of users are similar, the recommendations become homogeneous at a faster rate. In this example, they converge at approximately $\gamma = 0.6$.

Under Formulation 2—where the platform is subject to a penalty (Equation (7))—we perform the same experiments for romance- versus thriller-lovers. These experiments are illustrated in Figure 2 where we vary both η and γ. As before, the content distributions converge as γ increases. However, η serves to modulate the impact of γ on content recommendations. When η is small, the platform prefers to pay some tax to show more personalized content than they would under the hard constraint from Formulation 1. In fact, in Figure 2a, we see that even when $\gamma = 1$ (so the platform is penalized for any level of personalization), the platform prefers to pay some tax and personalize its recommendations, but for sufficiently large η (approximately $\eta \gtrsim 0.2$), the platform switches to obeying the γ constraint and paying no tax. For the other values of γ, the content recommendations change more gradually as η grows. However, after a certain point ($\eta \gtrsim 0.4$), only the value of γ leads to differences in the optimal policy.

6.3. Effect of the polarization tax on user utility

We next investigate the impact of the polarization penalty (Formulation 2) on the users' utility. We analyze the same setting from Section 6.2 where there is a polarized society consisting of romance- and thriller-lovers. Letting $(\boldsymbol{p}_i^{\gamma;\eta})_{i\in[n]}$ be the optimal policy under Equation (7) and $(\boldsymbol{p}_i^*)_{i\in[n]}$ policy with no penalty ($\eta = 0$), Figure 3a plots the ratio of the users' cumulative utilities under these two policies: $\sum \boldsymbol{\mu}_i \cdot \boldsymbol{p}_i^{\gamma;\eta} / \sum \boldsymbol{\mu}_i \cdot \boldsymbol{p}_i^*$. Figure 3b plots the same quantity under the homogenous society from Section 6.2 with only horror- and thriller-lovers.

In Figures 3a and 3b, utility decreases as γ and η grow, but there is a larger utility loss for the polarized group (Figure 3a) compared to the homogeneous group (Figure 3b). Interestingly, in the homogenous group (Figure 3b), utility continuously decreases as η increases, while in the polarized group (Figure 3a), the utility loss eventually flattens out. This is because when the population is homogeneous, as η increases the platform only recommends one genre rather than pay the tax, even when γ is small. However, when users are polarized (Figure 2), the platform recommends both genres and pays some tax for most values of γ. It is only when $\gamma = 1$ that the platform recommends only one genre.

Figure 3c plots the same quantity but without restricting the genres ($\boldsymbol{\mu}_i \in [0, 1]^{18}$). Since the users' preferences are more diverse, the users' cumulative utility suffers a larger but still minimal loss. This is because each user now sees a larger share of content they might not prefer since there are more groups on the platform. Finally, in the full version of the paper, we provide plots illustrating the additive utility loss (rather than multiplicative).

7. Conclusions and discussion

Our work proposes a flexible approach to disincentivizing filter bubbles that adapts to the interests of the individuals on the network. Under our model, if some users are shown a particular type of content, then all users see at least a small amount of that content. We show that our model incentivizes diversity in a way that is equitable to users on the platform and discuss algorithms for recommending content under our framework.

There remain many open questions around disincentivizing polarization in social networks. One might want to distinguish between the content of protected minority groups and that of hate-focused or troll groups. Our current formulation does not distinguish between these situations. One could consider a model where the penalties or cap might scale non-linearly with the size of the group, allowing for more effective moderation. In addition, there is more work to be done to understand the precise impacts of our constraints on the utility of the users and platform. Rewards could represent the profit of the platform or the utility of its users, and our current analysis does not address this distinction. However, the difference could be important when there is a wealth disparity between groups and differences in utility of the users might not easily map to differences in the platform's revenue. A related direction could be to extend our model to maximize popular and well-studied notions of fairness like *Nash social welfare*.

References

[1] R. Jiang, S. Chiappa, T. Lattimore, A. György, P. Kohli, Degenerate feedback loops in recommender systems, in: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, 2019, pp. 383–390.

[2] D. Krueger, T. Maharaj, J. Leike, Hidden incentives for auto-induced distributional shift, arXiv preprint arXiv:2009.09153 (2020).

[3] A.-A. Stoica, A. Chaintreau, Hegemony in social media and the effect of recommendations, in: Companion Proceedings of The 2019 World Wide Web Conference, 2019, pp. 575–580.

[4] A.-A. Stoica, C. Riederer, A. Chaintreau, Algorithmic glass ceiling in social networks: The effects of social recommendations on network diversity, in: Proceedings of the International World Wide Web Conference (WWW), 2018, pp. 923–932.

[5] F. Fabbri, M. L. Croci, F. Bonchi, C. Castillo, Exposure inequality in people recommender systems: The long-term effects, in: Proceedings of the International AAAI Conference on Web and Social Media, volume 16, 2022, pp. 194–204.

[6] D. Freelon, M. Bossetta, C. Wells, J. Lukito, Y. Xia, K. Adams, Black trolls matter: Racial and ideological asymmetries in social media disinformation, Social Science Computer Review 40 (2022) 560–578.

[7] L. E. Celis, S. Kapoor, F. Salehi, N. Vishnoi, Controlling polarization in personalization: An algorithmic framework, in: Proceedings of the conference on fairness, accountability, and transparency (FAT*), 2019, pp. 160–169.

[8] F. M. Harper, J. A. Konstan, The movielens datasets: History and context, ACM Transactions on Interactive Intelligent Systems (TIIS) 5 (2015) 1–19.

[9] N. Haghtalab, M. O. Jackson, A. D. Procaccia, Belief polarization in a complex world: A learning theory perspective, Proceedings of the National Academy of Sciences 118 (2021) e2010144118.

[10] C. A. Bail, L. P. Argyle, T. W. Brown, J. P. Bumpus, H. Chen, M. F. Hunzaker, J. Lee, M. Mann, F. Merhout, A. Volfovsky, Exposure to opposing views on social media can increase political polarization, Proceedings of the National Academy of Sciences 115 (2018) 9216–9221.

[11] A. Rychwalska, M. Roszczyńska-Kurasińska, Polarization on social media: when group dynamics leads to societal divides (2018).

[12] J. Su, A. Sharma, S. Goel, The effect of recommendations on network structure, in: Proceedings of the International World Wide Web Conference (WWW), 2016.

[13] E. Bozdag, J. Van Den Hoven, Breaking the filter bubble: democracy and design, Ethics and information technology 17 (2015) 249–265.

[14] M. Elahi, D. Jannach, L. Skjærven, E. Knudsen, H. Sjøvaag, K. Tolonen, Ø. Holmstad, I. Pipkin, E. Throndsen, A. Stenbom, et al., Towards responsible media recommendation, AI and Ethics (2022) 1–12.

[15] N. Helberger, K. Karppinen, L. D'acunto, Exposure diversity as a design principle for recommender systems, Information, Communication & Society 21 (2018) 191–207.

[16] P. Castells, N. Hurley, S. Vargas, Novelty and diversity in recommender systems, in: Recommender systems handbook, Springer, 2021, pp. 603–646.

[17] D. Halpern, A. D. Procaccia, I. Rahwan, I. Shapira, M. Wüthrich, Optimal engagement-diversity tradeoffs in social media (2023).

[18] W. R. Thompson, On the likelihood that one unknown probability exceeds another in view of the evidence of two samples, Biometrika 25 (1933) 285–294.

[19] A. Badanidiyuru, R. Kleinberg, A. Slivkins, Bandits with knapsacks, in: Symposium on Foundations of Computer Science (FOCS), 2013.

[20] A. Slivkins, Introduction to multi-armed bandits, arXiv preprint arXiv:1904.07272 (2019).

[21] S. Agrawal, N. R. Devanur, Bandits with concave rewards and convex knapsacks, in: ACM Conference on Economics and Computation (EC), 2014, pp. 989–1006.

[22] A. Pacchiano, M. Ghavamzadeh, P. Bartlett, H. Jiang, Stochastic bandits with linear constraints, in: International Conference on Artificial Intelligence and Statistics (AISTATS), PMLR, 2021, pp. 2827–2835.

[23] S. Amani, M. Alizadeh, C. Thrampoulidis, Linear stochastic bandits under safety constraints, in: Conference on Neural Information Processing Systems (NeurIPS), 2019.

[24] A. Moradipari, S. Amani, M. Alizadeh, C. Thrampoulidis, Safe linear thompson sampling with side information, IEEE Transactions on Signal Processing 69 (2021) 3755–3767.

[25] V. Dani, T. P. Hayes, S. M. Kakade, Stochastic linear optimization under bandit feedback, 2008.

[26] Y. Wu, R. Shariff, T. Lattimore, C. Szepesvári, Conservative bandits, in: International Conference on Machine Learning (ICML), 2016.

[27] H. Claure, Y. Chen, J. Modi, M. Jung, S. Nikolaidis, Multi-armed bandits with fairness constraints for distributing resources to human teammates, in: Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, 2020, pp. 299–308.

[28] S. Hossain, E. Micha, N. Shah, Fair algorithms for multi-agent multi-armed bandits, 2021, pp. 24005–24017.

[29] M. Joseph, M. Kearns, J. H. Morgenstern, A. Roth, Fairness in learning: Classic and contextual bandits, 2016.

[30] S. Bubeck, N. Cesa-Bianchi, G. Lugosi, Bandits with heavy tail, IEEE Transactions on Information Theory 59 (2013) 7711–7717.

[31] N. Alon, Y. Matias, M. Szegedy, The space complexity of approximating the frequency moments, in: Proceedings of the Annual Symposium on Theory of Computing (STOC), 1996, pp. 20–29.