Online Platforms and the Fair Exposure Problem Under Homophily

Jakob Schoeffer[1,*], **Alexander Ritchie**[2,*], **Keziah Naggita**[3,*], **Faidra Monachou**[4,*], **Jessie Finocchiaro**[5,*], and **Marc Juarez**[6]

[1]Karlsruhe Institute of Technology, jakob.schoeffer@kit.edu
[2]University of Michigan, aritch@umich.edu
[3]Toyota Technological Institute at Chicago, knaggita@ttic.edu
[4]Stanford University, monachou@stanford.edu
[5]University of Colorado Boulder, jessica.finocchiaro@colorado.edu
[6]University of Southern California, marc.juarez@usc.edu
[*]Equal contribution, listed in reverse-alphabetical order

ABSTRACT

In the wake of increasing political extremism, online platforms have been criticized for contributing to polarization. One line of criticism has focused on echo chambers and the recommended content served to users by these platforms. In this work, we introduce the *fair exposure problem*: given limited intervention power of the platform, the goal is to enforce balance in the spread of content (e.g., news articles) among two groups of users through constraints similar to those imposed by the *Fairness Doctrine* in the United States in the past. Groups are characterized by different affiliations (e.g., political views) and have different preferences for content. We develop a stylized framework that models intra- and inter-group content propagation under homophily, and we formulate the platform's decision as an optimization problem that aims at maximizing user engagement, potentially under fairness constraints. Our main notion of fairness requires that each group see a mixture of their preferred and non-preferred content, encouraging information diversity. Promoting such information diversity is often viewed as desirable and a potential means for breaking out of harmful echo chambers. We study the solutions to both the fairness-agnostic and fairness-aware problems. We prove that a fairness-agnostic approach inevitably leads to group-homogeneous targeting by the platform. This is only partially mitigated by imposing fairness constraints: we show that there exist optimal fairness-aware solutions which target one group with different types of content and the other group with only one type that is not necessarily the group's most preferred. Finally, using simulations with real-world data, we study the system dynamics and quantify the price of fairness.

1 Introduction

In the wake of increasing political extremism [US Department of Justice, 2021], online platforms (e.g., social media networks) have been extensively criticized for exacerbating political polarization in the United States [Boxell et al., 2017, Bail et al., 2018, Hawdon et al., 2020] and elsewhere. This phenomenon is often attributed to platform designs that aim to generate revenue by maximizing user engagement with promoted or shared content (e.g., news articles, opinions, ads). Motivated by the need to promote pluralism online, this paper focuses on understanding the spread of information under a limited platform intervention scheme, where the platform exposes (a subset of) users of the same affiliation to content of contrasting views. We introduce this problem and its study as the *fair exposure problem*.

From a historical perspective, parallels can be drawn between the fair exposure problem and the *Fairness Doctrine* [Ashford, 2021, Pickard, 2021], a past media policy which required that news media cover issues of public importance by presenting diverse, opposing perspectives in an attempt to ensure media diversity. Over the decades, the effectiveness and ethical use of this policy was questioned [Pickard, 2021]: for example, the doctrine enabled activists to help combat racist broadcasting, but it also helped promote the Anti-Equal Rights Amendment campaign [Pickard, 2021]. As history has shown, interventions aimed at balancing the exposure of the public to opposing views might

have ambiguous results. Thus, the goal of this paper is to shed light on the trade-offs that the adoption of such policies may introduce for online platforms.

Towards this goal, we develop a stylized model to understand the impact of platform interventions on the propagation of different articles over time. Our model considers two groups of users with different affiliations and different preferences for articles. Among two opposing articles, we assume that each group tends to like more the article that aligns with the group's views. Moreover, due to homophily in social networks, users in a given group see mostly articles shared by other users in the same group. In this framework, the platform wishes to maximize user exposure (measured through the aggregate number of clicks and likes), potentially due to fairness constraints. We only consider interventions where the platform chooses the articles that an *initial set* of users in each group sees. Our main fairness notion aims at approximately equalizing the relative exposure to a mixture of preferred and non-preferred articles across groups, by imposing certain lower and upper bounds. We analyze the platform's optimization problem and compare the solutions for its unconstrained (fairness-agnostic) version to the solutions for its constrained (fairness-aware) version. We prove that the fairness-agnostic solution always targets each group with one article. When the platform must abide by the fairness constraints, we show that at least one group will be targeted with a mixture of articles. However, depending on the model parameters, it may be optimal that the other group is targeted with only one article type; interestingly, the selected article may not be the group's preferred article. Thus, one group incurs the "cost of fairness," whereas the other one the "cost of maximizing engagement." When the content refers to high-stakes procedures (e.g., referendums, elections), such an outcome can be problematic.

We supplement our theoretical results with empirical results to gain additional insights by estimating our model parameters from real-world datasets collected from Twitter and Facebook [Garimella et al., 2017, Bakshy et al., 2015]. Moreover, we measure the *price of fairness*, i.e., the difference in the platform's utility between the fairness-aware and the fairness-agnostic settings. Using parameters estimated from Bakshy et al. [2015], we observe an optimal fairness-aware solution that heavily favors one group.

2 Related work

The spread of information in social networks is well-studied; the structure of these social networks tends to be homophilous [McPherson et al., 2001, Lazarsfeld and Merton, 1948]. Balancing information exposure has also been studied through several different technical methods; however, to our knowledge, the impact of platform interventions to ensure balanced exposure via fairness constraints has not been studied before. Celis et al. [2019] study a similar problem of controlling polarization in bandit settings, though our model differs by assuming that intervention is only possible at the first time step; their constrained problem is similar to our approximately fair average exposure constraint in (3). Our model is sequential like the social learning models of Banerjee [1992], Bikhchandani et al. [1992], which also study information spread, but without balancing content exposure. Papanastasiou [2020], Candogan and Drakopoulos [2020] study stylized models for fact-checking news articles in social networks when the platform can intervene to inspect the content or incentivize fact-checking by users through information design; Cisternas and Vásquez [2020] take a market design approach. Allon et al. [2021] further show that polarization arises due to uncertainty in content accuracy.

Starbird et al. [2018] demonstrate the emergence of echo chambers by a mixed methods analysis of perceptions of the White Helmets, particularly enabled by content sharing platforms such as Twitter, and Jeon et al. [2021] gamify the balance of seeking influence and reputability simultaneously on Twitter. Our setting is also similar to influence maximization literature [Kempe et al., 2003] in the sense that platform interventions are limited. However, our model is sequential and aims for balance in article exposure, while the influence maximization literature seeks to maximize information diffusion [Fish et al., 2019, Stoica et al., 2020, Ali et al., 2019]. Finally, balancing information propagation is well-studied in literature on recommender systems [Zoetekouw, 2019, Hu et al., 2012, Farajtabar et al., 2016] and the emergence of echo chambers [Barberá et al., 2015, Mukerjee et al., 2020, Dubois and Blank, 2018, Hosseinmardi et al., 2020]. Bakshy et al. [2015] and Garimella et al. [2017] empirically study the extent of disparity in intra-group exposure of ideas and do not aim to balance it. In general, although previous works [Bakshy et al., 2015, Garimella et al., 2017] investigate the empirics of information flow in similar models, they do not study the mechanisms that lead to (imbalanced) exposure; our model addresses this.

Many of the standard metrics of group fairness are not applicable in our setting as we work with heterogeneous preferences of outcomes: members of one group prefer seeing content that aligns with their group identity. Graph-based models of opportunity flow have considered similar, yet inherently different, fairness constraints and problems. For example, Liu et al. [2021] consider fair equality of opportunity in settings where flow of opportunity proceeds along an acyclic graph and everyone is striving for the same desired outcome. Similarly, Arunachaleswaran et al. [2021] approximately optimize social welfare in settings where opportunity flows along an acyclic graph. However,

neither of the approaches in this paper are directly applicable to our setting. The definitions of *fair exposure* presented in § 3 are stylized for this particular setting.

3 Model

3.1 General setup

We consider a platform with a finite mass M of users with affiliation group $g \in \{A, B\}$. Let $\pi_g \in (0, 1)$ denote the fraction of users from group g (at any time). We assume that $\pi_A = 1 - \pi_B = \pi$. Time is discrete with $t = 1, 2, \ldots, T$, $T \leq M$. All notation is summarized in Table 1.

Table 1: Overview of notation.

Symbol	Definition
M	Finite mass of users
$g \in \{A, B\}$	Affiliation group
$\pi_g \in (0, 1)$	Fraction of users in group g
$t \in \{1, \ldots, T\}$	Time step (discrete) with horizon $T \leq M$
$s \in \{a, b\}$	Article sources affiliated with groups A, B
$\theta_{g,s} \in [0, 1]$	Fraction of group g users who are shown article s by the platform at $t = 1$
$p_{g,s} \sim F_{g,s}$	Probability for users of group g to like an article of source s
$F_{g,s}$	Distribution with support $[0, 1]$
$c_{g,s} > 0$	Cost users in g occur when clicking on an article s
$v_{g,s} > 0$	Valuation of users in g when liking an article s
$q_g \in (1/2, 1)$	Probability of intra-group propagation
$l_{g,s}(t, \theta)$	Mass of users in g born at time t who have clicked and liked an article s
$e_{g,s}(t)$	Exposure of users in g to article s at time t
$\psi_{g,s} > 0$	Probability that user in g likes and clicks on article s
$\underline{\delta} < 1 < \overline{\delta}$	Fairness lower and upper bound parameters

Before time $t = 1$, the platform receives two articles representing different views a, b that are aligned with groups A, B, respectively (e.g., sponsored posts on Facebook or Twitter). For simplicity, we refer to the two articles as a and b, where a (resp. b) is the *in-group/preferred* (resp. *out-group/non-preferred*) article type of group A, and similarly for group B. At time $t = 1$, the platform decides how many users in group g to show an article s to. Let $\theta_{g,s}$ denote the fraction of users in group g who are shown article s by the platform at time $t = 1$.

Each user observes the source $s \in \{a, b\}$ of the article they are shown. Users of group g have a probability $p_{g,s} \sim F_{g,s}$ of "liking" an article of source s, where $F_{g,s}$ is a known distribution with support $[0, 1]$. Each user in group g knows their own realized probabilities $p_{g,s}$ for $s \in \{a, b\}$. Users from group A have a higher preference for articles of source a; the same holds for users of group B and articles of source b. To model this (stochastically) biased behavior of users in each group g, we assume that $p_{A,a} \succ_{FSD} p_{A,b}$ and $p_{B,b} \succ_{FSD} p_{B,a}$.[1]

At every time period $t > 0$, a unit mass of users arrives. At time t, each user in group g sees one article s and decides whether to click with probability dependent on $p_{g,s} \sim F_{g,s}$. If the user clicks on the article, they incur a constant cost $c_{g,s} > 0$ for reading the article. If they like it, they get valuation $v_{g,s} > 0$ (minus the cost $c_{g,s}$), so their final payoff is $v_{g,s} - c_{g,s} > 0$. If they do not like it, their final payoff is $-c_{g,s} < 0$.

At the next period $t + 1$, an equal mass of users arrives. This modeling choice reflects the general format of content sharing on social platforms, in which at different time steps, there are different batches of people on the online platform. We assume synchronicity in individual arrivals rather than time-step measures so we can reduce to a discretized time analysis without loss of generality. Specifically, each user from group g gives their position to a user[2] from the same group g with probability $q_g \in (1/2, 1)$, where the lower bound comes from homophily assumptions; we refer to this event as *intra-group propagation*. With probability $1 - q_g$, this user is replaced by a user in $g' \neq g$ (*inter-group propagation*). In order to ensure consistency with the fraction π_g of each group g over time, we require the parameters

[1] Recall that a random variable X with CDF F_X *first-order stochastically dominates* Y with CDF F_Y, that is $X \succ_{FSD} Y$, if $F_X(z) < F_Y(z)$ for all z.

[2] Our model and analysis can be directly extended to the case where a user in period t is replaced by $n_{t+1} > 1$ users in period $t + 1$. The current assumption is made for clarity of exposition.

q_A, q_B, and π satisfy $q_A\pi_A + (1 - q_B)\pi_B = \pi_A$.[3] If a user i arriving at time t liked the article, then the new user i', replacing user i at time $t + 1$, sees the same article as i. If user i did not like an article, then user i' is not shown any article at time $t + 1$.

For $t \geq 1$, let $l_{g,s}(t, \theta)$ denote the mass of users born at time t who belong to group g and have clicked and liked an article s. The objective of the platform is to maximize user exposure over time, i.e.,

$$\max_{\theta_{A,a},\theta_{B,a}\in[0,1]} \sum_{t=1}^{T} \sum_{g\in\{A,B\}} \sum_{s\in\{a,b\}} l_{g,s}(t, \theta), \tag{1}$$

potentially subject to fair exposure constraints. We measure *user exposure* in the number of users who click and like an article. The strengths of this metric are two-fold: first, because the platform has to plan for T time steps, ensuring an article is liked means it will continue to propagate in the next time step. Second, we assume that liking an article is a proxy for more *meaningful* engagement than simply clicking on it.

Discussion of model assumptions While our model makes many simplifying assumptions, this strengthens our negative results (e.g., Lemma 3) as they do not hold even in an oversimplified model. Moreover, while our model is not graph-based, it is an abstraction of the Erdős-Rényi random graph in expectation with different attachment parameters for each group. As many social networks generally closely resemble preferential attachment models rather than Erdős-Rényi graphs [Clauset and Larremore, 2021], we compare our model's performance to graph-based simulations in § E, and observe similar results.

3.2 Notions of fair exposure

Broadly speaking, we define *fair exposure* as a situation where users of different affiliation are similarly exposed to non-preferred content. Promoting such information diversity—as opposed to *selective exposure* [Freedman and Sears, 1965]—is often viewed as desirable and a potential means for breaking out of harmful echo chambers that are detrimental to "the quality, safety, and diversity of discourse online," as Gillani et al. [2018] put it. Garrett and Resnick [2011], among others, likewise suggest that "software designers ought to create tools that encourage and facilitate consumption of diverse news streams, making users, and society, better off." Diversity of perspectives might also help users to see things from novel perspectives or become aware that they might be already stuck in an echo chamber. We operationalize fair exposure through two types of constraints: first, we ask that exposure rates for both types of content be *equal at each point in time* ("constant fair exposure"). Acknowledging that this is a rather restrictive constraint, we also examine fair *average* exposure, where we further allow a certain deviation from equality ("approximately fair average exposure").

Constant fair exposure The rate of exposure of users to their preferred article s is constant at level $e \in [0, 1]$ at each time step and equal across groups, i.e.,

$$\frac{l_{A,s}(t, \theta)}{\pi_A} = \frac{l_{B,s'}(t, \theta)}{\pi_B} = e \quad \forall t \leq T,\ \forall s, s' \in \{a, b\}, s \neq s'. \tag{2}$$

Approximately fair average exposure The total exposure of users to their preferred article s (resp. non-preferred article s') is approximately equal across groups, i.e., for given parameters $\underline{\delta} < 1 < \overline{\delta}$,

$$\underline{\delta} \leq \frac{\sum_{t=1}^{T} l_{A,a}(t, \theta)}{\sum_{t=1}^{T} l_{B,b}(t, \theta)} \leq \overline{\delta} \quad \text{and} \quad \underline{\delta} \leq \frac{\sum_{t=1}^{T} l_{A,b}(t, \theta)}{\sum_{t=1}^{T} l_{B,a}(t, \theta)} \leq \overline{\delta}. \tag{3}$$

4 Theoretical analysis

4.1 Preliminaries

We begin with preliminaries. We define the users' decision problem, analytically describe the system dynamics, and finally transform them to a tractable non-recursive form.

[3] This is necessary for theoretical results, but it does not hold for the parameters used in § 5, and does not affect results there.

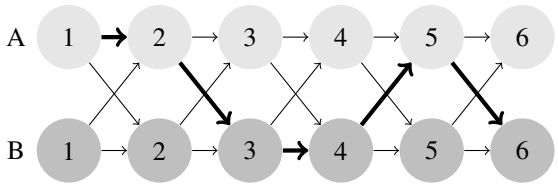

Figure 1: *Article sharing over time with $T = 6$:* The horizontal and diagonal edges represent intra-group and inter-group propagation, respectively. With thicker edges, we give an example of how an article s initially read by a user in A is propagated through the network.

Users' decision problem A user in group g with realized probability $p_{g,s}$ of liking an article shown to them clicks on the article if and only if their expected utility is non-negative, that is

$$v_{g,s}p_{g,s} \geq c_{g,s}. \tag{4}$$

Therefore, the fraction of users in g who click on article s shown to them is $1 - F_{g,s}(\frac{c_{g,s}}{v_{g,s}})$. Since $p_{A,a} \succ_{FSD} p_{A,b}$ and $p_{B,b} \succ_{FSD} p_{B,a}$, users tend to click more on their in-group articles.

Understanding system dynamics As a warm-up, we show how the different masses of users evolve in the first time period. We then generalize to any $t > 1$.

Time $t = 1$. Fix the fractions $\theta_{A,a}$ and $\theta_{B,a}$ of users in groups A and B, respectively, who are shown article a at time $t = 1$; recall that $\theta_{A,a}$ and $\theta_{B,a}$ are the platform's decision. Let L denote the Bernoulli random variable that a user likes the article after clicking on it. Then, the mass of users in g who clicked on the article from source a and liked it during period $t = 1$ is

$$\begin{aligned} l_{g,s}(1,\theta) =& \pi_g \theta_{g,s} \left(1 - F_{g,s}\left(\frac{c_{g,s}}{v_{g,s}}\right)\right) \int_0^1 \Pr[L = 1 \mid p_{g,s}] d\left(\frac{F_{g,s}(p_{g,s})}{1 - F_{g,s}\left(\frac{c_{g,s}}{v_{g,s}}\right)}\right) \\ =& \pi_g \theta_{g,s} \int_{c_{g,s}/v_{g,s}}^1 \Pr[L = 1 \mid p_{g,s}] dF_{g,s}(p_{g,s}) = \pi_g \theta_{g,s} \int_{c_{g,s}/v_{g,s}}^1 p dF_{g,s}(p). \end{aligned}$$

Symmetrically, the mass of users in g who clicked on article s but did *not* like it equals $\pi_g \theta_{g,s} \int_{c_{g,s}/v_{g,s}}^1 (1-p) dF_{g,s}(p)$. The rest of users in group g who were shown article s did not click on it; their mass equals $\pi_g \theta_{g,s} F_{g,s}\left(\frac{c_{g,s}}{v_{g,s}}\right)$.

Time $t > 1$. For general $t > 1$, recall that a user in group g who was shown article s is replaced by a user also in g in the next time period with probability q_g (and by a user in the opposite group $g' \neq g$ with probability $1 - q_g$). For brevity, we refer to the new user as the *replacing user*. Figure 1 illustrates how an article "travels" throughout the network via intra- and inter-group propagation.

Generalizing the system dynamics for $t > 1$, we obtain the following recursive formula:

$$\begin{aligned} l_{g,s}(t+1,\theta) &= (q_g l_{g,s}(t,\theta) + (1 - q_{g'}) l_{g',s}(t,\theta)) \int_{c_{g,s}/v_{g,s}}^1 p dF_{g,s}(p) \\ &= \psi_{g,s}(q_g l_{g,s}(t,\theta) + (1 - q_{g'}) l_{g',s}(t,\theta)), \end{aligned} \tag{5}$$

where we used the substitution $\psi_{g,s} \coloneqq \int_{c_{g,s}/v_{g,s}}^1 p dF_{g,s}(p)$.[4] For an exemplary visualization, see Figure 2. The next lemma follows from (5). All proofs can be found in § A.

Lemma 1. *The mass function can be written as $l_{g,s}(t,\theta) = \theta_{g,s} w_{g,s}(t) + \theta_{g',s} u_{g,s}(t)$, where $u_{g,s}(1) = 0$, $u_{g,s}(t) > 0$ for $t \geq 2$, and $w_{g,s}(t) > 0$ for $t \geq 1$.*

Lemma 1 says that $l_{g,s}$ is a strictly increasing linear function of $\theta_{g,s}$ and $\theta_{g',s}$ except at time $t = 1$, when $l_{g,s}$ is not a function of $\theta_{g',s}$. We note that $w_{g,s}(t)$ corresponds to the mass of intra-group propagation and $u_{g,s}(t)$ to that of inter-group propagation.

Unfortunately, the recursive expression for the mass function given in (5) is intractable. Thus, in Theorem 1 we derive an equivalent non-recursive expression using the one-sided $\mathcal{Z}$-transform.

[4] We assume $\psi_{g,s} > 0$, i.e., we do not consider the trivial case of $\psi_{g,s} = 0$.

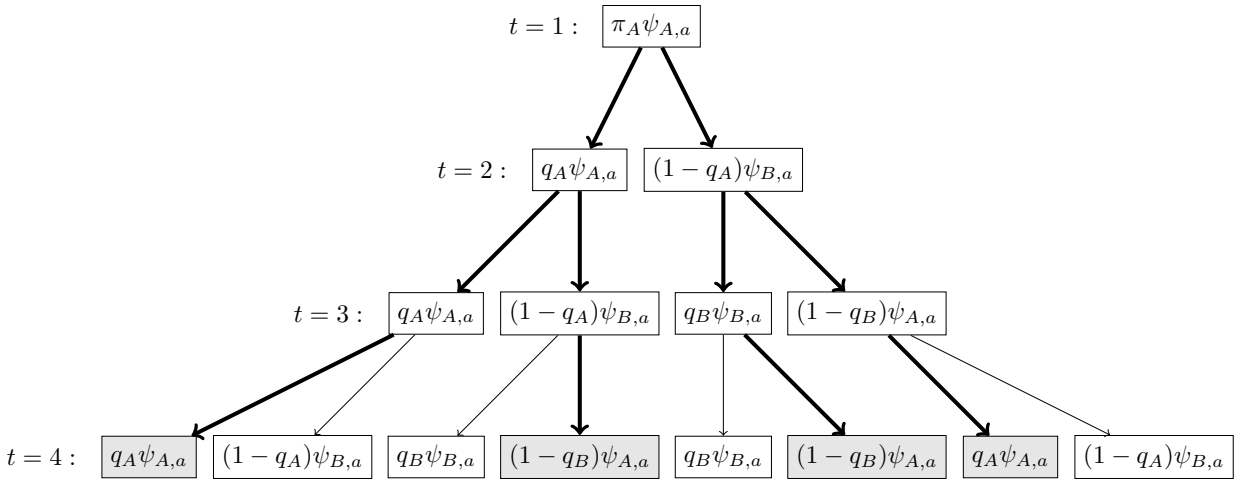

Figure 2: *Exemplary visualization of the network propagation for $l_{A,a}$, with $T = 4$, $\theta_{A,a} = 1$, $\theta_{B,a} = 0$*: The value of $l_{A,a}(t, \theta)$ is obtained by multiplying along the thick edges and adding up all the 2^{T-2} paths leading into leaves that include q_A (intra-group propagation within A) or $1 - q_B$ (inter-group propagation from B to A) as a factor (depicted in gray).

Theorem 1. *For all $t \geq 1$, regardless of group g and article s, we have*

$$w_{g,s}(t) = A^w_{1,g,s} a^{t-1}_{1,s} + A^w_{2,g,s} a^{t-1}_{2,s}, \quad t \geq 1, \tag{6}$$

$$u_{g,s}(t) = A^u_{g,s}(a^{t-1}_{1,s} - a^{t-1}_{2,s}), \quad t \geq 1, \tag{7}$$

with $w_{g,s}(t)$ and $u_{g,s}(t)$ as introduced in Lemma 1, and

$$\begin{aligned}
a_{1,s} &\coloneqq \frac{1}{2}\big(\psi_{g,s}q_g + \psi_{g',s}q_{g'} + \big[(\psi_{g,s}q_g + \psi_{g',s}q_{g'})^2 - 4\psi_{g,s}\psi_{g',s}(q_g + q_{g'} - 1)\big]^{\frac{1}{2}}\big) \\
a_{2,s} &\coloneqq \psi_{g,s}q_g + \psi_{g',s}q_{g'} - a_{1,s} \\
A^w_{1,g,s} &\coloneqq \frac{\pi_g\psi_{g,s} + \psi_{g',s}\left(\pi_{g'}\psi_{g',s}(1-q_{g'}) - \pi_g\psi_{g,s}q_{g'}\right)a^{-1}_{1,s}}{1 - a_{2,s}a^{-1}_{1,s}} \\
A^w_{2,g,s} &\coloneqq \frac{\pi_g\psi_{g,s} + \psi_{g',s}\left(\pi_{g'}\psi_{g',s}(1-q_{g'}) - \pi_g\psi_{g,s}q_{g'}\right)a^{-1}_{2,s}}{1 - a^{-1}_{2,s}a_{1,s}} \\
A^u_{g,s} &\coloneqq \frac{\psi_{g',s}\left(\pi_{g'}\psi_{g',s}(1-q_{g'}) - \pi_g\psi_{g,s}q_{g'}\right)a^{-1}_{1,s}}{1 - a_{2,s}a^{-1}_{1,s}}.
\end{aligned}$$

We note that all quantities in Theorem 1 are real numbers, which is shown in Lemma 5 in § B. An intuitive interpretation of these quantities is as follows. The terms $a_{1,s}$ and $a_{2,s}$ are the roots of a quadratic in $\mathcal{Z}$-space that roughly corresponds to a kinematic equation describing the homophilic sharing process. $A^u_{g,s}$ corresponds roughly to the difference between contributions to the mass that would have been realized if inter-group propagation had not occurred and those that would have been realized if intra-group propagation had not occurred. The quantities $A^w_{1,g,s}$ and $A^w_{2,g,s}$ correspond roughly to $A^u_{g,s}$ and $-A^u_{g,s}$, respectively, plus an additional term relating to the mass generated by propagation *within* group g and propagation *from* group g' to g.

4.2 Platform's optimization problem

Building upon our previous results, in this section we proceed to formulate the platform's problem, i.e., the maximization of user exposure, as a linear program subject to *approximately fair average exposure* constraints (see § 4.2.2 for constant fair exposure). More specifically, at time $t = 1$ the platform needs to decide the fraction of users in each group to show articles a and b. Recall that we denote the proportion of users in g that are shown article s by $\theta_{g,s}$. The platform wants to maximize the total number of users across all groups that click on and like the two articles, but also

faces a fair exposure constraint (see (3)). Thus, the platform's optimization problem becomes:

$$\max_{\theta_{A,a},\theta_{B,a}\in[0,1]} \sum_{t=1}^{T} \sum_{g\in\{A,B\}} \sum_{s\in\{a,b\}} l_{g,s}(t,\theta) \tag{P}$$

$$\text{s.t. } \underline{\delta} \leq \frac{\sum_{t=1}^{T} l_{A,a}(t,\theta)}{\sum_{t=1}^{T} l_{B,b}(t,\theta)} \leq \overline{\delta} \tag{C1}$$

$$\underline{\delta} \leq \frac{\sum_{t=1}^{T} l_{A,b}(t,\theta)}{\sum_{t=1}^{T} l_{B,a}(t,\theta)} \leq \overline{\delta}. \tag{C2}$$

Intuitively, to avoid the extreme, but feasible, case where each group g is only shown their preferred article, definition (3) introduces constraints (C1) and (C2). These constraints require that each group is exposed to their preferred article and their non-preferred article in a balanced way, i.e., the exposure ratio is similar for both articles within a group (within bounds $\underline{\delta} < 1 < \overline{\delta}$).

From Lemma 1, we know that, given $t \in \{1,\ldots,T\}$, $l_{g,s}(t,\theta)$ is a linear and strictly increasing function in $\theta_{g,s}$, $\theta_{g',s}$. Thus, the objective function of (P) is linear in two dimensions; similarly, the exposure constraints can also be transformed to linear inequalities. Consequently, (P) is a linear program.

4.2.1 Fairness-agnostic optimization problem

As a natural benchmark, we first consider the optimization problem (P) *without* exposure constraints (C1) and (C2), while retaining the constraint $\theta_{g,s} \in [0,1]$ for all g, s. We refer to this as the *fairness-agnostic* problem. We show that the exclusion of fairness constraints *always* results in all members of the same group being shown the same article by the platform at time $t = 1$. Specifically, the solution to the fairness-agnostic exposure problem is given in the following proposition.

Proposition 1. *The solution to the fairness-agnostic optimization problem is*

$$\theta^*_{A,a} = \mathbf{1}\left\{\sum_{t=1}^{T}\left((w_{A,a}(t) - w_{A,b}(t) + u_{B,a}(t) - u_{B,b}(t)\right) > 0\right\},$$

$$\theta^*_{B,a} = \mathbf{1}\left\{\sum_{t=1}^{T}\left(w_{B,a}(t) - w_{B,b}(t) + u_{A,a}(t) - u_{A,b}(t)\right) > 0\right\}.$$

From a theoretical perspective, this result follows from the linearity of (P). From a practical perspective, Proposition 1 suggests that targeting a group with their preferred article is not necessarily optimal for maximizing user engagement. Albeit counter-intuitive, it might be optimal for the platform to ignore group preferences and target the whole user network with a single article. Two additional implications of Proposition 1 are given in Corollary 1 and Lemma 2 below.

Corollary 1. *The feasible solution $\theta_{A,b} = 1$, $\theta_{B,a} = 1$ is never optimal for* (P).

Lemma 2. *Assume $\theta_{A,a} = 1$, $\theta_{B,b} = 1$. If*

$$\frac{q_A \pi_A}{(1-q_B)\pi_B} < 1, \quad \frac{\psi_{A,a}\psi_{B,a}}{\pi_B \psi_{A,b}\psi_{B,b}} < 1, \tag{8}$$

then group A is exposed more to article b than a over time, i.e., $e_{A,b}(T) = \frac{l_{A,b}(T,\theta)}{\pi_A} > e_{A,a}(T) = \frac{l_{A,a}(T,\theta)}{\pi_A}$ for any $T > 2$.

Under homophily, one might expect a group to be preferentially exposed to in-group articles. However, as Lemma 2 shows, this may not be the case if group sizes are radically different or if one group displays much lower levels of homophily than the other. Lemma 2 can shed light on several counter-intuitive possibilities for article exposure over time. More specifically, it suggests that, due to the network structure and the dynamics of propagation, targeting each group with their preferred article might *not always* bring the intended targeting and thus potentially lead to suboptimal outcomes for the platform. Even if the platform targets each group only with their preferred (in-group) article, one group may—after several rounds—be exposed to their non-preferred (out-group) article. For example, given a significantly larger group B, weak homophily for both groups ($q_A \simeq q_B \simeq 1/2$) and similar preferences for

compatible articles ($\psi_{A,a} \simeq \psi_{B,b}$), group A is exposed more to article b even if $\theta_{A,a} = 1$ and $\theta_{B,b} = 1$. A similar property holds when there is an extreme preference for article b in group B compared to moderate preference in group A, i.e., $\psi_{B,b} \gg \psi_{A,a}$.

In contrast to Lemma 2, Corollary 1 offers a quite intuitive insight, showing that the opposite strategy (i.e., targeting both groups with their out-group article) is *never* optimal. Indeed, depending on the model parameters, either the network will eventually favor the article with the largest sharing rate in total or the users in each group will start clicking more on their in-group article. In both cases, the platform's initial targeting $\theta_{A,b} = 1$, $\theta_{B,a} = 1$ would only manage to delay any of these events thus leading to a suboptimal number of clicks and likes at the initial stages of propagation.

4.2.2 Fairness-aware optimization problem: Constant fair exposure

In this section, we explore the feasibility of a natural but stricter fairness notion, i.e., constant fair exposure (2). As detailed in Lemma 3 below, we show that it is generally not possible to achieve equal and constant exposure at every time step unless certain restrictive conditions hold.

Lemma 3. *Let $e \in (0,1)$ be the platform's targeted fair exposure level. Achieving constant fair exposure is possible if and only if for both $s \in \{a,b\}$,*

$$\psi_{A,s}\left(q_A + \frac{1-\pi_A}{\pi_A}(1-q_B)\right) = \psi_{B,s'}\left(\frac{1-\pi_A}{\pi_A}q_B + (1-q_A)\right) \tag{9}$$

and the platform sets $\theta_{A,a} = 1 - \theta_{B,a} = e$ at time $t = 1$.

The conditions of Lemma 3 guarantee that the mass of users clicking on a given article will be the same across all groups and time steps. However, this will almost certainly never occur in practice due to differing preferences in content across groups. Therefore, we ask if average exposure over time can be equalized across groups, i.e., if $\frac{1}{T}\sum_{t=1}^{T}\frac{l_{A,s}(t,\theta)}{\pi_A} = \frac{1}{T}\sum_{t=1}^{T}\frac{l_{B,s'}(t,\theta)}{\pi_B} = e$ is possible. Lemma 4 shows that it is very difficult to achieve any desired average exposure rate.

Lemma 4. *For any $\pi_g \in (0,1)$, average exposure levels for group g to article s are achievable only in the range $0 \le e \le \frac{1}{T\pi_g}\sum_{t=1}^{T}(w_{g,s}(t) + u_{g,s}(t))$.*

Given the restrictive nature of constant fair exposure, we turn to a relaxed notion: *approximately fair average exposure*.

4.2.3 Fairness-aware optimization problem: Approximately fair average exposure

As our final step, we explore the feasibility of the optimization problem (P) *with* fairness constraints (C1) and (C2), and analytically describe the solution by deriving expressions for the extreme points of the constraint polytope. Let

$$\overline{m}_{g,s} \coloneqq \sum_{t=1}^{T} u_{g,s}(t) + \overline{\delta}\sum_{t=1}^{T} w_{g',s'}(t), \quad \underline{m}_{g,s} \coloneqq \sum_{t=1}^{T} u_{g,s}(t) + \underline{\delta}\sum_{t=1}^{T} w_{g',s'}(t)$$
$$\overline{n}_{g,s} \coloneqq \sum_{t=1}^{T} w_{g,s}(t) + \overline{\delta}\sum_{t=1}^{T} u_{g',s'}(t), \quad \underline{n}_{g,s} \coloneqq \sum_{t=1}^{T} w_{g,s}(t) + \underline{\delta}\sum_{t=1}^{T} u_{g',s'}(t)$$
$$m_{g,s} \coloneqq \sum_{t=1}^{T} u_{g,s}(t) + \sum_{t=1}^{T} w_{g,s}(t).$$

From constraints (C1) and (C2) and using Theorem 1, we can infer the feasible bounds on $\theta_{B,a}$ (dependent on $\theta_{A,a}$), in addition to $\theta_{A,a}, \theta_{B,a} \in [0,1]$. We state these bounds as well as the axes intersects of the hyperplanes that induce the half-spaces containing the feasible region in § C. Evaluating the relative positions of these hyperplanes, we can then infer when the fairness-aware optimization problem is infeasible, which is specified in Theorem 2.

Theorem 2. *The fairness-aware optimization problem is infeasible if and only if*

$$\begin{aligned}
\frac{\underline{\delta} m_{B,b}}{\underline{m}_{A,a}} > \frac{m_{A,b}}{\underline{m}_{A,b}} \quad &\textit{and} \quad \frac{\underline{\delta} m_{B,b}}{\underline{n}_{A,a}} > \frac{m_{A,b}}{\underline{n}_{A,b}}; \textit{ or} \\
\frac{m_{A,b}}{\overline{m}_{A,b}} > \frac{\overline{\delta} m_{B,b}}{\overline{m}_{A,a}} \quad &\textit{and} \quad \frac{m_{A,b}}{\overline{n}_{A,b}} > \frac{\overline{\delta} m_{B,b}}{\overline{n}_{A,a}}; \textit{ or} \\
\underline{\delta} \sum_{t=1}^{T} w_{B,b}(t) > \sum_{t=1}^{T} w_{A,a}(t) \quad &\textit{and} \quad \frac{\underline{\delta} m_{B,b}}{\underline{m}_{A,a}} > \frac{\underline{\delta} m_{B,b}}{\underline{\delta} m_{B,b} - \underline{n}_{A,a}}; \textit{ or} \\
\sum_{t=1}^{T} u_{A,b}(t) > \overline{\delta} \sum_{t=1}^{T} u_{B,a}(t) \quad &\textit{and} \quad \frac{m_{A,b}}{\overline{m}_{A,b}} > \frac{m_{A,b}}{m_{A,b} - \overline{n}_{A,b}}.
\end{aligned}$$

It follows that we can always make the problem feasible by setting $\underline{\delta}$ and $\overline{\delta}$ accordingly. As noted in § C, letting $\underline{\delta} \to 0$ and $\overline{\delta} \to \infty$, the fairness-agnostic problem is recovered. By the intermediate value theorem, there exist infinitely many values of $\underline{\delta}, \overline{\delta}$ that define a non-empty feasible region strictly contained in the unit box. Otherwise, if the problem is feasible, the fundamental theorem of linear programming states that an optimal solution will occur at a corner point of the feasible region, or on a line segment between two corner points.

Theorem 3. *For any admissible problem parameters, the optimal solution to the fairness-aware optimization problem is one of the following:*

$$\theta^1_{A,a} = \phi\left(\frac{m_{A,b}\overline{m}_{A,a} - \overline{\delta} m_{B,b}\underline{m}_{A,b}}{\underline{n}_{A,b}\overline{m}_{A,a} - \overline{n}_{A,a}\underline{m}_{A,b}}\right), \quad \theta^1_{B,a} \in \left\{0, 1, \frac{\overline{\delta} m_{B,b} - \theta^1_{A,a}\overline{n}_{A,a}}{\overline{m}_{A,a}}, \frac{m_{A,b} - \theta^1_{A,a}\underline{n}_{A,b}}{\underline{m}_{A,b}}\right\}$$

$$\theta^2_{A,a} = \phi\left(\frac{m_{A,b}\overline{m}_{A,a} - \overline{\delta} m_{B,b}\overline{m}_{A,b}}{\underline{n}_{A,b}\overline{m}_{A,a} - \overline{n}_{A,a}\overline{m}_{A,b}}\right), \quad \theta^2_{B,a} \in \left\{0, 1, \frac{\overline{\delta} m_{B,b} - \theta^2_{A,a}\overline{n}_{A,a}}{\overline{m}_{A,a}}, \frac{m_{A,b} - \theta^2_{A,a}\overline{n}_{A,b}}{\overline{m}_{A,b}}\right\}$$

$$\theta^3_{A,a} = \phi\left(\frac{m_{A,b}\underline{m}_{A,a} - \underline{\delta} m_{B,b}\underline{m}_{A,b}}{\underline{n}_{A,b}\underline{m}_{A,a} - \underline{n}_{A,a}\underline{m}_{A,b}}\right), \quad \theta^3_{B,a} \in \left\{0, 1, \frac{\underline{\delta} m_{B,b} - \theta^3_{A,a}\underline{n}_{A,a}}{\underline{m}_{A,a}}, \frac{m_{A,b} - \theta^3_{A,a}\underline{n}_{A,b}}{\underline{m}_{A,b}}\right\}$$

$$\theta^4_{A,a} = \phi\left(\frac{m_{A,b}\underline{m}_{A,a} - \underline{\delta} m_{B,b}\overline{m}_{A,b}}{\underline{n}_{A,b}\underline{m}_{A,a} - \underline{n}_{A,a}\overline{m}_{A,b}}\right), \quad \theta^4_{B,a} \in \left\{0, 1, \frac{\underline{\delta} m_{B,b} - \theta^4_{A,a}\underline{n}_{A,a}}{\underline{m}_{A,a}}, \frac{m_{A,b} - \theta^4_{A,a}\overline{n}_{A,b}}{\overline{m}_{A,b}}\right\}$$

$$\theta^5_{B,a} \in \{0,1\}, \quad \theta^5_{A,a} \in \left\{0, 1, \frac{\overline{\delta} m_{B,b} - \theta^5_{B,a}\overline{m}_{A,a}}{\overline{n}_{A,a}}, \frac{\underline{\delta} m_{B,b} - \theta^5_{B,a}\underline{m}_{A,a}}{\underline{n}_{A,a}}, \frac{m_{A,b} - \theta^5_{B,a}\overline{m}_{A,b}}{\overline{n}_{A,b}}, \frac{m_{A,b} - \theta^5_{B,a}\underline{m}_{A,b}}{\underline{n}_{A,b}}\right\},$$

where $\phi(\theta) \coloneqq \max(0, \min(1, \theta))$, $\theta \in [0, 1]$.

Theorem 3 gives the collection of possible solutions to the *fairness-aware* optimization problem. In particular, note that all of these solutions may not be feasible for a particular problem instance. Which of these solutions is feasible and optimal will depend on the true problem parameters. In particular, define $c_{g,s} \coloneqq \sum_{t=1}^{T}(w_{g,s}(t) - w_{g,s'}(t) + u_{g',s}(t) - u_{g',s'}(t))$ and write the objective as $\theta_{A,a}c_{A,a} + \theta_{B,a}c_{B,a}$. As in Proposition 1, the particular solution then depends on the signs and relative magnitudes of $c_{A,a}$ and $c_{B,a}$. For example, if $c_{A,a} \gg c_{B,a} > 0$, then the largest feasible value of $\theta^i_{A,a}$ and the corresponding $\theta^i_{B,a}$ will be the optimal solution; see Figure 9 in § C.2 for an illustration.

The main difference to Proposition 1 is that, due to the imposed fairness constraints, some of the optimal unconstrained solutions might be out of the feasible region. At a higher level, the more restrictive the bounds $\underline{\delta}, \overline{\delta}$ get, the further we move from the optimal binary solution of the fairness-agnostic problem. Thus, some solutions correspond to a mixture of articles shown to each group, and no group is targeted with one article type. However, others may correspond to cases where exactly one group is targeted with only one article, while the other sees both articles at unequal rates. Observe that it is still possible that a group is only shown their out-group article.

Our results offer novel insights for platform design. Even though satisfying (C1) and (C2) imposes a significant restriction on the platform and ostensibly seems to ensure a balanced exposure up to some extent, extreme solutions may still arise. Introducing fairness constraints does not automatically imply that the final outcome is *truly* fair—or even balanced. Furthermore, in any solution i where $\theta^i_{g,s} \in \{0,1\}$ while $0 < \theta^i_{g',s} < 1$, only one group incurs the "price of fairness" whereas the other group, which is targeted with only one article, serves the platform's major goal of maximizing clicks. (Note that, one can verify that half of the solutions in Theorem 3 have this property.) Thus, when

the content is related to sensitive or high-stakes procedures (e.g., a referendum), ensuring fair exposure is not just a technical challenge; if the interventions are not carefully designed (e.g., choosing $\underline{\delta}$, $\overline{\delta}$ thoughtfully), they can lead to unintended outcomes, potentially with legal consequences.

5 Simulations

We use our model to empirically study the effects of different model parameters from real-world click data. Tables 2 and 3 in § D describe the parameters used, such as number of runs, proportional representation of each group, among others. We evaluate four empirical datasets to estimate parameter values from Bakshy et al. [2015] and three examples from Garimella et al. [2017], deferred to § D. Parameter fitting is done by maximum likelihood estimation, and values are shown in Table 2 in § D. For p, we fit a beta distribution and present the parameters α and β here.

Both datasets are publicly available online, and Bakshy et al. [2015] require that no attempt at de-anonymization is made in the personal data. We make no such attempts, and our model does not produce any output that could aid in malicious attempts to de-anonymize the previous data.

5.1 Effect of population-independent parameters

Effect of balanced exposure parameters $\theta_{A,a}$ and $\theta_{B,a}$ We start by studying the effect of different model parameters on the platform's optimization and outcomes. In particular, we focus on the change of $\underline{\delta}$ and $\overline{\delta}$, and its impacts on exposure and click rates, both en masse and across groups.

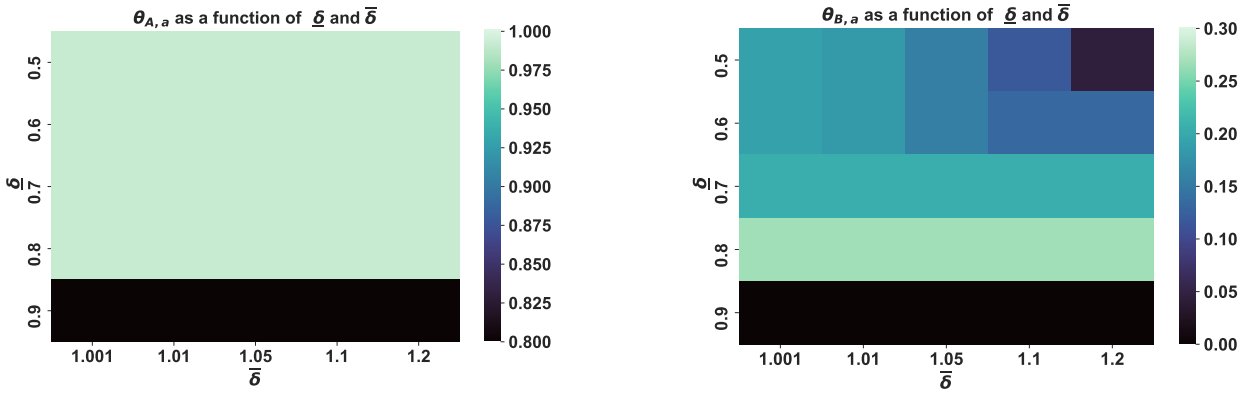

Figure 3: Calculating $\theta_{A,a}$ (left) and $\theta_{B,a}$ (right) as a function of $\underline{\delta}$ and $\overline{\delta}$ with parameters estimated from Bakshy et al. [2015]. Black cells in the bottom row indicate no feasible solution to the fairness-constrained problem.

Figure 3 illustrates that the optimal solution is to almost always show article a to members of group A, and fair exposure is then enforced by restricting how group B is shown articles. In this case, setting $\overline{\delta}$ closer to 1 (making the constraint more restrictive) generally increases the proportion of members of group B who are shown article a. It is helpful to understand when the fairness-aware problem is (i) feasible and (ii) restrictive; if $\underline{\delta}$ and $\overline{\delta}$ are too close to 1, the feasible region may be empty (as in the bottom row of Figure 3), but if they are too far from 1, they might not constrain the fairness-agnostic problem around the agnostic optimum. For intuition on how these parameters may affect the feasible region, see Figure 9 in § C.

Exposure disparity We are also interested in understanding how imposing balanced exposure constraints might affect disparity in expected exposure and clicks. Figure 4 highlights the disparity in exposure for different optimization policies θ; for disparities in engagement see § 5.3.

In Figure 4, we observe that a uniformly randomized policy (*half*) yields a large disparity in article exposure between article a and article b, while this disparity is lower in the fairness-agnostic (*opt*) and fairness-aware (*ratio*) settings. When evaluating differences in how often the articles get *liked* (right), this gap closes across all three policies, while the fairness-aware policy, on average, has the lowest disparity, followed closely by the fairness-agnostic policy.

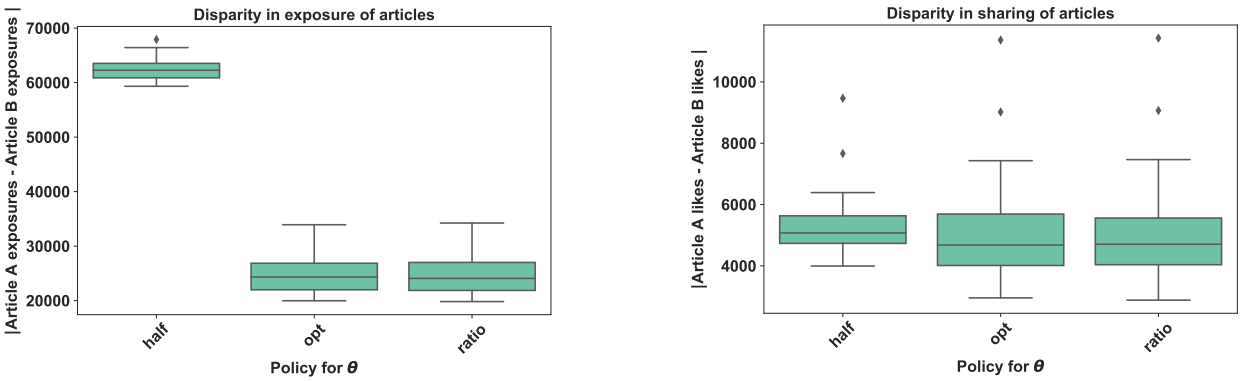

Figure 4: Disparity in exposure, both en masse and across groups.

5.2 Effect of population-based parameters

Population interaction with model: Varying values for clicking (ψ via proxy c/v) While many of the model parameters can be approximated via maximum likelihood estimation on a given dataset, it is difficult to estimate values and costs for clicking on an article. In Figure 5 (left), we show a heatmap for the click rates in the model as a function of $v_{g,s}$ and $v_{g',s}$, where $c_{g,s} = 1$ for all g and s, and preferences are symmetric. Figure 5 (right) shows inter-group click rates on the same axes. Results are averaged over 25 trials.

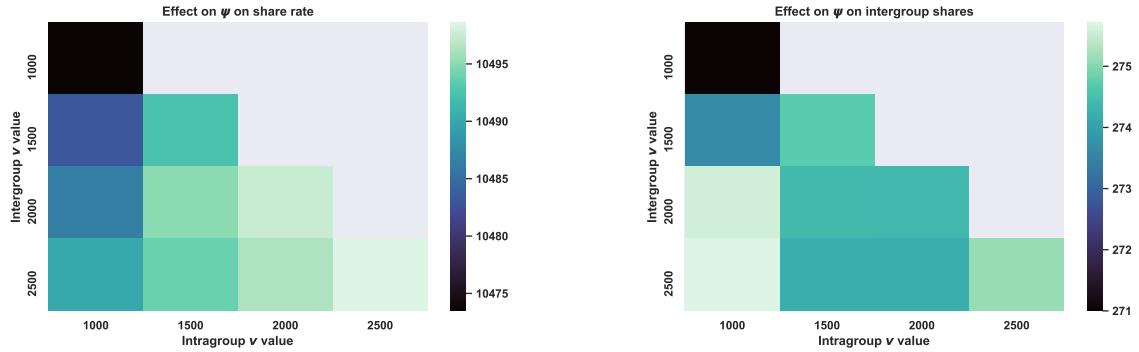

Figure 5: The effect of click valuation $\psi_{g,s}$ impacts the total number of clicks in the model.

In Figure 5 (right), we can see how increasing the inter-group value $v_{g,g'}$ unsurprisingly increases the number of shares, and intra-group $v_{g,g}$ seems to have little effect.

Population interaction with model: Varying homophily constraints Since homophily is such an important and well-studied aspect of information flow in a social network, we also consider the effect of estimating q_A and q_B from empirical datasets, it is worth observing that these variables might vary by topic, even on the same social network.[5] These results are unsurprising: increasing homophily increases total shares, yet decreases inter-group shares, even when increased along both groups simultaneously (Figure 6).

5.3 Engagement disparity

Perhaps unsurprisingly, we can see in Figure 7 (left) that inter-group exposure is significantly higher when randomizing exposure than when optimizing exposure as in the fairness-agnostic (*opt*) and fairness-aware (*ratio*) settings. When evaluating the number of *likes* across groups in Figure 7 (right), this arises as an artifact of the model more generally,

[5]For example, the Brexit dataset of Bakshy et al. [2015] is not very homophilous for one of the positions, and we suspect this might be because opinion on Brexit tends to be correlated with age, but social networks often have many connections across generations.

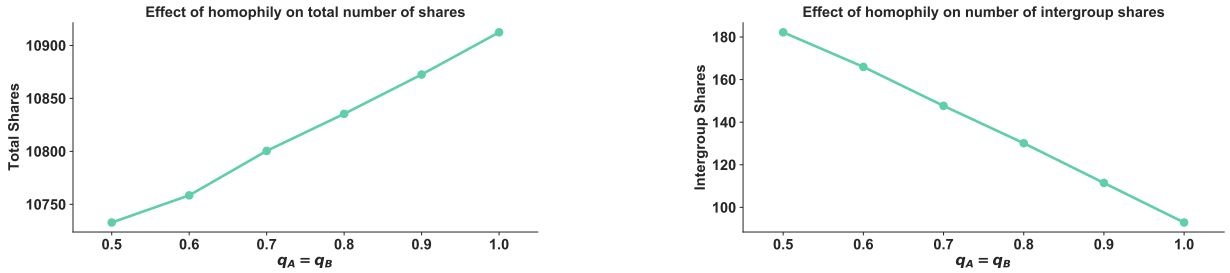

Figure 6: Changing the homophily parameter effects on the number of shares is perhaps very intuitive. Increasing homophily increases total shares, but decreases inter-group shares.

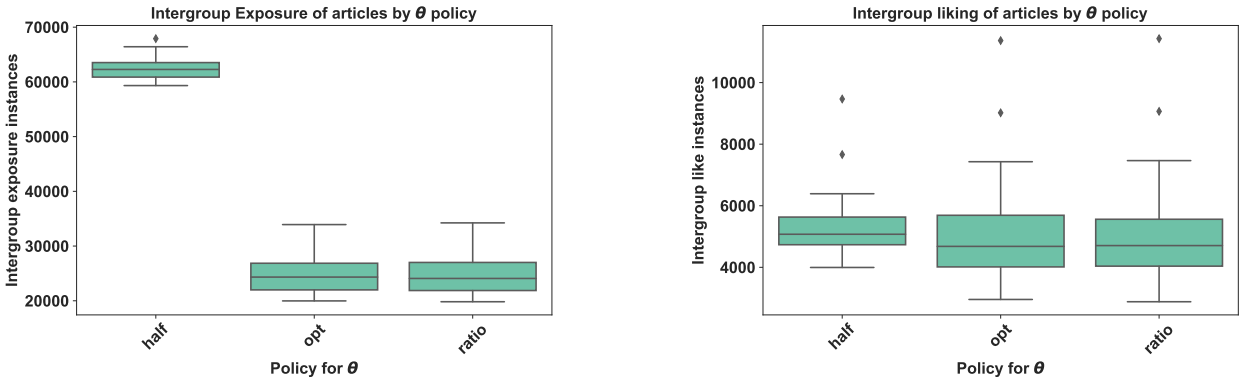

Figure 7: Inter-group clicks and exposure using model parameters from Bakshy et al. [2015].

though the gap significantly decreases. Moreover, we can see here optimizing in fairness-aware and -agnostic settings yield similar distributions of inter-group likes on articles.

Price of fairness We consider the price of fairness similar to that of Bertsimas et al. [2011], given in (10). Here, a lower price of fairness for a given policy is better, as it indicates being closer to the fairness-agnostic optimization problem.

$$POF(\theta) = \frac{\#clicks(\theta_{opt})}{\#clicks(\theta)} \tag{10}$$

We can see that the price of fairness for the fairness-aware optimization problem is close to 1 in most trials, which is observationally lower than the price of fairness for a uniformly randomized policy. Figures 10, 12, and 14 in § D show the price of fairness for adding constraints compared to a uniformly random policy using the parameters estimated from Bakshy et al. [2015]. Figure 8 suggests that our fairness-aware optimization problem yields approximately the same number of clicks as the fairness-agnostic solution.

6 Discussion and conclusion

Motivated by the concerning increase in polarization in social media platforms, this paper introduces the fair exposure problem and develops a theoretical dynamic model to study its implications. Albeit simple and intuitive, our model is highly stylized (as other models in the literature [Papanastasiou, 2020, Allon et al., 2021]). One simplification is the propagation scheme which aims at approximating article sharing and user exposure in a computationally tractable way. Thus, our framework offers novel insights about the propagation in expectation across groups (instead of propagation from individual to individual). Nevertheless, a theoretical analysis using an underlying graph structure would be a natural extension; we study this more realistic scenario through simulations in § E. Another assumption of our model is that each user can see only one article. We make this modeling choice merely for technical simplicity that offers tractability and clearer insights. However, a partial interpretation of this assumption would be that the platform has

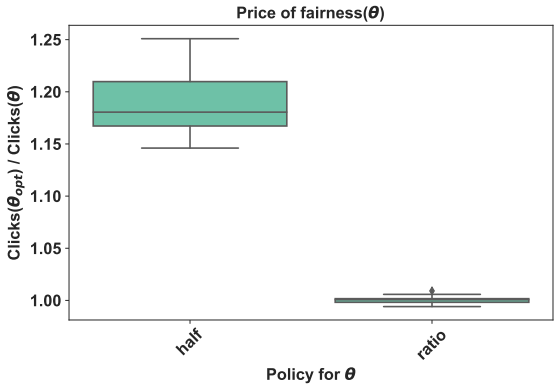

Figure 8: Price of fairness with parameters estimated from Garimella et al. [2017].

limited slots for promoted content or that users most likely click on the first article they see [Robertson and Belkin, 1978, Wang et al., 2013, Craswell et al., 2008].

As the Fairness Doctrine was introduced to ensure opinion diversity, a modern version of this policy could be similarly introduced in online platforms [Pickard, 2021]. Such platform interventions have the great potential to ensure diversity of viewpoints; however, the design of such policies entails the careful examination of any ethical concerns. A question that naturally arises is whether it is ethical for the platform to algorithmically control and potentially randomize the content that a user sees and, ultimately, who—if anyone—has the responsibility to ensure fair exposure in online spaces. This question has been under close scrutiny in interpreting Section 230 of the Communications Decency Act in the United States [Chip Law Group, 2021]. For example, given that news sharing and discussions in social media can determine important political outcomes and thus the passive or more restrictive role that the platform chooses to undertake matters (see, e.g., [British Broadcasting Corporation (BBC), 2020, Isaac and Frenkel, 2020]), it is unclear how a fair representation of content should be defined. Thus, we acknowledge that balancing exposure to different ideologies of content might not actually be fair in a given context. For instance, Bail et al. [2018] suggest that showing people opposing viewpoints makes them more polarized, whereas Becker et al. [2019] show that echo chambers do not necessarily increase polarization. Furthermore, considering the amount of disinformation and the technical challenges in identifying problematic content (e.g., fake news, hate speech) in platforms, the fair exposure constraints should not be applied to all content. Implementing fair exposure can thus become particularly challenging, and more interdisciplinary research is needed to understand where to draw the boundary. Our work is an initial step towards this broader goal.

Finally, our framework highlights how the introduction of fairness constraints can only partially mitigate group-homogeneous targeting and points to problematic outcomes, as sometimes only one group incurs the "price of fairness" while the other pays the "cost of user engagement." It also gives rise to a series of emerging, challenging directions for future research related to platforms and algorithmic fairness. These include the study of fair exposure notions, the design of dynamic interventions and more sophisticated targeting, ad pricing and revenue maximization under fair exposure constraints, and their implications on the competition among different platforms.

Acknowledgements

This project has been part of the MD4SG working group on Bias, Discrimination, and Fairness. The material is based on work supported by the National Science Foundation under Graduate Research Fellowship No. DGE-1650115 (Jessie Finocchiaro). Keziah Naggita was supported in part by the National Science Foundation under Grant No. CCF-1815011 and by the Simons Foundation under the Simons Collaboration on the Theory of Algorithmic Fairness. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding sources.

References

Junaid Ali, Mahmoudreza Babaei, Abhijnan Chakraborty, Baharan Mirzasoleiman, Krishna P Gummadi, and Adish Singla. On the fairness of time-critical influence maximization in social networks. *arXiv preprint arXiv:1905.06618*, 2019.

Gad Allon, Kimon Drakopoulos, and Vahideh Manshadi. Information inundation on platforms and implications. *Operations Research*, 69(6):1784–1804, 2021.

Eshwar Ram Arunachaleswaran, Sampath Kannan, Aaron Roth, and Juba Ziani. Pipeline interventions. In *12th Innovations in Theoretical Computer Science Conference*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2021.

Nicholas A. Ashford. Not on Facebook? You're still likely being fed misinformation. *The New York Times*, March 2021. URL `https://www.nytimes.com/2021/03/29/opinion/misinformation-television-radio.html`.

Christopher A. Bail, Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37):9216–9221, 2018.

Eytan Bakshy, Solomon Messing, and Lada A. Adamic. Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239):1130–1132, 2015.

Abhijit V. Banerjee. A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3):797–817, 1992.

Pablo Barberá, John T Jost, Jonathan Nagler, Joshua A Tucker, and Richard Bonneau. Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, 26(10):1531–1542, 2015.

Joshua Becker, Ethan Porter, and Damon Centola. The wisdom of partisan crowds. *Proceedings of the National Academy of Sciences*, 116(22):10717–10722, 2019.

Dimitris Bertsimas, Vivek F Farias, and Nikolaos Trichakis. The price of fairness. *Operations Research*, 59(1):17–31, 2011.

Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5):992–1026, 1992.

Levi Boxell, Matthew Gentzkow, and Jesse M Shapiro. Is the internet causing political polarization? Evidence from demographics. Technical report, National Bureau of Economic Research, 2017.

British Broadcasting Corporation (BBC). Facebook ad campaign helped Donald Trump win election, claims executive. *BBC News*, January 2020. URL `https://www.bbc.com/news/technology-51034641`.

Ozan Candogan and Kimon Drakopoulos. Optimal signaling of content accuracy: Engagement vs. misinformation. *Operations Research*, 68(2):497–515, 2020.

L Elisa Celis, Sayash Kapoor, Farnood Salehi, and Nisheeth Vishnoi. Controlling polarization in personalization: An algorithmic framework. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pages 160–169, 2019.

Chip Law Group. Understanding the controversy over section 230, January 2021. URL `https://www.lexology.com/library/detail.aspx?g=38784375-b9ba-4c31-8f06-5c30c37534dd`.

Gonzalo Cisternas and Jorge Vásquez. Fake news in social media: A supply and demand approach. *Available at SSRN 3698788*, 2020.

Aaron Clauset and Daniel B Larremore. Random graph models, 2021. URL `https://aaronclauset.github.io/courses/5352/csci5352_F21_L3.pdf`.

Nick Craswell, Onno Zoeter, Michael Taylor, and Bill Ramsey. An experimental comparison of click position-bias models. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, pages 87–94, 2008.

Elizabeth Dubois and Grant Blank. The echo chamber is overstated: The moderating effect of political interest and diverse media. *Information, Communication & Society*, 21(5):729–745, 2018.

Mehrdad Farajtabar, Xiaojing Ye, Sahar Harati, Le Song, and Hongyuan Zha. Multistage campaigning in social networks. In *Advances in Neural Information Processing Systems*, volume 29, pages 4725–4733, 2016.

Benjamin Fish, Ashkan Bashardoust, Danah Boyd, Sorelle Friedler, Carlos Scheidegger, and Suresh Venkatasubramanian. Gaps in information access in social networks? In *The World Wide Web Conference*, pages 480–490, 2019.

Jonathan L Freedman and David O Sears. Selective exposure. In *Advances in Experimental Social Psychology*, volume 2, pages 57–97. Elsevier, 1965.

Kiran Garimella, Aristides Gionis, Nikos Parotsidis, and Nikolaj Tatti. Balancing information exposure in social networks. In *Advances in Neural Information Processing Systems*, volume 30, pages 4663–4671, 2017.

R Kelly Garrett and Paul Resnick. Resisting political fragmentation on the internet. *Daedalus*, 140(4):108–120, 2011.

Nabeel Gillani, Ann Yuan, Martin Saveski, Soroush Vosoughi, and Deb Roy. Me, my echo chamber, and I: Introspection on social media polarization. In *Proceedings of the 2018 World Wide Web Conference*, pages 823–831, 2018.

James Hawdon, Shyam Ranganathan, Shane Bookhultz, and Tanushree Mitra. Social media use, political polarization, and social capital: Is social media tearing the US apart? In *International Conference on Human-Computer Interaction*, pages 243–260. Springer, 2020.

Homa Hosseinmardi, Amir Ghasemian, Aaron Clauset, David M Rothschild, Markus Mobius, and Duncan J Watts. Evaluating the scale, growth, and origins of right-wing echo chambers on YouTube. *arXiv preprint arXiv:2011.12843*, 2020.

Chuan Hu, C. Zhang, Tiejun Wang, and Qing Li. An adaptive recommendation system in social media. *2012 45th Hawaii International Conference on System Sciences*, pages 1759–1767, 2012.

Mike Isaac and Sheera Frenkel. Facebook braces itself for Trump to cast doubt on election results. *The New York Times*, August 2020. URL `https://www.nytimes.com/2020/08/21/technology/facebook-trump-election.html`.

Youngseung Jeon, Bogoan Kim, Aiping Xiong, Dongwon Lee, and Kyungsik Han. Chamberbreaker: Mitigating the echo chamber effect and supporting information hygiene through a gamified inoculation system. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–26, 2021.

David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 137–146, 2003.

Paul F Lazarsfeld and Robert King Merton. *Mass communication, popular taste and organized social action*. Bobbs-Merrill, College Division, 1948.

David M Liu, Zohair Shafi, Will Fleisher, Tina Eliassi-Rad, and Scott Alfeld. RAWLSNET: Altering Bayesian networks to encode Rawlsian fair equality of opportunity. *Available at SSRN 3816196*, 2021.

Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444, 2001.

Subhayan Mukerjee, Kokil Jaidka, and Yphtach Lelkes. The political landscape of the U.S. Twitterverse. *OSF Preprints*, 2020.

Yiangos Papanastasiou. Fake news propagation and detection: A sequential model. *Management Science*, 66(5): 1826–1846, 2020.

Victor Pickard. The Fairness Doctrine won't solve our problems — but it can foster needed debate. *The Washington Post*, February 2021. URL `https://www.washingtonpost.com/outlook/2021/02/04/fairness-doctrine-wont-solve-our-problems-it-can-foster-needed-debate/`.

Stephen E Robertson and Nicholas J Belkin. Ranking in principle. *Journal of Documentation*, 1978.

Kate Starbird, Ahmer Arif, Tom Wilson, Katherine Van Koevering, Katya Yefimova, and Daniel Scarnecchia. Ecosystem or echo-system? Exploring content sharing across alternative media domains. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 12, 2018.

Ana-Andreea Stoica, Jessy Xinyi Han, and Augustin Chaintreau. Seeding network influence in biased networks and the benefits of diversity. In *Proceedings of The Web Conference 2020*, pages 2089–2098, 2020.

US Department of Justice. Thirteen charged in federal court following riot at the United States Capitol, January 2021. URL `https://www.justice.gov/opa/pr/thirteen-charged-federal-court-following-riot-united-states-capitol`.

Chao Wang, Yiqun Liu, Min Zhang, Shaoping Ma, Meihong Zheng, Jing Qian, and Kuo Zhang. Incorporating vertical results into search click models. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 503–512, 2013.

K.F.A. Zoetekouw. A critical analysis of the negative consequences caused by recommender systems used on social media platforms, July 2019. URL `http://essay.utwente.nl/78500/`.

A Proofs for theoretical results in the main text

Proof for Lemma 1

Proof. The proof proceeds by induction. For the base case, let $t = 2$ and note that

$$l_{g,s}(2,\theta) = \theta_{g,s} q_g \pi_g \psi_{g,s}^2 + \theta_{g',s}(1-q_{g'})\pi_{g'}\psi_{g',s}\psi_{g,s}$$
$$l_{g',s}(2,\theta) = \theta_{g,s}(1-q_g)\pi_g\psi_{g',s}\psi_{g,s} + \theta_{g',s} q_{g'}\pi_{g'}\psi_{g',s}^2$$

are strictly increasing linear functions of $\theta_{g,s}$ and $\theta_{g',s}$.

Now suppose for some $t \geq 2$

$$l_{g,s}(t,\theta) = \theta_{g,s} w_1(t) + \theta_{g',s} w_2(t)$$
$$l_{g',s}(t,\theta) = \theta_{g,s} u_1(t) + \theta_{g',s} u_2(t),$$

where $w_1(t), w_2(t), u_1(t), u_2(t) > 0$. It follows from (5) that

$$\begin{aligned} l_{g,s}(t+1,\theta) = \theta_{g,s} &\left(q_g\psi_{g,s}w_1(t) + (1-q_{g'})\psi_{g',s}u_1(t)\right) \\ &+ \theta_{g',s}\left(q_g\psi_{g,s}w_2(t) + (1-q_{g'})\psi_{g',s}u_2(t)\right), \end{aligned}$$

which is a strictly increasing linear function of $\theta_{g,s}$ and $\theta_{g',s}$. We conclude that $l_{g,s}(t,\theta)$ is a strictly increasing linear function of $\theta_{g,s}$ and $\theta_{g',s}$ for all $t \geq 2$. Note that for $t = 1$, $l_{g,s}(1,\theta) = \theta_{g,s}\pi_g\psi_{g,s}$, which is a strictly increasing linear function of $\theta_{g,s}$ and does not depend on $\theta_{g',s}$. The case for $l_{g',s}(t,\theta)$ is similar. This completes the proof. □

Proof for Theorem 1

Proof. We will go through the proof for $w_{g,s}$, noting that the case for $u_{g,s}$ is similar.

We first mention that $w_{g,s}(t)$ is a right-tailed sequence, so we will be using the one-sided $\mathcal{Z}$-transform. This is important because (5) only holds for $t \geq 1$. We first rewrite (5) with different initial conditions so that the relation holds for all $t \geq 0$,

$$w_{g,s}(t) = \psi_{g,s}q_g w_{g,s}(t-1) + \psi_{g',s}(1-q_{g'})w_{g',s}(t-1) + \delta(t)\pi_g\psi_{g,s}. \tag{11}$$

Here, $\delta(t)$ is the Kronecker delta function and we take $w_{g,s}(-1) = w_{g',s}(-1) = 0$. Here we begin the time index at $t = 0$ to use the $\mathcal{Z}$-transform, and the index will later be shifted to start at $t = 1$.

Denote the one-sided $\mathcal{Z}$-transform of (11) as $\mathcal{W}_{g,s}(z)$, which is given by

$$\begin{aligned} \mathcal{W}_{g,s}(z) &= z^{-1}\psi_{g,s}q_g\mathcal{W}_{g,s}(z) + z^{-1}\psi_{g',s}(1-q_{g'})\mathcal{W}_{g',s}(z) + \pi_g\psi_{g,s} \\ &= \frac{\psi_{g',s}(1-q_{g'})z^{-1}}{1-\psi_{g,s}q_g z^{-1}}\mathcal{W}_{g',s}(z) + \frac{\pi_g\psi_{g,s}}{1-\psi_{g,s}q_g z^{-1}}. \end{aligned} \tag{12}$$

Similarly, we write the $\mathcal{Z}$-transform of $w_{g',s}$ using the definition of $w_{g',s}$ analogous to (11):

$$\begin{aligned} \mathcal{W}_{g',s}(z) &= z^{-1}\psi_{g',s}q_{g'}\mathcal{W}_{g',s}(z) + z^{-1}\psi_{g,s}(1-q_g)\mathcal{W}_{g,s}(z) + \pi_{g'}\psi_{g',s} \\ &= \frac{\psi_{g,s}(1-q_g)z^{-1}}{1-\psi_{g',s}q_{g'}z^{-1}}\mathcal{W}_{g,s}(z) + \frac{\pi_{g'}\psi_{g',s}}{1-\psi_{g',s}q_{g'}z^{-1}}. \end{aligned} \tag{13}$$

Substituting (13) into (12) yields

$$\begin{aligned} \mathcal{W}_{g,s}(z) &= \left(1 - \frac{\psi_{g',s}(1-q_{g'})z^{-1}}{1-\psi_{g,s}q_g z^{-1}}\frac{\psi_{g,s}(1-q_g)z^{-1}}{1-\psi_{g',s}q_{g'}z^{-1}}\right)^{-1} \\ &\quad \cdot \left(\frac{\psi_{g',s}(1-q_{g'})z^{-1}}{1-\psi_{g,s}q_g z^{-1}}\frac{\pi_{g'}\psi_{g',s}}{1-\psi_{g',s}q_{g'}z^{-1}} + \frac{\pi_g\psi_{g,s}}{1-\psi_{g,s}q_g z^{-1}}\right) \\ &= \frac{\pi_{g'}\psi_{g',s}^2(1-q_{g'})z^{-1} + \pi_g\psi_{g,s}(1-\psi_{g',s}q_{g'}z^{-1})}{(1-\psi_{g,s}q_g z^{-1})(1-\psi_{g',s}q_{g'}z^{-1}) - \psi_{g,s}\psi_{g',s}(1-q_g)(1-q_{g'})z^{-2}} \\ &= \frac{\pi_g\psi_{g,s} + \left(\pi_{g'}\psi_{g',s}^2(1-q_{g'}) - \pi_g\psi_{g,s}\psi_{g',s}q_{g'}\right)z^{-1}}{1-(\psi_{g,s}q_g + \psi_{g',s}q_{g'})z^{-1} + \psi_{g,s}\psi_{g',s}(q_g+q_{g'}-1)z^{-2}} \\ &= \frac{\pi_g\psi_{g,s} + \psi_{g',s}\left(\pi_{g'}\psi_{g',s}(1-q_{g'}) - \pi_g\psi_{g,s}q_{g'}\right)z^{-1}}{(1-a_{1,s}z^{-1})(1-a_{2,s}z^{-1})}, \end{aligned}$$

where $a_{1,s}$ and $a_{2,s}$ are as defined in the statement of the theorem. At this point, our goal is to take the inverse $\mathcal{Z}$-transform of the preceding expression. We will take the standard approach of inverting the partial fraction expansion of the above, which is given by

$$\mathcal{L}_{g,s} = \frac{A^w_{1,g,s}}{1 - a_{1,s}z^{-1}} + \frac{A^w_{2,g,s}}{1 - a_{2,s}z^{-1}},$$

where $A^w_{1,g,s}$ and $A^w_{2,g,s}$ are as defined in the statement of the theorem. The inverse transform is now easily computed by referencing a standard $\mathcal{Z}$-transform table, yielding

$$w_{g,s}(t) = A^w_{1,g,s}a^t_{1,s} + A^w_{2,g,s}a^t_{2,s}, \quad t \geq 0.$$

To transform back to the case where the sequence begins at $t = 1$, rather than $t = 0$ as we have used in this proof, we simply shift the sequence forward by a single time step, yielding

$$w_{g,s}(t) = A^w_{1,g,s}a^{t-1}_{1,s} + A^w_{2,g,s}a^{t-1}_{2,s}, \quad t \geq 1,$$

which completes the proof. □

Proof for Proposition 1

Proof. From Corollary 2 in § B we can immediately write the platform objective (P) as

$$\sum_{g\in\{A,B\}} \sum_{s\in\{a,b\}} \sum_{t=1}^{T} l_{g,s}(t,\theta) = c_{A,a}\theta_{A,a} + c_{B,a}\theta_{B,a},$$

where

$$c_{A,a} := \sum_{t=1}^{T}(w_{A,a}(t) - w_{A,b}(t) + u_{B,a}(t) - u_{B,b}(t))$$
$$c_{B,a} := \sum_{t=1}^{T}(u_{A,a}(t) - u_{A,b}(t) + w_{B,a}(t) - w_{B,b}(t)).$$

Since the objective (P) is linear in $\theta_{A,a}$ and $\theta_{B,a}$, the optimal $\theta_{A,a}$ is the largest allowable value if $c_{A,a} > 0$, and the smallest allowable value if $c_{A,a} < 0$. The case is similar for $\theta_{B,a}$. This completes the proof. □

Proof for Corollary 1

Proof. By Proposition 2 in § B, the solution to the fairness-agnostic optimization problem (P) can be written as

$$\theta^*_{g,a} = \mathbf{1}\left\{\frac{\psi_{g,a}(z_{1,a} + \psi_{g',a}(1 - q_{g'} - q_g)z_{2,a})}{\psi_{g,b}(z_{1,b} - \psi_{g',b}(1 - q_{g'} - q_g)z_{2,b})} > 1\right\}.$$

Assume that $\theta^*_{A,a} = 0$. Then, we will show that $\theta^*_{B,a} = 0$ as well.

Recall that, by assumption, $\psi_{A,a} > \psi_{A,b}$ and $\psi_{B,b} > \psi_{B,a}$. Therefore, given that $1 - q_A - q_B < 0$ and $z_{1,s}, z_{2,s} > 0$, we get that

$$\begin{aligned} 1 &> \frac{\psi_{A,a}(z_{1,a} + \psi_{B,a}(1 - q_A - q_B)z_{2,a})}{\psi_{A,b}(z_{1,b} - \psi_{B,b}(1 - q_A - q_B)z_{2,b})} = \frac{\psi_{A,a}(z_{1,a} + \psi_{B,a}(1 - q_A - q_B)z_{2,a})}{\psi_{A,b}(z_{1,b} - \psi_{B,b}(1 - q_A - q_B)z_{2,b})} \\ &> \frac{\psi_{B,a}(z_{1,a} + \psi_{A,a}(1 - q_A - q_B)z_{2,a})}{\psi_{B,b}(z_{1,b} - \psi_{A,b}(1 - q_A - q_B)z_{2,b})}. \end{aligned}$$

Thus, $\theta^*_{B,a} = 0$. □

Proof for Lemma 2

Proof. Consider the path representation of the mass computation problem as shown in Figure 1. Since $\theta_{A,a} = 1$ and $\theta_{B,a} = 0$, all A users at time step $t = 1$ are shown article a and all B users article B. As a result, all paths that constitute $l_{A,a}(t,\theta)$ start from node A_1 while all paths that constitute $l_{A,b}(t,\theta)$ start from node B_1; both end at node A_t.

There are two cases. Each such path P of length $t > 2$ can include either node A_2 or B_2.

Case: P includes A_2 Then,

$$\frac{l_{A,a}(t,\theta)}{l_{A,b}(t,\theta)} = \frac{\pi_A q_A}{\pi_B(1-q_B)} \cdot \frac{\psi_{A,a}}{\psi_{B,b}} \left(\frac{\psi_{A,a}}{\psi_{B,a}}\right)^i \left(\frac{\psi_{A,a}}{\psi_{B,a}}\right)^{t-1-i},$$

where $t-1 \geq i \geq 2$ is the number of A nodes in the path (*after* the first node). Observe that if the conditions in (8) hold, then the previous equation also holds.

Case: P includes B_2 Then,

$$\frac{l_{A,a}(t,\theta)}{l_{A,b}(t,\theta)} = \frac{\pi_A(1-q_A)}{\pi_B q_B} \cdot \frac{\psi_{A,a}}{\psi_{B,b}} \left(\frac{\psi_{A,a}}{\psi_{B,a}}\right)^i \left(\frac{\psi_{A,a}}{\psi_{B,a}}\right)^{t-1-i},$$

where $t-1 > i \geq 1$ is the number of A nodes in the path (*after* the first node). Observe that if $\frac{q_A \pi_A}{(1-q_B)\pi_B} < 1$ then $\frac{(1-q_A)\pi_A}{q_B \pi_B} < 1$ must also hold. Thus, if the conditions in (8) hold, then the previous equation also holds.

From the previous two cases, it follows that $l_{A,a}(t,\theta) < l_{A,b}(t,\theta)$. This argument applies to any t. Taking the sum gives the desired result. □

Proof for Lemma 3

Proof. If $\theta_{A,a} \neq 1-\theta_{B,a}$, then the lemma trivially holds. Assume now that the platform manages to target equal fractions $\theta_{A,a} = 1-\theta_{B,a} = e$ at time $t=1$. Thus,

$$\frac{l_{A,s}(1,\theta)}{\pi_A} = \frac{l_{B,s'}(1,\theta)}{\pi_B} = e.$$

The rest of the proof is by induction.

Assume that the platform has achieved equal, constant exposure up to time $t-1$. To achieve the same level of exposure e at time $t \geq 2$, it must hold that

$$\frac{l_{A,s}(t,\theta)}{\pi_A} = \frac{l_{B,s'}(t,\theta)}{\pi_B} = \frac{l_{B,s'}(t,\theta)}{1-\pi_A} = e. \tag{14}$$

Using (14) for $t-1$ (due to our induction hypothesis) in (5), we get that

$$l_{A,s}(t,\theta) = l_{A,s}(t-1,\theta)\left(q_A + \frac{1-\pi_A}{\pi_A}(1-q_B)\right)\psi_{A,s}$$
$$l_{B,s'}(t,\theta) = l_{A,s}(t-1,\theta)\left(\frac{1-\pi_A}{\pi_A}q_B + (1-q_A)\right)\psi_{B,s'}.$$

Thus, (14) is equivalent to

$$\psi_{A,s}\left(q_A + \frac{1-\pi_A}{\pi_A}(1-q_B)\right) = \psi_{B,s'}\left(\frac{1-\pi_A}{\pi_A}q_B + (1-q_A)\right).$$

Therefore, we conclude that equal, constant exposure over time is possible if and only if (9) holds and $\theta_{A,a} = \theta_{B,a}$. □

Proof for Lemma 4 The proof makes use of Lemma 5 in § B, which simply states that the relevant quantities in Theorem 1 are real numbers and that $a_{1,s} > a_{2,s} > 0$.

Proof. Our goal is to show that there exist $\theta_{g,s}, \theta_{g',s}, \pi_g \in (0,1)$ such that, for any $e \in (0,1)$,

$$e = \frac{1}{T}\sum_{t=1}^{T}\frac{l_{g,s}(t,\theta)}{\pi_g} = \frac{1}{T\pi_g}\left(\theta_{g,s}\sum_{t=1}^{T}w_{g,s}(t) + \theta_{g',s}\sum_{t=1}^{T}u_{g,s}(t)\right),$$

where the rightmost expression comes from Corollary 2 in § B.

From Lemma 1 we know that $w_{g,s}(t) > 0$ for all $t \geq 1$, $u_{g,s}(t) > 0$ for all $t \geq 2$, and $u_{g,s}(1) = 0$. Therefore, $\sum_{t=1}^{T} w_{g,s}(t) > 0$, $\sum_{t=1}^{T} u_{g,s}(t) > 0$ and, from the constraint $\theta_{g,s} \in [0,1]$, it is clear that

$$0 \leq e \leq \frac{1}{T\pi_g}\sum_{t=1}^{T}\left(w_{g,s}(t) + u_{g,s}(t)\right). \tag{15}$$

This completes the proof. □

Proof for Theorem 2

Proof. Consider the hyperplanes y_1 to y_4, as introduced in § C.2, which induce the half-spaces containing the feasible region of the fairness-aware optimization problem (P). Note that y_1 and y_4 are upper bounds, and y_2 and y_3 are lower bounds. Additionally, we have $\theta_{A,a}, \theta_{B,a} \in [0, 1]$. Intuitively, the problem is infeasible if upper bounds are smaller than lower bounds, or if lower bounds are greater than upper bounds—each over the unit box. Or, in other words, if the half-spaces induced by two bounding hyperplanes do not intersect over the unit box. Note that y_1 to y_4 all have negative slope, as can be seen in § C.2. Hence the fairness-aware problem will be infeasible in one of the following cases:

Case: y_1 **below** y_2 This is equivalent to $y_1(\theta_{A,a} = 0) < y_2(\theta_{A,a} = 0)$ and $\theta_{A,a}(y_1 = 0) < \theta_{A,a}(y_2 = 0)$. But this cannot happen, as shown in (21) in § C.2.

Case: y_4 **below** y_2 This is equivalent to $y_4(\theta_{A,a} = 0) < y_2(\theta_{A,a} = 0)$ and $\theta_{A,a}(y_4 = 0) < \theta_{A,a}(y_2 = 0)$. Using the formulas from § C.2, we know that this is the case if and only if $\frac{m_{A,b}}{\underline{m}_{A,b}} < \underline{\delta}\frac{m_{B,b}}{\underline{m}_{A,a}}$ and $\frac{m_{A,b}}{\underline{n}_{A,b}} < \underline{\delta}\frac{m_{B,b}}{\underline{n}_{A,a}}$.

Case: y_1 **below** y_3 This is equivalent to $y_1(\theta_{A,a} = 0) < y_3(\theta_{A,a} = 0)$ and $\theta_{A,a}(y_1 = 0) < \theta_{A,a}(y_3 = 0)$. Using the formulas from § C.2, we know that this is the case if and only if $\overline{\delta}\frac{m_{B,b}}{\overline{m}_{A,a}} < \frac{m_{A,b}}{\overline{m}_{A,b}}$ and $\overline{\delta}\frac{m_{B,b}}{\overline{n}_{A,a}} < \frac{m_{A,b}}{\overline{n}_{A,b}}$.

Case: y_4 **below** y_3 This is equivalent to $y_4(\theta_{A,a} = 0) < y_3(\theta_{A,a} = 0)$ and $\theta_{A,a}(y_4 = 0) < \theta_{A,a}(y_3 = 0)$. But this cannot happen, as shown in (20) in § C.2.

Case: y_1 **below** 0 Since y_1 is a line with negative slope, this is the case if and only if $y_1(\theta_{A,a} = 0) < 0$. But this cannot happen because $y_1(\theta_{A,a} = 0) = \overline{\delta}\frac{m_{B,b}}{\overline{m}_{A,a}} \geq 0$ due to $\overline{\delta}, m_{B,b}, \overline{m}_{A,a} \geq 0$ (see Lemma 1 and the definitions of $m_{g,s}$ and $\overline{m}_{g,s}$).

Case: y_4 **below** 0 With the similar argument to the previous case, this cannot happen.

Case: 1 **below** y_2 Note that y_2 can only be "above" 1, i.e., north-east of the unit box, if $\theta_{A,a}(y_2 = 0) > 1$. This is the case if and only if $\underline{\delta} m_{B,b} > \underline{n}_{A,a}$, which in turn is equivalent to $\underline{\delta}\sum_{t=1}^{T} w_{B,b}(t) > \sum_{t=1}^{T} w_{A,a}(t)$. Additionally, for y_2 to be outside the unit box, it must be that $y_2(\theta_{A,a} = 0)$ is large enough such that y_2 runs just above $(\theta_{A,a}, \theta_{B,a}) = (1, 1)$. To determine this critical value, note that a line with $\theta_{A,a}$-axis intercept x and running through $(1, 1)$ has a $\theta_{B,a}$-axis intercept of $\frac{x}{x-1}$. Hence, the second requirement for y_2 to be "above" 1 is that

$$y_2(\theta_{A,a} = 0) = \underline{\delta}\frac{m_{B,b}}{\underline{m}_{A,a}} > \frac{\theta_{A,a}(y_2 = 0)}{\theta_{A,a}(y_2 = 0) - 1} = \underline{\delta}\frac{m_{B,b}}{\underline{\delta} m_{B,b} - \underline{n}_{A,a}}.$$

Case: 1 **below** y_3 With the similar argument to the previous case, we obtain the requirements $\sum_{t=1}^{T} u_{A,b}(t) > \overline{\delta}\sum_{t=1}^{T} u_{B,a}(t)$ and $\frac{m_{A,b}}{\overline{m}_{A,b}} > \frac{m_{A,b}}{m_{A,b} - \overline{n}_{A,b}}$.

This completes the proof as all possible cases for which the fairness-aware problem is infeasible are covered. □

Proof for Theorem 3 We restate the theorem for convenience.

Theorem. *For any admissible problem parameters, the optimal solution to the fairness-aware optimization problem is one of the following:*

$$\theta^1_{A,a} = \phi\left(\frac{m_{A,b}\overline{m}_{A,a} - \overline{\delta} m_{B,b}\underline{m}_{A,b}}{\underline{n}_{A,b}\overline{m}_{A,a} - \overline{n}_{A,a}\underline{m}_{A,b}}\right), \quad \theta^1_{B,a} \in \left\{0, 1, \frac{\overline{\delta} m_{B,b} - \theta^1_{A,a}\overline{n}_{A,a}}{\overline{m}_{A,a}}, \frac{m_{A,b} - \theta^1_{A,a}\underline{n}_{A,b}}{\underline{m}_{A,b}}\right\}$$

$$\theta^2_{A,a} = \phi\left(\frac{m_{A,b}\overline{m}_{A,a} - \overline{\delta} m_{B,b}\overline{m}_{A,b}}{\underline{n}_{A,b}\overline{m}_{A,a} - \overline{n}_{A,a}\overline{m}_{A,b}}\right), \quad \theta^2_{B,a} \in \left\{0, 1, \frac{\overline{\delta} m_{B,b} - \theta^2_{A,a}\overline{n}_{A,a}}{\overline{m}_{A,a}}, \frac{m_{A,b} - \theta^2_{A,a}\overline{n}_{A,b}}{\overline{m}_{A,b}}\right\}$$

$$\theta^3_{A,a} = \phi\left(\frac{m_{A,b}\underline{m}_{A,a} - \underline{\delta} m_{B,b}\underline{m}_{A,b}}{\underline{n}_{A,b}\underline{m}_{A,a} - \underline{n}_{A,a}\underline{m}_{A,b}}\right), \quad \theta^3_{B,a} \in \left\{0, 1, \frac{\underline{\delta} m_{B,b} - \theta^3_{A,a}\underline{n}_{A,a}}{\underline{m}_{A,a}}, \frac{m_{A,b} - \theta^3_{A,a}\underline{n}_{A,b}}{\underline{m}_{A,b}}\right\}$$

$$\theta^4_{A,a} = \phi\left(\frac{m_{A,b}\underline{m}_{A,a} - \underline{\delta} m_{B,b}\overline{m}_{A,b}}{\underline{n}_{A,b}\underline{m}_{A,a} - \underline{n}_{A,a}\overline{m}_{A,b}}\right), \quad \theta^4_{B,a} \in \left\{0, 1, \frac{\underline{\delta} m_{B,b} - \theta^4_{A,a}\underline{n}_{A,a}}{\underline{m}_{A,a}}, \frac{m_{A,b} - \theta^4_{A,a}\overline{n}_{A,b}}{\overline{m}_{A,b}}\right\}$$

$$\theta^5_{B,a} \in \{0,1\}, \quad \theta^5_{A,a} \in \left\{0, 1, \frac{\overline{\delta} m_{B,b} - \theta^5_{B,a}\overline{m}_{A,a}}{\overline{n}_{A,a}}, \frac{\underline{\delta} m_{B,b} - \theta^5_{B,a}\underline{m}_{A,a}}{\underline{n}_{A,a}}, \frac{m_{A,b} - \theta^5_{B,a}\overline{m}_{A,b}}{\overline{n}_{A,b}}, \frac{m_{A,b} - \theta^5_{B,a}\underline{m}_{A,b}}{\underline{n}_{A,b}}\right\},$$

where $\phi(\theta) \coloneqq \max(0, \min(1, \theta))$, $\theta \in [0, 1]$.

Proof. The solution to a linear program with linear inequality constraints will lie on the vertices of the feasible region, which is a polytope. As referenced in the proof of Theorem 2, and discussed in § C.2, the constraints (C1) and (C2), along with the unit box constraint, can be rewritten as the system of linear inequalities

$$\begin{aligned} y_3(\theta_{A,a}) = \frac{m_{A,b}}{\overline{m}_{A,b}} - \theta_{A,a}\frac{\overline{n}_{A,b}}{\overline{m}_{A,b}} \leq & \theta_{B,a} \leq \frac{\overline{\delta} m_{B,b}}{\overline{m}_{A,a}} - \theta_{A,a}\frac{\overline{n}_{A,a}}{\overline{m}_{A,a}} = y_1(\theta_{A,a}), \\ y_2(\theta_{A,a}) = \frac{\underline{\delta} m_{B,b}}{\underline{m}_{A,a}} - \theta_{A,a}\frac{\underline{n}_{A,a}}{\underline{m}_{A,a}} \leq & \theta_{B,a} \leq \frac{m_{A,b}}{\underline{m}_{A,b}} - \theta_{A,a}\frac{\underline{n}_{A,b}}{\underline{m}_{A,b}} = y_4(\theta_{A,a}), \\ 0 \leq & \theta_{A,a}, \theta_{B,a} \leq 1, \end{aligned}$$

where we have used notation introduced in the main paper. The first two inequalities define four half-spaces. The intersection of these half-spaces and the unit box is a polytope defining the feasible region, which we call the constraint polytope. Assume the problem is feasible according to the conditions of Theorem 2. There are two possibilities: (1) the intersection of the half-spaces defined by y_1, y_2, y_3, y_4 is contained in the unit box, or (2) it is only partially contained in the unit box. Since $y_1 > y_2$ and $y_4 > y_3$ on $0 \leq \theta_{A,a} \leq 1$ (as noted in § C.2), in the first case there will be at most four vertices of the constraint polytope. These vertices are solutions to the equations $y_1(\theta_{A,a}) = y_4(\theta_{A,a})$, $y_1(\theta_{A,a}) = y_3(\theta_{A,a})$, $y_2(\theta_{A,a}) = y_4(\theta_{A,a})$, and $y_2(\theta_{A,a}) = y_3(\theta_{A,a})$, respectively given by

$$\theta^1_{A,a} = \frac{m_{A,b}\overline{m}_{A,a} - \overline{\delta} m_{B,b}\underline{m}_{A,b}}{\underline{n}_{A,b}\overline{m}_{A,a} - \overline{n}_{A,a}\underline{m}_{A,b}}, \quad \theta^1_{B,a} = \frac{\overline{\delta} m_{B,b}}{\overline{m}_{A,a}} - \theta^1_{A,a}\frac{\overline{n}_{A,a}}{\overline{m}_{A,a}} = y_1(\theta^1_{A,a}) = y_4(\theta^1_{A,a})$$

$$\theta^2_{A,a} = \frac{m_{A,b}\overline{m}_{A,a} - \overline{\delta} m_{B,b}\overline{m}_{A,b}}{\underline{n}_{A,b}\overline{m}_{A,a} - \overline{n}_{A,a}\overline{m}_{A,b}}, \quad \theta^2_{B,a} = \frac{\overline{\delta} m_{B,b}}{\overline{m}_{A,a}} - \theta^2_{A,a}\frac{\overline{n}_{A,a}}{\overline{m}_{A,a}} = y_1(\theta^2_{A,a}) = y_3(\theta^2_{A,a})$$

$$\theta^3_{A,a} = \frac{m_{A,b}\underline{m}_{A,a} - \underline{\delta} m_{B,b}\underline{m}_{A,b}}{\underline{n}_{A,b}\underline{m}_{A,a} - \underline{n}_{A,a}\underline{m}_{A,b}}, \quad \theta^3_{B,a} = \frac{\underline{\delta} m_{B,b}}{\underline{m}_{A,a}} - \theta^3_{A,a}\frac{\underline{n}_{A,a}}{\underline{m}_{A,a}} = y_2(\theta^3_{A,a}) = y_4(\theta^3_{A,a})$$

$$\theta^4_{A,a} = \frac{m_{A,b}\underline{m}_{A,a} - \underline{\delta} m_{B,b}\overline{m}_{A,b}}{\underline{n}_{A,b}\underline{m}_{A,a} - \underline{n}_{A,a}\overline{m}_{A,b}}, \quad \theta^4_{B,a} = \frac{\underline{\delta} m_{B,b}}{\underline{m}_{A,a}} - \theta^4_{A,a}\frac{\underline{n}_{A,a}}{\underline{m}_{A,a}} = y_2(\theta^4_{A,a}) = y_3(\theta^4_{A,a}),$$

where $(\theta^i_{A,a}, \theta^i_{B,a})$ represents the coordinate of a vertex.

In the second case, if one or more of these vertices lie outside of the unit box, possible vertices exist where the y_i intersect the boundary of the unit box, i.e., where $\theta_{A,a} \in \{0,1\}$ and/or $\theta_{B,a} \in \{0,1\}$. This set might also include vertices of the unit box as vertices of the feasible region. This set of possible solutions is given by

$$\begin{aligned} &\widehat{\theta}_{A,a} \in \{0,1\}, \quad \widehat{\theta}_{B,a} \in \{0, 1, y_1(\widehat{\theta}_{A,a}), y_2(\widehat{\theta}_{A,a}), y_3(\widehat{\theta}_{A,a}), y_4(\widehat{\theta}_{A,a})\}, \\ &\widetilde{\theta}_{B,a} \in \{0,1\}, \quad \widetilde{\theta}_{A,a} \in \{0, 1, y_1^{-1}(\widetilde{\theta}_{B,a}), y_2^{-1}(\widetilde{\theta}_{B,a}), y_3^{-1}(\widetilde{\theta}_{B,a}), y_4^{-1}(\widetilde{\theta}_{B,a})\}. \end{aligned}$$

Collectively the six sets of possible vertices given to this point represent every possible vertex of the constraint set. We can simplify the solution classes by including the cases where one or both of $\theta_{A,a}$ and $\theta_{B,a}$ are 0 or 1. This yields

$$\theta^1_{A,a} = \phi\left(\frac{m_{A,b}\overline{m}_{A,a} - \overline{\delta}m_{B,b}\underline{m}_{A,b}}{\underline{n}_{A,b}\overline{m}_{A,a} - \overline{n}_{A,a}\underline{m}_{A,b}}\right), \quad \theta^1_{B,a} \in \{0, 1, y_1(\theta^1_{A,a}), y_4(\theta^1_{A,a})\}$$
$$\theta^2_{A,a} = \phi\left(\frac{m_{A,b}\overline{m}_{A,a} - \overline{\delta}m_{B,b}\overline{m}_{A,b}}{\underline{n}_{A,b}\overline{m}_{A,a} - \overline{n}_{A,a}\overline{m}_{A,b}}\right), \quad \theta^2_{B,a} \in \{0, 1, y_1(\theta^2_{A,a}), y_3(\theta^2_{A,a})\}$$
$$\theta^3_{A,a} = \phi\left(\frac{m_{A,b}\underline{m}_{A,a} - \underline{\delta}m_{B,b}\underline{m}_{A,b}}{\underline{n}_{A,b}\underline{m}_{A,a} - \underline{n}_{A,a}\underline{m}_{A,b}}\right), \quad \theta^3_{B,a} \in \{0, 1, y_2(\theta^3_{A,a}), y_4(\theta^3_{A,a})\}$$
$$\theta^4_{A,a} = \phi\left(\frac{m_{A,b}\underline{m}_{A,a} - \underline{\delta}m_{B,b}\overline{m}_{A,b}}{\underline{n}_{A,b}\underline{m}_{A,a} - \underline{n}_{A,a}\overline{m}_{A,b}}\right), \quad \theta^4_{B,a} \in \{0, 1, y_2(\theta^4_{A,a}), y_3(\theta^4_{A,a})\},$$
$$\theta^5_{B,a} \in \{0, 1\}, \quad \theta^5_{A,a} \in \{0, 1, y_1^{-1}(\theta^5_{B,a}), y_2^{-1}(\theta^5_{B,a}), y_3^{-1}(\theta^5_{B,a}), y_4^{-1}(\theta^5_{B,a})\}.$$

Plugging in the expressions for y_i and y_i^{-1} gives the result:

$$\theta^1_{A,a} = \phi\left(\frac{m_{A,b}\overline{m}_{A,a} - \overline{\delta}m_{B,b}\underline{m}_{A,b}}{\underline{n}_{A,b}\overline{m}_{A,a} - \overline{n}_{A,a}\underline{m}_{A,b}}\right), \quad \theta^1_{B,a} \in \left\{0, 1, \frac{\overline{\delta}m_{B,b} - \theta^1_{A,a}\overline{n}_{A,a}}{\overline{m}_{A,a}}, \frac{m_{A,b} - \theta^1_{A,a}\underline{n}_{A,b}}{\underline{m}_{A,b}}\right\}$$
$$\theta^2_{A,a} = \phi\left(\frac{m_{A,b}\overline{m}_{A,a} - \overline{\delta}m_{B,b}\overline{m}_{A,b}}{\underline{n}_{A,b}\overline{m}_{A,a} - \overline{n}_{A,a}\overline{m}_{A,b}}\right), \quad \theta^2_{B,a} \in \left\{0, 1, \frac{\overline{\delta}m_{B,b} - \theta^2_{A,a}\overline{n}_{A,a}}{\overline{m}_{A,a}}, \frac{m_{A,b} - \theta^2_{A,a}\overline{n}_{A,b}}{\overline{m}_{A,b}}\right\}$$
$$\theta^3_{A,a} = \phi\left(\frac{m_{A,b}\underline{m}_{A,a} - \underline{\delta}m_{B,b}\underline{m}_{A,b}}{\underline{n}_{A,b}\underline{m}_{A,a} - \underline{n}_{A,a}\underline{m}_{A,b}}\right), \quad \theta^3_{B,a} \in \left\{0, 1, \frac{\underline{\delta}m_{B,b} - \theta^3_{A,a}\underline{n}_{A,a}}{\underline{m}_{A,a}}, \frac{m_{A,b} - \theta^3_{A,a}\underline{n}_{A,b}}{\underline{m}_{A,b}}\right\}$$
$$\theta^4_{A,a} = \phi\left(\frac{m_{A,b}\underline{m}_{A,a} - \underline{\delta}m_{B,b}\overline{m}_{A,b}}{\underline{n}_{A,b}\underline{m}_{A,a} - \underline{n}_{A,a}\overline{m}_{A,b}}\right), \quad \theta^4_{B,a} \in \left\{0, 1, \frac{\underline{\delta}m_{B,b} - \theta^4_{A,a}\underline{n}_{A,a}}{\underline{m}_{A,a}}, \frac{m_{A,b} - \theta^4_{A,a}\overline{n}_{A,b}}{\overline{m}_{A,b}}\right\}$$
$$\theta^5_{B,a} \in \{0, 1\}, \quad \theta^5_{A,a} \in \left\{0, 1, \frac{\overline{\delta}m_{B,b} - \theta^5_{B,a}\overline{m}_{A,a}}{\overline{n}_{A,a}}, \frac{\underline{\delta}m_{B,b} - \theta^5_{B,a}\underline{m}_{A,a}}{\underline{n}_{A,a}}, \frac{m_{A,b} - \theta^5_{B,a}\overline{m}_{A,b}}{\overline{n}_{A,b}}, \frac{m_{A,b} - \theta^5_{B,a}\underline{m}_{A,b}}{\underline{n}_{A,b}}\right\},$$

□

B Additional theoretical results

We first state a corollary of Theorem 1:

Corollary 2.

$$\sum_{t=1}^{T} l_{g,s}(t, \theta) = \theta_{g,s} \sum_{t=1}^{T} w_{g,s}(t) + \theta_{g',s} \sum_{t=1}^{T} u_{g,s}(t)$$

where $w_{g,s}(t)$ and $u_{g,s}(t)$ are as defined in Theorem 1 and do not depend on $\theta_{g,s}$ for any (g, s).

Proof. The proof follows immediately from Theorem 1 by adding over $t = 1, \ldots, T$ and factoring out $\theta_{g,s}$ and $\theta_{g',s}$. □

Lemma 5. *The quantities*

$$\begin{aligned}
a_{1,s} &\coloneqq \frac{1}{2}\big(\psi_{g,s}q_g + \psi_{g',s}q_{g'} + \big[(\psi_{g,s}q_g + \psi_{g',s}q_{g'})^2 - 4\psi_{g,s}\psi_{g',s}(q_g + q_{g'} - 1)\big]^{\frac{1}{2}}\big),\\
a_{2,s} &\coloneqq \psi_{g,s}q_g + \psi_{g',s}q_{g'} - a_{1,s},\\
A^w_{1,g,s} &\coloneqq \frac{\pi_g\psi_{g,s} + \psi_{g',s}\left(\pi_{g'}\psi_{g',s}(1-q_{g'}) - \pi_g\psi_{g,s}q_{g'}\right)a_{1,s}^{-1}}{1 - a_{2,s}a_{1,s}^{-1}},\\
A^w_{2,g,s} &\coloneqq \frac{\pi_g\psi_{g,s} + \psi_{g',s}\left(\pi_{g'}\psi_{g',s}(1-q_{g'}) - \pi_g\psi_{g,s}q_{g'}\right)a_{2,s}^{-1}}{1 - a_{2,s}^{-1}a_{1,s}},\\
A^u_{g,s} &\coloneqq \frac{\psi_{g',s}\left(\pi_{g'}\psi_{g',s}(1-q_{g'}) - \pi_g\psi_{g,s}q_{g'}\right)a_{1,s}^{-1}}{1 - a_{2,s}a_{1,s}^{-1}}
\end{aligned}$$

are real numbers for all $q_g, q_{g'} \in (\frac{1}{2}, 1)$, $\psi_{g,s}, \psi_{g',s} \in (0,1)$. *Furthermore,* $a_{1,s} > a_{2,s} > 0$.

Proof. Recall that

$$\begin{aligned}
a_{1,s} &\coloneqq \frac{1}{2}\big(\psi_{g,s}q_g + \psi_{g',s}q_{g'} + \big[(\psi_{g,s}q_g + \psi_{g',s}q_{g'})^2 - 4\psi_{g,s}\psi_{g',s}(q_g + q_{g'} - 1)\big]^{\frac{1}{2}}\big),\\
a_{2,s} &\coloneqq \psi_{g,s}q_g + \psi_{g',s}q_{g'} - a_{1,s}.
\end{aligned}$$

These values are simply the roots of a quadratic that appears in the proof of Theorem 1. We can rewrite the discriminant

$$(\psi_{g,s}q_g + \psi_{g',s}q_{g'})^2 - 4\psi_{g,s}\psi_{g',s}(q_g + q_{g'} - 1) = q_{g'}(\underbrace{\psi_{g',s} + 2q_g - 4}_{>-3}) + q_g(\underbrace{\psi_{g,s} - 1}_{>-1}) + 4 > 0,$$

where the lower bounds on intermediate quantities follow from allowable values of problem parameters. Therefore, $a_{1,s}$ and $a_{2,s}$ are real. Since the discriminant is strictly positive, we have $a_{1,s} > \frac{1}{2}(\psi_{g,s}q_g + \psi_{g',s}q_{g'})$. By substitution, we immediately have $a_{2,s} < \frac{1}{2}(\psi_{g,s}q_g + \psi_{g',s}q_{g'})$. Therefore, $a_{1,s} > a_{2,s}$. To show positivity, note that

$$\begin{aligned}
a_{1,s} &= \frac{1}{2}\big(\psi_{g,s}q_g + \psi_{g',s}q_{g'} + \big[(\psi_{g,s}q_g + \psi_{g',s}q_{g'})^2 - 4\psi_{g,s}\psi_{g',s}(q_g + q_{g'} - 1)\big]^{\frac{1}{2}}\big)\\
&< \frac{1}{2}\big(\psi_{g,s}q_g + \psi_{g',s}q_{g'} + \big[(\psi_{g,s}q_g + \psi_{g',s}q_{g'})^2\big]^{\frac{1}{2}}\big)\\
&= \psi_{g,s}q_g + \psi_{g',s}q_{g'}.
\end{aligned}$$

Again, by substitution into the definition of $a_{2,s}$ we have $a_{2,s} > 0$. This completes the proof. □

Lemma 6. *For all* $t \geq 1$*, regardless of group* g *and article* s*,*

$$l_{g,s}(t,\theta) = A_{1,g,s}a_{1,s}^{t-1} + A_{2,g,s}a_{2,s}^{t-1}, \quad t \geq 1, \tag{16}$$

where

$$\begin{aligned}
a_{1,s} &\coloneqq \frac{1}{2}\big(\psi_{g,s}q_g + \psi_{g',s}q_{g'} + \big[(\psi_{g,s}q_g + \psi_{g',s}q_{g'})^2 - 4\psi_{g,s}\psi_{g',s}(q_g + q_{g'} - 1)\big]^{\frac{1}{2}}\big),\\
a_{2,s} &\coloneqq \psi_{g,s}q_g + \psi_{g',s}q_{g'} - a_{1,s},\\
A_{1,g,s} &\coloneqq \frac{\theta_{g,s}\pi_g\psi_{g,s} + \psi_{g,s}\psi_{g',s}\left(\theta_{g',s}\pi_{g'}(1-q_{g'}) - \theta_{g,s}\pi_g q_{g'}\right)a_{1,s}^{-1}}{1 - a_{2,s}a_{1,s}^{-1}},\\
A_{2,g,s} &\coloneqq \frac{\theta_{g,s}\pi_g\psi_{g,s} + \psi_{g,s}\psi_{g',s}\left(\theta_{g',s}\pi_{g'}(1-q_{g'}) - \theta_{g,s}\pi_g q_{g'}\right)a_{2,s}^{-1}}{1 - a_{2,s}^{-1}a_{1,s}}.
\end{aligned}$$

Proof. We first mention that $l_{g,s}(t,\theta)$ is a right-tailed sequence, so we will be using the one-sided $\mathcal{Z}$-transform. This is important because (5) only holds for $t \geq 1$. We first rewrite (5) with different initial conditions so that the relation holds for all $t \geq 0$:

$$l_{g,s}(t,\theta) = \psi_{g,s}\left(q_g l_{g,s}(t-1,\theta) + (1-q_{g'})l_{g',s}(t-1,\theta)\right) + \delta(t)\theta_{g,s}\pi_g\psi_{g,s}. \tag{17}$$

Here, $\delta(t)$ is the Kronecker delta function and we take $l_{g,s}(-1,\theta) = l_{g',s}(-1,\theta) = 0$ for all θ.

Denote the one-sided $\mathcal{Z}$-transform of (17) as $\mathcal{L}_{g,s}(z)$, which is given by

$$\begin{aligned}\mathcal{L}_{g,s}(z) &= z^{-1}\psi_{g,s}q_g\mathcal{L}_{g,s}(z) + z^{-1}\psi_{g,s}(1-q_{g'})\mathcal{L}_{g',s}(z) + \theta_{g,s}\pi_g\psi_{g,s} \\ &= \frac{\psi_{g,s}(1-q_{g'})z^{-1}}{1-\psi_{g,s}q_g z^{-1}}\mathcal{L}_{g',s}(z) + \frac{\theta_{g,s}\pi_g\psi_{g,s}}{1-\psi_{g,s}q_g z^{-1}}. \end{aligned} \tag{18}$$

Similarly, we write the $\mathcal{Z}$-transform of $l_{g',s}$ using the definition of $l_{g',s}$ analogous to (17):

$$\begin{aligned}\mathcal{L}_{g',s}(z) &= z^{-1}\psi_{g',s}q_{g'}\mathcal{L}_{g',s}(z) + z^{-1}\psi_{g',s}(1-q_g)\mathcal{L}_{g,s}(z) + \theta_{g',s}\pi_{g'}\psi_{g',s} \\ &= \frac{\psi_{g',s}(1-q_g)z^{-1}}{1-\psi_{g',s}q_{g'}z^{-1}}\mathcal{L}_{g,s}(z) + \frac{\theta_{g',s}\pi_{g'}\psi_{g',s}}{1-\psi_{g',s}q_{g'}z^{-1}}. \end{aligned} \tag{19}$$

Substituting (19) into (18) yields

$$\begin{aligned}\mathcal{L}_{g,s}(z) &= \left(1 - \frac{\psi_{g,s}(1-q_{g'})z^{-1}}{1-\psi_{g,s}q_g z^{-1}}\frac{\psi_{g',s}(1-q_g)z^{-1}}{1-\psi_{g',s}q_{g'}z^{-1}}\right)^{-1} \\ &\quad \cdot \left(\frac{\psi_{g,s}(1-q_{g'})z^{-1}}{1-\psi_{g,s}q_g z^{-1}}\frac{\theta_{g',s}\pi_{g'}\psi_{g',s}}{1-\psi_{g',s}q_{g'}z^{-1}} + \frac{\theta_{g,s}\pi_g\psi_{g,s}}{1-\psi_{g,s}q_g z^{-1}}\right) \\ &= \frac{\theta_{g',s}\pi_{g'}\psi_{g',s}\psi_{g,s}(1-q_{g'})z^{-1} + \theta_{g,s}\pi_g\psi_{g,s}(1-\psi_{g',s}q_{g'}z^{-1})}{(1-\psi_{g,s}q_g z^{-1})(1-\psi_{g',s}q_{g'}z^{-1}) - \psi_{g,s}\psi_{g',s}(1-q_g)(1-q_{g'})z^{-2}} \\ &= \frac{\theta_{g,s}\pi_g\psi_{g,s} + (\theta_{g',s}\pi_{g'}\psi_{g',s}\psi_{g,s}(1-q_{g'}) - \theta_{g,s}\pi_g\psi_{g,s}\psi_{g',s}q_{g'})\,z^{-1}}{1-(\psi_{g,s}q_g + \psi_{g',s}q_{g'})z^{-1} + \psi_{g,s}\psi_{g',s}(q_g+q_{g'}-1)z^{-2}} \\ &= \frac{\theta_{g,s}\pi_g\psi_{g,s} + \psi_{g,s}\psi_{g',s}\,(\theta_{g',s}\pi_{g'}(1-q_{g'}) - \theta_{g,s}\pi_g q_{g'})\,z^{-1}}{(1-a_{1,s}z^{-1})(1-a_{2,s}z^{-1})}, \end{aligned}$$

where $a_{1,s}$ and $a_{2,s}$ are as defined previously. At this point, our goal is to take the inverse $\mathcal{Z}$-transform of the preceding expression. We will take the standard approach of inverting the partial fraction expansion of the above, which is given by

$$\mathcal{L}_{g,s} = \frac{A_{1,g,s}}{1-a_{1,s}z^{-1}} + \frac{A_{2,g,s}}{1-a_{2,s}z^{-1}},$$

where $A_{1,g,s}$ and $A_{2,g,s}$ are as previously defined. The inverse transform is now easily computed by referencing a standard $\mathcal{Z}$-transform table, yielding

$$l_{g,s}(t,\theta) = A_{1,g,s}a_{1,s}^t + A_{2,g,s}a_{2,s}^t, \quad t \geq 0.$$

To transform back to the case where the sequence begins at $t=1$, rather than $t=0$ as we have used in this proof, we simply shift the sequence forward by a single time step, yielding

$$l_{g,s}(t,\theta) = A_{1,g,s}a_{1,s}^{t-1} + A_{2,g,s}a_{2,s}^{t-1}, \quad t \geq 1,$$

which completes the proof. □

Proposition 2. *The solution to the fairness-agnostic optimization problem is*

$$\theta^*_{g,a} = \mathbf{1}\left\{\frac{\psi_{g,a}(z_{1,a} + \psi_{g',a}(1-q_{g'}-q_g)z_{2,a})}{\psi_{g,b}(z_{1,b} - \psi_{g',b}(1-q_{g'}-q_g)z_{2,b})} > 1\right\}$$

for $g \in \{A, B\}$, where

$$\begin{aligned} z_{1,s} &\coloneqq \sum_{t=1}^{T}\sum_{j=0}^{t-1} a_{1,s}^{t-j-1}a_{2,s}^j = \frac{(a_{2,s}-1)a_{1,s}^{T+1} - a_{1,s}a_{2,s}^{T+1} + a_{1,s} + a_{2,s}(a_{2,s}^T - 1)}{(a_{1,s}-1)(a_{2,s}-1)(a_{1,s}-a_{2,s})}, \\ z_{2,s} &\coloneqq \sum_{t=1}^{T}\sum_{j=0}^{t-2} a_{1,s}^{t-j-2}a_{2,s}^j = \frac{(a_{2,s}-1)a_{1,s}^{T} - a_{1,s}a_{2,s}^{T} + a_{1,s} + a_{2,s}(a_{2,s}^{T-1} - 1)}{(a_{1,s}-1)(a_{2,s}-1)(a_{1,s}-a_{2,s})}. \end{aligned}$$

Proof. Using Lemma 6, we find the derivative of the objective $g(\theta_{A,a}, \theta_{B,a}) \coloneqq \sum_{t=1}^{T} \sum_{g\in\{A,B\}} \sum_{s\in\{a,b\}} l_{g,s}(t,\theta)$ with respect to $\theta_{g,a}$:

$$
\begin{aligned}
\frac{\partial g}{\partial \theta_{g,a}} &= \sum_{t=1}^{T} \frac{a_{1,a}^{t-1}}{a_{1,a}-a_{2,a}} \left(\pi_g \psi_{g,a} a_{1,a} - \psi_{g,a}\psi_{g',a}\pi_g q_{g'}\right) + \frac{a_{2,a}^{t-1}}{a_{2,a}-a_{1,a}} \left(\pi_g \psi_{g,a} a_{2,a} - \psi_{g,a}\psi_{g',a}\pi_g q_{g'}\right) \\
&\quad - \frac{a_{1,b}^{t-1}}{a_{1,b}-a_{2,b}} \left(\pi_g \psi_{g,b} a_{1,b} - \psi_{g,b}\psi_{g',b}\pi_g q_{g'}\right) - \frac{a_{2,a}^{t-1}}{a_{2,a}-a_{1,a}} (\pi_g \psi_{g,a} a_{2,a} - \psi_{g,a}\psi_{g',a}\pi_g q_{g'}) \\
&\quad + \frac{a_{1,a}^{t-1}}{a_{1,a}-a_{2,a}} \psi_{g',a}\psi_{g,a}\pi_g(1-q_g) + \frac{a_{2,a}^{t-1}}{a_{2,a}-a_{1,a}} \psi_{g',a}\psi_{g,a}\pi_g(1-q_g) \\
&\quad - \frac{a_{1,b}^{t-1}}{a_{1,b}-a_{2,b}} \psi_{g',b}\psi_{g,b}\pi_g(1-q_g) - \frac{a_{2,b}^{t-1}}{a_{2,b}-a_{1,b}} \psi_{g',b}\psi_{g,b}\pi_g(1-q_g) \\
&= \sum_{t=1}^{T} \frac{\pi_g \psi_{g,a}}{a_{1,a}-a_{2,a}} \left(a_{1,a}^{t} - a_{2,a}^{t} - \psi_{g',a} q_{g'} (a_{1,a}^{t-1} - a_{2,a}^{t-1})\right) - \frac{\pi_g \psi_{g,b}}{a_{1,b}-a_{2,b}} \left(a_{1,b}^{t} - a_{2,b}^{t} - \psi_{g',b} q_{g'} (a_{1,b}^{t-1} - a_{2,b}^{t-1})\right) \\
&\quad + \frac{a_{1,a}^{t-1} - a_{2,a}^{t-1}}{a_{1,a}-a_{2,a}} \psi_{g',a}\psi_{g,a}\pi_g(1-q_g) - \frac{a_{1,b}^{t-1} - a_{2,b}^{t-1}}{a_{1,b}-a_{2,b}} \psi_{g',b}\psi_{g,b}\pi_g(1-q_g) \\
&= \sum_{t=1}^{T} \frac{\pi_g \psi_{g,a}}{a_{1,a}-a_{2,a}} \left(a_{1,a}^{t} - a_{2,a}^{t} + \psi_{g',a}(1-q_{g'}-q_g)(a_{1,a}^{t-1} - a_{2,a}^{t-1})\right) \\
&\quad - \frac{\pi_g \psi_{g,b}}{a_{1,b}-a_{2,b}} \left(a_{1,b}^{t} - a_{2,b}^{t} - \psi_{g',b}(1-q_{g'}-q_g)(a_{1,b}^{t-1} - a_{2,b}^{t-1})\right) \\
&= \sum_{t=1}^{T} \pi_g \psi_{g,a} \left[\sum_{j=0}^{t-1} a_{1,a}^{t-j-1} a_{2,a}^{j} + \psi_{g',a}(1-q_{g'}-q_g) \sum_{j=0}^{t-2} a_{1,a}^{t-j-2} a_{2,a}^{j}\right] \\
&\quad - \pi_g \psi_{g,b} \left[\sum_{j=0}^{t-1} a_{1,b}^{t-j-1} a_{2,b}^{j} - \psi_{g',b}(1-q_{g'}-q_g) \sum_{j=0}^{t-2} a_{1,b}^{t-j-2} a_{2,b}^{j}\right].
\end{aligned}
$$

Given the fact that $\frac{\partial g}{\partial \theta_{g,a}}$ is constant (and thus independent of $\theta_{g,a}, \theta_{g',a}$), it holds that, if $\frac{\partial g}{\partial \theta_{g,a}} > 0$, then $\theta_{g,a} = 1$ is optimal; otherwise, $\theta_{g,a} = 0$ is optimal.

Using the definitions of $z_{1,s}$, $z_{2,s}$, we get that $\frac{\partial g}{\partial \theta_{g,a}} > 0$ if and only if

$$\psi_{g,a}(z_{1,a} + \psi_{g',a}(1-q_{g'}-q_g) z_{2,a}) > \psi_{g,b}(z_{1,b} - \psi_{g',b}(1-q_{g'}-q_g) z_{2,b}).$$

Given that $z_{1,b} - \psi_{g',b}(1-q_{g'}-q_g) z_{2,b} > 0$, the previous condition is equivalent to

$$\frac{\psi_{g,a}(z_{1,a} + \psi_{g',a}(1-q_{g'}-q_g) z_{2,a})}{\psi_{g,b}(z_{1,b} - \psi_{g',b}(1-q_{g'}-q_g) z_{2,b})} > 1.$$

Consequently, the optimal $\theta_{g,a}^{*}$ to the fairness-agnostic optimization problem equals

$$\theta_{g,a}^{*} = \mathbf{1}\left\{\frac{\psi_{g,a}(z_{1,a} + \psi_{g',a}(1-q_{g'}-q_g) z_{2,a})}{\psi_{g,b}(z_{1,b} - \psi_{g',b}(1-q_{g'}-q_g) z_{2,b})} > 1\right\}.$$

This completes the proof. □

C Addendum to the platform optimization problem

C.1 Fairness-aware optimization problem

According to Theorem 1, we have

$$
\begin{aligned}
l_{A,a}(t,\theta) &= \theta_{A,a} w_{A,a}(t) + \theta_{B,a} u_{A,a}(t) \\
l_{A,b}(t,\theta) &= \theta_{A,b} w_{A,b}(t) + \theta_{B,b} u_{A,b}(t) \\
l_{B,a}(t,\theta) &= \theta_{B,a} w_{B,a}(t) + \theta_{A,a} u_{B,a}(t) \\
l_{B,b}(t,\theta) &= \theta_{B,b} w_{B,b}(t) + \theta_{A,b} u_{B,b}(t).
\end{aligned}
$$

Define

$$c_{A,a} := \sum_{t=1}^{T}(w_{A,a}(t) - w_{A,b}(t) + u_{B,a}(t) - u_{B,b}(t))$$
$$c_{B,a} := \sum_{t=1}^{T}(u_{A,a}(t) - u_{A,b}(t) + w_{B,a}(t) - w_{B,b}(t)).$$

Then the fairness-aware optimization problem can be written as

$$\begin{aligned}
\max_{\theta_{A,a},\theta_{B,a}} \quad & c_{A,a}\theta_{A,a} + c_{B,a}\theta_{B,a} \\
\text{s.t.} \quad & \frac{\theta_{A,a}\sum_{t=1}^{T} w_{A,a}(t) + \theta_{B,a}\sum_{t=1}^{T} u_{A,a}(t)}{\theta_{A,b}\sum_{t=1}^{T} u_{B,b}(t) + \theta_{B,b}\sum_{t=1}^{T} w_{B,b}(t)} && \leq \overline{\delta} \\
& \frac{-\theta_{A,a}\sum_{t=1}^{T} w_{A,a}(t) - \theta_{B,a}\sum_{t=1}^{T} u_{A,a}(t)}{\theta_{A,b}\sum_{t=1}^{T} u_{B,b}(t) + \theta_{B,b}\sum_{t=1}^{T} w_{B,b}(t)} && \leq -\underline{\delta} \\
& \frac{\theta_{A,b}\sum_{t=1}^{T} w_{A,b}(t) + \theta_{B,b}\sum_{t=1}^{T} u_{A,b}(t)}{\theta_{A,a}\sum_{t=1}^{T} u_{B,a}(t) + \theta_{B,a}\sum_{t=1}^{T} w_{B,a}(t)} && \leq \overline{\delta} \\
& \frac{(\theta_{A,a}-1)\sum_{t=1}^{T} w_{A,b}(t) + (\theta_{B,a}-1)\sum_{t=1}^{T} u_{A,b}(t)}{\theta_{A,a}\sum_{t=1}^{T} u_{B,a}(t) + \theta_{B,a}\sum_{t=1}^{T} w_{B,a}(t)} && \leq -\underline{\delta} \\
& \theta_{A,a}, \theta_{B,a} && \leq 1 \\
& \theta_{A,a}, \theta_{B,a} && \geq 0.
\end{aligned}$$

This is equivalent to:

$$\begin{aligned}
\max_{\theta_{A,a},\theta_{B,a}} \quad & c_{A,a}\theta_{A,a} + c_{B,a}\theta_{B,a} \\
\text{s.t.} \quad & \theta_{A,a}\left(\sum_{t=1}^{T} w_{A,a}(t) + \overline{\delta}\sum_{t=1}^{T} u_{B,b}(t)\right) + \theta_{B,a}\left(\sum_{t=1}^{T} u_{A,a}(t) + \overline{\delta}\sum_{t=1}^{T} w_{B,b}(t)\right) && \leq \overline{\delta}\sum_{t=1}^{T} u_{B,b}(t) + \overline{\delta}\sum_{t=1}^{T} w_{B,b}(t) \\
& \theta_{A,a}\left(-\underline{\delta}\sum_{t=1}^{T} u_{B,b}(t) - \sum_{t=1}^{T} w_{A,a}(t)\right) + \theta_{B,a}\left(-\underline{\delta}\sum_{t=1}^{T} w_{B,b}(t) - \sum_{t=1}^{T} u_{A,a}(t)\right) && \leq -\underline{\delta}\sum_{t=1}^{T} u_{B,b}(t) - \underline{\delta}\sum_{t=1}^{T} w_{B,b}(t) \\
& \theta_{A,a}\left(-\sum_{t=1}^{T} w_{A,b}(t) - \overline{\delta}\sum_{t=1}^{T} u_{B,a}(t)\right) + \theta_{B,a}\left(-\sum_{t=1}^{T} u_{A,b}(t) - \overline{\delta}\sum_{t=1}^{T} w_{B,a}(t)\right) && \leq -\sum_{t=1}^{T} w_{A,b}(t) - \sum_{t=1}^{T} u_{A,b}(t) \\
& \theta_{A,a}\left(\sum_{t=1}^{T} w_{A,b}(t) + \underline{\delta}\sum_{t=1}^{T} u_{B,a}(t)\right) + \theta_{B,a}\left(\sum_{t=1}^{T} u_{A,b}(t) + \underline{\delta}\sum_{t=1}^{T} w_{B,a}(t)\right) && \leq \sum_{t=1}^{T} w_{A,b}(t) + \sum_{t=1}^{T} u_{A,b}(t) \\
& \theta_{A,a}, \theta_{B,a} && \leq 1 \\
& \theta_{A,a}\theta_{B,a} && \geq 0.
\end{aligned}$$

C.2 Fairness constraints

From constraints (C1) and (C2) and using Theorem 1, we infer the following feasible bounds on $\theta_{B,a}$ (dependent on $\theta_{A,a}$), in addition to $\theta_{A,a}, \theta_{B,a} \in [0,1]$:

$$\theta_{B,a} \leq \frac{\overline{\delta}\sum_{t=1}^{T} u_{B,b}(t) + \overline{\delta}\sum_{t=1}^{T} w_{B,b}(t)}{\sum_{t=1}^{T} u_{A,a}(t) + \overline{\delta}\sum_{t=1}^{T} w_{B,b}(t)} - \theta_{A,a}\frac{\overline{\delta}\sum_{t=1}^{T} u_{B,b}(t) + \sum_{t=1}^{T} w_{A,a}(t)}{\sum_{t=1}^{T} u_{A,a}(t) + \overline{\delta}\sum_{t=1}^{T} w_{B,b}(t)} =: y_1$$

$$\theta_{B,a} \geq \frac{\underline{\delta}\sum_{t=1}^{T} u_{B,b}(t) + \underline{\delta}\sum_{t=1}^{T} w_{B,b}(t)}{\sum_{t=1}^{T} u_{A,a}(t) + \underline{\delta}\sum_{t=1}^{T} w_{B,b}(t)} - \theta_{A,a}\frac{\underline{\delta}\sum_{t=1}^{T} u_{B,b}(t) + \sum_{t=1}^{T} w_{A,a}(t)}{\sum_{t=1}^{T} u_{A,a}(t) + \underline{\delta}\sum_{t=1}^{T} w_{B,b}(t)} =: y_2$$

$$\theta_{B,a} \geq \frac{\sum_{t=1}^{T} u_{A,b}(t) + \sum_{t=1}^{T} w_{A,b}(t)}{\sum_{t=1}^{T} u_{A,b}(t) + \overline{\delta}\sum_{t=1}^{T} w_{B,a}(t)} - \theta_{A,a}\frac{\overline{\delta}\sum_{t=1}^{T} u_{B,a}(t) + \sum_{t=1}^{T} w_{A,b}(t)}{\sum_{t=1}^{T} u_{A,b}(t) + \overline{\delta}\sum_{t=1}^{T} w_{B,a}(t)} =: y_3$$

$$\theta_{B,a} \leq \frac{\sum_{t=1}^{T} u_{A,b}(t) + \sum_{t=1}^{T} w_{A,b}(t)}{\sum_{t=1}^{T} u_{A,b}(t) + \underline{\delta}\sum_{t=1}^{T} w_{B,a}(t)} - \theta_{A,a}\frac{\underline{\delta}\sum_{t=1}^{T} u_{B,a}(t) + \sum_{t=1}^{T} w_{A,b}(t)}{\sum_{t=1}^{T} u_{A,b}(t) + \underline{\delta}\sum_{t=1}^{T} w_{B,a}(t)} =: y_4.$$

Using the substitutions introduced in the main paper, we then obtain

$$y_1 = \overline{\delta}\frac{m_{B,b}}{\overline{m}_{A,a}} - \theta_{A,a}\frac{\overline{n}_{A,a}}{\overline{m}_{A,a}}$$

$$y_2 = \underline{\delta}\frac{m_{B,b}}{\underline{m}_{A,a}} - \theta_{A,a}\frac{\underline{n}_{A,a}}{\underline{m}_{A,a}}$$

$$y_3 = \frac{m_{A,b}}{\overline{m}_{A,b}} - \theta_{A,a}\frac{\overline{n}_{A,b}}{\overline{m}_{A,b}}$$

$$y_4 = \frac{m_{A,b}}{\underline{m}_{A,b}} - \theta_{A,a}\frac{\underline{n}_{A,b}}{\underline{m}_{A,b}}.$$

Note that the bounding hyperplanes y_1 to y_4 all have negative slope. In a 2-dimensional plot, the axes intersects of those hyperplanes are given as follows:

$$y_1(\theta_{A,a}=0) = \frac{\sum_{t=1}^{T} u_{B,b}(t) + \sum_{t=1}^{T} w_{B,b}(t)}{\frac{1}{\overline{\delta}}\sum_{t=1}^{T} u_{A,a}(t) + \sum_{t=1}^{T} w_{B,b}(t)}, \quad \theta_{A,a}(y_1=0) = \frac{\sum_{t=1}^{T} u_{B,b}(t) + \sum_{t=1}^{T} w_{B,b}(t)}{\sum_{t=1}^{T} u_{B,b}(t) + \frac{1}{\overline{\delta}}\sum_{t=1}^{T} w_{A,a}(t)}$$

$$y_2(\theta_{A,a}=0) = \frac{\sum_{t=1}^{T} u_{B,b}(t) + \sum_{t=1}^{T} w_{B,b}(t)}{\frac{1}{\underline{\delta}}\sum_{t=1}^{T} u_{A,a}(t) + \sum_{t=1}^{T} w_{B,b}(t)}, \quad \theta_{A,a}(y_2=0) = \frac{\sum_{t=1}^{T} u_{B,b}(t) + \sum_{t=1}^{T} w_{B,b}(t)}{\sum_{t=1}^{T} u_{B,b}(t) + \frac{1}{\underline{\delta}}\sum_{t=1}^{T} w_{A,a}(t)}$$

$$y_3(\theta_{A,a}=0) = \frac{\sum_{t=1}^{T} u_{A,b}(t) + \sum_{t=1}^{T} w_{A,b}(t)}{\sum_{t=1}^{T} u_{A,b}(t) + \overline{\delta}\sum_{t=1}^{T} w_{B,a}(t)}, \quad \theta_{A,a}(y_3=0) = \frac{\sum_{t=1}^{T} u_{A,b}(t) + \sum_{t=1}^{T} w_{A,b}(t)}{\overline{\delta}\sum_{t=1}^{T} u_{B,a}(t) + \sum_{t=1}^{T} w_{A,b}(t)}$$

$$y_4(\theta_{A,a}=0) = \frac{\sum_{t=1}^{T} u_{A,b}(t) + \sum_{t=1}^{T} w_{A,b}(t)}{\sum_{t=1}^{T} u_{A,b}(t) + \underline{\delta}\sum_{t=1}^{T} w_{B,a}(t)}, \quad \theta_{A,a}(y_4=0) = \frac{\sum_{t=1}^{T} u_{A,b}(t) + \sum_{t=1}^{T} w_{A,b}(t)}{\underline{\delta}\sum_{t=1}^{T} u_{B,a}(t) + \sum_{t=1}^{T} w_{A,b}(t)}.$$

As a sanity check, note that as we let $\overline{\delta} \to \infty$ and $\underline{\delta} \to 0$ (and using Lemma 1), the axes intersects of y_1 and y_4 will become greater or equal to 1, whereas the axes intersects of y_2 and y_3 will go to 0, rendering constraints (C1) and (C2) redundant, as desired.

We further see that

$$\begin{aligned} &y_4(\theta_{A,a}=0) > y_3(\theta_{A,a}=0) \\ &\theta_{A,a}(y_4=0) > \theta_{A,a}(y_3=0) \end{aligned} \tag{20}$$

as well as

$$\begin{aligned} &y_1(\theta_{A,a}=0) > y_2(\theta_{A,a}=0) \\ &\theta_{A,a}(y_1=0) > \theta_{A,a}(y_2=0). \end{aligned} \tag{21}$$

Figure 9 depicts exemplary non-empty feasible regions of the fairness-aware optimization problem for all eight possible qualitative positions of lower-bounding hyperplanes y_2 and y_3 in combination with the upper-bound constraint induced by y_1. Given y_1, y_2, y_3, the second upper-bounding hyperplane y_4 can take on any position such that (20) is satisfied—otherwise, the feasible set will be empty. Figure 9 also shows in dotted red the trajectory of the optimal solution to the fairness-aware problem for the given example and the case of $c_{A,a} > c_{B,a} > 0$, as a function of the position of y_4. Note that the optimal solution in this case will involve the largest feasible value of $\theta_{A,a}$.

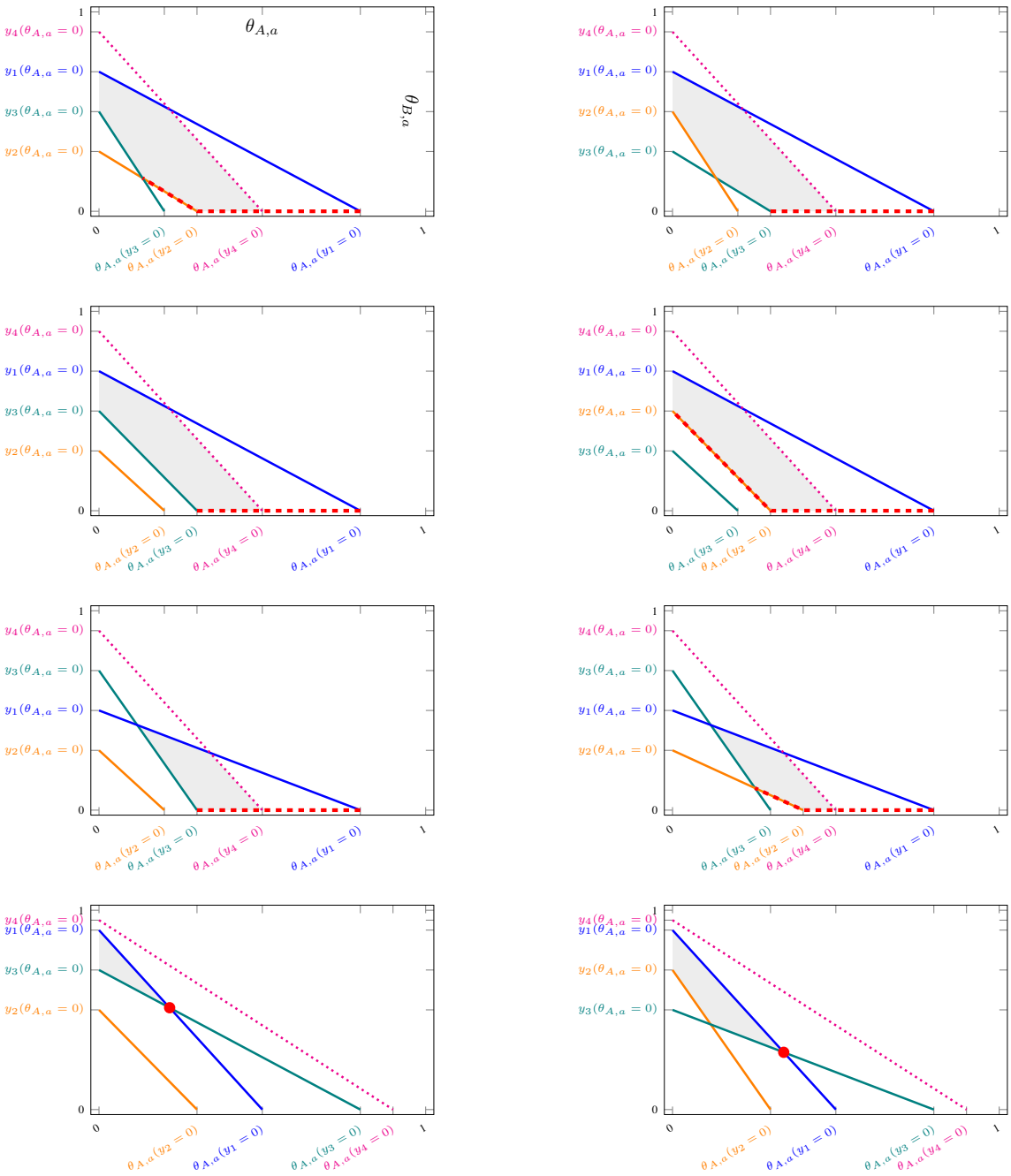

Figure 9: Exemplary feasible regions (shaded gray) of the fairness-aware optimization problem for all (except ones with empty feasible regions) possible qualitative positions of lower-bound constraints (in orange and teal) in combination with the blue upper-bound constraint. Indicated by the red dashed line (or dot) is the trajectory of the optimal solution if $c_{A,a} > c_{B,a} > 0$, depending on the position of the second upper-bound constraint in dotted magenta.

Table 2: Estimated model parameters from empirical datasets.

Variable	TWITTER: US ELECTIONS	TWITTER: BREXIT	TWITTER: ABORTION	FACEBOOK
(π_A, π_B)	(0.432, 0.567)	(0.480, 0.520)	(0.623, 0.370)	(0.500, 0.500)
(q_A, q_B)	(0.9877, 1.0000)	(0.6800, 0.3840)	(0.5500, 0.8200)	(0.7200, 0.6800)
$p_{A,a}$ params: (α, β)	(41.46, 556.87)	(1.64, 62.92)	(2.30, 27.59)	(0.95, 1.35)
$p_{A,b}$ params: (α, β)	(0.75, 413.47)	(1.72, 380.14)	(0.16, 50.83)	(0.18, 2.76)
$p_{B,a}$ params: (α, β)	(6.10, 1519.85)	(1.48, 27.40)	(0.25, 7.40)	(0.10, 3.09)
$p_{B,b}$ params: (α, β)	(2153.00, 23467.67)	(39.60, 505.90)	(2.20, 53.70)	(0.88, 1.62)

Table 3: Default network-independent parameters for experiments, unless otherwise changed.

Variable	Value
Number of trials	25
T	10
n	10^5
$c_{g,s}$	1
$v_{g,g}$	2000
$v_{g,g'}$	200
$\underline{\delta}$	$1/4$
$\overline{\delta}$	2

D Parameter values from datasets

Estimated parameters from empirical datasets are shown in Table 2. Moreover, some of our model parameters are network-independent. In that case, some of those default parameters are in Table 3.

D.1 Deferred experimental results

Garimella et al. [2017] studies user engagement with articles both inside and outside their preferred group across 14 different topics. They found a stronger preference for like-minded articles in political topics than non-political (e.g., rooting for a football team to win the Super Bowl), and we estimate our model parameters by maximum likelihood estimation from their dataset. We provide the code at this link: `https://github.com/jfinocchiaro/fair-exposure`.

Garimella et al. [2017] collected data on publicly available Tweets via the Twitter streaming API. Consent is culturally assumed by Tweets being publicly available, though it is worth noting that this dataset being publicly available requires effort on the public's behalf to (i) ascertain knowledge of their presence in the dataset, and (ii) a request to be removed from the dataset does not prevent previous users of the dataset from having their Tweets. Part of the dataset is available at `https://github.com/gvrkiran/BalancedExposure#readme` without license. We only use publicly available data.

D.1.1 Twitter: US Elections [Garimella et al., 2017]

The first topic of engagement studied is on articles about presidential candidates in the 2016 US presidential election. Figure 10 shows the price of fairness for a uniformly randomized policy and fairness-aware optimized policy respectively; the price of fairness for the fairness-aware optimization is close to 1, meaning that performance (measured by number of clicks) of this policy is nearly optimal. Figure 11 shows inter-group and intra-group disparity in exposure and clicks on articles. In line with the price of fairness, we can observe that the fairness-aware and fairness-agnostic optimization problems yield similar exposure and clicks. This might be an artifact of the fairness parameters $\overline{\delta}$ and $\underline{\delta}$ being too non-restrictive.

D.1.2 Twitter: Brexit [Garimella et al., 2017]

The next topic studied is on articles about Brexit[6]. Figure 12 shows the price of fairness for a uniformly randomized policy and fairness-aware optimized policy respectively; here, a randomized policy performs closer to optimal than a

[6] `https://www.government.nl/topics/brexit/question-and-answer/what-is-brexit`

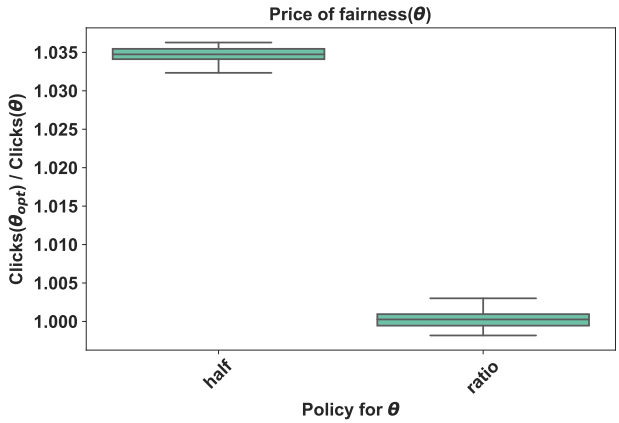

Figure 10: Price of fairness with parameters estimated from Garimella et al. [2017]: US Elections.

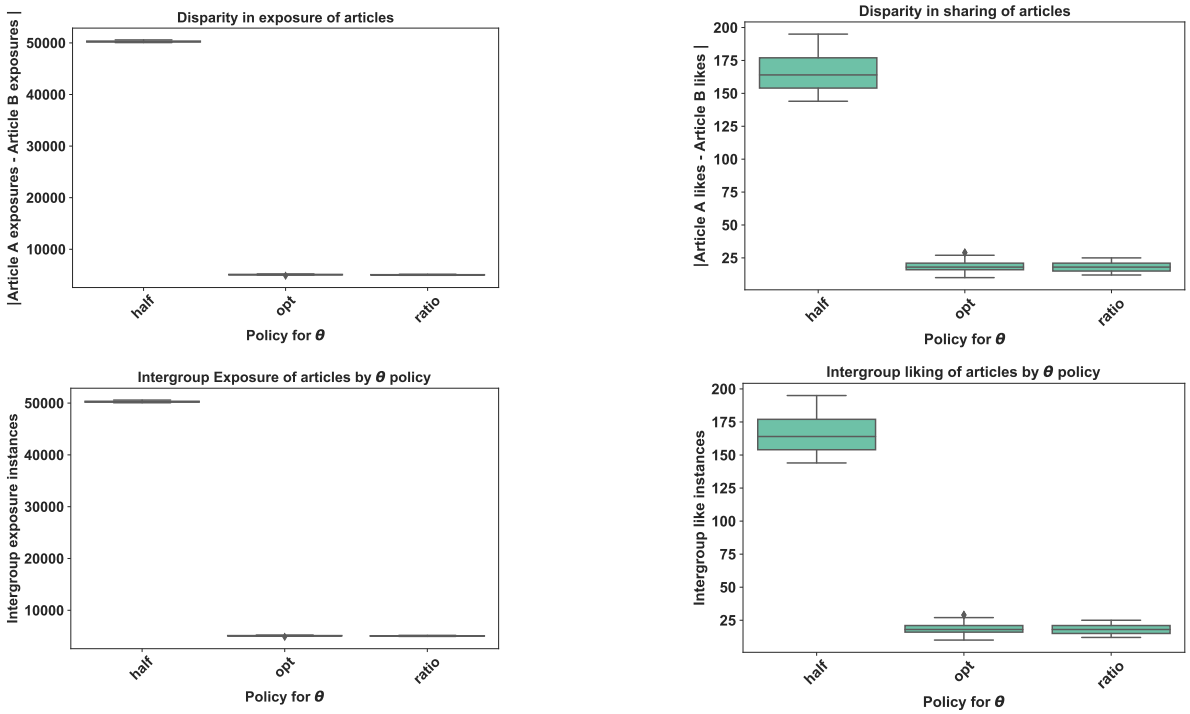

Figure 11: Disparity in clicks and exposure, both en masse and across groups using model parameters from Garimella et al. [2017]: US Elections.

fairness-aware policy. Figure 13 shows inter-group and intra-group disparity in exposure and clicks on articles. We conjecture these results might be this way because of the generational divide guiding many opinions on Brexit instead of a solely political one, so the homophily variable q_B (see Table 2) is too low for one of the groups, and therefore groups change more often than not.

D.1.3 Twitter: Abortion [Garimella et al., 2017]

The final topic of engagement studied is on articles about access to abortion services. Figure 14 shows the price of fairness for a uniformly randomized policy and fairness-aware optimized policy respectively; the price of fairness for the fairness-aware optimization is close to 1, though noticably above it, meaning that performance (measured by number of clicks) of this policy is nearly optimal but still constrained. Figure 15 shows inter-group and intra-group

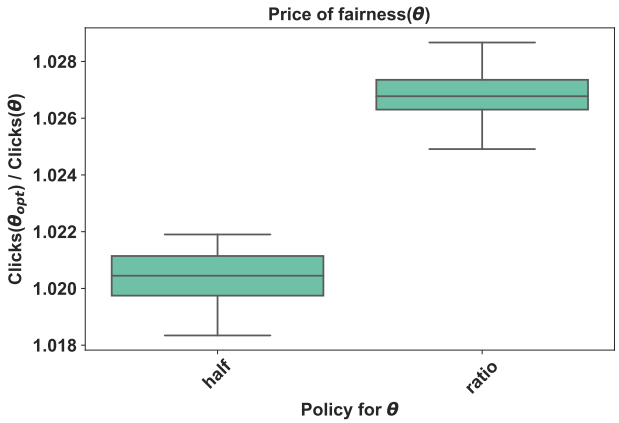

Figure 12: Price of fairness with parameters estimated from Garimella et al. [2017]: Brexit.

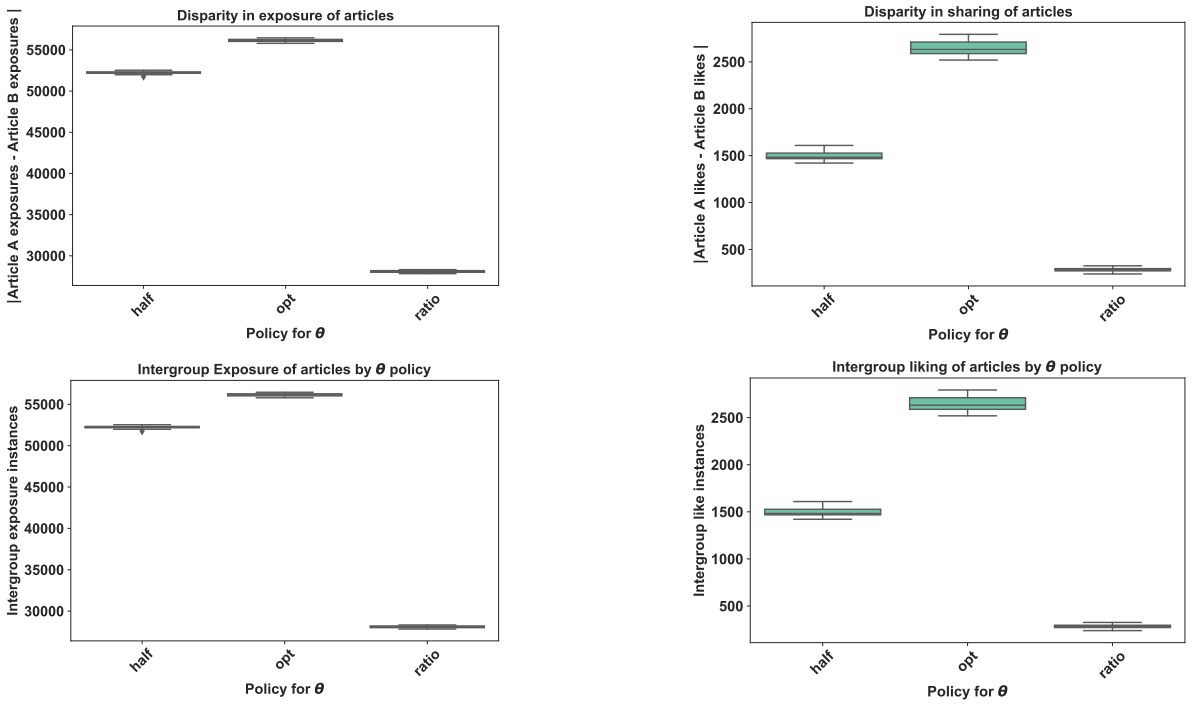

Figure 13: Disparity in clicks and exposure, both en masse and across groups, for Garimella et al. [2017]: Brexit.

disparity in exposure and clicks on articles. In line with the price of fairness, we can observe that the fairness-aware and fairness-agnostic optimization problems yield similar exposure and clicks.

E Comparing our model with a graph model

Even though most social network platforms allow users to broadcast content to a large subset of their *friends*, our model imposes that a user can only share an article with one friend at a time. To understand the implications of this assumption, we compare the propagation of articles in our model with that of a network where the assumption does not hold.

We used the Abortion dataset [Garimella et al., 2017] for this experiment—the median number of friends of a user in this dataset is 27. In the network setting, we allow every user in the dataset to share the article that they read with all of their friends (adjacent nodes). In the model setting, we uniformly select one friend with whom to share the article, while controlling for all the other parameters of the model (e.g., π_g, $p_{g,s}$, $\theta_{g,s}$, M). We set these parameters to the

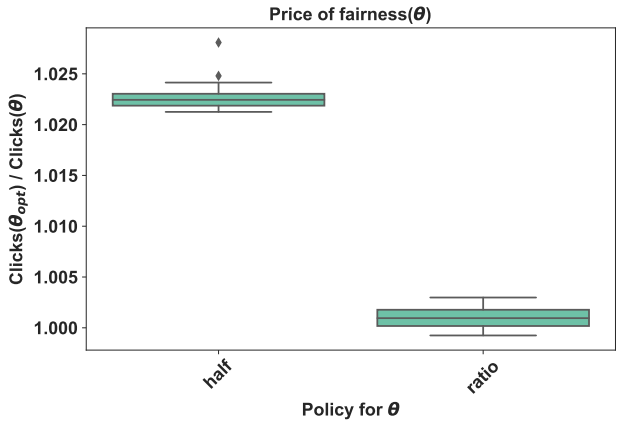

Figure 14: Price of fairness with parameters estimated from Garimella et al. [2017]: Abortion.

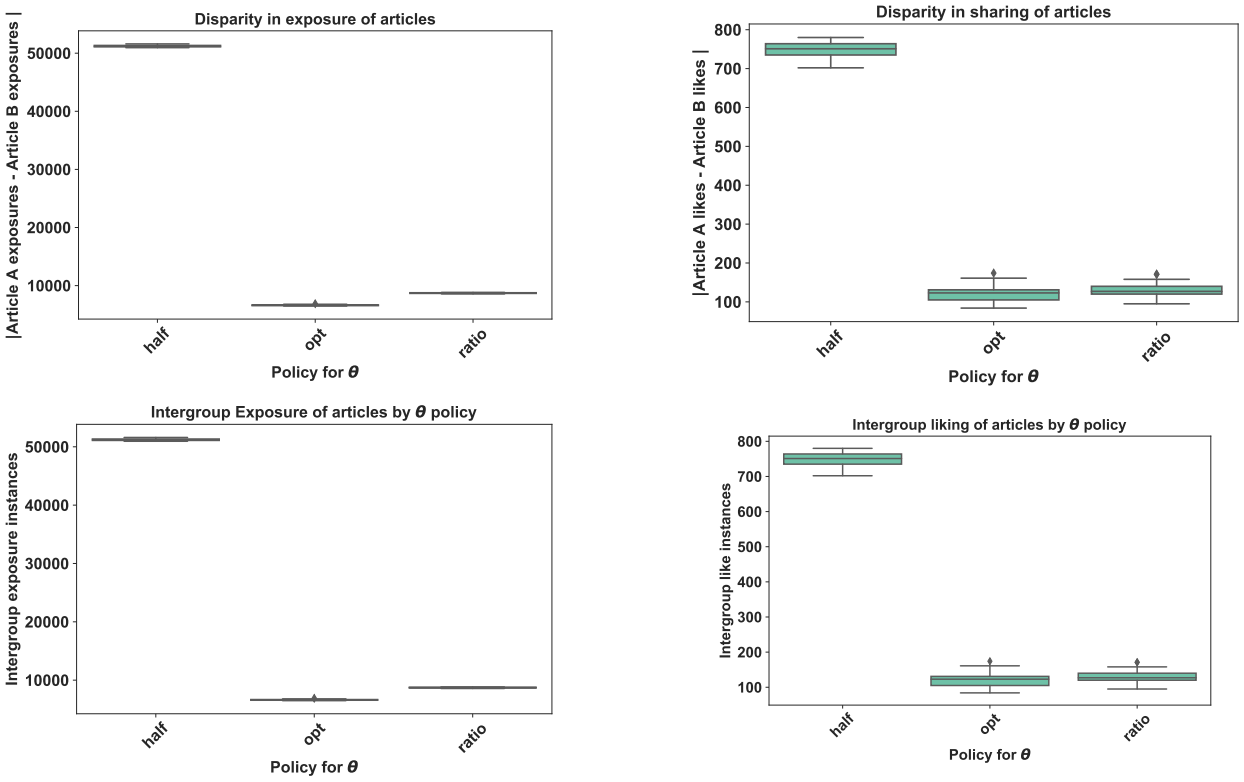

Figure 15: Disparity in clicks and exposure, both en masse and across groups, for Garimella et al. [2017]: Abortion.

values in Table 3 and propagate two contrasting articles for both groups of the dataset. We repeat this experiment 100 times.

Figure 16 shows the number of users in the system over time in the network model (first row) compared to our model (second row) for different inital exposures (columns). It is not surprising that the number of users in the network model does not decrease monotonically as, for each user, we may have several friends sharing the article. By contrast, in our model, the article can only be shared—or not shared—by only one friend. As a result, the article campaigns die out much faster in our model than in the network.

Even though for specific time steps, we see large differences between the network and our model, some of the relative trends that we observe for our model in § 5 hold for the network model. In the top row, we see that the initial exposure levels found by solving the three optimization problems also apply to the network: the groups have similar numbers

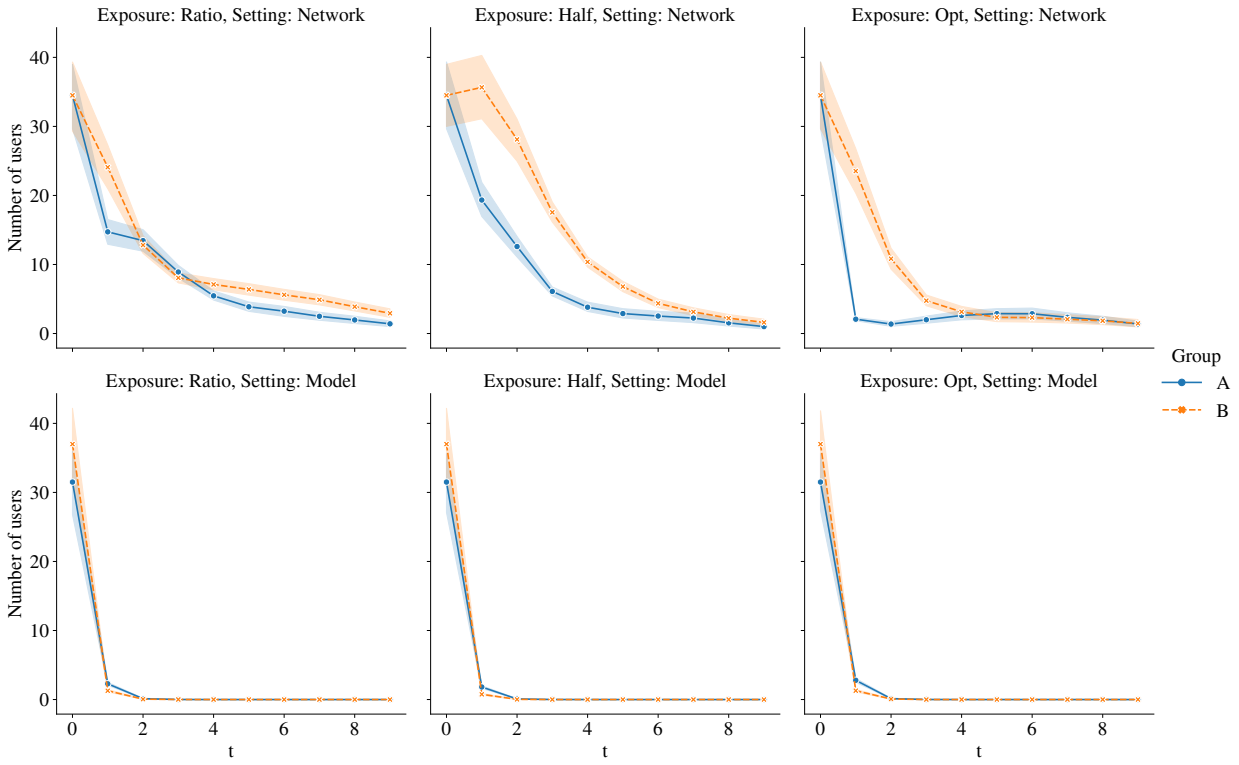

Figure 16: Number of users sharing an article at each time step t in the network model (top) and our model (bottom). Each column is a different exposure level: the first, second and third columns are the solutions for the *ratio*, *half*, and *opt* optimization problems, respectively. The error bands are the 95% confidence intervals of the number of users.

of users for all time steps in the *ratio* exposure (first column), while optimizing for engagement (third column) leads to the lowest number of users for the least homophilous group (A).

Furthermore, in aggregate, we find fewer users that share articles in group A than B for any given exposure level, which is a consequence of the higher homophily of group B in this dataset. However, homophily has a greater effect on the network model, as shown by the greater number of users in group B for the *half* and *opt* exposures. Finally, in § D.1.3, we see that the total expected number of shares for *half* exposure is significantly higher than the other exposure levels, and so is in the network model.

To conclude, even though we do not accurately model the propagation of the articles for each propagation time step, our model may be helpful to study the relative effects of homophily in expectation. In addition, the initial exposure obtained by solving the fair exposure optimization problem (*ratio*) in our model also achieves balanced exposure in the network model.