UNIVERSITAT DE BARCELONA
SCHOOL OF ECONOMICS

Working Papers

Col·lecció d'Economia E24/462

THE DATA ECONOMY AND POLARIZATION ON SOCIAL MEDIA

Goonj Mohan

ISSN 1136-8365

UB Economics *Working Paper No. 462*

Title: The Data Economy and Polarization on Social Media

Abstract: This paper studies the incentive of a social media platform (SMP) to increase polarization of its user network. I propose a two-group network model where the SMP earns revenue from user-data driven personalization. The objective of the SMP is to maximize the amount of valuable data generated. To this end, it relies on an algorithm that, at a cost, encourages users to form new links. Within a microfounded model, I show that two opposite forces impinge on the SMP. 1) The relative-size effect incentivizes the SMP to increase polarization since this increases amount of data it gathers. 2) The diversification effect incentivizes the SMP to decrease polarization since this increases value from data. Balancing these two forces, the SMP decides the optimal level of polarization it induces. Overall, the result provides a rationalization for opposite empirical results concerning the effect of an SMP on polarization. Further, if users prefer interacting with same-group users, the SMP internalizes this heterogeneity and has a greater incentive to increase polarization. Finally, increase in polarization can be tempered by levying a tax on the revenue of the SMP.

JEL Codes: D83, D85, I31, L86, Z13

Keywords: platforms, polarization, data, network

Authors:

Goonj Mohan

University of Barcelona

Email:

goonj.mohan@ub.edu

Date: July 2024

Acknowledgements:

I would like to thank Fernando Vega-Redondo for his invaluable advice throughout my PhD. I am grateful to Christoph Carnehl, Olga Chiappinelli, Nenad Kos, Massimo Morelli and Nicolas Serrano- Velarde for their constant guidance. I would also like to thank Antonio Cabrales, Francesco Decarolis, Katrin G¨odker, Benjamin Golub, Basile Grassi, Massimo Marinacci, Clement Mazet-Sonilhac, Marco Ottaviani, Fabrizio Panebianco, Eleonora Patacchini, Silvio Petriconi, Paolo Pin, Carlo Schwarz, Jakob Sørensen, David Yang and Noam Yuchtman for their helpful comments..

The Data Economy and Polarization on Social Media

Latest version: click here

Goonj Mohan*

Abstract

This paper studies the incentive of a social media platform (SMP) to increase polarization of its user network. I propose a two-group network model where the SMP earns revenue from user-data driven personalization. The objective of the SMP is to maximize the amount of valuable data generated. To this end, it relies on an algorithm that, at a cost, encourages users to form new links. Within a microfounded model, I show that two opposite forces impinge on the SMP. 1) The *relative-size effect* incentivizes the SMP to increase polarization since this increases amount of data it gathers. 2) The *diversification effect* incentivizes the SMP to decrease polarization since this increases value from data. Balancing these two forces, the SMP decides the optimal level of polarization it induces. Overall, the result provides a rationalization for opposite empirical results concerning the effect of an SMP on polarization. Further, if users prefer interacting with same-group users, the SMP internalizes this heterogeneity and has a greater incentive to increase polarization. Finally, increase in polarization can be tempered by levying a tax on the revenue of the SMP.

JEL: D83, D85, I31, L86, Z13

*University of Barcelona; goonj.mohan@ub.edu
I would like to thank Fernando Vega-Redondo for his invaluable advice throughout my PhD. I am grateful to Christoph Carnehl, Olga Chiappinelli, Nenad Kos, Massimo Morelli and Nicolas Serrano-Velarde for their constant guidance. I would also like to thank Antonio Cabrales, Francesco Decarolis, Katrin Gödker, Benjamin Golub, Basile Grassi, Massimo Marinacci, Clement Mazet-Sonilhac, Marco Ottaviani, Fabrizio Panebianco, Eleonora Patacchini, Silvio Petriconi, Paolo Pin, Carlo Schwarz, Jakob Sørensen, David Yang and Noam Yuchtman for their helpful comments.

1 Introduction

The effect of social media platforms (SMPs) on polarization is an issue of urgent public concern. Since extreme polarization weakens democracy (see McCoy et al. (2018)), it is important to understand how a social media affects polarization. A Facebook whistleblower has argued that the SMP's algorithms increases polarization.[1] However, Facebook's study (Bakshy et al. (2015)) suggests that user choice is a stronger driver of selective exposure.

It is unclear how social media affects polarization. A growing literature examines this question (see Sunstein (2018) and Dwivedi et al. (2021)). The paper by Allcott, Braghieri, Eichmeyer and Gentzkow (2020) suggests that Facebook increases polarization in the USA context. But the paper by Asimovic, Nagler, Bonneau, and Tucker (2021) suggests that Facebook decreases polarization in the Bosnia and Herzegovina context. Despite the urgency of this issue, there is no existing framework that analyzes how the profit motive of an SMP affects polarization of its user network. This paper is the first to link polarization to profit incentive of an SMP.

The model consists of a social media platform that maximizes its revenue by maximizing the amount of data generated on its user network. Users are linked via an initial two-group network model (for e.g., liberals form one group and conservatives form another group). Linked users generate data by consuming each other's content. The SMP earns revenue by using this data to advertise products to users (see also Bergemann et al. (2022)).

The SMP's revenue is rooted in a microfoundation outlined in Mohan et al. (2023). The microfoundation assumes that an SMP uses data to a) enhance user-product matching (documented by Ali et al. (2019) and Kozyreva et al. (2021)) and b) personalize product prices in advertisements (documented by Zuiderveen Borgesius and Poort (2017) and Bourreau and De Streel (2018)). Ultimately, revenue optimization of an SMP hinges on maximizing the amount of data generated on the user network.[2] In

[1]https://www.reuters.com/technology/facebook-accountable-no-one-whistleblower-will-say-testimony-2021-10-04/
https://www.theguardian.com/technology/2021/oct/25/facebook-whistleblower-frances-haugen-calls-for-urgent-external-regulation

[2]This is consistent with the reduced form payoff formulation in Acemoglu et al. (2019), Choi et al.

our model this is equivalent to maximizing the number of links on the user network.

The SMP relies on an algorithm that, at a cost, encourages users to form new links. The users in turn passively accept links suggested by the SMP (as evidenced by Bucher (2013) and Eslami et al. (2015)). The objective of the SMP is to maximize its profit (data revenue - algorithm cost). The SMP is said to increase polarization if it *disproportionately* encourages links between users of the same-type relative to users of different-types. This measure of polarization is akin to polarization of opinions and beliefs (see Mutz (2002), Pettigrew and Tropp (2006) and Levy (2021) for evidence).

Two opposite forces impinge on the profit-maximizing SMP. On the one hand, the *relative-size effect* incentivizes the SMP to increase polarization. When the SMP has more users of one type (for e.g., Twitter has more liberals than conservatives), it can encourage more links by searching for users of the same-type than by searching for users of different-types. This *relative-size effect* incentivizes the SMP to increase polarization. On the other hand, the *diversification effect* drives the SMP to decrease polarization. When initial polarization of the user network is high, a user is more likely to link with another user of the same-type. This reduces the pool of same-type users that the SMP can choose from when encouraging new links.[3] Consequently, it is sub-optimal for the SMP to only find users of the same-type (see Ioannides and Datcher Loury (2004), Ostrom (2008) and Currarini et al. (2009) for analogous frameworks) and it has an incentive to diversify. This *diversification effect* incentivizes the SMP to encourage links between users of different-types, thereby decreasing polarization. Therefore, an SMP *always* influences a user network. When initial polarization is low, *diversification effect* is weaker and the *relative-size effect* incentivizes the SMP to increase polarization. When initial polarization is high, *diversification effect* is stronger and incentivizes the SMP to decrease polarization.

There are three main implications of this result. First, different levels of initial polarization in the USA and Bosnia and Herzegovina rationalizes why Facebook in-

(2019).

[3]The potential exposure options for any user is finite as SMPs like Facebook and Twitter limit their content search for any user. See https://about.fb.com/news/2021/01/how-does-news-feed-predict-what-you-want-to-see/ for details about how the algorithm works. Also see Perra and Rocha (2019) for empirical evidence reiterating that the exposure market for any user is thin, which makes the relative-size effect and the diversification effect relevant.

creases polarization in USA (Allcott, Braghieri, Eichmeyer and Gentzkow (2020)) but decreases polarization in ethnically divided Bosnia and Herzegovina (Asimovic, Nagler, Bonneau, and Tucker (2021)). Second, when users prefer interacting with same-type linked users, the SMP internalizes this heterogeneity in preferences via the *relative-size effect*. Consequently, as users increasingly prefer interacting with same-type users, the incentive of the SMP to increase polarization becomes greater. Third, whenever initial polarization is low enough for the SMP to increase polarization, the final polarization induced can be tempered by levying a tax on the SMP's revenue. The SMP internalizes this taxation by forming fewer new links. Since initial polarization is low the reduction disproportionately affects links between different-types users, thereby reducing final polarization induced.

Related Literature: This paper is related to three strands of literature, namely, social media platforms, networks and polarization.

The recent work on social media platforms by Choi et al. (2019), Acemoglu et al. (2019) and Bergemann et al. (2022) has studied the value of data for a social media platform. These papers focus on how value of user data affects user privacy and this paper studies how the value of user data affects polarization of the user network. Interestingly, the microfounded revenue function of the SMP used in this paper is consistent with the reduced form payoff function of the SMP modelled in Acemoglu et al. (2019). A more related paper is by Acemoglu, Ozdaglar and Siderius (2021), where the SMP amplifies misinformation to maximize engagement of a low reliability content piece. This paper shows that the SMP has an incentive to increase polarization irrespective of the reliability of a content piece. Specifically, the SMP increases polarization when the relative-size of the two groups is sufficiently high or when initial polarization of the user network is sufficiently low.

This paper leans on the existing networks literature, where the work by Ioannides and Datcher Loury (2004), Ostrom (2008), Currarini et al. (2009) and Currarini et al. (2016) guides the network model formulated in this paper. The recent work by Fainmesser and Galeotti (2016), Chen et al. (2018) and Fainmesser and Galeotti (2020) analyzes how information generated on a social network affects pricing strategies of a seller. Since these papers focus on a network good, their analysis is complementary to this paper but the common thread of inquiry, which is the value of data generated on a social network, ties all these papers together.

Finally, this paper contributes to the literature on polarization. Recent work has shown that there isn't a one-size-fits-all answer to how SMPs affect polarization. As noted by Halberstam and Knight (2016) and Törnberg et al. (2021) the answer depends on the ambient environment. In particular, it should depend on the user network that exists before the execution of the algorithm of an SMP (as seen in Sîrbu et al. (2019) and Cinelli et al. (2021)). Additionally, the work by Barberá, (2014) and (2020) has cautioned that the effect of an SMP on polarization is nuanced. This paper reiterates that, albeit in a different context. This paper also provides a rationalization for opposite empirical findings of Allcott et al. (2020) and Asimovic et al. (2021) by highlighting the heterogeneous initial conditions that are driving these seemingly opposite results.

The rest of the paper is organized follows. Section 2 describes the model, Section 3 characterizes the decision problem of the SMP and the effect of the decision on polarization. Section 4 analyzes the implications of the SMP decision. Section 5 concludes and proofs are in Appendix.

2 Model

The model comprises of a social media platform where users are linked on an initial network. Each link between a pair of users represents an interaction which generates information or *data* about both users. The SMP earns revenue from user-data driven personalization. To that end, the SMP uses its algorithm to encourage new links on the network. This paper studies if the algorithm results in increased polarization, that is, whether the SMP disproportionately encourages links between users of the same-type relative to users of different-types.

2.1 A Network

Let us define the network structure which links users on the SMP.

We define an islands network (as in Jackson and Rogers (2005) and Jackson (2010)), which is a special case of the multi-type random network. Given a set of n users $N = \{1, ..., n\}$, a network is represented via its adjacency matrix: a symmetric n-by-n matrix $\mathbf{G}$ with entries in $\{0, 1\}$. If $G_{ij} = G_{ji} = 1$ then users i and j are linked, and the assumption of symmetry restricts attention to undirected networks. Assume that users does not link with themselves, that is, set $G_{ii} = 0$ for each i.

Users have "types", which are the distinguishing features that affect their propensities to connect to each other. Types might be based on any characteristics that influence users' probabilities of linking to each other, including age, race, gender, profession, education level, and political leaning.

Suppose there are m different types in the society. Let $N_k \subset N$ denote the users of type k, so the society is partitioned into m groups, $(N_1, ..., N_m)$. Let $n_k = |N_k|$ denote the size of group k.

A *multi-type random network* is defined by the cardinality vector $\mathbf{n}$ together with a symmetric m-by-m matrix $\mathbf{P}$, whose entries in $[0, 1]$ describe the probabilities of links between various types. The entry P_{kl} captures the probability that a user of type k links to a user of type l.

The islands network is the special case of the multi-type random networks model, such that, each user only distinguishes between users of one's own group and users of a different group. Moreover, all users are symmetric in how they do this. Formally, the multi-type random network $\mathbf{G(n, P)}$ is an *islands network* $\mathbf{G(n, p_n, q_n)}$ if:

- there are two groups N_1 and N_2 with $n = n_1 + n_2$
- $P_{ii} = p_n$ for $i \in \{1, 2\}$
- $P_{ij} = q_n$ for $i \neq j$.

Consequently, the expected number of links of any user $l_i \in N_i$ is k_{in} where

$$k_{1n} = \sum_{k \in N} G_{l_1 k} = p_n(n_1 - 1) + q_n n_2$$
$$k_{2n} = \sum_{k \in N} G_{l_2 k} = p_n(n_2 - 1) + q_n n_1$$

A user in N_i links with one of the other $n_i - 1$ users in the same group with probability p_n and links with one of the n_{-i} users in the other group with probability q_n. This gives us the above expressions for k_{1n} and k_{2n}. To obtain the network structure that links users we extend these formulations to a large n in the section below.

2.2 A Large Network

Users are linked on the SMP on a large network. A large network is essentially an islands network $G(n, p_n, q_n)$ with n *large*.

Consider a set of islands networks $\{G(n, p_n, q_n)\}_{n \in \mathbb{N}}$ where n, p_n and q_n satisfy

$$n_2 = \frac{\alpha}{1+\alpha} n \quad n_1 = \frac{1}{1+\alpha} n \quad p_n = \frac{p}{n} \quad q_n = \frac{q}{n}$$

for some $\alpha > 0$ and $p, q \in [0, 1]$. We say a large network $G(\alpha, p, q)$ is obtained when n becomes large (goes to infinity in the limit). Therefore, a large network is defined using three parameters - relative-size α of the two groups, *same-group density* p that a user of type i links to another user of type i and *different-group density* q that a user of type i links to another user of type j, $j \neq i$.

A large network imbibes two essential properties of SMPs like Facebook and Twitter.

1. The network has infinitely many users since n is large and
2. The number of links of a user remains finite. Specifically, a user in N_i has k_i links in expectation where

$$k_1 = \lim_{n \to \infty} k_{1n} = \lim_{n \to \infty} \frac{p}{n}(n_1 - 1) + \frac{q}{n} n_2 = \frac{p + \alpha q}{1 + \alpha}$$
$$k_2 = \lim_{n \to \infty} k_{2n} = \lim_{n \to \infty} \frac{p}{n}(n_2 - 1) + \frac{q}{n} n_1 = \frac{\alpha p + q}{1 + \alpha}$$

I make two assumptions about a large network.

1. N_2 is weakly larger than N_1 *or* $\alpha \geq 1$
2. the network exhibits weak homophily *or* $p \geq q$.

The first assumption is clearly without loss of generality. The second assumption is only used to give a sharper characterisation of the diversification effect in Lemma 3. It is not required for the main result of the paper stated in Proposition 2.

2.3 Data Generated

Having described the structure of a large network on which users are linked, we now describe the data (or information) generated on a large network. As before, we first consider an islands network and then extend the formulation to a large network. Given an islands network, data points about a user are generated through links of the user. We assume that any two linked users i and j

- interact with each other and
- each interaction generates *two data points*; one about each user.

Discussion - A link between two users can be interpreted as: each user sees a post by the other user. Interaction with the post seen (a user may 'like', 'comment', scroll quickly or slowly past the post etc.) generates a data point about each user.

From the assumptions above we see that the expected number of data points generated about a user is equal to the expected number of links of the user. By a slight abuse of notation we use k_{in} to denote both, which gives us

$$k_{1n} = p_n(n_1 - 1) + q_n n_2$$
$$k_{2n} = p_n(n_2 - 1) + q_n n_1$$

Therefore, k_{1n} data points are generated about each user in N_1 and k_{2n} data points are generated about each user in N_2. This formulation extends to any large network $G(\alpha, p, q)$. The expected number of data points generated about a user is equal to the expected number of links of the user. We use k_i to denote both, so for any user in N_i

$$k_1 = \frac{p + \alpha q}{1 + \alpha}$$
$$k_2 = \frac{\alpha p + q}{1 + \alpha}$$

Therefore, for any large network $G(\alpha, p, q)$, k_1 data points are generated about any user in N_1 and k_2 data points are generated about any user in N_2. The SMP maximizes its revenue by maximising the total amount of data generated about the users. Consequently, for any initial large network the SMP uses its algorithm, at a cost, to form new links and generate more data. The decision problem faced by the SMP is described in detail in the next section.

3 Decision Problem of the SMP

The decision problem faced by the SMP is as follows. The SMP *knows* that users will be linked via an initial large network $G(\alpha, p_0, q_0)$. The SMP chooses (p^*, q^*) such that its payoff under the final network $G(\alpha, p^*, q^*)$ is maximum. The payoff of the SMP has two components.

1. Revenue attained from data generated on the final network

2. Cost incurred to form the final network.

The revenue attained and cost incurred by the SMP are described below.

3.1 Revenue from Data

The SMP attains revenue by analysing the data generated about the users on the network. The SMP uses data about a user to

1. provide better product recommendations to the user (documented by Ali et al. (2019) and Kozyreva et al. (2021)) and

2. personalize the price of the recommended product (documented by Zuiderveen Borgesius and Poort (2017) and Bourreau and De Streel (2018)).

Using a Bayesian updating model, Mohan et al. (2023) shows that if the SMP attains x data points about a user then revenue attained from that user is

$$g(x) = \left(\frac{x}{x+1}\right)\left(\frac{x+2}{x+3}\right)$$

The detailed derivation of this expression can be seen in Mohan et al. (2023). Crucially, revenue attained by the SMP increases as it attains more data points about a user. Let us see how. Intuitively, the expression $\frac{x}{x+1}$ captures the effect of better product recommendations. As the SMP attains more data about a user or as x increases, the probability that the SMP matches the taste of the user increases. This probability is captured by $\frac{x}{x+1}$, which lies in $[0, 1]$ and is increasing in x. Additionally, the expression takes value 1 and the SMP definitely matches the taste of a user as x goes to infinity or when the SMP has infinitely many data points about a user. The expression $\frac{x+2}{x+3}$ captures the effect of personalized pricing. As the SMP attains more data about a user or as x increases, the SMP improves its estimate of user's willingness to pay and extracts higher rent from a user. This is captured by $\frac{x+2}{x+3}$, which is increasing in x and takes value 1 (assumed to be expected value of user's willingness to pay) as x goes to infinity. Therefore, the SMP provides better product recommendation and extracts more surplus from a user as it attains more data about the user.
The lemma below extends the result to any large network. Crucially, the property that revenue of the SMP is increasing in number of data points generated is preserved. Assume that the revenue attained by the SMP from a large network is the sum of revenues attained from all the users on the network.

Lemma 1. *Given a large network $G(\alpha, p, q)$ the revenue $u_P^d(p, q)$ attained by the SMP from the network is*

$$\begin{aligned} u_P^d(p, q) &= \frac{2(p + \alpha q)}{3(1 + \alpha)^2} + \frac{2\alpha(\alpha p + q)}{3(1 + \alpha)^2} \\ &= \frac{2k_1}{3(1 + \alpha)} + \frac{2\alpha k_2}{3(1 + \alpha)} \end{aligned}$$

Discussion - As the number of data points generated on a large network increases, the revenue attained by the SMP from the network increases. This is because revenue attained from a network is the sum of revenues attained from all the users on the network. Consequently, if the SMP incurred no cost when increasing links between users then it would choose to form a final network that exhibited maximum linkage. However, the cost of the algorithm deters the SMP from doing so. The cost incurred by the SMP is described in the next section.

3.2 Cost of the Algorithm

Consider an initial large network $G(\alpha, p_0, q_0)$. The SMP uses its algorithm to increase interaction between users by forming new links. Suppose the resulting final network formed is $G(\alpha, p, q)$. Then the cost $c(p, q)$ incurred by the SMP is

$$c(p, q) = c(p - p_0)^2 + c(q - q_0)^2$$

where $c > 0$ is an arbitrary fixed parameter.

Discussion - The formulation of the cost function is a consequence of how the algorithm of an SMP generally works. An SMP finds content for a user within a limited engagement pool.[4] It has been observed that content from users of the same-type is substitutable (two people in the same company share similar information about job openings) and that content from users of different-types is independent (two people in different companies give different information about the job openings in their respective companies). The work by Ioannides and Datcher Loury (2004), Ostrom (2008) and Currarini et al. (2009) chronicles these effects. The property of substitution-and-independence is assumed in this model. Consequently, forming links with users who are in the same group has an increasing marginal cost due to the substitution effect and forming links with users in different groups has a fixed marginal cost due to the independence effect. The resulting cost function $c(p, q)$ is formulated above. We have assumed a squared functional form for $c(p, q)$ and the results of the paper remain qualitatively unchanged for any functional form with power greater than one.

3.3 Decision of the SMP

Based on the revenue attained and the cost incurred by the SMP the decision problem faced by the SMP is as follows.

[4]For details about how the algorithm works see https://about.fb.com/news/2021/01/how-does-news-feed-predict-what-you-want-to-see/. Also see Perra and Rocha (2019) for empirical evidence reiterating that the exposure market for any user is limited.

The SMP knows that users will form an initial large network $G(\alpha, p_0, q_0)$. It chooses p^* and q^* such that payoff of the SMP is maximized from the final network $G(\alpha, p^*, q^*)$. For an initial network $G(\alpha, p_0, q_0)$ and final network $G(\alpha, p, q)$ payoff of the SMP $u_P(p, q)$ is defined as the revenue attained from the final network deducted by the cost incurred to form the final network.

$$\begin{aligned} u_P(p,q) &:= u_P^d(p,q) - c(p,q) \\ &= \frac{2(p+\alpha q)}{3(1+\alpha)^2} + \frac{2\alpha(\alpha p + q)}{3(1+\alpha)^2} - c(p-p_0)^2 - c(q-q_0)^2 \end{aligned}$$

For any initial network $G(\alpha, p_0, q_0)$ the SMP maximizes its payoff by choosing (p^*, q^*) that satisfies

$$\begin{aligned} (p^*, q^*) &= \arg\max_{(p,q)} \quad u_P(p,q) \\ &= \arg\max_{(p,q)} \quad \frac{2(p+\alpha q)}{3(1+\alpha)^2} + \frac{2\alpha(\alpha p + q)}{3(1+\alpha)^2} - c(p-p_0)^2 - c(q-q_0)^2 \end{aligned}$$

For any initial network we get a unique final network that maximizes the payoff of the SMP, as seen in the proposition below.

Proposition 1. *For any initial network $G(\alpha, p_0, q_0)$ the payoff of the SMP is maximized from the final network $G(\alpha, p^*, q^*)$ where*

$$p^* = p_0 + \frac{1+\alpha^2}{3c(1+\alpha)^2}$$

$$q^* = q_0 + \frac{2\alpha}{3c(1+\alpha)^2}$$

The final network formed by the SMP exhibits some interesting properties that are stated in the corollaries below.

Corollary 1. *Given any initial network $G(\alpha, p_0, q_0)$ the final network formed by the SMP satisfies $p^* > p_0$ and $q^* > q_0$.*

Corollary 2. *If an initial network $G(\alpha, p_0, q_0)$ exhibits weak homophily then the final network formed by the SMP also exhibits weak homophily.*

$$p_0 \geq q_0 \implies p^* \geq q^*$$

These results tell us that the SMP always increases links between users and that weak homophily is preserved under the influence of the SMP. To see how these properties interact with polarization let us define polarization of a network.

4 Polarization and Effect of SMP Decision

Definition 4.1. *For a network $G(\alpha, p, q)$ polarization is defined as $P(p, q) = \frac{p}{q}$.*

This measure of polarization is akin to polarization of opinions and beliefs (see Mutz (2002), Pettigrew and Tropp (2006) and Levy (2021) for evidence). The effect of the SMP on polarization is understood by comparing polarization of the initial network to that of the final network.

- Polarization of an initial network $G(\alpha, p_0, q_0)$ is $P(p_0, q_0) = \frac{p_0}{q_0}$ and shall henceforth be called initial polarization.
- Polarization of the corresponding final network $G(\alpha, p^*, q^*)$ is $P(p^*, q^*) = \frac{p^*}{q^*}$ and shall henceforth be called final polarization.

Definition 4.2. *The SMP is said to increase polarization if and only if final polarization is weakly higher than initial polarization.*

$$\frac{p^*}{q^*} \geq \frac{p_0}{q_0}$$

By Corollary 1 we know that the SMP forms a final network which has more links than the initial network. However, since $p^* > p_0$ and $q^* > q_0$ can result in both $\frac{p^*}{q^*} \geq \frac{p_0}{q_0}$ and $\frac{p^*}{q^*} < \frac{p_0}{q_0}$ this corollary does not tell us how the SMP affects polarization.

Corollary 2 does gives us some insight. Rewriting the corollary in terms of polarization tells us that

$$P(p_0, q_0) \geq 1 \implies P(p^*, q^*) \geq 1$$

Therefore, if users are more likely to link within their own group on an initial network then users are also more likely to be linked within their own group on the final network. However, this still does not tell us whether the SMP increases or decreases polarization.

In order to get a deeper insight of how the SMP affects polarization let us consider a specific example. We take a fixed α, vary p_0 and q_0 such that $P(p_0, q_0)$ increases and calculate the corresponding $P(p^*, q^*)$ using the p^* and q^* obtained in Proposition 1. We plot the initial polarization and final polarization below.

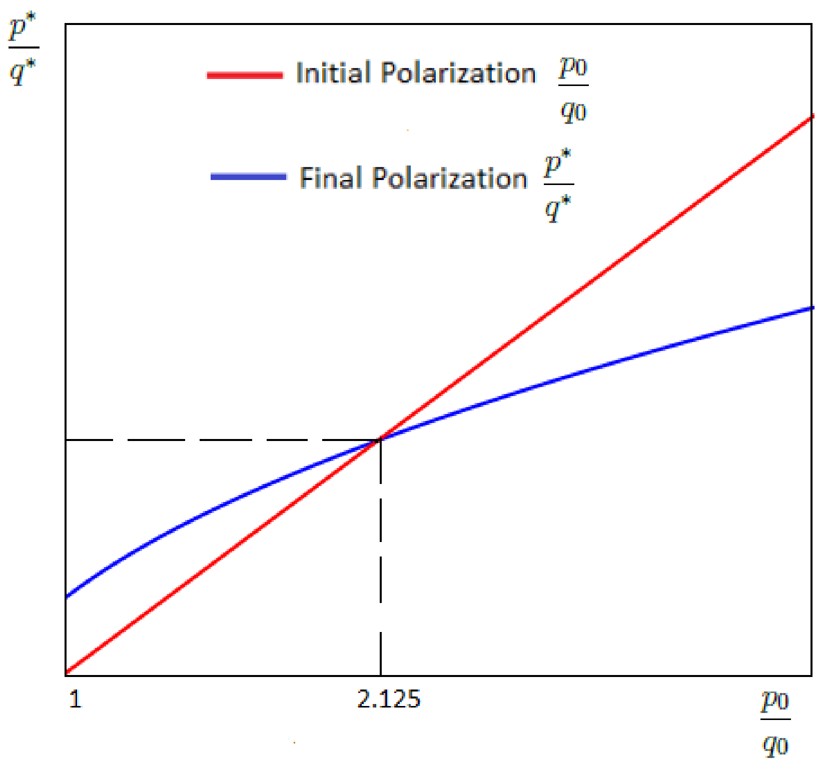

Figure 1: This figure shows the effect of the SMP on polarization for a fixed $\alpha = 4$.

The figure suggests that the SMP does not always increase (or decrease) polarization. It seems to indicate that the SMP increases polarization when initial polarization is low and decreases polarization when initial polarization is high. We formalise this intuition in the rest of this section, starting with two crucial lemmas.

Lemma 2. *Consider an initial network $G(\alpha, p_0, q_0)$ with $p_0 = q_0$. Then the SMP increases polarization of the network,*

$$P(p^*, q^*) \geq P(p_0, q_0)$$

and strict inequality holds when $\alpha > 1$.

This lemma shows how relative-size α of the two groups drives the SMP to increase polarization. When $p_0 = q_0$ the cost incurred by increasing p or by increasing q is the same, so the decision of the SMP depends only on the revenue side of the payoff. When $\alpha \geq 1$, data generated by increasing p is weakly higher than data generated by increasing q. Therefore, the SMP chooses $p^* \geq q^*$ which implies

$$P(p^*, q^*) \geq 1 = P(p_0, p_0)$$

and strict inequality holds for $\alpha > 1$.
Henceforth, Lemma 2 shall be referred to as the *relative-size effect*. Thus, the *relative-size effect* drives the SMP to increase polarization.

Lemma 3. *Consider an initial network $G(\alpha, p_0, q_0)$ with $\alpha = 1$. Then the SMP decreases polarization of the network,*

$$P(p^*, q^*) \leq P(p_0, q_0)$$

and strict inequality holds when $\frac{p_0}{q_0} > 1$.

This lemma shows how the diversification of links $\frac{p_0}{q_0}$ between same-types and different-types on the initial network drives the SMP to decrease polarization. When $\alpha = 1$ the revenue attained by increasing p or by increasing q is the same, so the decision of the SMP depends only on the cost side of the payoff. Consequently, the marginal cost of increasing p is equal to the marginal cost of increasing q, which implies

$p^* - p_0 = q^* - q_0 = a$ for some arbitrary $a > 0$. Then,

$$P(p^*, q^*) \leq P(p_0, q_0) \iff \frac{p^*}{q^*} = \frac{p_0 + a}{q_0 + a} \leq \frac{p_0}{q_0} \iff q_0 \leq p_0$$

Since the initial network exhibits weak homophily the inequality holds and the strict inequality holds when $p_0 > q_0$.
Henceforth, Lemma 3 shall be referred to as the *diversification effect*. Thus, the *diversification effect* drives the SMP to decrease polarization.

Ultimately, the *relative-size effect* captured by α incentivizes the SMP to increase polarization and the *diversification effect* captured by $\frac{p_0}{q_0}$ incentivizes the SMP to decrease polarization. The effect of the SMP decision on polarization depends on the relative strength of these two effects, as seen below in the main result of the paper.

Proposition 2. *For an initial network $G(\alpha, p_0, q_0)$ the SMP increases polarization if and only if*

$$\frac{1 + \alpha^2}{2\alpha} \geq \frac{p_0}{q_0} \tag{TC}$$

Therefore, the SMP always affects polarization of the initial network. If initial polarization is low then the *diversification effect* is weaker than the *relative-size effect* and the SMP increases polarization. If initial polarization is high then the *diversification effect* is stronger than the *relative-size effect* and the SMP decreases polarization. The threshold condition TC encapsulates the comparison of these two forces and we can now understand Fig. 1 through the lens of this comparison.

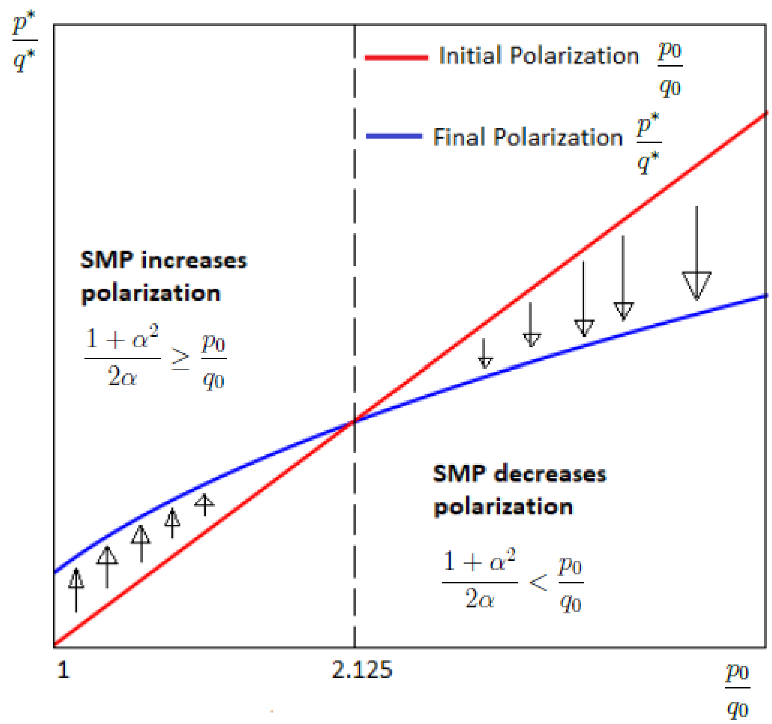

Figure 2: The figure shows the effect of the SMP for $\alpha = 4$. The SMP increases polarization when $\frac{1+\alpha^2}{2\alpha} = 2.125 \geq \frac{p_0}{q_0}$. When initial polarization is lower than 2.125 the SMP increases polarization, so the final polarization (in blue) is higher than the initial polarization (in red). When the initial polarization is higher than 2.125 the SMP decreases polarization, so the final polarization is lower than the initial polarization.

5 Implications of the SMP Decision

Having captured the effect of the SMP decision on polarization we turn our attention to other implications of the SMP decision.

5.1 Rationalizing opposite empirical findings

Proposition 2 provides a rationalization that resolves seemingly opposite empirical findings of "The Welfare Effects of Social Media" by Allcott et al. (2020) and "Testing the effects of Facebook usage in an ethnically polarized setting" by Asimovic et al.

(2021).

Alcott et. al. (2020) find that Facebook increases polarization in USA and Asimovic et. al. (2021) find that Facebook decreases polarization in Bosnia and Herzegovina. However, this does not necessarily imply that the two results contradict each other. Instead, it is possible that differential ambient conditions, namely different levels of initial polarization in USA versus Bosnia and Herzegovina drives these results. The example below illustrates this.

Low initial polarization: Consider initial network $G(\alpha = 3, p_0 = .3, q_0 = .2)$. The network has low initial polarization and is a representation of USA.

$p_0 = .3$	$q_0 = .2$	$\frac{p_0}{q_0} = 1.5$
TC: $1.66 \geq 1.5$	$\rightarrow$	satisfied
$p^* = .98$	$q^* = .52$	$\frac{p^*}{q^*} = 1.89$

The chosen values of p_0 and q_0 result in low (lower than threshold 1.66) initial polarization. Consequently, the threshold condition is satisfied and the SMP increases polarization from 1.5 to 1.66, as has been evidenced in USA.

High initial polarization: Consider initial network $G(\alpha = 3, p_0 = .28, q_0 = .12)$. The network has high initial polarization and is a representation of Bosnia and Herzegovina.

$p_0 = .28$	$q_0 = .12$	$\frac{p_0}{q_0} = 2.33$
TC: $1.66 \geq 2.33$	$\rightarrow$	violated
$p^* = .96$	$q^* = .44$	$\frac{p^*}{q^*} = 2.18$

The chosen values of p_0 and q_0 result in high (higher than threshold 1.66) initial polarization. Consequently, the threshold condition is not satisfied and the SMP decreases polarization from 2.33 to 2.18, as has been evidenced in Bosnia and Herzegovina.

The above example has provided a rationalization of opposite empirical findings, wherein Facebook increases polarization in USA since *initial polarization is low* and decreases polarization in Bosnia and Herzegovina *since initial polarization is high.*

Notably, evidence suggests that initial polarization in Bosnia and Herzegovina is indeed higher than in USA. Firstly, the conflict in Bosnia and Herzegovina (1992-1995) *ethnically polarized the society* (as the title suggests), which is consistent with high initial polarization. Secondly, Asimovic et al. (2021) investigate why Facebook usage decreases polarization. They find that the effect of Facebook is concentrated among those who live in ethnically homogeneous environments. This suggests that Facebook usage decreases polarization precisely because initial polarization of the network is high. Therefore, Proposition 2 and supporting empirical evidence provides a rationalization for seemingly opposite findings in the literature.

5.2 Heterogeneous Preferences for Interactions

This section analyzes how heterogeneous preferences for interactions affects the incentives of the SMP. Bakshy et al. (2015) find that users who are exposed to SMP-influenced content, *clicked through to 70% less of cross-cutting content.* For e.g., suppose a liberal user views a news article posted by another liberal friend and a news article posted by a conservative friend on Facebook. Then Bakshy et al. (2015) tells us that the liberal user is more likely to click on (and further interact with) the article posted by the liberal friend. Consequently, a link between users of the same-type generates more data than a link between users of different-types. This heterogeneity is modelled as follows.

Definition 5.1. *Users on a network $G(\alpha, p, q)$ are said to have heterogeneous preferences for interactions, denoted by $\lambda \geq 1$, if an interaction between users of the same-type generates λ data points as opposed to one data point generated by an interaction between users of different-types.*

For an arbitrary λ, the number of data points $k_{i\lambda}$ generated about a user in N_i is,

$$k_{1\lambda} = \frac{\boldsymbol{\lambda} p + \alpha q}{1+\alpha}$$
$$k_{2\lambda} = \frac{\boldsymbol{\lambda} \alpha p + q}{1+\alpha}$$

Consequently, payoff of the SMP from a network $G(\alpha, p, q)$ is $u_P^\lambda(p, q)$, where

$$u_P^\lambda(p,q) = \frac{2(\lambda p + \alpha q)}{3(1+\alpha)^2} + \frac{2\alpha(\lambda\alpha p + q)}{3(1+\alpha)^2} - c(p-p_0)^2 - c(q-q_0)^2$$

As before, the SMP forms a final network $G(\alpha, p^*, q^*)$ that maximizes its payoff. This affects polarization as follows.

Proposition 3. *For an initial network $G(\alpha, p_0, q_0)$ and heterogeneity λ, the SMP increases polarization of the network if and only if*

$$\lambda \left(\frac{1+\alpha^2}{2\alpha}\right) \geq \frac{p_0}{q_0} \qquad (\mathrm{TC}^\lambda)$$

The SMP knows that an interaction between users of the same-type generates more data than an interaction between users of different-types. This increases the incentive of the SMP to form links between users of the same-type. Thus, as users' preference λ for interacting with other same-type users increases, the incentive of the SMP to increase polarization becomes greater. This modifies the threshold condition TC to TC^λ, wherein the incentive to increase polarization is amplified by a factor of λ.

5.3 Policy Intervention: Taxation

The TC elucidates the condition under which the SMP increases polarization and this section analyzes how *taxation* can mitigate the extent of this increase.

Suppose that $\frac{p_0}{q_0} < \frac{1+\alpha^2}{2\alpha}$. Absent any intervention, the SMP chooses to increase polarization (see TC). Suppose taxation is introduced and revenue of the SMP is taxed by an arbitrary and fixed rate $0 \leq t \leq 1$.
For revenue r generated by the SMP, tr is paid as tax and $(1-t)r$ is retained by the SMP. For any final network $G(\alpha, p, q)$ the modified utility of the SMP is

$$u_{P,t}(p,q) = (1-t)\left[\frac{2(p+\alpha q)}{3(1+\alpha)^2} + \frac{2\alpha(\alpha p + q)}{3(1+\alpha)^2}\right] - c(p-p_0)^2 - c(q-q_0)^2$$

Clearly, revenue of the SMP is different (weakly lower) under taxation than under no taxation. Consequently, the SMP chooses (p_t^*, q_t^*) that maximizes this modified utility.

Proposition 4. *For an initial network* $G(\alpha, p_0, q_0)$ *and taxation rate* t*, the SMP forms the final network* $G(\alpha, p_t^*, q_t^*)$ *where*

$$p_t^* = p_0 + \frac{(1+\alpha^2)}{3c(1+\alpha)^2}(1-t)$$

$$q_t^* = q_0 + \frac{2\alpha}{3c(1+\alpha)^2}(1-t)$$

Given any initial network, the SMP attains higher revenue by increasing the links on the network. However, taxation weakens the incentive to increase links as the SMP obtains only a part of the revenue at the same cost. This can be seen in the expression above - the SMP increases same-group and different-group density, but these increases are dampened by $(1-t)$. This also shows that higher the taxation rate t, lesser is the increase in links by the SMP.

Since taxation dampens the increase in same-group and different-group density, its effect on polarization is not apparent. The two corollaries below characterize the effect of taxation on polarization.

Corollary 3. *For an initial network* $G(\alpha, p_0, q_0)$ *and taxation rate* t*, the SMP increases polarization if and only if the Eq.* (TC) *holds.*

Thus, the condition which determines whether the SMP increases or decreases polarization remains unchanged under taxation. Although surprising, the following intuition makes this result clear. Taxation dampens the increase in same-group and different-group density by the same factor $(1-t)$. This leaves the *relative-size effect* unchanged. The *diversification effect* also remains unchanged as the initial same-group density p_0 and different-group density q_0 are fixed. Therefore, for any initial network $G(\alpha, p_0, q_0)$, the *relative-size effect* and *diversification effect* remain unchanged under taxation, leaving the threshold condition unchanged. Consequently, the SMP increases polarization when initial polarization is low, irrespective of the taxation rate t.

We now know that when initial polarization is low, final polarization is higher under taxation and under no taxation. The next corollary compares final polarization under taxation and final polarization under no taxation.

Corollary 4. *For an initial network $G(\alpha, p_0, q_0)$ and taxation rate t, final polarization is lower under taxation than under no taxation. Specifically,*

$$\frac{p_t^*}{q_t^*} < \frac{p_{t=0}^*}{q_{t=0}^*}$$

Since $p_t^* \leq p_{t=0}^*$ and $q_t^* \leq q_{t=0}^*$ the comparison between final polarization under taxation and final polarization under no taxation is not obvious. However, since initial polarization is low, the dampened increase in links affects different-group density more than same-group density. As a result, final polarization under taxation is lower than final polarization under no taxation.

Therefore, taxation dampens (but does not eradicate) the incentive of the SMP to increase polarization. As a result, final polarization under taxation is still higher than the initial polarization but it is lower than the final polarization that would be attained under no taxation. Consequently, taxation can be used as a policy tool to mitigate the increase in polarization by the SMP.

6 Conclusion

This paper analyzes the incentive of a social media platform (SMP) to increase polarization of its user network. We consider a two-group network model of an SMP, where the SMP earns revenue from user-data driven personalization. Under a microfounded model, the payoff of the SMP is maximized when the amount of data generated on the network is maximized. The SMP relies on an algorithm that, at a cost, encourages users to form new links and generate new data on the SMP. The final of objective of the SMP is to maximize the amount of valuable data generated on the user network. In doing so, the SMP increases polarization of the user network if it disproportionately increases links between users of the same-type relative to links between users of different-types. Two opposite forces impinge on the profit motive of the firm. On the one hand, the *relative-size effect* incentivizes the SMP to increase polarization, as this increases the amount of data it gathers. On the other hand, the *diversification effect* incentivizes the SMP to decrease polarization since this increases value from data. Balancing these two forces, the SMP decides the optimal level of polarization it induces. Overall, the model provides a rationalization for the discrepancy between empirical results, where an SMP seems to both increase polarization (Allcott, Braghieri, Eichmeyer and Gentzkow) and decrease polarization (Asimovic, Nagler, Bonneau, and Tucker). Second, when users prefer interacting with same-type linked users, the SMP internalizes the heterogeneity in preferences via the *relative-size effect*. Subsequently, as users' preference for interacting with same-type linked users increases, the incentive of the SMP to increase polarization becomes greater. Finally, the decision of the SMP typically aggravates inefficiencies, i.e. it either increases polarization excessively or decreases it insufficiently, relative to the level that maximizes user welfare.

References

D. Acemoglu, A. Makhdoumi, A. Malekian, and A. Ozdaglar. Too much data: Prices and inefficiencies in data markets. Technical report, National Bureau of Economic Research, 2019.

D. Acemoglu, A. Ozdaglar, and J. Siderius. Misinformation: Strategic sharing, homophily, and endogenous echo chambers. Technical report, National Bureau of Economic Research, 2021.

M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, and A. Rieke. Discrimination through optimization: How facebook's ad delivery can lead to biased outcomes. *Proceedings of the ACM on human-computer interaction*, 3(CSCW):1–30, 2019.

H. Allcott, L. Braghieri, S. Eichmeyer, and M. Gentzkow. The welfare effects of social media. *American Economic Review*, 110(3):629–76, 2020.

N. Asimovic, J. Nagler, R. Bonneau, and J. A. Tucker. Testing the effects of facebook usage in an ethnically polarized setting. *Proceedings of the National Academy of Sciences*, 118(25):e2022819118, 2021.

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

P. Barberá. How social media reduces mass political polarization. evidence from germany, spain, and the us. *Job Market Paper, New York University*, 46:1–46, 2014.

P. Barberá. Social media, echo chambers, and political polarization. *Social media and democracy: The state of the field, prospects for reform*, 34, 2020.

D. Bergemann, A. Bonatti, and T. Gan. The economics of social data. *The RAND Journal of Economics*, 2022.

M. Bourreau and A. De Streel. The regulation of personalised pricing in the digital era. 2018.

T. Bucher. The friendship assemblage: Investigating programmed sociality on facebook. *Television & New Media*, 14(6):479–493, 2013.

Y.-J. Chen, Y. Zenou, and J. Zhou. Competitive pricing strategies in social networks. *The RAND Journal of Economics*, 49(3):672–705, 2018.

J. P. Choi, D.-S. Jeon, and B.-C. Kim. Privacy and personal data collection with information externalities. *Journal of Public Economics*, 173:113–124, 2019.

M. Cinelli, G. De Francisci Morales, A. Galeazzi, W. Quattrociocchi, and M. Starnini. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9):e2023301118, 2021.

S. Currarini, M. O. Jackson, and P. Pin. An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, 77(4):1003–1045, 2009.

S. Currarini, J. Matheson, and F. Vega-Redondo. A simple model of homophily in social networks. *European Economic Review*, 90:18–39, 2016.

Y. K. Dwivedi, L. Hughes, E. Ismagilova, G. Aarts, C. Coombs, T. Crick, Y. Duan, R. Dwivedi, J. Edwards, A. Eirug, et al. Artificial intelligence (ai): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 57:101994, 2021.

M. Eslami, A. Rickman, K. Vaccaro, A. Aleyasen, A. Vuong, K. Karahalios, K. Hamilton, and C. Sandvig. " i always assumed that i wasn't really that close to [her]" reasoning about invisible algorithms in news feeds. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pages 153–162, 2015.

I. P. Fainmesser and A. Galeotti. Pricing network effects. *The Review of Economic Studies*, 83(1):165–198, 2016.

I. P. Fainmesser and A. Galeotti. Pricing network effects: Competition. *American Economic Journal: Microeconomics*, 12(3):1–32, 2020.

Y. Halberstam and B. Knight. Homophily, group size, and the diffusion of political information in social networks: Evidence from twitter. *Journal of public economics*, 143:73–88, 2016.

Y. M. Ioannides and L. Datcher Loury. Job information networks, neighborhood effects, and inequality. *Journal of economic literature*, 42(4):1056–1093, 2004.

M. O. Jackson. *Social and economic networks*. Princeton university press, 2010.

M. O. Jackson and B. W. Rogers. The economics of small worlds. *Journal of the European Economic Association*, 3(2-3):617–627, 2005.

A. Kozyreva, P. Lorenz-Spreen, R. Hertwig, S. Lewandowsky, and S. M. Herzog. Public attitudes towards algorithmic personalization and use of personal data online: Evidence from germany, great britain, and the united states. *Humanities and Social Sciences Communications*, 8(1):1–11, 2021.

R. Levy. Social media, news consumption, and polarization: Evidence from a field experiment. *American economic review*, 111(3):831–70, 2021.

J. McCoy, T. Rahman, and M. Somer. Polarization and the global crisis of democracy: Common patterns, dynamics, and pernicious consequences for democratic polities. *American Behavioral Scientist*, 62(1):16–42, 2018.

G. Mohan et al. Polarization, regulation and networks in a data market. 2023.

D. C. Mutz. Cross-cutting social networks: Testing democratic theory in practice. *American Political Science Review*, 96(1):111–126, 2002.

E. Ostrom. The difference: How the power of diversity creates better groups, firms, schools, and societies. by scott e. page. princeton: Princeton university press, 2007. 448p. 19.95 paper. *Perspectives on Politics*, 6(4):828–829, 2008.

N. Perra and L. E. Rocha. Modelling opinion dynamics in the age of algorithmic personalisation. *Scientific reports*, 9(1):1–11, 2019.

T. F. Pettigrew and L. R. Tropp. A meta-analytic test of intergroup contact theory. *Journal of personality and social psychology*, 90(5):751, 2006.

A. Sîrbu, D. Pedreschi, F. Giannotti, and J. Kertész. Algorithmic bias amplifies opinion fragmentation and polarization: A bounded confidence model. *PloS one*, 14(3): e0213246, 2019.

C. R. Sunstein. *# Republic: Divided democracy in the age of social media*. Princeton University Press, 2018.

P. Törnberg, C. Andersson, K. Lindgren, and S. Banisch. Modeling the emergence of affective polarization in the social media society. *Plos one*, 16(10):e0258259, 2021.

F. Zuiderveen Borgesius and J. Poort. Online price discrimination and eu data privacy law. *Journal of consumer policy*, 40(3):347–366, 2017.

Proof of Lemma 1:

Proof. Payoff attained by the SMP from a network $G^n(\alpha, p_n, q_n)$ is

$$\begin{aligned} u_P^d(p_n, q_n) &= n_1 g\left(\frac{k_{1n}}{n}\right) + n_2 g\left(\frac{k_{2n}}{n}\right) \\ &= n_1\left(\frac{\left(\frac{k_{1n}}{n}\right)}{\left(\frac{k_{1n}}{n}\right)+1}\frac{\left(\frac{k_{1n}}{n}\right)+2}{\left(\frac{k_{1n}}{n}\right)+3}\right) + n_2\left(\frac{\left(\frac{k_{2n}}{n}\right)}{\left(\frac{k_{2n}}{n}\right)+1}\frac{\left(\frac{k_{2n}}{n}\right)+2}{\left(\frac{k_{2n}}{n}\right)+3}\right) \end{aligned}$$

$$\begin{aligned} \lim_{n\to\infty} u_P^d(p, q) &= \frac{2n_1}{3}\frac{\left(\frac{k_{1n}}{n}\right)}{\left(\frac{k_{1n}}{n}\right)+1} + \frac{2n_2}{3}\frac{\left(\frac{k_{2n}}{n}\right)}{\left(\frac{k_{2n}}{n}\right)+1} \\ &= \frac{2}{3(1+\alpha)}k_{1n} + \frac{2\alpha}{3(1+\alpha)}k_{2n} \end{aligned}$$

As $n \to \infty$ we know that there exists $p, p_0,\ q, q_0 \in (0, \infty)$ such that

$$p_n = \frac{p}{n}, p_{0n} = \frac{p_0}{n}, q_n = \frac{q}{n}, q_{0n} = \frac{q_0}{n}$$

Therefore

$$\begin{aligned} \lim_{n\to\infty} k_{1n} &= \lim_{n\to\infty} p\frac{n_1 - 1}{n} + q\frac{n_2}{n} = \frac{p+\alpha q}{1+\alpha} \\ \lim_{n\to\infty} k_{2n} &= \lim_{n\to\infty} p\frac{n_2 - 1}{n} + q\frac{n_1}{n} = \frac{\alpha p+q}{1+\alpha} \end{aligned}$$

Substituting the above terms into SMP payoff we get

$$\lim_{n\to\infty} u_P^d\left(\frac{p}{n}, \frac{q}{n}\right) = \frac{2(p+\alpha q)}{3(1+\alpha)^2} + \frac{2\alpha(\alpha p+q)}{3(1+\alpha)^2}$$

□

Proof of Proposition 1:

Proof. The SMP maximizes its payoff by choosing the optimal p and q.

The payoff of the SMP is concave and increasing in (p, q), equivalently, in (p, q). Therefore, the optimal p and q are obtained by taking the first order derivative w.r.t. p and q.

$$\frac{\partial}{\partial p}\left[\frac{2(p+\alpha q)}{3(1+\alpha)^2}+\frac{2\alpha(\alpha p+q)}{3(1+\alpha)^2}-c(p-p_0)^2-c(q-q_0)^2\right]=0$$

$$\frac{2}{3(1+\alpha)^2}+\frac{2\alpha^2}{3(1+\alpha)^2}-2c(p-p_0)=0$$

$$p=\frac{1+\alpha^2}{3c(1+\alpha)^2}+p_0$$

Similarly,

$$\frac{\partial}{\partial q}\left[\frac{2(p+\alpha q)}{3(1+\alpha)^2}+\frac{2\alpha(\alpha p+q)}{3(1+\alpha)^2}-c(p-p_0)^2-c(q-q_0)^2\right]=0$$

$$\frac{2\alpha}{3(1+\alpha)^2}+\frac{2\alpha}{3(1+\alpha)^2}-2c(q-q_0)=0$$

$$q=\frac{2\alpha}{3c(1+\alpha)^2}+q_0$$

The optimal p^* and q^* are then

$$p^*=p_0+\frac{1+\alpha^2}{3c(1+\alpha)^2}$$

$$q^*=q_0+\frac{2\alpha}{3c(1+\alpha)^2}$$

□

Proof of Corollary 1:

Proof. From Proposition 1,

$$p^* - p_0 = \frac{1+\alpha^2}{3c(1+\alpha)^2}$$
$$q^* - q_0 = \frac{2\alpha}{3c(1+\alpha)^2}$$

Since relative size $\alpha > 0$ and the cost parameter $c > 0$ it is clear that $p^* - p_0 > 0$ and $q^* - q_0 > 0$, thereby proving the corollary. □

Proof of Corollary 2:

Proof. From Proposition 1 we know that

$$p^* - q^* = p_0 - q_0 + \frac{(\alpha - 1)}{3c(1+\alpha)^2}$$

If the initial network exhibits weak homophily, that is, if $p_0 \geq q_0$ then $p_0 - q_0 \geq 0$. Also, $\frac{(\alpha-1)}{3c(1+\alpha)^2} \geq 0$ since relative-size $\alpha \geq 1$. Therefore, if $p_0 \geq q_0$ then $p^* \geq q^*$, thereby proving the corollary. □

Proof of Lemma 2:

Proof. When $p_0 = q_0$, corresponding p^* ad q^* are

$$p^* = p_0 + \frac{1+\alpha^2}{3c(1+\alpha)^2}$$

$$q^* = p_0 + \frac{2\alpha}{3c(1+\alpha)^2}$$

The SMP increases polarization if and only if

$$\frac{p^*}{q^*} \geq 1$$

$$p_0 + \frac{1+\alpha^2}{3c(1+\alpha)^2} \geq p_0 + \frac{2\alpha}{3c(1+\alpha)^2}$$

$$(\alpha-1)^2 \geq 0$$

Since $\alpha \geq 1$ the above condition holds and the SMP increases polarization. □

Proof of Lemma 3:

Proof. When $\alpha = 1$, the corresponding p^* ad q^* are

$$p^* = p_0 + \frac{2}{12c}$$

$$q^* = q_0 + \frac{2}{12c}$$

The SMP increases polarization if and only if

$$\frac{p^*}{q^*} \geq \frac{p_0}{q_0}$$

$$p_0 q_0 + \frac{2q_0}{12c} \geq p_0 q_0 + \frac{2p_0}{12c}$$

$$q_0 \geq p_0$$

Since the network exhibits weak homophily the above condition never holds and the SMP decreases polarization. □

Proof of Proposition 2:

Proof. The final polarization is

$$\frac{p^*}{q^*} = \frac{\frac{1+\alpha^2}{3c(1+\alpha)^2} + p_0}{\frac{2\alpha}{3c(1+\alpha)^2} + q_0}$$

The SMP increases polarization if and only if $P(p^*, q^*) \geq P(p_0, q_0)$. Let us evaluate this inequality.

$$\frac{(1+\alpha^2)/(3c(1+\alpha)^2) + p_0}{2\alpha/3c(1+\alpha)^2 + q_0} \geq \frac{p_0}{q_0}$$

$$q_0 \left(\frac{1+\alpha^2}{3c(1+\alpha)^2} \right) + p_0 q_0 \geq p_0 \left(\frac{2\alpha}{3c(1+\alpha)^2} \right) + p_0 q_0$$

$$q_0 \left(\frac{1+\alpha^2}{3c(1+\alpha)^2} \right) \geq p_0 \left(\frac{2\alpha}{3c(1+\alpha)^2} \right)$$

$$\frac{1+\alpha^2}{2\alpha} \geq \frac{p_0}{q_0}$$

□

Proof of Proposition 3:

Proof. Payoff of the SMP is

$$\begin{aligned} u_P^\lambda(p,q) &= n_1 g\left(\frac{k_{1n\lambda}}{n}\right) + n_2 g\left(\frac{k_{2n\lambda}}{n}\right) - c(p-p_0)^2 - c(q-q_0)^2 \\ &= n_1 \left(\frac{\left(\frac{k_{1n\lambda}}{n}\right)}{\left(\frac{k_{1n\lambda}}{n}\right)+1} \frac{\left(\frac{k_{1n\lambda}}{n}\right)+2}{\left(\frac{k_{1n\lambda}}{n}\right)+3} \right) + n_2 \left(\frac{\left(\frac{k_{2n\lambda}}{n}\right)}{\left(\frac{k_{2n\lambda}}{n}\right)+1} \frac{\left(\frac{k_{2n\lambda}}{n}\right)+2}{\left(\frac{k_{2n\lambda}}{n}\right)+3} \right) - c(p-p_0)^2 - c(q-q_0)^2 \end{aligned}$$

$$\lim_{n\to\infty} u_P^\lambda(p,q) = \lim_{n\to\infty} \frac{2n_1}{3}\frac{\left(\frac{k_{1n\lambda}}{n}\right)}{\left(\frac{k_{1n\lambda}}{n}\right)+1} + \frac{2n_2}{3}\frac{\left(\frac{k_{2n\lambda}}{n}\right)}{\left(\frac{k_{2n\lambda}}{n}\right)+1} - c(p-p_0)^2 - c(q-q_0)^2$$
$$= \frac{2k_{1\lambda}}{3(1+\alpha)} + \frac{2\alpha k_{2\lambda}}{3(1+\alpha)} - c(p-p_0)^2 - c(q-q_0)^2$$

As n becomes large we know that there exists $p, p_0,\ q, q_0 \in [0,1]$ such that

$$p_n = \frac{p}{n}, p_{0n} = \frac{p_0}{n}, q_n = \frac{q}{n}, q_{0n} = \frac{q_0}{n}$$

Therefore

$$\lim_{n\to\infty} k_{1n\lambda} = \lim_{n\to\infty} \lambda p\frac{n_1-1}{n} + q\frac{n_2}{n} = \frac{\lambda p + \alpha q}{1+\alpha}$$
$$\lim_{n\to\infty} k_{2n\lambda} = \lim_{n\to\infty} \lambda p\frac{n_2-1}{n} + q\frac{n_1}{n} = \frac{\lambda\alpha p + q}{1+\alpha}$$

Substituting the above terms into SMP payoff we get

$$\lim_{n\to\infty} u_P^\lambda\left(\frac{p}{n},\frac{q}{n}\right) = \frac{2(\lambda p+\alpha q)}{3(1+\alpha)^2} + \frac{2\alpha(\lambda\alpha p + q)}{3(1+\alpha)^2} - c(p-p_0)^2 - c(q-q_0)^2$$

The SMP maximizes its payoff by choosing the optimal p and q. The payoff of the SMP is concave and increasing in (p,q). Therefore, the optimal p and q are obtained by taking the first order derivative w.r.t. p and q.

$$\frac{\partial}{\partial p}\left[\frac{2(\lambda p+\alpha q)}{3(1+\alpha)^2} + \frac{2\alpha(\lambda\alpha p + q)}{3(1+\alpha)^2} - c(p-p_0)^2 - c(q-q_0)^2\right] = 0$$

$$\frac{2\lambda}{3(1+\alpha)^2} + \frac{2\alpha^2\lambda}{3(1+\alpha)^2} - 2c(p-p_0) = 0$$

$$p = \frac{\lambda(1+\alpha^2)}{3c(1+\alpha)^2} + p_0$$

Similarly,

$$\frac{\partial}{\partial q}\left[\frac{2(p+\alpha q)}{3(1+\alpha)^2}+\frac{2\alpha(\alpha p+q)}{3(1+\alpha)^2}-c(p-p_0)^2-c(q-q_0)^2\right]=0$$

$$\frac{2\alpha}{3(1+\alpha)^2}+\frac{2\alpha}{3(1+\alpha)^2}-2c(q-q_0)=0$$

$$q=\frac{2\alpha}{3c(1+\alpha)^2}+q_0$$

The optimal p^*_λ and q^*_λ are then

$$p^*_\lambda=\frac{\lambda(1+\alpha^2)}{3c(1+\alpha)^2}+p_0$$

$$q^*_\lambda=\frac{2\alpha}{3c(1+\alpha)^2}+q_0$$

The final polarization is

$$\frac{p^*_\lambda}{q^*_\lambda}=\frac{\frac{\lambda(1+\alpha^2)}{3c(1+\alpha)^2}+p_0}{\frac{2\alpha}{3c(1+\alpha)^2}+q_0}$$

The SMP increases polarization if and only if $\frac{p^*_\lambda}{q^*_\lambda}\geq\frac{p_0}{q_0}$.

$$\frac{\frac{\lambda(1+\alpha^2)}{3c(1+\alpha)^2}+p_0}{\frac{2\alpha}{3c(1+\alpha)^2}+q_0}\geq\frac{p_0}{q_0}$$

$$q_0\left(\frac{\lambda(1+\alpha^2)}{3c(1+\alpha)^2}\right)+p_0q_0\geq p_0\left(\frac{2\alpha}{3c(1+\alpha)^2}\right)+p_0q_0$$

$$q_0\left(\frac{\lambda(1+\alpha^2)}{3c(1+\alpha)^2}\right)\geq p_0\left(\frac{2\alpha}{3c(1+\alpha)^2}\right)$$

$$\lambda\left(\frac{1+\alpha^2}{2\alpha}\right)\geq\frac{p_0}{q_0}$$

□

Proof of Proposition 4:

Proof. The proof follows the steps similar to the proof of Proposition 1. □

Proof of Corollary 3:

Proof. The proof follows the steps similar to the proof of Proposition 2. □

Proof of Corollary 4:

Proof. We compare the final polarization $P(p_t^*, q_t^*)$ under tax rate t and final polarization $P(p_{t=0}^* q_{t=0}^*)$ under no taxation.

$$P(p_t^*, q_t^*) \leq P(p_{t=0}^* q_{t=0}^*)$$

$$p_t^* q_{t=0}^* \leq p_{t=0}^* q_t^*$$

$$\left[p_0 + \frac{(1+\alpha^2)}{3c(1+\alpha)^2}(1-t)\right]\left[q_0 + \frac{2\alpha}{3c(1+\alpha)^2}\right] \leq \left[p_0 + \frac{1+\alpha^2}{3c(1+\alpha)^2}\right]\left[q_0 + \frac{2\alpha}{3c(1+\alpha)^2}(1-t)\right]$$

$$p_0q_0 + \frac{2\alpha p_0}{3c(1+\alpha)^2} + \frac{(1+\alpha^2)q_0}{3c(1+\alpha)^2}(1-t) + \frac{(1+\alpha^2)}{3c(1+\alpha)^2}\frac{2\alpha}{3c(1+\alpha)^2}(1-t)$$

$$\leq p_0q_0 + \frac{2\alpha p_0}{3c(1+\alpha)^2}(1-t) + \frac{(1+\alpha^2)q_0}{3c(1+\alpha)^2} + \frac{(1+\alpha^2)}{3c(1+\alpha)^2}\frac{2\alpha}{3c(1+\alpha)^2}(1-t)$$

$$\frac{2\alpha p_0}{3c(1+\alpha)^2}t \leq \frac{(1+\alpha^2)q_0}{3c(1+\alpha)^2}t$$

$$\frac{p_0}{q_0} \leq \frac{(1+\alpha)^2}{2\alpha}$$

Since initial polarization is low the final inequality holds the corollary has been proven.

□