Universiteit
Antwerpen

This item is the archived peer-reviewed author-version of:

What's not to like? Facebook page likes reveal limited polarization in lifestyle preferences

Reference:
Praet Stiene, Guess Andrew M., Tucker Joshua A., Bonneau Richard, Nagler Jonathan.- What's not to like? Facebook page likes reveal limited polarization in lifestyle preferences
Political communication - ISSN 1091-7675 - 39:3(2022), p. 311-338
Full text (Publisher's DOI): https://doi.org/10.1080/10584609.2021.1994066
To cite this reference: https://hdl.handle.net/10067/1837440151162165141

uantwerpen.be
Institutional repository IRUA

What's Not to Like? Facebook Page Likes Reveal Limited Polarization in Lifestyle Preferences*

Stiene Praet†[a,e], Andrew M. Guess[b,e], Joshua A. Tucker[c,e], Richard Bonneau[d,e], and Jonathan Nagler[c,e]

[a]Department of Engineering Management, University of Antwerp
[b]Department of Politics, Princeton University
[c]Wilf Family Department of Politics, New York University
[d]Department of Biology, New York University
[e]Center for Social Media and Politics, New York University

December 5, 2021

Abstract

Increasing levels of political animosity in the United States invite speculation about whether polarization extends to aspects of daily life. However, empirical study about the relationship between political ideologies and lifestyle choices is limited by a lack of comprehensive data. In this research, we combine survey and Facebook Page "likes" data from more than 1,200 respondents to investigate the extent of polarization in lifestyle domains. Our results indicate that polarization is present in page categories that are somewhat related to politics — such as opinion leaders, partisan news sources, and topics related to identity and religion — but, perhaps surprisingly, it is mostly not evident in other domains, including sports, food, and music. On the individual level, we find that people who are higher in political news interest and have stronger ideological predispositions have a greater tendency to "like" ideologically homogeneous pages across categories. Our evidence, drawn from rare digital trace data covering more than 5,000 pages, adds nuance to the narrative of widespread polarization across lifestyle sectors and it suggests domains in which cross-cutting preferences are still observed in American life.

***Keywords* —** Polarization, Lifestyle, Facebook Likes, Digital trace data

*This study has been approved by the New York University Institutional Review Board (IRB-12-9058, IRB-FY2017-150). We are grateful to Samantha Luks at YouGov for facilitating survey data collection and Jona Ronen for development and technical assistance. Thanks to Jen Rosiere Reynolds, lab manager at the Center for Social Media and Politics at New York University, Kira Schumm, and Shraman Sen for excellent research assistance. We also thank Jan Zilinsky and Cody Buntain for helpful feedback. Praet's research was funded by University of Antwerp [FFB160362 BOF GOA]; the Center for Social Media and Politics at New York University is generously supported by funding from the U.S. National Science Foundation, the John S. and James L. Knight Foundation, the Charles Koch Foundation, the Hewlett Foundation, Craig Newmark Philanthropies, the Siegel Family Endowment, and NYU's Global Institute for Advanced Study.

†Corresponding author: stiene.praet@uantwerp.be

1 Introduction

During the last few U.S. presidential election campaigns, figures across diverse sectors of society used their platforms on social media to persuade or mobilize their fans. For inattentive or first-time voters, this may have offered a rare encounter with political advocacy as well as the authenticity that can be gained when mostly apolitical actors or entities take a stand (Zilinsky et al. 2020).

Fans of L.A. Lakers basketball player LeBron James (with over 23 million followers on Facebook) may be familiar with this dynamic — as well as its pitfalls. As James took a more prominent role supporting voter registration efforts and criticizing President Trump in the midst of the pandemic and a summer of roiling protests over racial justice, he found himself the target of attacks by the president himself. In a period of intense affective polarization at the mass level, politics threatened to ensnare one of the most popular celebrities in America.

Athletes and sports leagues have again found themselves at the crosshairs of political controversy in the United States, as symbolic demonstrations of racial solidarity in the form of kneeling protests have become increasingly commonplace, sometimes revealing divides between players and their fans. These episodes are an especially vivid illustration of how seemingly apolitical domains — including sports but also food, artistic and cultural preferences, and consumer decisions — can become caught in the partisan currents of the larger society, producing a mix that is not always appealing to citizens who would rather "keep politics out of sports" (Thorson and Serazio 2018).

More specifically for the focus of this study, they also demonstrate the importance of social media as an arena where lifestyle preferences in all their dimensions can intersect with politics. Given growing concern that polarization at its current historic level in U.S. politics is permeating society, this study asks *to what extent* this phenomenon is reflected in other realms. Widespread polarization across lifestyle domains would have serious implications because cross-cutting pressures in formally apolitical spheres may be critical for maintaining social harmony in otherwise highly polarized political systems (Mutz 2006; Pettigrew 1998).

In this study, we take rare advantage of comprehensive revealed preference information from page "likes" on Facebook — a social platform used by nearly 70% of Americans[1] — to understand whether preference sorting across various lifestyle categories follows the pattern established in partisan politics. We

[1] Source: https://www.pewresearch.org/fact-tank/2019/05/16/facts-about-americans-and-facebook/

contribute to the literature on "lifestyle politics," or the idea that political and ideological divisions extend to leisure activities, consumption choices, aesthetic taste, personal morality and other aspects of daily life (Bennett 1998; Giddens 1991, 2013; Khan et al. 2013; DellaPosta et al. 2015; Hetherington and Weiler 2018). By combining survey and Facebook "likes" data from more than 1,200 respondents in the United States, we directly test whether pages belonging to more "political" categories will be liked by more politically homogenous audiences and whether individual-level characteristics are associated with liking pages in categories exhibiting greater page-level polarization. We find, in contrast to some existing work, a clear divide between more polarized page categories that are commonly understood to be related to politics — such as opinion leaders, partisan news sources, and topics related to identity and religion — and other categories, including sports, with relatively low levels of ideological homogeneity within pages. On the individual level, we find that people with higher political news interest and stronger ideological predispositions are especially likely to endorse such ideologically homogeneous pages across categories.

This paper proceeds as follows. First, we briefly review related work on affective polarization, how this polarization may be reflected in lifestyle preferences and consumer choices, and the role of social media in facilitating these divides. This discussion leads to a summary of our main research questions and hypotheses. We then describe our data, measures, and research approach and present our results in several steps. We begin with a macro-level description of polarization across pages and page categories. We then zoom in to the individual level, examining characteristics associated with liking more polarized page categories. A discussion then concludes.

2 Polarization, Lifestyle, and Social Media

There is a longstanding debate about the nature and extent of polarization in American society, defined as party differences in issue positions or attitudes (Evans et al. 2001; Fiorina et al. 2008; Hetherington 2009; Baldassarri and Gelman 2008; Abramowitz and Saunders 2008). To synthesize this large body of research, it can be argued that today's high level of observed partisan polarization in ideological affiliations and issue positions is largely a reflection of increased sorting among the parties over time (e.g., conservatives into the Republican Party and liberals into the Democratic Party), and that evidence of increasing extremity over time is mixed, especially among the mass public (e.g., Lelkes 2016). More recently, scholarly attention has shifted to the affective dimension of polarization, rooted in an understanding of partisanship as a social identity

(Mason 2018b; Iyengar et al. 2019; Finkel et al. 2020). This aspect of polarization, distinct from specific attitudes, is particularly important for understanding how personal and emotional attachments formed in the political arena could potentially carry over to other domains. Such a process is suggested by the "oil spill" model of polarization, in which clusters of initially disparate issues — including cultural and moral issues — become connected in a belief system (DellaPosta 2020). The potential for polarization to spread beyond strictly political settings is also suggested in a conception of partisan attachment as reflecting a shared understanding of constituent social groups (Green et al. 2004). Related conceptions have likewise been proposed for understanding ideologies as a basis for group identification (Conover and Feldman 1981; Mason 2018a).

The literature on assortativity documents how political identities have begun to structure social behavior in everything from online dating behavior (Huber and Malhotra 2017) to marriage (Alford et al. 2011; Iyengar et al. 2012), to cultural consumption (Toff 2021; Bertrand and Kamenica 2018), while debate continues on the extent to which people choose to locate geographically in areas with like-minded partisans (e.g., Bishop 2009; Mummolo and Nall 2017). It is important to note that while average partisan differences have been clearly documented in these areas, gaps can be exaggerated relative to overall levels. Such a dynamic has been observed in persistent exaggerations of partisan differences in demographic composition and other common partisan stereotypes, for example (e.g., Ahler and Sood 2018).

Research similarly suggests a nuanced understanding of how social media reflects these patterns. A large study of retweet networks found much more ideological homogeneity between users tweeting about political topics than about nonpolitical topics, such as the 2013 Boston Marathon bombing in its initial aftermath and the 2014 Super Bowl (Barberá et al. 2015). But the marathon bombings themselves became politicized over time, and polarization in these retweet networks increased as a result. How topics can come to be seen as "political" or not is itself a challenging question, as Settle (2018) argues in the context of Facebook's News Feed. In Settle's theory, politically inattentive Facebook users come to make inferences about their more political friends by observing their posts and endorsements (including via likes). Through this process, associations come to form between political identities and lifestyle preferences. This is vividly illustrated in the book by the example of Chick Fil-A, which became a flashpoint in America's culture wars over the issue of same-sex marriage, which the chain's owner publicly opposed. To some, the choice of fast-food chain for a quick meal may not reflect political preferences, but such decisions can nonetheless take on a larger symbolic meaning to outside observers.

Settle's argument raises the question of whether social media, and Facebook specifically, is accelerating the process of politicizing lifestyle choices and preferences so that they more closely map onto the partisan political divide. Since our data provide a snapshot in time, this study cannot specifically answer this question, though it sheds light on the baseline levels of polarization across different areas of society. However, evidence is accumulating for the specific mechanisms likely at play, namely inferences due to apolitical cues (e.g., Lee 2020).

3 Hypothesis and Research Questions

A large amount of research on social media, including Twitter (Barberá 2015; Boutyline and Willer 2017; Eady et al. 2019) and Facebook (Bakshy et al. 2015; Bond and Messing 2015), focuses on retweeting, following or liking of political figures, news media or other political content. Despite Facebook's importance for our understanding of social media and political communication, average users of the social platform likely do not primarily use it to follow news or engage in politics. This study focuses on a key affordance: the ability to "like" public pages. This behavior has been conceptualized as a form of social endorsement, since these likes are publicly observable by friends in one's network (Bond and Messing 2015). Liking a page is also a meaningful indicator of affinity, since it signals a user's desire to see more posts from the page's publisher. There is some tension between interpreting Facebook page liking as endorsement behavior and as revealed preferences. We acknowledge that the social observability of likes may introduce an expressive component, which implies that they are not always pure indicators of private preferences. At the same time, such behavior is likely to be more indicative of genuine preferences than overt engagement (and certainly more credible than self-reported survey responses), not least because it can be seen as a costly signal (Barberá 2015; Bode 2017). We propose an "expressiveness continuum" of online behavior with private web browsing on one end (Guess 2021), sharing on the other (Guess et al. 2019), and likes (or perhaps follows in the case of Twitter) in between (Bond and Messing 2015). Further, whether people perceive likes to be "expressive" might depend on their own awareness of the observability of their actions on Facebook. This awareness is not necessarily natural and is an important component of internet skill (Hargittai and Micheli 2019), which varies considerably in the population. Ultimately, while keeping these considerations in mind, we argue that Facebook page likes constitute an ecologically valid, approximate behavioral indicator of social and lifestyle preferences.

We analyze Facebook likes to gain insights into which aspects of people's social lives appear to reflect strong ideological divides. We construct measures of the ideological *homogeneity* of users who like individual pages, allowing us in turn to draw conclusions about the overall polarization of liking patterns across "political" and "non-political" categories. At the page level, we first ask: *How ideologically homogeneous are political and lifestyle categories on Facebook? (RQ1)* We expect political pages to exhibit relatively high levels of ideological homogeneity since online endorsement of political figures is a reflection of political ideology and opinion (Bond and Messing 2015). Prior research has found that motivations to "like" a political page include self-reflection, showing political interest, and seeking engagement with others (Macafee 2013). Although there are various reasons to refrain from liking the Facebook pages of political candidates (such as social anxiety and audience diversity; see Marder 2018), it is relatively uncommon to like the Facebook page of an opposing candidate or party.

We may similarly expect divides in the audiences of news sources' Facebook pages, though more so for partisan news than hard news sources. The most recent evidence using behavioral data on individuals' online media consumption suggests nontrivial overlap in Democrats' and Republicans' news diets, which consist to a large extent of relatively centrist mainstream outlets and large portals (Guess 2021). Data on people's follow networks on Twitter similarly suggest a meaningful amount of exposure to cross-cutting news (Eady et al. 2019). Twitter allows users to follow accounts without this necessarily being visible to the public (e.g., by creating a "private" list). However, the more observable nature of people's page preferences on Facebook suggests a potentially greater role for political identities to shape self-presentation, which could affect conscious decisions to like sources perceived as politically congenial. In this way, we expect patterns of Facebook news likes to more closely mirror survey evidence (e.g., Newman et al. 2019; Jurkowitz et al. 2020), which scholars have argued can be distorted by biases in recall, partisan bias, and the difficulty of the task itself (e.g., Prior 2009, 2013).

Finally, even non-political lifestyle categories (arts and culture, food, sports, etc.) have been found in studies based on survey data to map onto ideological divides (DellaPosta et al. 2015; Hetherington and Weiler 2018), though we would expect ideological diversity to be higher for lifestyle domains than for more explicitly political categories. To the best of our knowledge, the only social media study that has explicitly focused on the partisan divide in non-political domains in the United States was conducted by Shi et al. (2017), who analyze Twitter co-following networks.[2]

[2]Facebook conducted an informal, proprietary analysis in 2014 focusing on the music, TV,

Next, we turn to the individual level and ask whether traditional predictors of strong political engagement (e.g., Saunders and Abramowitz 2004) also predict liking relatively polarized *non-political* pages. This would indicate that the predictors of polarized endorsements of traditionally political content carry over into lifestyle and related domains. Therefore, we propose the following hypothesis: *People with greater political interest and stronger ideological affiliations are more likely to like politically homogeneous Facebook pages in non-political categories (H1).* Thus for both political and non-political pages we would expect "very liberal" and "very conservative" ideological affiliations and high levels of political news interest to be significantly related to high page homogeneity scores, all else equal. High page homogeneity scores tell us that the audience of a Facebook page is similar in terms of their (self-reported) ideology.

To further understand how individual-level characteristics relate to preferences for lifestyles and habits shared among relatively homogeneous groups, we build upon research from political psychology to explore which other characteristics are associated with higher levels of political homophily,[3] and ask a second research question: *Is left vs. right* ideology associated with a greater likelihood of "liking" ideologically homogeneous lifestyle pages? (RQ2) We focus on how liking patterns for lifestyle pages differ from political pages and news and media pages. There are conflicting expectations in the literature regarding ideological differences. Jost et al. (2009) argue that conservatives and liberals differ in their need for certainty, making intolerance of ambiguity more typical of the political right as compared to liberals' greater openness to new experiences and cognitive complexity. On the other side of the asymmetry debate, Greenberg and Jonas (2003) show that individuals on *either* ideological extreme possess greater preference for certainty than more moderate ones (though see Guay and Johnston 2020). (Debate on these questions continues with a recent exchange between Ditto et al. 2019 and Baron and Jost 2019.) This debate intersects with questions about lifestyle preferences, since research on ideological asymmetries often explores differences in personality characteristics, which can also be reflected in aesthetic preferences and openness to new experiences and diversity (e.g., Carney et al. 2008; Bachrach et al. 2012). To the extent that asymmetries manifest in differing cognitive styles and political preferences, they may be correlated with lifestyle preferences as well. At the same time, there is some evidence that liberals prefer more homogeneous content: Bakshy et al. (2015) find

and other cultural preferences of users who liked official political pages during the midterm election campaign. See `https://www.facebook.com/notes/facebook-data-science/politics-and-culture-on-facebook-in-the-2014-midterm-elections/10152598396348859/`.

[3]Political homophily refers to the tendency to associate with others who are similar in political ideology (for a review, see Boutyline and Willer 2017).

that liberals tend to be connected to fewer ideologically dissimilar friends on Facebook than conservatives, while Eady et al. (2019) find that liberals are less likely to follow media and political accounts classified as right-leaning than vice versa.

4 Data and Methods

4.1 Data Collection

A panel survey on (social) media use was conducted during the 2016 U.S. presidential election ($N = 3,500$) over three waves. The respondents were asked to complete a survey and to indicate their ideology on a 5-point scale (see results in Table 1). In Wave 1 (April 9–May 1, 2016), we also asked respondents to indicate social networking sites for which they had accounts (options: Twitter, Facebook, Instagram, LinkedIn, Snapchat, and other).

Table 1: Sample details

Ideology	Respondents
Very liberal	189 (16%)
Liberal	218 (18%)
Moderate	360 (30%)
Conservative	175 (14%)
Very conservative	92 (8%)
Not sure	51 (4%)
No answer	126 (10%)

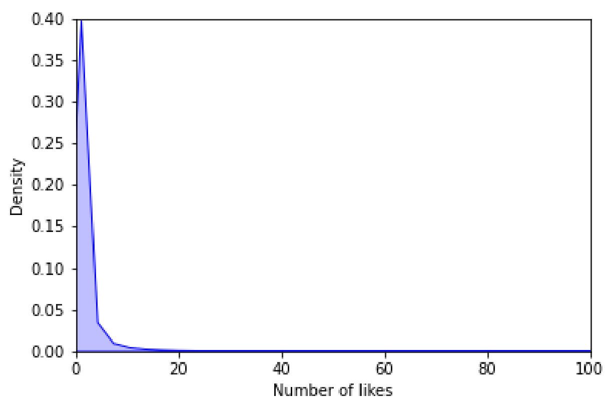
After the election we asked respondents if they would be willing to supply information about their own past Facebook activity. This was done via a separate survey question that sent respondents to a web application facilitating an authenticated link to the Facebook API. $1,331$ respondents consented to let us retrieve their Facebook information[4]. Specifically, we requested their public profile information, Timeline posts (including text and links if available), page likes, and what Facebook saves as religious and political views. If a respondent chose to log into Facebook after the survey prompt, they were asked what specific pieces of information they were willing to share. They could approve sharing all of the given types of

[4]Respondents who consented to provide their Facebook data were compensated with an additional $5 in YouGov "points" above what they received for taking the survey

information, selectively approve only some of these types of information, or approve none of them.[5] Of the $1,331$ respondents (comprising 45% of the 2,711 respondents who reported having a Facebook account) who agreed to share Facebook profile data, $1,230$ could be successfully linked back to the survey for our study. Based on observable characteristics, the subgroup for which we have Facebook data is a fairly representative cross section of the overall sample (see Table SA1 in Appendix A). However, those who shared data were slightly more liberal on average, more likely to participate in elections, and more politically engaged.

The Facebook profile data used in this study consists of *public page likes*, i.e. the public Facebook pages that a user likes and that show up as being liked in the "About" section of that person's profile. When someone likes a page, they are showing support for the page and indicating that they want to see content from it.[6] This dataset consists of 387,671 unique Facebook pages. The majority of these pages are liked by fewer than 5 respondents in our dataset (see Figure 1).

Figure 1: Page "likes" distribution



To ensure that our results are not being impacted by small numbers of people liking particular pages, we restrict our analysis to pages that are liked by at least 30 respondents each.[7] Two independent coders were trained to categorize all Facebook pages into 24 predefined categories (e.g. politics, news, sports, or food).[8] Since the categories are not mutually exclusive (for example, LeBron James is included in **Sports** as well

[5]No data on News Feed content or exposure was shared with researchers. Data access was temporary and lasted only 2 months after permission was granted. All respondents who agreed to share information consented to a privacy policy that specified, in part, "This application will not access the profile information of any friends, groups, or other information associated with your profile page."

[6]see `https://www.facebook.com/help/171378103323792`

[7]This way, the number of pages included in our analysis is reduced to 5,155, still accounting for almost 25% of total page likes. In Appendix D.1, we show that the individual-level "like" distributions of all pages and pages with a minimum of 30 likes are highly similar. We acknowledge that the results and conclusions in this study are based on relatively popular Facebook pages and that we cannot analyze polarization on smaller pages.

[8]Details on the coding task can be found in Appendix B

as **Public Figures**), the coders coders could assign a maximum of 3 categories per page; they agreed on at least one category for 70% of the Facebook pages. The final categories and coding results can be found in Table 2.

At the highest level, we distinguish between three groups: (1) Political pages, (2) News & Media, and (3) all other "Lifestyle" categories.[9] We further break the second group down to (2a) Partisan news and (2b) Hard news. This was done manually by the authors. "Hard news" refers to non-partisan news sources. "Partisan news" refers to partisan news sources such as HuffPost. News outlets that combine "hard news" with clearly partisan opinion desks (e.g., Fox News) are included in both categories. News categories related to lifestyle, celebrities, sports, and science are included in group (3) Lifestyle.

[9]Civil Society, Public Figures, Individual opinion leaders, Research & Education Arts & Culture, Tv Shows, Entertainment, Movies, Interests, Music, Sports, Beauty & Health, Food & Beverage, Shopping & retail, Travel, Cars and transportation.

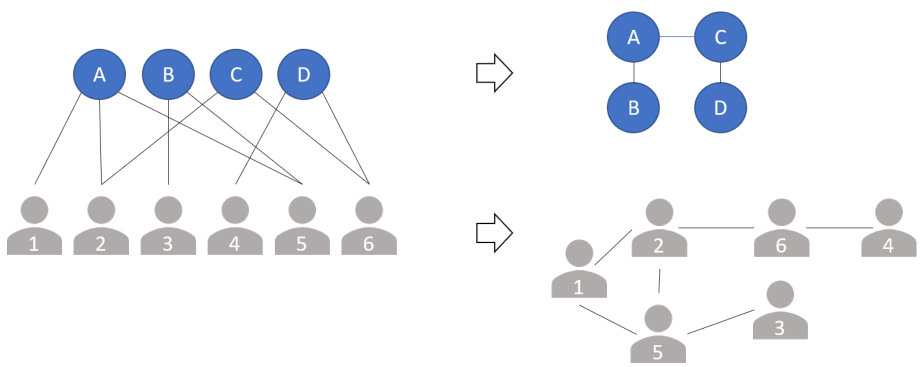
Table 2: Description of the Facebook categories and the number of pages per category. The categories are not mutually exclusive, which is why the number of pages per category do not sum up to the total number of pages.

Category	Description	# pages
Shopping & retail	Apparel, accessories, clothing, fashion, consumer electronics, home decoration, stores, shopping mall, wholesale, etc.	1858
Public Figures	Public figures	876
Food & Beverage	Food, cooking, restaurants, drinks, spirits, breweries etc.	837
Entertainment	Entertainment, games, humor, amusement, comedy etc.	532
Music	Music, bands, producers, record labels, albums, awards, concerts, music festivals etc.	499
Tv Shows	TV shows, episodes, channels, TV awards	463
Political	Politicians, political parties, political content, political communities and government organizations	353
Movies	Movies, actors, directors, movie characters, cinema and awards	341
Services	Marketing, advertising, legal, finance, consulting, etc.	339
Beauty & Health	Cosmetics, healthcare, medical	337
Civil Society	Nonprofit organizations and labor unions (formal organizations)	211
Partisan news	News and media about politics	188
Interests	Interests, communities (informal) and hobbies	166
Arts & Culture	Arts, culture, photography, museums, artists, musicals, theater, literature, libraries, writers, etc.	110
Hard news	Factual reportage of events which are socially or politically significant and of a serious nature	103
Sports	Sports, teams, athletes, leagues, games, gym	99
Cars and transportation	Car brands, automotive, airlines, boats, etc.	74
Identity & Religion	Pages referring to home country, region, ethnic or cultural groups, religious pages, religious organizations	72
Travel	Travel, tour agencies and tourism	70
Individual opinion leaders	Individual influencers, bloggers, commentators, etc.	51
Research & Education	Schools, universities, student organizations, educational programs, (non-)scientific research	43
TOTAL		5155

4.2 Methods and Measures

We start with network analysis and community detection to study the extent of political polarization on political and lifestyle-related pages on Facebook (RQ1). We first create a bipartite network in which Facebook users are the bottom nodes and Facebook pages are the top nodes. An edge exists between a user and a page when the user has liked the page. Next, we project the bipartite network to a homogeneous unigraph of the bottom nodes, where Facebook users are linked if they have liked a common Facebook page (see Figure 2).

Figure 2: Bipartite network with like relationships between users (bottom nodes) and Facebook pages (top nodes), and the top- and bottom-node projections.



The weight of the edge between each pair of nodes captures the number of shared Facebook pages. The Louvain algorithm (Blondel et al. 2008) is applied on this bottom node projection to detect communities of highly connected users. The algorithm is based on modularity optimization. Modularity (see Equation 2 in Appendix C) is the relative density of edges inside communities with respect to edges outside of communities and measures the extent to which a network is divided into different clusters or communities. Networks with high modularity have dense connections between the nodes within clusters but sparse connections between nodes in different clusters. In practice, a modularity value above 0.3 is a good indicator of significant community structure in the network (Clauset et al. 2004). The existence of ideologically different communities would indicate polarization in the network, since ideologically similar individuals are more strongly connected to each other than to ideologically dissimilar ones.

Separately, we perform a top node projection of the bipartite network, where the weights of the edges between Facebook pages reflects the audience overlap, and we repeat the analysis.

Next, we analyze the ideological homogeneity[10] of different lifestyle categories and users on Facebook

[10]We focus on ideology, though the analyses presented in this study could be done using party identification instead.

in more detail. We measure the ideology and homogeneity of individual Facebook pages based on the liking behavior of our respondents and average this over categories and individuals. In the following discussion, consider Facebook page Z, self-reported individual ideology score k for $k = 0, ..., 4$ (where 0 = very liberal and 4 = very conservative), and ideology class c that groups these ideology scores k into three groups, with $c = 0, 1, 2$ (where 0 = liberal, 1 = moderate, and 2 = conservative).

Measuring page ideology Using like behavior and self-reported ideology of the respondents in our sample, we map the ideologies of Facebook pages. Several behavioral approaches can be found in the literature. For example, Bakshy et al. (2015) estimate the ideology of news media by calculating the difference in the proportion of self-reported liberals and conservatives who share links to such media on Facebook, while Messing et al. (2017) calculate media ideology by averaging the NOMINATE scores of members of Congress who share news media URLs. Similar to these approaches, we average the self-reported ideology scores (k) of respondents who liked Facebook page Z to calculate the page ideology score of Facebook page Z, which ranges from 0 to 1. To adjust for uneven partisan distribution, we add a correction factor of 0.06 to each page's ideology score.[11]

Measuring page homogeneity Page homogeneity tells us how diverse the audience of a Facebook page is in terms of their ideology. To assess page homogeneity, we use the chi-square statistic.[12] This statistic is used by authors such as Desmet et al. (2017) to measure overlap between ethnicity and culture and Selway (2011) to measure cross-cuttingness. It has the advantage that it takes into account the prior distribution of ideology in our sample when calculating homogeneity.

We consider three ideology groups (c), i.e., $liberal$, $moderate$, and $conservative$. The fraction of likes from users with ideology c is equal to p_c (see Table SA2). The chi-square statistic is based on comparing

In fact, page homogeneity scores based on ideology and 3-point party identification are almost perfectly correlated in our data.

[11] As liberals outnumber conservatives in our dataset (see Table SA2 in the Appendix), the average ideology score across all page likes turns out to be less than 0.5 (0.44). As a result, a Facebook page that is liked at the same rate by liberals, moderates, and conservatives in our sample would have a page ideology of 0.44. Therefore, we add a correction factor of 0.06 to each page's ideology score. For example, the page ideology of the Facebook page "Independent Voter" is 0.46 without the correction factor and becomes 0.52 when applying the correction. Note that this correction factor shifts the distribution of page ideology to center around 0.5 but does not affect the relative distance between different pages' ideologies. Shi et al. (2017) apply a similar method to adjust for uneven partisan distribution.

[12] As there are several other metrics could be used to measure homogeneity or audience diversity (Bhadani et al. 2020), we compare results using entropy and variance in Appendix D.2.

the distribution of ideology groups across users who have liked Facebook page Z to the distribution across users who have not liked Z. If both distributions are the same, then knowing whether a user liked Facebook page Z or not conveys no information about their ideology. If instead the distributions are distinct, then the audience of Facebook page Z is more ideologically homogeneous than the overall population.

Let N_c^1 be the count of Facebook users who liked Facebook page Z and belong to ideology group c and N_k^0 be the count of users in ideology group c that did not like Z. Under independence, the expected number of individuals that belong to ideology group c and like ($i = 1$) or not like ($i = 0$) page Z is $N^i \times p_c$, while the observed frequency is N_c^i. The statistic for page Z is equal to the deviation of observed values and expected values and is given by:

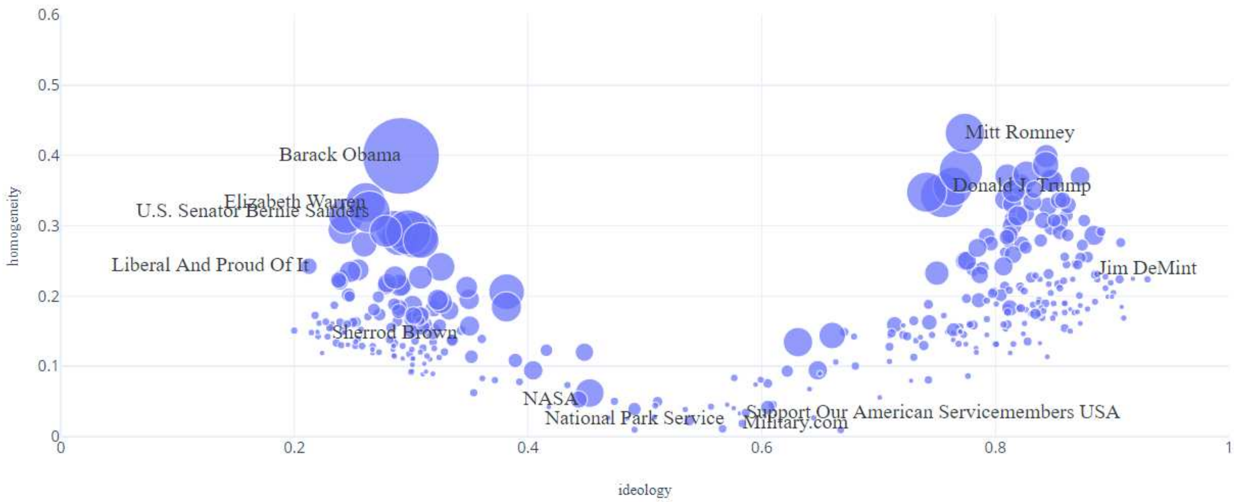
$$\chi^2(Z) = \sum_{i=0}^{1} \sum_{c=0}^{3} \frac{(N_c^i - N^i \times p_c)^2}{N^i \times p_c} \tag{1}$$

To ensure that this value lies between 0 and 1 we use Cramer's normalization (Equation 3 in Appendix C). We will consider political pages as a baseline category, and discuss homogeneity of the other categories relative to this reference category.

Finally, the ideology and homogeneity scores in each category are calculated by averaging the page ideology and homogeneity scores of all pages per category, weighted by the total number of likes per page. The average homogeneity scores per Facebook category provide us with an answer to RQ1. As an example, we show the ideology and homogeneity scores of political pages in Figure 3. We map the ideology score per Facebook page on the x-axis, and the homogeneity score on the y-axis. Note that page ideology and page homogeneity are related analytically. Pages with a very homogeneous liberal or conservative audience will accordingly have a very low or high page ideology score, while pages with a diverse audience will have a moderate page ideology score, hence the U-shaped graphs. A moderate page ideology score combined with high homogeneity score indicates a predominantly moderate audience (e.g. Serena Williams in Figure SA5e), but these pages are rare. For homogeneous categories the majority of pages will be located at both ends of the U-shaped graph, while for more heterogeneous categories the majority of pages will be located in the bottom-center of the graph.

Similarly, for each user, we average the homogeneity scores of all Facebook pages they have liked. A high homogeneity score indicates that the user tends to like more ideologically homogeneous pages. We will build a regression model to provide an observational portrait of the individual-level characteristics related

Figure 3: Ideology and homogeneity scores for the Facebook pages in the category political pages. The magnitude of the circle represents the total number of likes of the Facebook page.



to high homogeneity scores. To test H1 and answer RQ2 and RQ3, we include include seven-point party identification, and political news interest. As additional control variables, we include a mix of relevant sociodemographic variables including age, race, gender, family income, and educational attainment. The dependent variable is the individual homogeneity score measured by Cramer's V. Because of the nature of the dependent variable (between 0 and 1), beta regressions are used.

Our approach to measure ideology and homogeneity, while straightforward, may seem to introduce a concern about circularity of measures: we seek to explain the relationship between individuals' ideological self-placements and preferences for ideologically homogeneous pages, but we use the former as a way to estimate the latter. In order to scale Facebook pages in a way analogous to Barberá (2015) on Twitter or Bond and Messing (2015) on Facebook, we would require comprehensive data on pages and the users who like them to estimate the desired ideal points. Since such data at the scale of the Facebook population are impossible for outside researchers to acquire, the next best alternative is the data we have, which enable averaging of ideological self-placement of users who like a page (similar to Bakshy et al. 2015). The challenge is that this is the same sample we need to test hypotheses at the individual level explaining variation in our dependent variable.

We address this concern in three ways. First, it is important to note that our dependent variable is ideological homogeneity, which is a measure of the diversity (or variance) of the audience of a Facebook page in terms of their ideology (which are indeed based on self-reported ideology). Thus, we are using self-reported ideology to explain individual tendency to like Facebook pages with ideologically non-diverse audiences, not to explain the mean ideology of Facebook pages. Second, we replicate our analyses using party identification rather than ideological self-placement as a predictor and obtain qualitatively similar results (see Table SA7). Third, we validate our page ideology estimates using an external measure of ideology. Ideally, we could obtain existing page ideology scores matching our set. While there is no such data set of which we are aware, we are fortunate that 78 of our pages correspond to news domains found in Bakshy et al. (2015). When we correlate the page ideology scores derived entirely from our own data with the "alignment score" estimates of Bakshy et al. (2015), we find a remarkably high correlation of $r = 0.969$.

5 Results

We begin this section with some preliminary insights based on network analysis and an investigation of the page ideology distributions. We then turn to addressing our research questions and hypotheses.

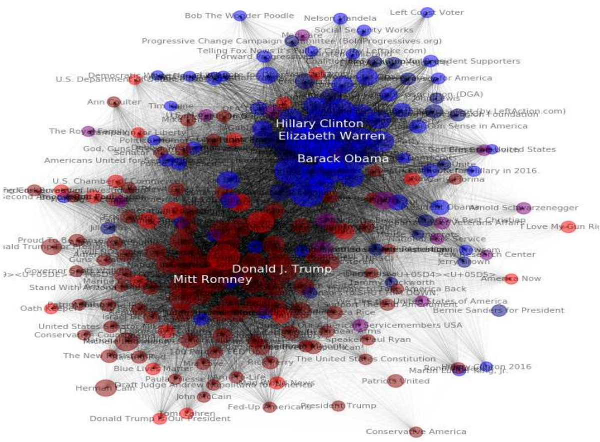
The results of the network analysis and community detection reflect the extent of political polarization on political and lifestyle-related pages on Facebook (RQ1). The top node projection[13] (see Figure 4) reveals a clear ideological divide for the political pages based on Democratic and Republican presidential candidates. The community detection algorithm reveals distinct liberal and conservative communities with high modularity (see Table SA9 in Appendix E.1). For partisan news, there is an overlap between partisan liberal and mainstream news outlets, and they are concentrated around outlets such as *The New York Times*, HuffPost, and CNN; Fox News and Conservative Daily are clearly separated from the others. This seeming asymmetry recalls the analyses in Benkler et al. (2018), who describe links between conservative media and extreme partisan sites that frequently publish misleading content. For the hard news subcategory a similar pattern can be observed, except that the mainstream news outlets have a more central location in the network and form a community of their own. Strikingly, in contrast to these patterns, no ideologically distinct communities appear to emerge within the lifestyle category.

Related to our first research question, these results provide some preliminary evidence that political and partisan news pages are relatively polarized, while we find no indication of polarization in lifestyle domains.

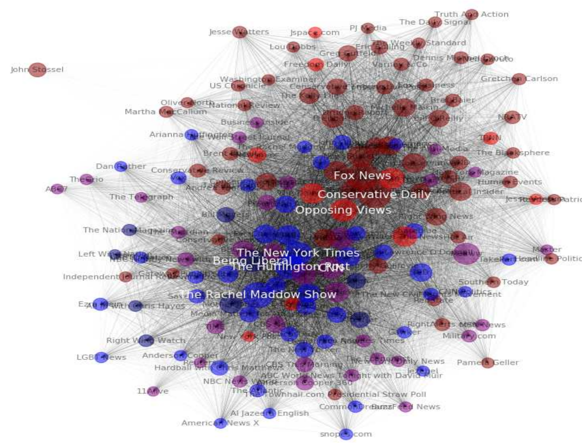
Next, we look at the page ideology distribution among liberal, moderate, and conservative respondents (see Figure 5). We use the overlap of these distributions as a measure of the degree to which users like ideo-

[13]Results for the bottom node projection can be found in Appendix E.1.

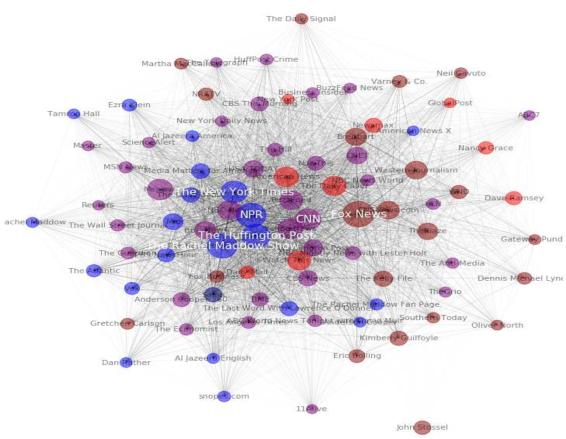
Figure 4: Network visualization and modularity (M) for the top node projection of (a) political pages, (b) partisan news, (c) hard news, and (d) lifestyle pages. The size of the bubble represents the total number of likes of the page, and the color represents the average ideology of the audience.



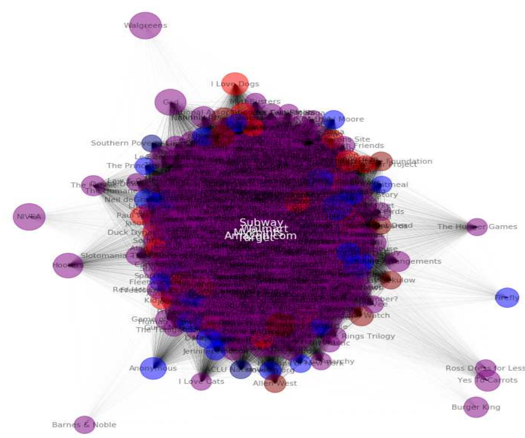
(a) Political ($M = 0.40$)



(b) Partisan news ($M = 0.28$)



(c) Hard news ($M = 0.17$)



(d) Lifestyle ($M = 0.13$)

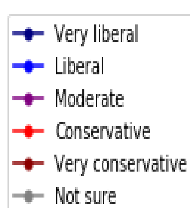
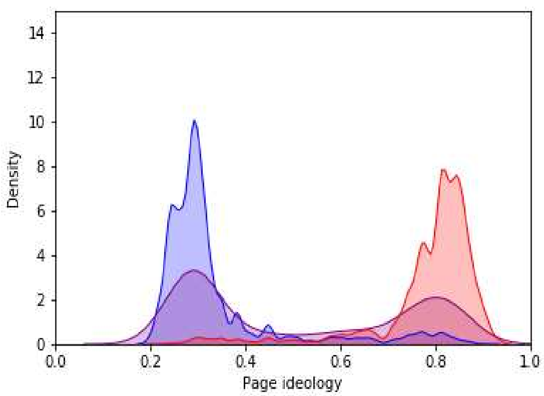
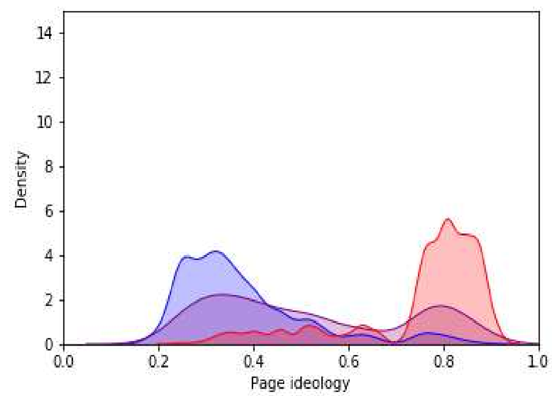


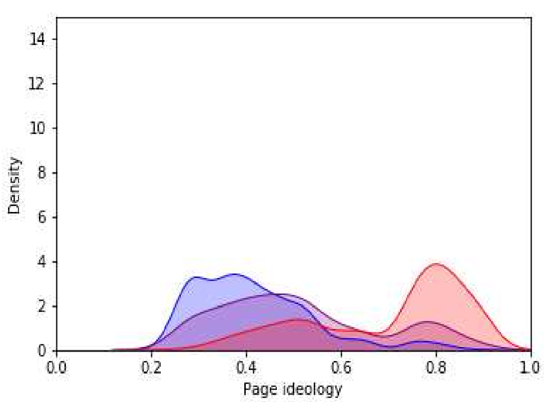
Figure 5: Page ideology distribution for liberals (blue), moderates (purple) and conservatives (red) when taking into account (a) political pages, (b) partisan news, (c) hard news, and (d) lifestyle pages; and the overlapping coefficient (OVL) for the liberal and conservative distribution.



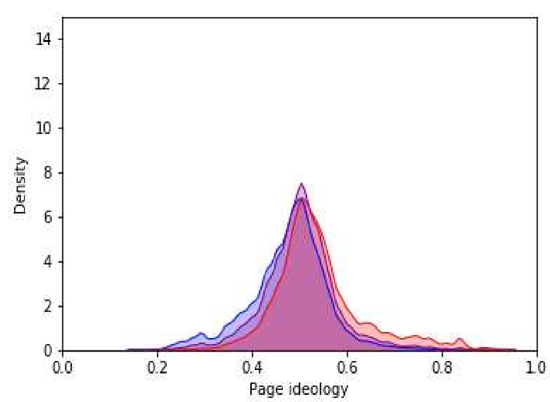
(a) Political ($OVL = 0.15$)



(b) Partisan news ($OVL = 0.24$)



(c) Hard news ($OVL = 0.36$)



(d) Lifestyle ($OVL = 0.75$)

logically similar pages (see Eady et al. 2019).[14] The overlap is smallest between liberals and conservatives for political pages; it increases for partisan news and hard news, though it is still low to moderate. The low overlapping coefficient for our news categories is especially notable in its contrast to estimates of the same statistic for news consumption through website visits on desktop, laptop and mobile devices (Guess 2021). A likely source of this divergence is the relative absence on social media of the potentially moderating influence of news portals, aggregators, and popular mainstream website homepages — a reflection not only of differing affordances but of distinct uses and gratifications, which on Facebook may include motivations for identity signaling and affirmation in addition to simply seeking out information (Settle 2018).

Meanwhile for lifestyle-related pages on Facebook, the distributions almost completely overlap, which again confirms our observation that lifestyle pages do not exhibit a strong ideological divide. Finally, for all types of pages, moderate participants have a larger overlap with liberals than with conservatives (see Table SA10). In the following subsection, we address the first research question in more detail by analyzing the ideological homogeneity of different lifestyle categories.The average page homogeneity is a quantitative metric capturing the extent of polarization across all pages in a certain category.

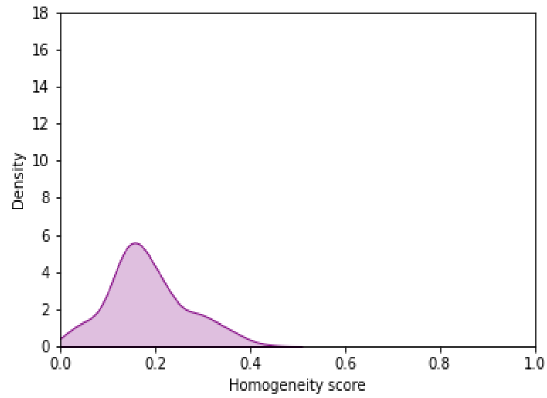
5.1 Page Homogeneity Across Facebook Categories

To address **RQ1**, we first show the page homogeneity distribution for political, hard news, partisan news and lifestyle pages in Figure 6. Visually, we observe that page homogeneity scores are highest in the political category and lowest in the lifestyle category. In other words, the majority of lifestyle pages has low homogeneity scores. To analyze to what extent this finding is driven by the fact that the vast majority of traffic to pages is to the popular and diverse ones, in the Supplemental Appendix we represent the cumulative distribution of page likes from low to high homogeneity pages (see Figure SA4). We find that for lifestyle pages, indeed the number of likes accumulates faster for pages with low homogeneity scores, whereas for political pages the most diverse pages are less popular.

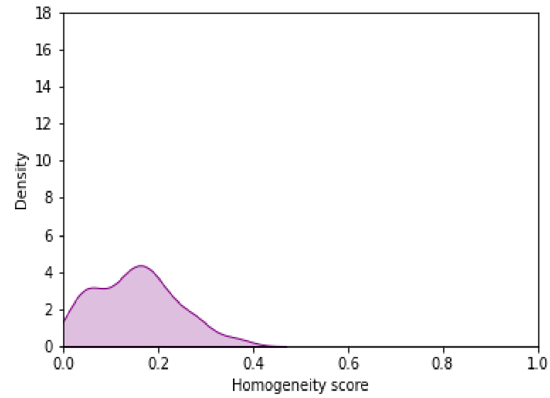
Secondly, Table 3 summarizes the average ideology and homogeneity scores of the Facebook pages in each category. **Political** and **Partisan news** pages have the highest homogeneity scores and serve as a benchmark for polarization. Several other categories that are somewhat related to politics are relatively polarized as well: **Hard news**, **Civil society**, **Identity & religion**, **Individual opinion leaders**, and **Public figures**.

[14]We use the `overlap` package in R (Meredith and Ridout 2014) to calculate the area lying under both of the density curves.

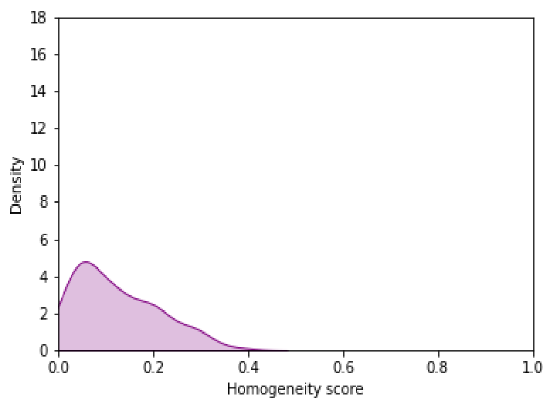
Figure 6: Page homogeneity distribution for (a) political, (b) partisan news, (c) hard news, and (d) lifestyle pages.



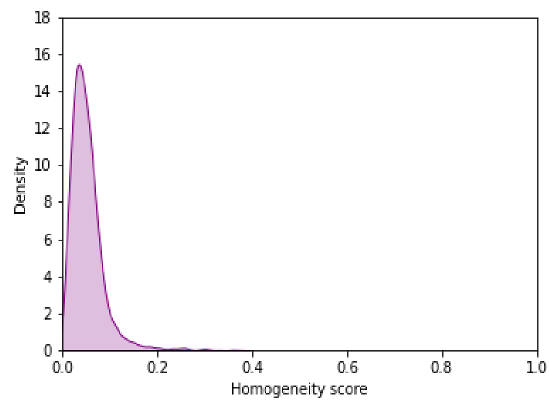
(a) Political



(b) Partisan news



(c) Hard news



(d) Lifestyle

Although they are around half as homogeneous as political pages, these categories are more homogeneous than the database average. Most other lifestyle categories, however, show much less polarization across ideological lines. Among the least polarized categories we find are **Shopping & retail**, **Food & beverage**, and **Cars & transportation**. These pages are more then 5 times less polarized than political pages, and are liked at almost equal rates by liberals, moderates, and conservatives.

Table 3: Weighted average (and standard deviation) for homogeneity, measured by Cramer's V, and page ideology for all Facebook categories in the U.S., ordered from high to low homogeneity.

Category	Homogeneity	Ideology
Political	**0.22 (0.09)**	**0.57 (0.25)**
Partisan news	**0.18 (0.10)**	**0.57 (0.23)**
Hard news	0.15 (0.10)	0.55 (0.20)
Civil Society	0.12 (0.09)	0.51 (0.19)
Identity & Religion	0.12 (0.05)	0.66 (0.13)
Individual opinion leaders	0.12 (0.11)	0.60 (0.17)
Public Figures	0.10 (0.09)	0.51 (0.16)
Arts & Culture	0.07 (0.06)	0.48 (0.13)
Tv Shows	0.07 (0.05)	0.48 (0.11)
Entertainment	0.07 (0.05)	0.51 (0.10)
Research & Education	0.06 (0.04)	0.46 (0.09)
Music	0.06 (0.03)	0.50 (0.10)
Interests	0.06 (0.03)	0.49 (0.10)
Movies	0.06 (0.04)	0.48 (0.10)
Sports	0.05 (0.03)	0.50 (0.08)
Services	0.05 (0.04)	0.51 (0.06)
Beauty & Health	0.04 (0.02)	0.50 (0.05)
Travel	0.04 (0.02)	0.52 (0.06)
Shopping & retail	0.04 (0.02)	0.51 (0.05)
Food & Beverage	0.04 (0.02)	0.50 (0.05)
Cars and transportation	0.04 (0.02)	0.52 (0.06)
Total	0.07 (0.07)	0.51 (0.11)

A more detailed analysis of the individual pages per category (see Appendix E.2) sheds light on the most and least polarizing pages. For political pages (see Figure SA5a in the Supplemental Appendix), the Democratic and Republican political candidates hold the highest homogeneity scores (e.g., $V(BarackObama) = 0.39$ and $V(MittRomney) = 0.43$). Some government organizations such as NASA ($V = 0.05$) and the National Park Service ($V = 0.01$) are liked by a heterogeneous audience, but overall the number of pages with low homogeneity in this category is low. Next to political pages, also **Partisan news** and **Hard news** show fairly high homogeneity scores, which is consistent with our network analysis results. Figure SA5b and SA5c show that Facebook audiences for news outlets are often heavily right- (Fox News [$V = 0.38$, $I = 0.77$] and Conservative Daily [$V = 0.34$, $I = 0.76$]) or left-leaning (*The New York Times* [$V = 0.19$, $I = 0.36$] and NPR [$V = 0.28$, $I = 0.32$]), with relatively few outlets attracting people with different ideologies (CNN [$V = 0.06$, $I = 0.45$], Meaww [$V = 0.01$, $I = 0.51$] and *The Los Angeles Times* [$V = 0.00$, $I = 0.51$]).

Even within the least-polarized categories, individual pages with high homogeneity scores do exist. Looking within the **Food & beverage** category (see Figure SA5f), we see that, as the discussion in Settle (2018) suggests, Chick-fil-A ($V = 0.17$, $I = 0.61$) does have a relatively high homogeneity score in addition to its more conservative ideology rating. As Settle (2018) recounts, the chain encountered controversy in 2012 about its owner's (and charitable arm's) support for anti-gay organizations, after which activists (mainly liberals) announced a boycott of the restaurant, and others (mainly conservatives) began a counter-boycott. In this way, Chick-fil-A became a politicized topic such that, apparently, by the time of our data collection in 2016, liking the Facebook page of the fast-food chain could be seen as an endorsement of the political views of the company. In the opposite sense, the ice-cream brand Ben & Jerry's openly promotes progressive values and expresses support for social and environmental justice initiatives around the country. Though homogeneity is low ($V = 0.09$, $I = 0.36$), it is relatively high compared to other pages in the food category, and the brand is predominantly liked by liberal users.

Likewise, **Sports** (see Figure SA5e) can also become caught in the partisan currents of the larger society as a result of symbolic actions and outspoken statements of its players. For example, Tim Tebow ($V = 0.11$, $I = 0.71$) has a predominantly conservative following, and some NASCAR ($V = 0.11$, $I = 0.57$) drivers have publicly supported Republican candidates. The Olympics ($V = 0.09$, $I = 0.33$) and the Pittsburgh Steelers ($V = 0.10$, $I = 0.43$) have a more liberal audience, while Serena Williams (0.13, 0.47) fans are more moderate on average. Still, the majority of sports pages appear to unite people with different ideologies, and have very low homogeneity scores including the Boston Red Sox ($V = 0.00$, $I = 0.48$), New England Patriots ($V = 0.02$, $I = 0.48$), and the New York Mets ($V = 0.04$, $I = 0.47$). In 2016, LeBron James ($V = 0.04$, $I = 0.43$) was popular across the ideological spectrum,[15] though we suspect that this might have shifted in the period after our data collection given his subsequent criticisms of President

[15] Note that even though the average ideology of LeBron James is equal to that of the Pittsburgh Steelers, the homogeneity score of the first is much lower and thus his audience is more diverse.

Table 4: Average number of page likes per ideology

	Politics	Political News	Hard News	Lifestyle	All pages
Liberals	21.79 (8%)	12.45 (5%)	7.37 (3%)	234.46 (85%)	276.07
Moderates	17.83 (6%)	10.66 (4%)	7.03 (2%)	252.03 (88%)	287.55
Conservatives	41.62 (13%)	22.64 (7%)	11.51 (3%)	253.32 (77%)	329.09

Trump.

Our results show both similarities and contrasts with Shi et al. (2017), who study partisan divisions in the U.S. by analyzing Twitter co-following networks. They too find cultural dimensions other than religion to be substantially less polarized than political domains. A striking difference from our study, however, is that they find news and media among the dimensions that cut across the political divide. Such interpretations illustrate the difficulty of establishing empirical benchmarks for polarization in addition to comparing estimates of magnitude across different measures.

5.2 Predictors for Individual Homogeneity

We now turn to the individual level analysis, and start with some descriptive statistics before we turn to our empirical assessments of our hypotheses and research question. On average, conservative users like more pages on Facebook, and a slightly higher percentage of the pages they like are political (Table 4).

Similar to Eady et al. (2019), we also look at the proportion of liberals and conservatives whose page likes include at least 5% of pages at the right and left ends of the spectrum, respectively. For each group of pages, we consider "left-leaning" pages as pages with a page ideology score that is lower than the 70th percentile of all pages liked by liberal participants, and "right-leaning" pages are pages with a page ideology score higher than the 30th percentile of all pages liked by conservative participants. Examples of pages at these percentiles of political, news, and lifestyle pages can be found in Table 5.

We find that 9% of liberals have at least 5% of pages to the right of Mitt Romney among their political page likes (see Table 6). Similarly, the political page likes of 8% of conservatives consist of at least 5% pages to the left of Chelsea Clinton. For news pages, conservatives are more likely to like "left-leaning" pages than the other way around, a finding that corresponds to the results of Eady et al. (2019). For lifestyle pages, almost all liberals and conservatives like at least 5% opposite-leaning pages.

It is possible that politically engaged individuals consciously choose to like political or news pages that

Table 5: Examples and ideology score of 70th percentile of all pages liked by liberals and 30th percentile of all pages liked by conservatives per group of pages.

	70th percentile liberal page likes	30th percentile conservative page likes
Politics	Chelsea Clinton (0.32)	Mitt Romney (0.78)
Political news	BBC News (0.40)	Conservative Daily (0.76)
Hard news	Washington Post (0.46)	Fox 5 New York (0.62)
Lifestyle	NIVEA (0.52)	Amazon (0.49)

Table 6: Proportion of liberals with at least 5% right-leaning page likes (page ideology higher than 30th percentile of all pages liked by conservatives) and of conservatives with at least 5% left-leaning page likes (page ideology lower than 70th percentile of all pages liked by liberals) per group of pages.

	Liberals - right likes	Conservatives - left likes
Political	9%	8%
Partisan news	7%	19%***
Hard news	15%	28%***
Lifestyle	97%	97%

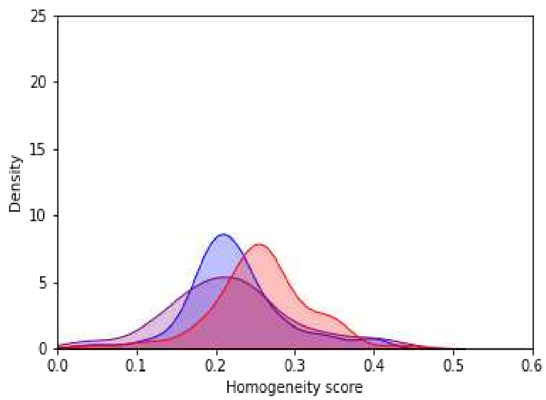
$^{*}p < .1$; $^{**}p < .05$; $^{***}p < .01$
Two-tailed Z-test for two population proportions

have opposing views in order to stay informed about "the other side." To test this, we use a two-tailed t-test to compare the average number of political page likes by liberals and conservatives that like opposing pages, to the average number of political page likes by liberals and conservatives overall (Table 7). Liberals that like news pages with opposing views have more political page likes on average,[16] and this is also the case for conservatives when we consider hard news pages. In general, more politically engaged individuals are thus more likely to follow news pages that contain opposing views, but not more likely to follow opposing political candidates.

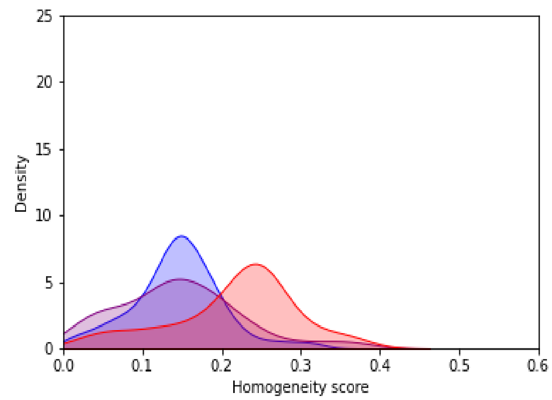
In Figure 7, we zoom in on the individual homogeneity scores of our participants. Individual homogeneity is calculated as the average of the homogeneity scores of all Facebook pages the user has "liked" and functions as a measure for individual tendency to prefer homogeneous spaces. For political and news pages, conservatives have slightly higher homogeneity scores than liberals and moderates. For lifestyle pages, the majority of homogeneity scores are low (below 0.1) regardless of ideology.

[16]We do not find significant results for political and lifestyle pages.

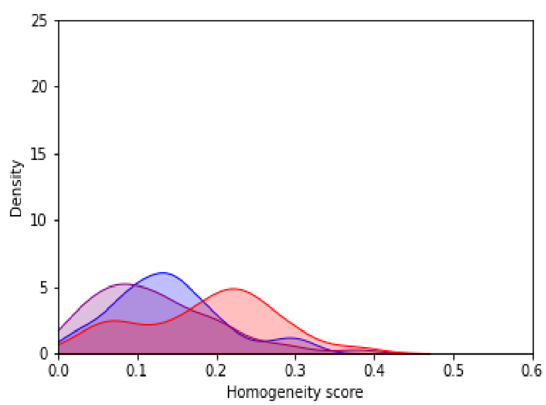
Figure 7: Individual homogeneity distribution for liberals (blue), moderates (purple) and conservatives (red) when taking into account (a) political pages, (b) partisan news, (c) hard news, and (d) lifestyle pages.



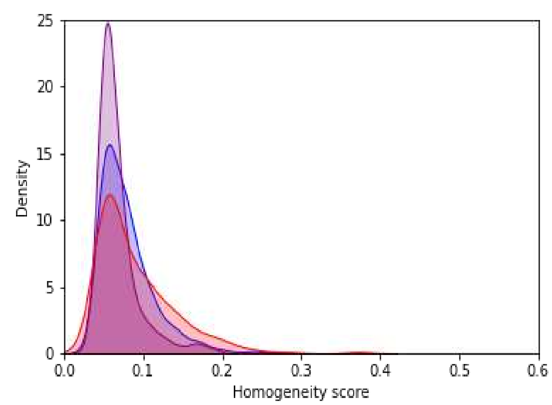
(a) Political



(b) Partisan news



(c) Hard news



(d) Lifestyle

Table 7: Average number of political page likes for liberals with at least 5% right-leaning page likes and for conservatives with at least 5% left-leaning page likes per group of pages.

	Liberals - right likes	Conservatives - left likes
Political	26.23	23.77
Partisan news	38.11***	57.64
Hard news	42.68***	59.25**
Lifestyle	22.14	41.18

*p < .1; **p < .05; ***p < .01
Two-tailed T-test for the means of two independent samples
Mean political page likes for liberals is 21.79
Mean political page likes for conservatives is 41.62

To test **Hypothesis 1**, we perform a beta regression that includes ideology and political news interest as predictors. As additional control variables, we include a mix of relevant sociodemographic variables including age, race, gender, family income, and educational attainment. Total number of page likes (in logarithmic form) is also included as a control variable to account for individual-level differences in engagement with the platform. The results for the four page categories are shown in Table 8. We find that non-moderate ideology is significantly related to liking more homogeneous lifestyle pages.[17] The magnitude of the relationship can be explained in terms of the logarithm of the ratio of the homogeneity score over one minus the homogeneity score (similar to the odds ratio). For example, a liberal ideology increases this ratio by 8% compared to a moderate ideology, while conservative ideology increases this ratio by almost 17% compared to a moderate ideology.

Secondly, for liberal ideology, the effect is larger for strong ideology (very liberal), which is not the case for conservative ideology.[18] Higher political news interest is also associated with higher individual homogeneity scores across all categories.

These results suggest that people with greater political news interest are more likely to like politically homogeneous Facebook pages in non-political categories, which partially confirms H1. The results for ideological affiliation strength are a bit more nuanced. People with non-moderate ideology are more likely to like homogeneous Facebook pages in non-political categories. For liberals we find an effect for stronger ideological affiliation, but not for conservatives. Moreover, when we include party identification instead

[17] Moderate is the reference category for party ideology.

[18] When including “Liberal" as reference category “Very liberal" has a significant positive effect. However, with “conservative" as reference category “very conservative" is not significantly positive.

of ideology (see Table SA7 in the Supplemental Appendix), we find Republicans to be more likely to like homogeneous pages, though partisanship strength does not seem to matter. For lifestyle pages, we find no relationship with party identification. Our results thus suggest that polarization in lifestyle pages is predominantly tied to political interest rather than ideological strength and strong partisanship.

With respect to **RQ2**, we find that conservatives (whether strong or not) are more likely than moderates to like homogeneous pages regardless of category. For liberals this is true for those who are the strongest liberals, except for the lifestyle category, where liberals (whether strong or not) are also more likely to like homogeneous pages. Still we find that conservatives are more likely than liberals to like homogeneous pages in this category. We conclude that conservative ideology is associated with a greater likelihood of "liking" ideologically homogeneous pages in all categories.

The analysis in Table 8 also reveals other several other relationships worth exploring in future research. First, we find that older age is predictive of greater homogeneity in page likes, but only for news and lifestyle pages. Our results for news pages are consistent with Guess et al. (2019), who find — using the same underlying data source as this analysis — that conservatives and people over the age of 65 were more likely to share "fake news" on Facebook in 2016, all else equal. This suggests that page liking patterns may be part of a process in which online misinformation reaches people's social media feeds, thereby increasing the likelihood of engaging with it and sharing it with one's social connections (e.g., Grinberg et al. 2019).

Second, individuals with higher educational attainment are more likely to like more homogeneous lifestyle pages but not more likely to like homogeneous political or news pages, a pattern consistent with research finding that highly educated people are more likely to make consumer decisions that reflect their political leanings (Newman and Bartels 2011). Similarly, gender only has a significant effect within lifestyle pages, but we speculate that the effect could vary depending on the lifestyle subcategories.

Liking explicit political content on Facebook is a form of political participation or endorsement, while liking lifestyle pages may seem apolitical at first sight. Our findings suggest that the characteristics of individuals who exhibit high levels of ideological homogeneity are different for explicitly and implicitly political pages. While most research on polarization and echo-chamber dynamics has focused on networks around explicit political content, an analysis of these lifestyle categories reveals a subtler form of political homophily.

Table 8: *Determinants of Facebook Page Like Homogeneity per category*

	Politics	Political news	Hardnews	Lifestyle
Age: 30-44	-0.086	0.128	-0.029	-0.010
	(0.058)	(0.084)	(0.093)	(0.045)
Age: 45-65	-0.065	0.152^{*}	0.020	0.029
	(0.054)	(0.080)	(0.087)	(0.042)
Age: Over 65	-0.017	0.270^{***}	0.194^{**}	0.254^{***}
	(0.061)	(0.087)	(0.095)	(0.048)
Black	0.002	-0.339^{***}	-0.358^{***}	-0.149^{***}
	(0.055)	(0.075)	(0.086)	(0.042)
Hispanic	-0.088	-0.184^{**}	-0.272^{***}	0.011
	(0.066)	(0.088)	(0.097)	(0.050)
Other Race	-0.134^{**}	-0.340^{***}	-0.419^{***}	-0.108^{**}
	(0.065)	(0.086)	(0.098)	(0.049)
Female	0.030	-0.014	-0.061	-0.111^{***}
	(0.032)	(0.041)	(0.047)	(0.025)
Income	0.006	-0.006	0.002	0.001
	(0.004)	(0.005)	(0.006)	(0.003)
Education	-0.023^{**}	0.011	0.026	0.020^{**}
	(0.011)	(0.015)	(0.017)	(0.009)
Very Liberal	0.088^{*}	0.051	0.177^{**}	0.175^{***}
	(0.048)	(0.064)	(0.072)	(0.037)
Liberal	0.026	-0.035	0.062	0.078^{**}
	(0.045)	(0.061)	(0.069)	(0.036)
Conservative	0.209^{***}	0.283^{***}	0.303^{***}	0.156^{***}
	(0.044)	(0.057)	(0.066)	(0.034)
Very Conservative	0.151^{***}	0.488^{***}	0.546^{***}	0.184^{***}
	(0.057)	(0.069)	(0.078)	(0.044)
Political news interest	0.095^{***}	0.182^{***}	0.191^{***}	0.134^{***}
	(0.022)	(0.031)	(0.034)	(0.017)
Log Number of likes	-0.061^{***}	-0.052^{***}	-0.035^{**}	-0.043^{***}
	(0.011)	(0.015)	(0.017)	(0.008)
Constant	-0.677^{***}	-1.248^{***}	-1.480^{***}	-2.159^{***}
	(0.112)	(0.151)	(0.170)	(0.081)
N	826	774	740	1,085
Pseudo R^2	0.096	0.189	0.194	0.285

$^{*}p < .1$; $^{**}p < .05$; $^{***}p < .01$

Beta regressions with survey weights applied. Reference categories are Age: 18-29, White race, Male gender and Moderate ideology.

Income ranges from 1 to 31, Education from 1 to 6, and political news interest from 1 to 4.

6 Conclusion

In light of increasing discussions about political divides in the United States, we explore polarization in political and traditionally "non-political" domains on social media. Our results from analyzing Facebook "likes" data suggest that ideological divides are large in relatively political domains such as news and media, civil society, and religion but much less pronounced in areas such as culture, food, and sports. Our findings show that polarization does not permeate society as a whole: Lifestyle endeavors still offer cross-cutting spaces, and polarization, when it does emerge, seems limited to a narrow set of politicized examples. Considering that Facebook users primarily engage with non-political Facebook likes, our findings add nuance to debates about the divisive nature of social platforms.

At the individual level, we find that polarization in page liking patterns is more associated with individuals with higher political interest. Furthermore, when we distinguish between polarization in political and non-political domains, we find that individuals who exhibit high levels of political homophily are different. For example, highly educated people are more likely to make lifestyle choices that are related to their political views, while they are less likely to be in echo chambers reflecting the political pages that they follow. This finding has potentially important implications for research on online polarization.

Given our findings, then, why do narratives of enduring political divides in non-political domains persist? One explanation is that people draw inferences on the basis of vivid but unrepresentative examples, as our analysis of Chick-fil-A and prominent sports figures suggests.[19] Similarly, people have exaggerated perceptions of the differences between the parties, both in terms of demographic composition and lifestyle tendencies (Ahler and Sood 2018).[20] Social media itself may drive these misperceptions by fueling cycles of engagement with content that promotes disparagement of partisan outgroups (Rathje et al. 2021; Wojcieszak et al. 2021). Future research should consider how users' online social endorsements interact with these dynamics over time, especially as a possible window into the politicization of figures and brands. As Settle (2018) illustrates, the process by which one's political preferences come to influence seemingly distinct consumer and lifestyle choices can emerge unexpectedly as a result of both elite actions and mass

[19] This is also borne out in polling, which tends to emphasize these vivid examples in addition to brands known to be polarizing, such as media organizations. See `https://morningconsult.com/polarizing-brands-2018/`.

[20] *The New York Times*' recent feature asking readers to guess people's vote preferences from the contents of their refrigerators illustrates the limited predictive value of partisan stereotypes. See `https://www.nytimes.com/interactive/2020/10/27/upshot/biden-trump-poll-quiz.html`.

mobilization. Even though we show these cases to be the exception, they demonstrate how the coexistence of political and other identities on social media leaves users vulnerable to mechanisms of social polarization.

Of course, our study is not without limitations that could affect the generalizability of our conclusions. First, while our reliance on a large online sample from YouGov enables the data linkage necessary to conduct our analyses, the resulting inferences may be subject to qualification on the basis of known differences likely arising from the sampling frame of internet users (Ansolabehere and Schaffner 2014) — namely, respondents in these samples tend to be more knowledgeable about, interested in and engaged with politics than the adult population as a whole. The subset of these respondents who opt in to sharing private data may differ in additional ways, though Table SA1 in the Appendix indicates similar levels of news interest across the samples, suggesting that any substantively relevant differences may be minor.[21] We think it likely that such biases mean that our findings on page liking patterns are an upper bound on polarization. Since we find that polarization in liking is correlated with political interest, and that polarization among lifestyle-related pages is greater among the more highly educated, such patterns would be attenuated — and polarization in page liking muted overall — in a sample less skewed toward the politically engaged. Second, for privacy reasons we only include pages liked by a minimum of 30 respondents, which implies that we are analyzing only a small subset of the Facebook pages in the dataset. We are able to show that the distribution of individual page ideology and homogeneity scores based on all pages is in fact highly similar to those we keep in the analysis, yet we acknowledge that our results are based on relatively more popular Facebook pages and that we cannot analyze polarization patterns involving smaller pages.

Social media data offer a rich source of information about individuals' revealed social and lifestyle preferences, at a resolution that would be difficult to attain with traditional survey techniques. At the same time, the collection of online behavioral data comes with its own set of ethical and privacy challenges (Salganik 2019; Stier et al. 2019). Drawing inferences from online data should be performed with caution since ignoring offline behavior may leave us with a distorted view. Still, linking digital trace data with survey data helps us to understand the relationship between lifestyle preferences and politics and to map the landscape of political culture — both its fault lines and its areas of overlap.

[21] Relatedly, Guess et al. (2020) find limited differences on observables between participants who opt to share private web consumption data and the larger YouGov population.

References

Abramowitz, A. I. and K. L. Saunders (2008). Is polarization a myth? *The Journal of Politics 70*(2), 542–555.

Adamic, L. A. and N. Glance (2005). The political blogosphere and the 2004 us election: divided they blog. *Proceedings of the 3rd international workshop on Link discovery*, 36–43.

Ahler, D. J. and G. Sood (2018). The parties in our heads: Misperceptions about party composition and their consequences. *The Journal of Politics 80*(3), 964–981.

Alford, J. R., P. K. Hatemi, J. R. Hibbing, N. G. Martin, and L. J. Eaves (2011). The politics of mate choice. *The Journal of Politics 73*(2), 362–379.

Ansolabehere, S. and B. F. Schaffner (2014). Does survey mode still matter? findings from a 2010 multi-mode comparison. *Political Analysis 22*(3), 285–303.

Bachrach, Y., M. Kosinski, T. Graepel, P. Kohli, and D. Stillwell (2012). Personality and patterns of facebook usage. In *Proceedings of the 4th annual ACM web science conference*, pp. 24–32.

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

Baldassarri, D. and A. Gelman (2008). Partisans without constraint: Political polarization and trends in american public opinion. *American Journal of Sociology 114*(2), 408–446.

Barberá, P. (2015). Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data. *Political Analysis 23*(1), 76–91.

Barberá, P., J. T. Jost, J. Nagler, J. A. Tucker, and R. Bonneau (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science 26*(10), 1531–1542. PMID: 26297377.

Baron, J. and J. T. Jost (2019). False equivalence: Are liberals and conservatives in the united states equally biased? *Perspectives on Psychological Science 14*(2), 292–303.

Benkler, Y., R. Faris, and H. Roberts (2018). *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press.

Bennett, W. L. (1998). The uncivic culture: Communication, identity, and the rise of lifestyle politics. *PS: Political Science & Politics 31*(4), 741–761.

Bertrand, M. and E. Kamenica (2018). Coming apart? cultural distances in the united states over time. Technical report, National Bureau of Economic Research.

Bhadani, S., S. Yamaya, A. Flammini, F. Menczer, G. Ciampaglia, and B. Nyhan (2020). Political audience diversity and news reliability in algorithmic ranking. *ArXiv abs/2007.08078*.

Bishop, B. (2009). *The big sort: Why the clustering of like-minded America is tearing us apart*. Houghton Mifflin Harcourt.

Blondel, V. D., J.-L. Guillaume, R. Lambiotte, and E. Lefebvre (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment 2008*(10), P10008.

Bode, L. (2017). Gateway political behaviors: The frequency and consequences of low-cost political engagement on social media. *Social Media+ Society 3*(4), 2056305117743349.

Bond, R. and S. Messing (2015). Quantifying social media's political space: Estimating ideology from publicly revealed preferences on facebook. *American Political Science Review 109*(1), 62–78.

Boutyline, A. and R. Willer (2017). The social structure of political echo chambers: Variation in ideological homophily in online networks. *Political Psychology 38*(3), 551–569.

Carney, D. R., J. T. Jost, S. D. Gosling, and J. Potter (2008). The secret lives of liberals and conservatives: Personality profiles, interaction styles, and the things they leave behind. *Political psychology 29*(6), 807–840.

Clauset, A., M. E. Newman, and C. Moore (2004). Finding community structure in very large networks. *Physical review E 70*(6), 066111.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Conover, P. J. and S. Feldman (1981). The Origins and Meaning of Liberal/Conservative Self-Identifications. *American Journal of Political Science*, 617–645.

DellaPosta, D. (2020). Pluralistic collapse: The “oil spill" model of mass opinion polarization. *American Sociological Review*, 0003122420922989.

DellaPosta, D., Y. Shi, and M. Macy (2015). Why do liberals drink lattes? *American Journal of Sociology 120*(5), 1473–1511.

Desmet, K., I. Ortuño-Ortín, and R. Wacziarg (2017). Culture, ethnicity, and diversity. *American Economic Review 107*(9), 2479–2513.

Ditto, P. H., B. S. Liu, C. J. Clark, S. P. Wojcik, E. E. Chen, R. H. Grady, J. B. Celniker, and J. F. Zinger (2019). At least bias is bipartisan: A meta-analytic comparison of partisan bias in liberals and conservatives. *Perspectives on Psychological Science 14*(2), 273–291.

Eady, G., J. Nagler, A. Guess, J. Zilinsky, and J. A. Tucker (2019). How many people live in political bubbles on social media? Evidence from linked survey and Twitter data. *Sage Open 9*(1), 2158244019832705.

Evans, J. H., B. Bryson, and P. DiMaggio (2001). Opinion polarization: Important contributions, necessary limitations. *American Journal of Sociology 106*(4), 944–959.

Finkel, E. J., C. A. Bail, M. Cikara, P. H. Ditto, S. Iyengar, S. Klar, L. Mason, M. C. McGrath, B. Nyhan, D. G. Rand, et al. (2020). Political sectarianism in america. *Science 370*(6516), 533–536.

Fiorina, M. P., S. A. Abrams, and J. C. Pope (2008). Polarization in the american public: Misconceptions and misreadings. *The Journal of Politics 70*(2), 556–560.

Giddens, A. (1991). *Modernity and self-identity: Self and society in the late modern age*. Stanford University Press.

Giddens, A. (2013). *The consequences of modernity*. John Wiley & Sons.

Green, D. P., B. Palmquist, and E. Schickler (2004). *Partisan hearts and minds: Political parties and the social identities of voters*. Yale University Press.

Greenberg, J. and E. Jonas (2003). Psychological motives and political orientation–the left, the right, and the rigid: comment on jost et al. (2003). *Psychological Bulletin 129*(3), 376–382.

Grinberg, N., K. Joseph, L. Friedland, B. Swire-Thompson, and D. Lazer (2019). Fake news on twitter during the 2016 us presidential election. *Science 363*(6425), 374–378.

Guay, B. and C. D. Johnston (2020). Ideological asymmetries and the determinants of politically motivated reasoning. *American Journal of Political Science*.

Guess, A., K. Munger, J. Nagler, and J. Tucker (2019). How accurate are survey responses on social media and politics? *Political Communication 36*(2), 241–258.

Guess, A., J. Nagler, and J. Tucker (2019). Less than you think: Prevalence and predictors of fake news dissemination on facebook. *Science advances 5*(1), eaau4586.

Guess, A. M. (2021). (Almost) Everything in Moderation: New Evidence on Americans' Online Media Diets. *American Journal of Political Science*.

Guess, A. M., B. Nyhan, and J. Reifler (2020). Exposure to untrustworthy websites in the 2016 US election. *Nature human behaviour 4*(5), 472–480.

Hargittai, E. and M. Micheli (2019). Internet skills and why they matter. *Society and the internet: How networks of information and communication are changing our lives 109*.

Hetherington, M. and J. Weiler (2018). *Prius or pickup?: How the answers to four simple questions explain America's great divide*. Houghton Mifflin.

Hetherington, M. J. (2009). Putting polarization in perspective. *British Journal of Political Science*, 413–448.

Huber, G. A. and N. Malhotra (2017). Political homophily in social relationships: Evidence from online dating behavior. *The Journal of Politics 79*(1), 269–283.

Iyengar, S., Y. Lelkes, M. Levendusky, N. Malhotra, and S. J. Westwood (2019). The origins and consequences of affective polarization in the united states. *Annual Review of Political Science 22*, 129–146.

Iyengar, S., G. Sood, and Y. Lelkes (2012). Affect, not ideologya social identity perspective on polarization. *Public opinion quarterly 76*(3), 405–431.

Jost, J. T., C. M. Federico, and J. L. Napier (2009). Political ideology: Its structure, functions, and elective affinities. *Annual review of psychology 60*, 307–337.

Jurkowitz, M., A. Mitchell, E. Shearer, and M. Walker (2020). Us media polarization and the 2020 election: A nation divided. Technical report.

Khan, R., K. Misra, and V. Singh (2013). Ideology and brand consumption. *Psychological science 24*(3), 326–333.

Lee, A. H.-Y. (2020). How the politicization of everyday activities affects the public sphere: The effects of partisan stereotypes on cross-cutting interactions. *Political Communication*, 1–20.

Lelkes, Y. (2016). Mass polarization: Manifestations and measurements. *Public Opinion Quarterly 80*(S1), 392–410.

Macafee, T. (2013). Some of these things are not like the others: Examining motivations and political predispositions among political facebook activity. *Computers in Human Behavior 29*(6), 2766–2775.

Magurran, A. E. (2013). *Measuring biological diversity*. John Wiley & Sons.

Marder, B. (2018). Trumped by context collapse: Examination of 'liking'political candidates in the presence of audience diversity. *Computers in Human Behavior 79*, 169–180.

Mason, L. (2018a, April). Ideologues without Issues: The Polarizing Consequences of Ideological Identities. *Public Opinion Quarterly 82*(S1), 866–887.

Mason, L. (2018b). *Uncivil Agreement: How Politics Became Our Identity*. University of Chicago Press.

Meredith, M. and M. Ridout (2014). Overview of the overlap package. *R. Proj*, 1–9.

Messing, S., P. van Kessel, and A. Hughes (2017, December). Sharing the news in a polarized congress. Pew Research Center.

Mummolo, J. and C. Nall (2017). Why partisans do not sort: The constraints on political segregation. *The Journal of Politics 79*(1), 45–59.

Mutz, D. C. (2006). *Hearing the other side: Deliberative versus participatory democracy*. Cambridge University Press.

Newman, B. J. and B. L. Bartels (2011). Politics at the checkout line: Explaining political consumerism in the united states. *Political Research Quarterly 64*(4), 803–817.

Newman, M. (2010). *Networks: An introduction*. Oxford university press.

Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the national academy of sciences 103*(23), 8577–8582.

Newman, N., R. Fletcher, A. Kalogeropoulos, and R. Nielsen (2019). *Reuters institute digital news report 2019*, Volume 2019. Reuters Institute for the Study of Journalism.

Panzeri, S. and A. Treves (1996). Analytical estimates of limited sampling biases in different information measures. *Network: Computation in neural systems 7*(1), 87–107.

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual review of psychology 49*(1), 65–85.

Prior, M. (2009). Improving media effects research through better measurement of news exposure. *The Journal of Politics 71*(3), 893–908.

Prior, M. (2013). The challenge of measuring media exposure: Reply to dilliplane, goldman, and mutz. *Political Communication 30*(4), 620–634.

Provost, F. and T. Fawcett (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media, Inc.

Rathje, S., J. J. Van Bavel, and S. van der Linden (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences 118*(26).

Salganik, M. J. (2019). *Bit by bit: Social research in the digital age*. Princeton University Press.

Saunders, K. L. and A. I. Abramowitz (2004). Ideological realignment and active partisans in the american electorate. *American Politics Research 32*(3), 285–309.

Selway, J. S. (2011). The measurement of cross-cutting cleavages and other multidimensional cleavage structures. *Political Analysis 19*(1), 48–65.

Settle, J. E. (2018). *Frenemies: How Social Media Polarizes America*. Cambridge University Press.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal 27*(3), 379–423.

Shi, Y., K. Mast, I. Weber, A. Kellum, and M. Macy (2017). Cultural fault lines and political polarization. In *Proceedings of the 2017 ACM on Web Science Conference*, pp. 213–217.

Stier, S., J. Breuer, P. Siegers, and K. Thorson (2019). Integrating survey data and digital trace data: Key issues in developing an emerging field. *Social Science Computer Review 38*(5), 503–516.

Thorson, E. A. and M. Serazio (2018). Sports fandom and political attitudes. *Public Opinion Quarterly 82*(2), 391–403.

Toff, B. (2021). Prime time television in red and blue: Partisan divides in entertainment media. https://benjamintoff.com/prime-time-television-in-red-and-blue/.

Wojcieszak, M., A. Casas, X. Yu, J. Nagler, and J. A. Tucker (2021). Echo chambers revisited: The (overwhelming) sharing of in-group politicians, pundits and media on twitter.

Zilinsky, J., C. Vaccari, J. Nagler, and J. A. Tucker (2020). Don't republicans tweet too? using twitter to assess the consequences of political endorsements by celebrities. *Perspectives on Politics 18*(1), 144–160.

A Sample details

Table SA1: Comparison of samples.

	Full Sample	Sample 2[a]	Sample 3[b]	p^c	Sample 4[d]
% Democrat	31	32	40	0.17	46
Mean ideology (5-point)	2.98	2.89	2.76	0.01	2.79
Mean news interest (1-4)	3.28	3.27	3.32	0.10	3.33
Mean age	51	49	49	0.16	48
% High school or less	23	20	22	0.17	26
% Female	54	57	57	0.67	58
% Less than $20,000	14	14	14	0.82	15
% White	75	77	76	0.49	76
Median number of page likes					374
Median number of page likes with minimum 30 likes					115
N	3,500	2,711	1,331		1,211

[a]Respondents who said in the survey that they have a Facebook account (i.e., they selected "Facebook" from the list of response options to the question "Do you have accounts on any of the following social media sites?").

[b]Respondents (regardless of their answer in the previous question) who consented to share Facebook profile information with the researchers.

[c]p-values are computed from t-tests of the difference in means between the sample of respondents who reported having a Facebook account and those who consented to provide access to their profile data.

[d]The final column subsets to those who shared any Facebook data at all that we were able to link back to the survey and that have liked at least one of the pages included in our analysis.

Table SA2: Total number of page likes (for the 5155 pages included in our study) per ideology

Ideology score	Number of likes	Ideology	Number of likes
Very liberal	47,440 (17%)	Liberal	105,473 (38%)
Liberal	58,033 (21%)		
Moderate	96,837 (34%)	Moderate	96,837 (34%)
Conservative	41,384 (15%)	Conservative	78,466 (28%)
Very conservative	37,082 (13%)		

B Coding categories

Starting from the initial Facebook categories[22], we extracted a list of 24 categories. Two independent coders received the following instructions and a short training session to assign Facebook pages to the corresponding category.

Instructions The goal of this coding task is to divide Facebook pages into 24 categories, based on the content of the Facebook page and your background knowledge about the subject. The 24 categories can be found in the Codebook (Table SA3), together with a description and some illustrative examples. Some things to keep in mind while coding:

- Select the correct category from the drop-down list. You can assign multiple categories to one Facebook page, but try to indicate the primary category first. For example, The Daily Show is a TV Show in the first place, but could also be classified as 'Entertainment' or even 'News & Media'.
- Make sure to visit the Facebook page (by following the link provided) before you assign a category. For example, Harry Potter could refer to both the books or the movies and you will need to visit the page in order to assign the correct category.
- If the link we provided is broken, please try to find the page manually. If that does not work, the page is likely removed from Facebook. We ask that you indicate this in the 'Error' column and that you try to assign a category based on the name of the page alone.
- If a page does not belong to any of the provided categories you can assign the category 'Other'. In this case, we ask you to specify in your own words which category you would assign to this page.

[22] see `https://www.facebook.com/pages/category/`

After coding, the smallest categories were grouped together to result in the final categorization that was shown in Table 2. **Religion** and **Identity/affinity groups** were merged to **Identity & religion**, **Books** were added to **Arts & Culture** and **Bars & Nightlife** to **Entertainment**. Thereafter, inter coder reliability was calculated. For the final results, the codings of one of the two coders were selected (after a shallow quality inspection), to ensure internal consistency of the coded categories. Finally, pages in the category **Shows & Events** were manually reassigned by the authors to a corresponding category (e.g. Music, TV Shows, Movies or Entertainment) and also pages from the category **Other** were assigned to a suitable category. (Non-) scientific research was added to **Education** and the category was renamed to **Research & Education**.

Table SA3: Categories codebook

Category	Description	Examples
Government & Politics	Politicians, political parties, political content, political communities and government organizations	Barack Obama, Being Conservative, U.S. Army
News & Media	News, media, radio, magazines, journalists, etc.	Fox News , The Economist
Individual opinion leaders	Individual influencers, bloggers, commentators, etc.	Michelle Malkin, Michael Moore
Civil Society	Nonprofit organizations and labor unions (formal organizations)	Human rights campaign, AFL-CIO
Religion	Religious pages, religious organizations	Jezus loves you, Franklin Graham,
Identity/affinity groups	Pages reffering to home country, region, ethnic or cultural groups	Israel is my heart, Africans-In-America
Books	Books, libraries, publishers, writers, poetry, thematical magazines	Barnes & Noble, Lord of the Rings books, Stephen King
Tv Shows	TV shows, episodes, channels	The Big Bang Theory, The Daily Show, National geographic channel
Music	Music, bands, producers, record labels, albums, etc.	The Beatles, Gibson, Warner Music Group
Movies	Movies, actors, directors, movie characters, cinema and	Harry Potter (movie), Regal cinemas, Alfred Hitchcoc
Food & Beverage	Food, cooking, restaurants, drinks, spirits, breweries etc.	Starbucks, Pepsi, Tasty
Sports	Sports, teams, athletes, leagues, games, gym	New York Yankees, NFL, Road Runner sports
Beauty & Health	Cosmetics, healthcare, medical	MinuteClinic, NIVEA, Bayer Aspirin
Arts & Culture	Arts, culture, photography, museums, artists, musicals, theater, etc.	American museum of natural history, Andy Warhol, WICKED the musical
Education	Schools, universities, student organizations and education	LeapFrog USA, New York University, VINCI Schools
Travel	Travel, tour agencies and tourism	Southwest Airlines, Hilton Hotels & Resorts, Love GREAT Brittain
Bars & Nightlife	Bars, cafes, pubs, clubs etc.	House of Yes, Smalls Jazz Club (also music), The Wayland
Shows & events	One-off or limited occurrences, such as festivals, shows, performances and concerts	Honda Stage, Ultra Music Festival, TomorrowWorld
Entertainment	Entertainment, games, humor, amusement, comedy etc.	Larry the cable guy, Grumpy cat memes, Candy Crush
Public Figures	Public figures	Ellen DeGeneres, Michelle Obama, Dave Ramsey
Interests	Interests, communities (informal) and hobbies	Dogs, Hippie Peace Freaks, Humans of New York
Shopping & retail	Apparel, accessories, clothing, fashion, consumer electronics, home decaration, stores, shopping mall, wholesale, etc.	Converse, Amazon.com, Bose
Cars and transportation	Car brands, automotive, airlines, boats, etc.	Hyundai, American Airlines, GasBuddy
Services	Marketing, advertising, legal, finance, consulting, etc.	BFAds - Black Friday Ads, PayPal, Facebook Business
Other	Everything that does not fit in the other categories. Please specify in your own words which category you would assign to this page	

C Methods and measures

Modularity is the relative density of edges inside communities with respect to edges outside of communities (Newman 2010):

$$Q = \frac{1}{2m} \sum_{ij} (A_{ij} - \frac{k_i k_j}{2m}) \delta(c_i, c_j), \tag{2}$$

where m is the number of edges, A_{ij} represents the weight of the edge between node i and j, k_i is the number of edges adjacent to node i, and $\delta(ci, cj)$ is 1 if i and j are in the same community and 0 otherwise. A modularity of zero indicates that the fraction of within-community edges is no different from what we would expect for a randomized network, while a value above about 0.3 is found to be a good indicator of significant community structure in a network (Clauset et al. 2004). By modularity optimization networks can be divided into communities or clusters. For example, the modularity of a network of 105 books on American politics on Amazon.com is $Q = 0.52$ (Newman 2006). In this network four communities were identified: one consisting almost entirely of liberal books and one almost entirely of conservative books, and two containing most of the centrist books. Similarly, a network of political 2004 U.S election blogs (Adamic and Glance 2005) was cleanly divided into a conservative and a liberal community with an optimal modularity of $Q = 0.43$ (Newman 2006).

Cramer's V is a normalized version of the Chi-square statistic and determines the effect size. It is defined by (Cohen 2013):

$$V(Z) = \sqrt{\frac{\chi^2(Z)/N^1}{min(r-1, c-1)}} \tag{3}$$

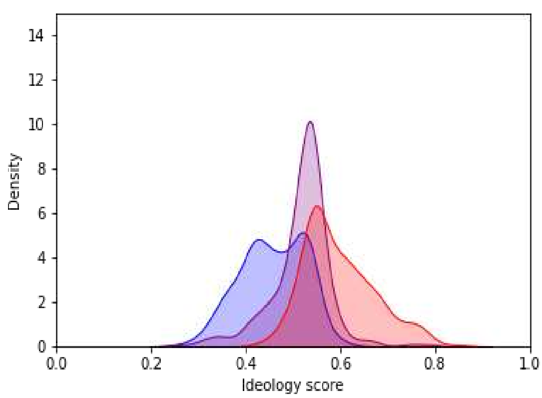
where r is the number of rows (2) and c the number of columns (3) in the contingency table.

D Robustness checks

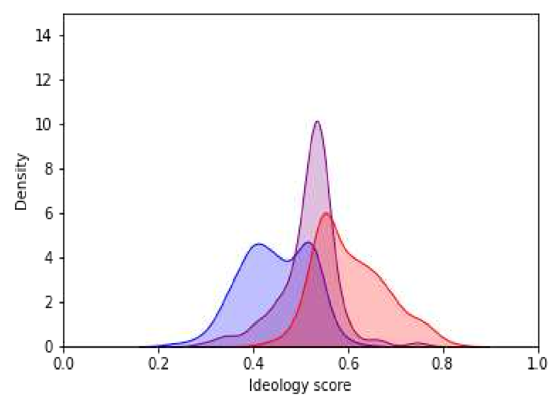
D.1 Including all Facebook pages

We only include pages that are liked by a minimum of 30 respondents in our analysis, to ensure that the calculated page ideology and homogeneity scores are reliable[23]. A very limited amount of pages –5155– has a minimum of 30 likes, which implies that almost 99% of the Facebook pages in our dataset can not be used for analysis. Per user, on average only one third of their likes consists of pages with a minimum of 30 likes. Although this seems like a huge loss of information, in Figure SA1 and Figure SA2 we show that the distribution of individual page ideology and homogeneity scores based on all pages and based on only pages with a minimum of 30 likes is in fact highly similar. The mean individual homogeneity score for all pages (0.097) is only slightly lower than for the pages with 30 likes (0.105). Yet, we do acknowledge that the results and conclusions in this paper our based on widespread Facebook pages and that we cannot analyze polarization in smaller pages.

Figure SA1: Page ideology distribution for liberals (blue), moderates (purple) and conservatives(red) when taking into account (a) Facebookpages with a minimum of 30 likes and (b) all Facebook pages in our dataset.



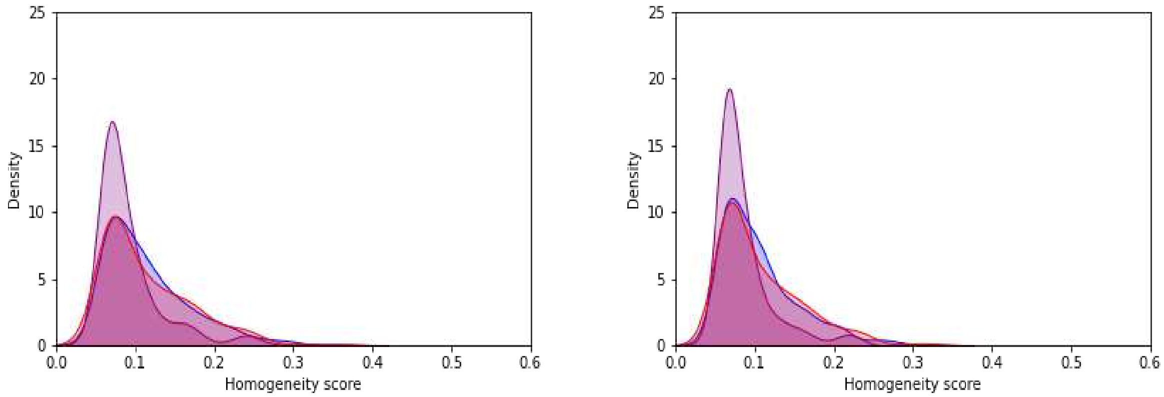
(a) Facebook pages with minimum 30 likes



(b) All Facebook pages

[23] With a sample size of 30 likes, we can estimate page ideology with 95% confidence and a margin of error no larger than 0.4.

Figure SA2: Homogeneity distribution for liberals (blue), moderates (purple) and conservatives (red) when taking into account (a) Facebookpages with a minimum of 30 likes and (b) all Facebook pages in our dataset.



(a) Facebook pages with minimum 30 likes *(b) All Facebook pages*

D.2 Homogeneity metrics

We acknowledge that several other metrics could be used to measure homogeneity (see Bhadani et al. 2020). As a robustness check, we compare our results using Cramer's V to two other measures of homogeneity: Entropy and Variance.

Entropy is a measure of disorder that captures how mixed (impure) a set is with respect to the properties of interest. Originally introduced by Shannon in 1994 to quantify uncertainty in strings of text (Shannon 1948), it has become a basic quantity in information theory and has found its way into many applications. It is used as a basis for machine learning methods (Provost and Fawcett 2013), for studying neuronal activity (Panzeri and Treves 1996) or even as a measurement of biological diversity (Magurran 2013). The Shannon entropy of Facebook page Z is defined as (Shannon 1948):

$$S(Z) = -\sum_{c=0}^{3} \mathrm{p}(c) \log_2 \mathrm{c}(k) \tag{4}$$

The probability $p(c)$ that a user that liked page Z has ideology c is estimated using maximum likelihood, or $p(c) = N_c^1/N$.

With three classes, entropy ranges from close to zero at minimal disorder (the page is liked by almost exclusively members of the same class) to 1.58 at maximum disorder (the classes are balanced with 33% class $liberal$, 33% class $moderate$ and 33% class $conservative$). In other words, the lower the entropy, the more homogeneous the respondents that liked the page are in terms of their self-reported ideology. We will re-scale this value between 0 and 1.

Variance can also be used to measure audience homogeneity:

$$\sigma^2(Z) = \sum \frac{(s_i - I(Z))^2}{N} \tag{5}$$

with s_i the ideology score (0-4) of the i-th individual and I(Z) the average ideology of page Z.

The ranking of categories (see Table SA4) is very similar for the three different metrics[24]. Apart from some slight individual differences, the overall conclusion stays the same. For the beta regressions (Table SA5), the results for the dependent variables Entropy and Cramer's V are highly similar[25] but differ slightly from the regression using Variance.

[24]For Entropy and Variance, smaller values indicate higher homogeneity.

[25]Note that for Entropy and Variance the relation with homogeneity is negative, while for Cramer's V this is positive, hence the opposite sign for Cramer's V

Table SA4: Entropy. Cramer's V and Variance for the Facebook categories.

Category	Entropy	Cramer's V	Variance
Politics	0.66	0.22	0.05
Political news	0.72	0.18	0.06
Hardnews	0.79	0.15	0.07
Civil Society	0.80	0.12	0.08
Identity & Religion	0.86	0.12	0.09
Individual opinion leaders	0.81	0.12	0.08
Public Figures	0.84	0.10	0.08
Arts & Culture	0.87	0.07	0.09
Tv Shows	0.89	0.07	0.08
Entertainment	0.91	0.07	0.09
Research & Education	0.90	0.06	0.09
Music	0.91	0.06	0.09
Interests	0.91	0.06	0.09
Movies	0.90	0.06	0.09
Sports	0.91	0.05	0.08
Services	0.95	0.05	0.09
Beauty & Health	0.94	0.04	0.08
Travel	0.95	0.04	0.09
Shopping & retail	0.95	0.04	0.09
Food & Beverage	0.95	0.04	0.09
Cars and transportation	0.96	0.04	0.09
TOTAL	0.90	0.07	0.08

Table SA5: Determinants of individual homogeneity for different homogeneity metrics.

	Cramer's V	Entropy	Variance
Age: 30-44	-0.001	0.083	-0.036^{***}
	(0.048)	(0.070)	(0.013)
Age: 45-65	0.071	0.058	-0.064^{***}
	(0.045)	(0.067)	(0.012)
Age: Over 65	0.329^{***}	-0.318^{***}	-0.128^{***}
	(0.051)	(0.076)	(0.015)
Black	-0.149^{***}	0.140^{**}	0.007
	(0.044)	(0.066)	(0.012)
Hispanic	0.010	-0.046	-0.009
	(0.052)	(0.079)	(0.015)
Other Race	-0.143^{***}	0.206^{***}	0.025^{*}
	(0.052)	(0.077)	(0.014)
Female	-0.114^{***}	0.158^{***}	0.031^{***}
	(0.026)	(0.040)	(0.008)
Income	0.0002	-0.002	-0.0002
	(0.003)	(0.005)	(0.001)
Education	0.020^{**}	-0.061^{***}	-0.003
	(0.009)	(0.014)	(0.003)
Very Liberal	0.187^{***}	-0.409^{***}	-0.017
	(0.039)	(0.058)	(0.011)
Liberal	0.089^{**}	-0.260^{***}	-0.020^{*}
	(0.037)	(0.056)	(0.011)
Conservative	0.173^{***}	-0.114^{**}	-0.010
	(0.036)	(0.056)	(0.010)
Very Conservative	0.215^{***}	-0.152^{**}	0.007
	(0.046)	(0.072)	(0.014)
Political news interest	0.159^{***}	-0.247^{***}	-0.036^{***}
	(0.018)	(0.027)	(0.005)
Log Number of likes	-0.037^{***}	-0.024^{*}	0.017^{***}
	(0.008)	(0.013)	(0.002)
Constant	-2.085^{***}	2.097^{***}	-2.525^{***}
	(0.085)	(0.130)	(0.025)
N	1,087	1,087	1,087
Pseudo R^2	0.306	0.281	0.217

$^{*}p < .1$; $^{**}p < .05$; $^{***}p < .01$

Beta regressions with survey weights applied. Reference categories are Age: 18-29, White race, Male gender, and Moderate ideology.

Income ranges from 1 to 31, Education from 1 to 6, and political news interest from 1 to 4.

D.3 Regressions

As a robustness check, we include the regression results weighted for the number of likes in Table SA6. The results are congruent with the regression results for Cramer's V in Table SA5. Secondly we include party identification strength in stead of ideology in Table SA7. We do not find a significant relation for party identification strength and liking more homogeneous pages for any of the groups. The explanation could be that users with moderate ideology are less likely than both liberals and conservatives to like homogeneous pages and since moderates more often identify as democrats than as republicans we find republicans to be more likely to like homogeneous pages, independent of party identification strength.

Table SA6: Determinants of individual homogeneity when regression is weighted for individual number of page likes

	Homogeneity
Age: 30-44	-0.060^{***}
	(0.022)
Age: 45-65	0.081^{***}
	(0.021)
Age: Over 65	0.207^{***}
	(0.024)
Black	-0.095^{***}
	(0.019)
Hispanic	0.042^{*}
	(0.022)
Other Race	-0.037^{**}
	(0.017)
Female	-0.196^{***}
	(0.012)
Income	0.002
	(0.001)
Education	0.032^{***}
	(0.004)
Very Liberal	0.097^{***}
	(0.017)
Liberal	0.017
	(0.015)
Conservative	0.116^{***}
	(0.016)
Very Conservative	0.320^{***}
	(0.017)
Political news interest	0.112^{***}
	(0.007)
Log Number of likes	-0.069^{***}
	(0.004)
Constant	-1.951^{***}
	(0.042)
N	1,087
Pseudo R^2	0.270

$^{*}p < .1$; $^{**}p < .05$; $^{***}p < .01$

Beta regression weighted for individual page likes. Reference categories are Age: 18-29, White race, Male gender, and Moderate ideology. Income ranges from 1 to 31, Education from 1 to 6, and political news interest from 1 to 4.

Table SA7: Determinants of individual homogeneity with party identification (PID) strength

	Politics	Political news	Hardnews	Lifestyle
Age: 30-44	-0.102^{*}	0.196^{**}	-0.016	-0.038
	(0.056)	(0.084)	(0.093)	(0.044)
Age: 45-65	-0.088^{*}	0.211^{***}	0.032	0.0003
	(0.053)	(0.080)	(0.087)	(0.042)
Age: Over 65	-0.048	0.319^{***}	0.216^{**}	0.220^{***}
	(0.059)	(0.086)	(0.095)	(0.047)
Black	0.027	-0.246^{***}	-0.261^{***}	-0.145^{***}
	(0.053)	(0.073)	(0.084)	(0.041)
Hispanic	-0.070	-0.109	-0.196^{**}	0.032
	(0.065)	(0.088)	(0.098)	(0.050)
Other Race	-0.078	-0.294^{***}	-0.372^{***}	-0.094^{*}
	(0.064)	(0.086)	(0.098)	(0.049)
Female	0.028	-0.024	-0.078^{*}	-0.113^{***}
	(0.031)	(0.041)	(0.047)	(0.025)
Income	0.005	-0.013^{**}	-0.003	-0.00001
	(0.004)	(0.005)	(0.006)	(0.003)
Education	-0.017	0.021	0.034^{**}	0.023^{**}
	(0.011)	(0.014)	(0.017)	(0.009)
PID strength	0.028	0.042	0.042	0.013
	(0.020)	(0.026)	(0.030)	(0.016)
Democrat	0.088	-0.156^{*}	-0.188^{*}	-0.017
	(0.069)	(0.090)	(0.102)	(0.053)
Republican	0.288^{***}	0.303^{***}	0.174^{*}	0.071
	(0.067)	(0.086)	(0.098)	(0.051)
Political news interest	0.086^{***}	0.198^{***}	0.234^{***}	0.145^{***}
	(0.021)	(0.029)	(0.033)	(0.016)
Log Number of likes	-0.062^{***}	-0.053^{***}	-0.034^{*}	-0.042^{***}
	(0.011)	(0.015)	(0.017)	(0.008)
Constant	-0.834^{***}	-1.309^{***}	-1.350^{***}	-2.077^{***}
	(0.115)	(0.156)	(0.175)	(0.085)
N	847	793	760	1,117
Pseudo R^2	0.121	0.211	0.203	0.261

$^{*}p < .1$; $^{**}p < .05$; $^{***}p < .01$

Beta regressions with survey weights applied. Reference categories are Age: 18-29, White race, Male gender, and Independent.

Income ranges from 1 to 31, Education from 1 to 6, PID strength from 1 to 4, and political news interest from 1 to 4.

E Results

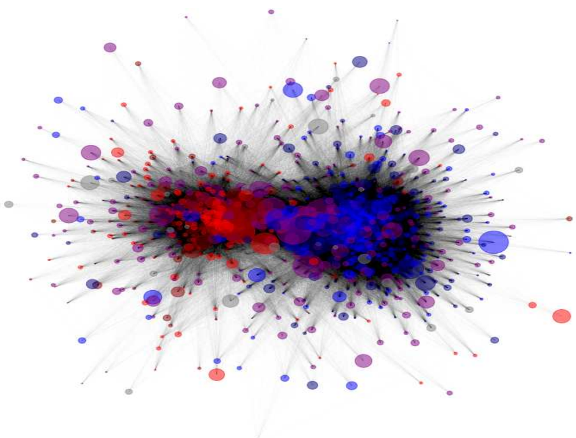
E.1 Network analysis

Network analysis of the bottom node projection points in the direction of political polarization on Facebook (see Figure SA3). This is most outspoken when we consider only political pages and the community detection algorithm reveals two communities with a distinctly different ideological composition (Table SA8). For partisan news pages and hard news a third (and even a fourth in case of the latter) community exists that is predominantly composed of moderate voters. For lifestyle pages however, the three detected communities show hardly any ideological differences, and the modularity of these communities is much lower than for the political pages (Table SA8).

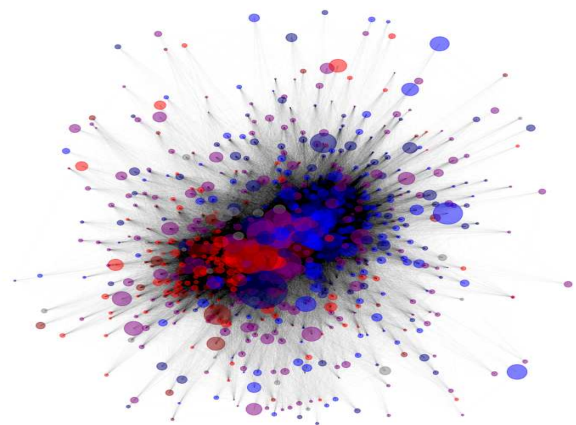
Table SA8: Size, density and hierarchy of the communities in the bottom node projection.

Group	Modularity	Community	Ideology	Size	Density	Hierarchy
Politics	0.38	1	0.28	456	0.65	0.34
		2	0.7	344	0.52	0.45
Political news	0.30	1	0.29	389	0.45	0.53
		2	0.71	271	0.63	0.37
		3	0.52	86	0.45	0.39
Hardnews	0.25	1	0.28	306	0.56	0.40
		2	0.71	223	0.60	0.39
		3	0.49	138	0.39	0.60
		4	0.46	46	0.69	0.27
Lifestyle	0.15	1	0.46	786	0.63	0.34
		2	0.54	286	0.87	0.13
		3	0.40	5	0.70	0.30
All	0.16	1	0.34	555	0.74	0.26
		2	0.53	288	0.91	0.09
		3	0.73	241	0.81	0.19

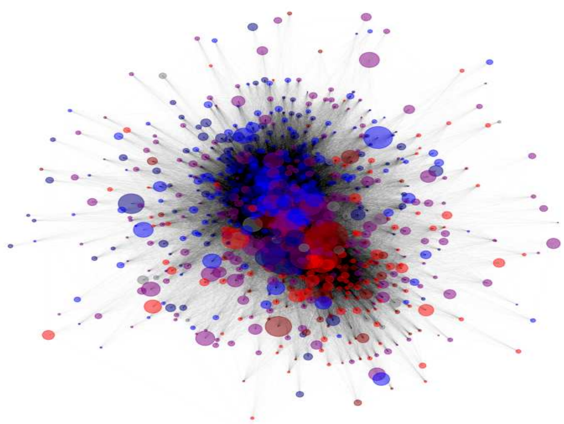
Figure SA3: Network visualization for the bottom node projection of (a) Political pages, (b) Partisan news, (c) Hard news, and (d) Lifestyle pages. The size of the bubble represents the total number of likes of the user, the color represents the ideology of the user.



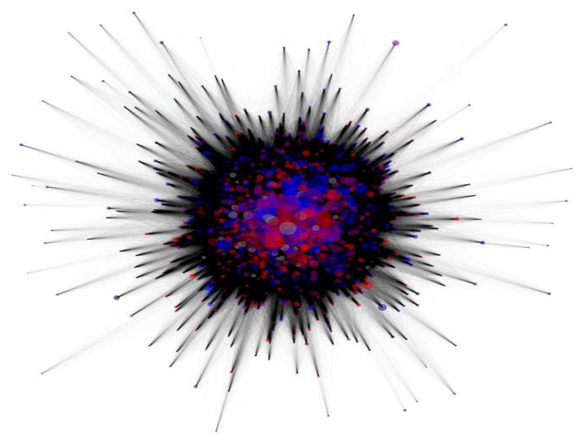
(a) Political



(b) Partisan news



(c) Hard news



(d) Lifestyle

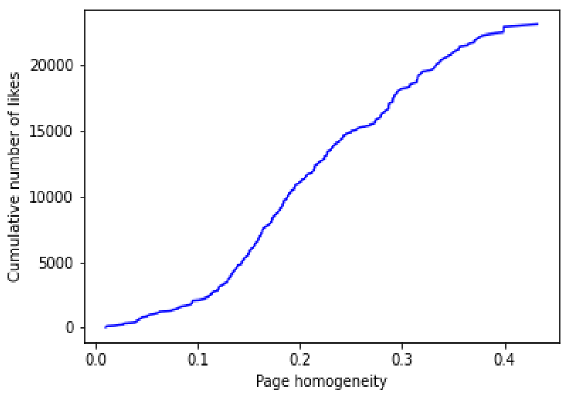
Table SA9: Mean ideology, size, density and hierarchy of the communities in the top node projection.

Group	Modularity	Community	Ideology	Size	Density	Hierarchy
Politics	0.40	1	0.31	102	0.98	0.02
		2	0.80	131	0.98	0.02
Political news	0.28	1	0.80	72	0.98	0.02
		2	0.38	87	1.00	0.00
Hardnews	0.17	1	0.40	49	1.00	0.00
		2	0.56	9	1.00	0.00
		3	0.77	30	0.93	0.07
Lifestyle	0.13	1	0.50	1912	0.90	0.10
		2	0.50	1734	0.99	0.01
All	0.15	1	0.51	2429	0.89	0.11
		2	0.50	1986	0.99	0.01

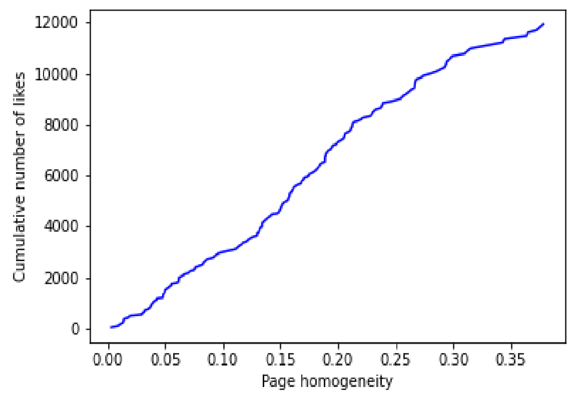
E.2 Facebook categories

We represent the cumulative distribution of page likes from low to high homogeneity pages in Figure SA4. We find that for lifestyle pages, the slope is steepest for low homogeneity scores, which indicates that lifestyle pages that attract most people (highest number of likes) also attract people with different ideologies. In contrast, for political pages, the most diverse pages are not the most popular.

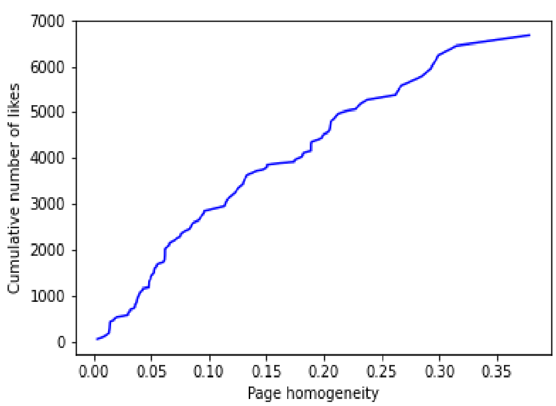
Figure SA4: Cumulative distribution of page likes from low to high page homogeneity for (a) Political pages, (b) Partisan news, (c) Hard news, and (d) Lifestyle pages.



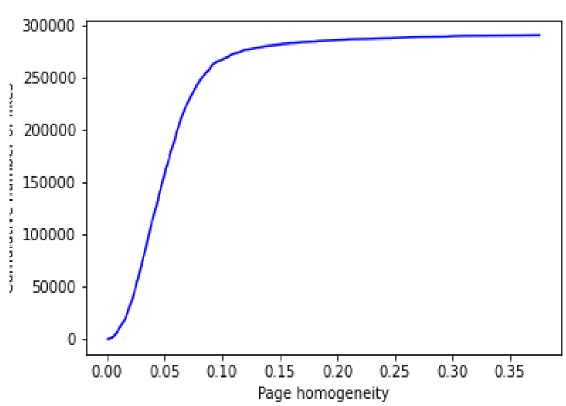
(a) Political



(b) Partisan news



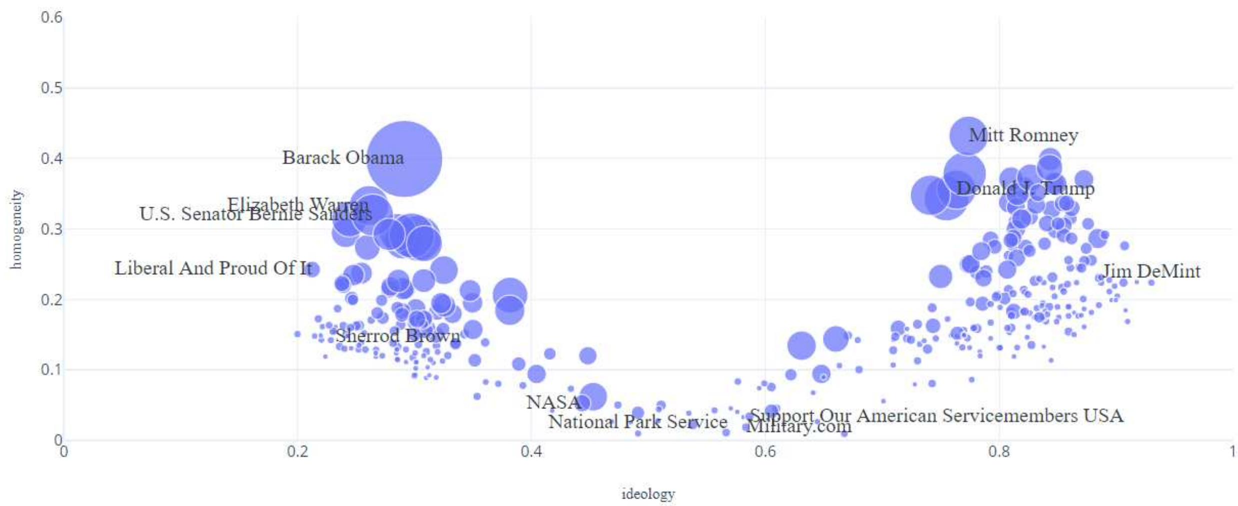
(c) Hard news



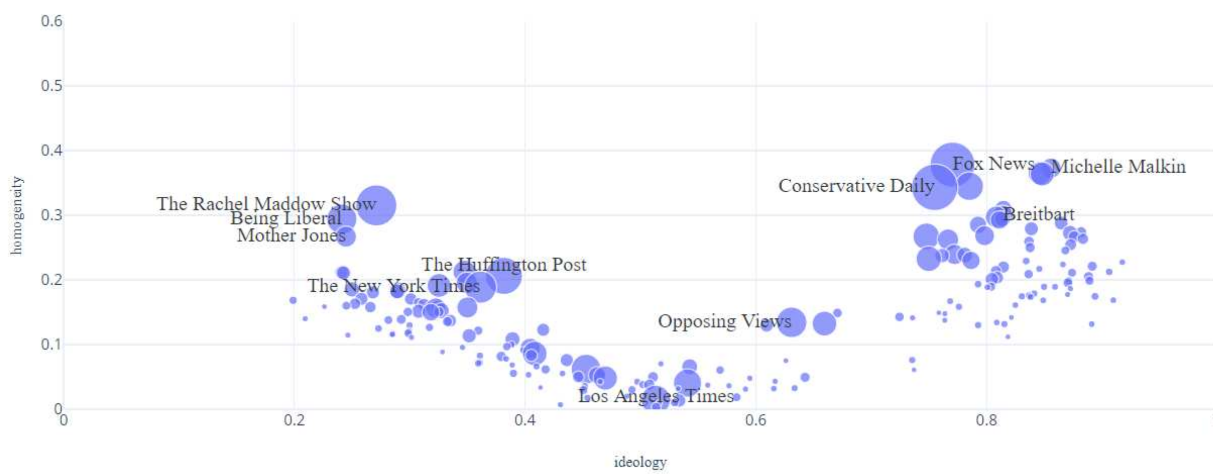
(d) Lifestyle

In Figure SA5, we map the ideology score per Facebook page on the x-axis, and the homogeneity score on the y-axis. Note that page ideology and page homogeneity are related analytically. Pages with a very homogeneous liberal or conservative audience will accordingly have a very low or high page ideology score, while pages with a diverse audience will have a moderate page ideology score, hence the U-shaped graphs. A moderate page ideology score combined with high homogeneity score indicates a predominantly moderate audience (e.g. Serena Williams in Figure SA5e), but these pages are rare. For homogeneous categories (e.g. Figure SA5a) the majority of pages are located at both ends of the U-shaped graph, while for more heterogeneous categories (e.g. Figure SA5f) the pages are located in the bottom-center of the graph.

Figure SA5: Ideology and homogeneity scores for the Facebook pages per category. The magnitude of the circle represents the total number of likes of the Facebook page.



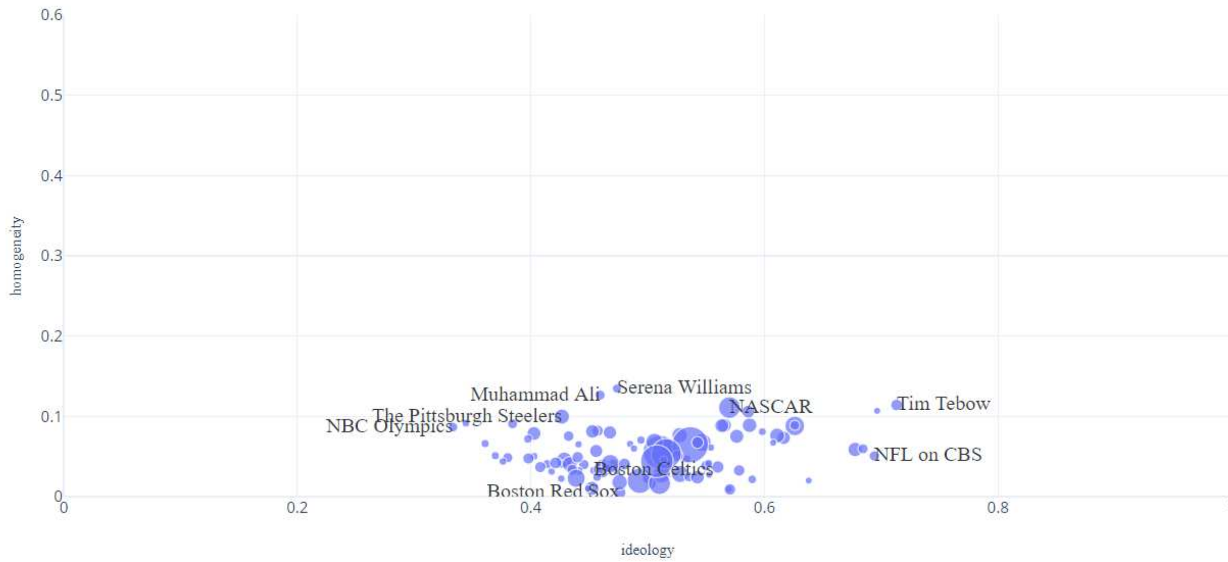
(a) Political



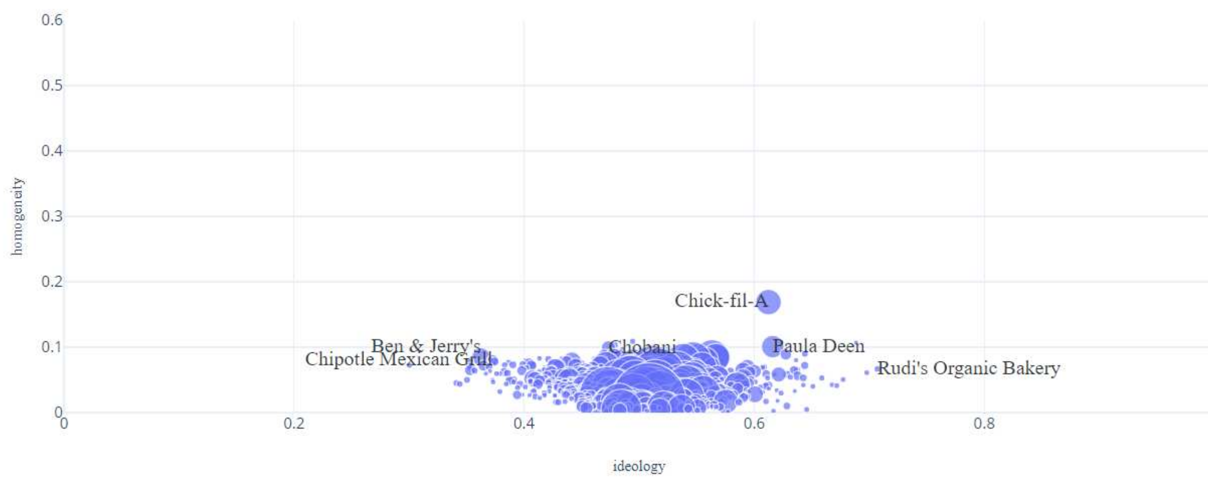
(b) Partisan news

(c) Hard news

(d) Arts & Culture



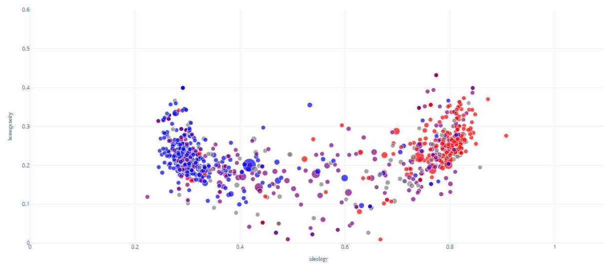
(e) Sports



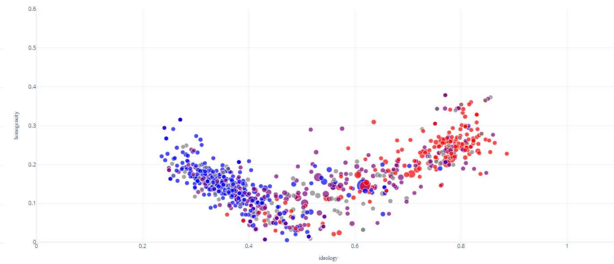
(f) Food & Beverages

E.3 Facebook users

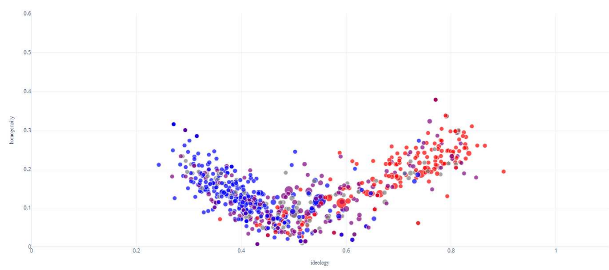
Figure SA6: Page ideology and homogeneity of the users when taking into account (a) Political pages, (b) Partisan news, (c) Hard news, and (d) Lifestyle pages. The size of the bubble represents the total number of likes of the user, the color represents the self-reported ideology of the user.



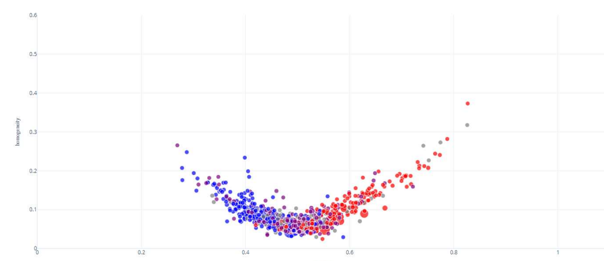
(a) Political



(b) Partisan news



(c) Hard news



(d) Lifestyle

Table SA10: Overlapping coefficient (bootstrap mean) for the probability density curves of liberals, moderates and conservatives.

	Liberal-Conservative	Liberal-Moderate	Conservative-Moderate
Political	0.15	0.64	0.47
Partisan news	0.24	0.69	0.48
Hard news	0.36	0.75	0.54
Lifestyle	0.75	0.89	0.85