

Who is Curating My Political Feed?

Characterizing Political Exposure of Registered U.S. Voters on Twitter

Assaf Shamir¹, Jennifer Oser², and Nir Grinberg^{1,*}

¹ Department of Software and Information Systems Engineering, Ben-Gurion University, Israel

² Department of Politics and Government, Ben-Gurion University, Israel

Abstract

Social media platforms offer people a variety of options to engage with politics, from directly following elected officials to discussing politics with social peers. Despite major advances in recent research on online political exposure through the lens of selective exposure, filter bubbles, and ideological echo chambers, little is known about the fundamental questions of what types of political actors people are exposed to on social media, and how these distinctive types vary across socio-demographic groups. We address this gap in the literature by analyzing unique panel data on more than 600,000 registered U.S. voters on Twitter during the 2020 U.S. Presidential campaign. We analyze this dataset to identify distinct types of political consumers and how they vary in terms of socio-demographics. Our findings suggest that the bulk of the population has a meaningful share of political content available from social peers, that the majority of this content originates from traditional sources of political information (media organizations, journalists, and politicians), and that media organizations are the dominant and direct source of political information on Twitter for nearly 20% of the sample population. These results advance our understanding of the way citizens learn about politics in new media, and pave the way for next-step research to identify the causal effect of exposure to distinct curators of political content on individuals' political attitudes and political behavior.

Keywords: political exposure, social media, online media diets, Twitter, curated flows

Biography: Assaf Shamir is an M.Sc. student at the Department of Software and Information Systems Engineering. Jennifer Oser is an Associate Professor in the Department of Politics and Government who focuses on the relationships between public opinion, political participation, and policy outcomes. Nir Grinberg is an Assistant Professor in the Department of Software and Information Systems Engineering who focuses on computational social science.

* Corresponding author. POB 653, Beer-Sheva, 8410501, Israel. Tel. +972-8-6479347, Email nirgrn@bgu.ac.il.

Introduction

The tectonic shifts in the media environment and the rise of social media platforms over the past two decades significantly changed the ways in which people are exposed to news and political information worldwide (Fletcher and Nielsen, 2018). This trend has been particularly swift in the United States, as Americans are now exposed to news more often on social media than in print, and for younger generations social media has become the dominant channel for news (Shearer, 2018).

In this increasingly networked media environment, the information that populates one's feed is an amalgamation of curation decisions taken by others, including social peers, journalists, politicians, advertisers, and proprietary ranking algorithms (Thorson and Wells, 2016). For example, Bakshy et al. (2015) showed how selective exposure on Facebook is partially determined by Facebook's Newsfeed ranking algorithm and determined more dominantly by the individual's choice of whom to follow. Of course, the effects of social media and digital media use writ large extend beyond the online world, and a growing body of research shows mobilization effects, where digital media use is associated with more traditional forms of political participation offline (Oser and Boulianne, 2020; Vaccari et al., 2015; Vaccari and Valeriani, 2021). Therefore, it is no surprise that issues of power and control (Barzilai-Nahon, 2008), limits of free speech (Morrow et al., 2021), and individual choice (Bakshy et al., 2015; Robertson et al., 2023) in political exposure on social media are some of the most contested topics of our time.

Despite the clear importance of advancing scholarly and real-world knowledge about political exposure on social media, we know relatively little about two key parameters of political exposure, namely: the prevalence of different types of actors in the stream of political

content people get from their ego-networks, and how political exposure varies across socio-demographic groups. Currently, no social media platform provides precise individual-level or comprehensive aggregate-level information about exposure, which poses a key obstacle to the field for advancing research on these crucial topics. The 'Social Science One' initiative (King and Persily, 2020) does provide aggregate information about viewership, but this information is limited in several important ways, namely: it is currently limited to Facebook data, includes only URLs and not all political content, does not distinguish eligible from non-eligible voters, and does not provide information about the person who posted the content. In lieu of more precise measurement, researchers have made recent contributions on these topics by relying on self-reported measures of political consumption and general-purpose web tracking data, while acknowledging the serious selection bias challenges that are inherent to this approach (Guess, 2021; Wojcieszak et al., 2022b).

In this study, we build on the contributions of this extant literature on political exposure by using a research design that leverages a large panel of U.S. registered voters and their activity on Twitter. The combination of these two data sources – i.e., Twitter data and registered voter data – creates the opportunity to ask and answer research questions that correspond to the two key parameters of political exposure noted above as requiring scholarly attention, namely: What are the types of political exposure on social media from different types of actors (RQ1)? And how do these types of political exposure vary across socio-demographic groups (RQ2)? To answer these research questions, we build on the curated flows theoretical framework developed by Thorson and Wells (2016) to identify the political content that is available to registered voters on Twitter and curated by different actors, including media organizations, journalists, politicians, opinion leaders, and social peers. We do so by using clustering methods that identify the

prototypical modes of political exposure in terms of the breakdown of the actors responsible for this content exposure, and by identifying the socio-demographic covariates of each distinctive cluster.

Using this novel methodological approach, our contributions are twofold. First, we provide new empirical evidence about the prototypical modes of political exposure – both in terms of quantity of political content, and composition of different actors who curate this content – by a large and representative sample of registered U.S. voters on Twitter. Second, we present findings on the varying levels and compositions of political exposure by different socio-demographic groups of registered U.S. voters on Twitter. Taken together, our contributions begin to address some of the most basic, yet unanswered, questions at the heart of the curated flow framework and social media communications: who are the most significant curators in political communication and for whom.

The Importance of Political Exposure Online and on Social Media

Numerous studies show that online political exposure and information consumption on social media are related to political attitudes and behaviors, both online and offline. For example, Valeriani and Vaccari (2016) found that accidental information on social media is positively associated with online political participation in multiple national contexts. A recent meta-analysis concluded that incidental exposure, an unintended form of exposure that is common on social media, is positively associated with a variety of pro-democratic attitudes and behaviors including news use, political knowledge, political participation, expressive engagement, and political discussion (Nanz and Matthes, 2022). In contrast, overreliance on the news to find you on social networks is negatively associated with important socio-political

indicators of political knowledge, political interest, and voting (Zúñiga and Diehl, 2019). Weeks et al. (2017) further find that counter-attitudinal incidental exposure on social media drives processes of selective exposure among stronger partisans, which subsequently leads to greater political information sharing.

A stream of recent studies informed by Thorson and Wells's curated flows framework has shown that the impact of political messaging also depends on the type of actor who is delivering it, as the same political message received from different types of sources may have a different impact on attitudes and behavior. For example, recent research indicates that statements by celebrities and online influencers seem to affect the public's real-world beliefs compared to similar statements by non-celebrities (Alatas et al., 2019; Arababa'h et al., 2021; Suuronen et al., 2021). In the realm of media sources, research has shown that high levels of exposure to media outlets with high levels of political content shape political knowledge and behavior, including the propensity to vote (de Vreese and Boomgaarden, 2006). Turning to the domain of peer networks, research by Graham et al. (2015) showed that over half of the political discussions in online forums in the U.K. led to at least one political action. The importance of the clear identification of actors is evident in Taylor et al.'s (2022) large-scale longitudinal field experiment which showed that content provided by anonymous sources is less impactful on viewers' opinions and behaviors compared to content shared by identified individuals with known reputations. Taken together, this emerging research indicates that the messenger's identity may be as important as the message itself.

Who is Heard and by Whom in Political Communication

A central element in democratic theory is the expression of ideas to allow public information-sharing and deliberation (Habermas, 1984). While research on the ways citizens construct their information diets certainly precedes the digital era (Katz and Blumler, 1974; Sears and Freedman, 1967), the shift to online media – accompanied by the weakening of traditional gatekeepers, and the context collapse that is common on social media (Davis and Jurgenson, 2014) – calls for renewed attention to the fundamental question of who is being heard in modern political communication. Addressing this question is important for advancing our understanding of the extent to which social media, and information systems more broadly, fulfill their egalitarian potential (Allen, 2015) or reinforce old political structure and power as the weapon of the strong (Hindman, 2009).

As noted, the theoretical and empirical importance of examining who is being heard is highlighted by Thorson and Wells' (2016) discussion of the role of individual-level "curation" for understanding media exposure and its effects. While individuals choose whom to follow, the notion of curation emphasizes the agency of external actors over the composition of one's social media feed. In particular, the curated flows framework lists a number of key actors including social peers, journalists, politicians, advertisers, and proprietary ranking algorithms. Merten (2021) explored the decisions (e.g. follow, block, or hide) users report taking in response to news curation by others. However, there is little empirical work that shows the relative prevalence of different actors in individuals' political exposure (Wells and Thorson, 2017). Two notable exceptions are the recent work by Wojcieszak et al. (2022b), which sheds new light on the channels (search engines, social media, aggregators, etc.) that lead people to news, and the work by Jürgens and Stark (2022), which measured the diversity of news accessed through different

channels. Nevertheless, we need to follow once more Prior's (2009) call for better measurement of news exposure to advance our understanding of the media effects of social media and gain better understanding of the ways political learning takes place on such social platforms (Bode, 2016). Currently, little is known about the amount of political content people are exposed to on social media, and the different kinds of actors in conveying this information. Therefore, our first research question is the following:

A key element in the composition of political exposure is political ideology and the range of ideas being represented. Some recent work indicates that exposure to political content through online social networks may serve to increase political polarization (Bail et al., 2018; Garrett et al., 2014; Shmargad and Klar, 2020). Yet, other studies indicate that social media exposure through weak ties and the visibility of social endorsements reduce polarization by offering diversity of exposure (Barberá, 2015; Messing and Westwood, 2014). People follow more frequently media and politician accounts that align with their ideology (Eady et al., 2019; Wojcieszak et al., 2022a), but there is still substantial overlap in people's news diets (Guess, 2021). Particularly because there is no consensus about the polarizing effects of media or social media (Prior, 2013; Zhuravskaya et al., 2020), it is important that we refine our understanding of political exposure on social media and consider it jointly with political ideology.

(RQ1) What are the prototypical types of political exposure on social media in terms of overall quantity, composition from different types of actors, and political ideology?

Socio-demographic characteristics are also linked to political consumption. Consistent with Vaccari and Valeriani's (2021) call to move beyond the "one-effect-fits-all" fallacy, we draw on prior literature to assess how key socio-demographic characteristics relate to distinctive types of political exposure. Firstly, there is a well-documented age gradient observed in the level of

interest in politics and the self-efficacy of individuals (Verba et al., 1995). As younger generations increasingly get their news on social media (Shearer, 2018), it is important to study the types of political content they are getting. In general, research shows that those with traditionally advantaged socio-demographic backgrounds (e.g., male, older) are more active politically, including efforts to seek out political content (Schlozman et al., 2018). Yet, research suggests that social media and online participation may have differential mobilization effects that recruit younger groups and women more actively into politics (Oser et al., 2013; Oser and Boulianne, 2020). A possible reason for that is that publics and counterpublics pay attention to different issues on social media (Jackson and Foucault Welles, 2015; Shugars et al., 2021). Therefore, our second research question is the following:

(RQ2) How do the prototypical types of political exposure on social media vary for distinctive socio-demographic groups?

We now turn to the methodological challenges and opportunities for making robust inferences about the political exposure of citizens on social media.

Measuring Political Exposure in the Digital Age: Challenges and Opportunities

Although survey data have long been a leading source of information about habits of political consumption, researchers are actively looking for ways to improve their accuracy (Berinsky, 2017; Guess, 2015). In the context of social media, prior work showed that there could be large discrepancies between actual and reported frequency of posting about politics (Guess et al., 2019; Henderson et al., 2021).

Digital trace data provide new and complementary ways to measure individuals' behavior directly, often collected through dedicated software installed by participants (Flaxman et al., 2016). For example, Guess (2021) uses web browsing data combined with survey responses to characterize Americans' media consumption habits and examine whether internet use indeed facilitates selective exposure to like-minded views. While this approach provides the most comprehensive picture of both objective and subjective measures of political engagement, it is often limited to a few thousand participants who are willing to volunteer their data. In addition to selection issues, the sample quickly becomes statistically underpowered for obtaining accurate descriptions of subgroups and heterogeneity of activity (Hughes et al., 2021). This challenge of directly measuring political exposure for the field as a whole is clearly identified in Amsalem and Zoizner's (2022) observation in their comprehensive meta-analysis of learning about politics on social media that most relevant studies do not include any direct measure of political exposure, and also lack sufficient sample size to estimate heterogeneous effects.

A recently developed alternative approach for directly gathering data on individuals' behavior is to use publicly available social media data. Despite meaningful changes in Twitter's leadership and policies beginning in 2022 (Anderson, 2023), it has been a uniquely important social media platform for investigating exposure to political content of a large sample of users due to the active engagement of media outlets and political figures on the platform up through and including the observation period of the current study (Bail et al., 2018; Barberá, 2015; Eady et al., 2019; Guess, 2021). In 2021, around one-in-five (23%) of Americans reported using Twitter (Odabaş, 2022), and almost seven in ten of them said they receive their news regularly through the platform (Mitchell et al., 2021). While Twitter users in the U.S. were found to be younger and more likely to be Democrats in comparison to the general public (Wojcik and

Hughes, 2019), prior work has shown that differences between Twitter users and non-users are due mostly to the demographic composition of social media users, which can be addressed by controlling for a few demographic variables (Mellon and Prosser, 2017). Importantly, rigorous empirical work on the representativeness of Twitter users shows some modest demographic differences between Twitter users and the general population that can be accounted for analytically (Hughes et al., 2021).

This recently developed methodological approach of analyzing publicly available social media data is an important contribution to extant literature, as no social media platform currently offers public access to data about individuals' exposure to distinctive types of political content, and as a result hardly any research has directly measured it. An increasingly prominent approach for approximating exposure involves the collection of content posted by accounts followed by the focal user on social media (Eady et al., 2019; Grinberg et al., 2019). As described in Grinberg et al. (2019) this approach does not guarantee exposure, i.e. that an individual actually saw a particular post, but it does directly speak to the content available to people in their social feeds from their ego-network.

Building on this discussion of the current state of the art in research on political exposure, the following section details how the current study applies this recently developed novel methodological approach to measuring potential political exposure.

Data and Methods

Twitter Panel and Political Exposure

The foundation of this research is a sampling frame of over 1.5 million Twitter users that were successfully matched to public U.S. voter registration records. Following the same approach

described in prior work (see Grinberg et al., 2019 and Shugars et al., 2021 for more details), the matching used the Twitter Decahose, a 10% random sample of all tweets, to identify 290 million profiles who posted content between January 2014 and March 2017. The profiles were then matched against voter records provided by TargetSmart in October 2017 for all 50 U.S. states and the District of Columbia. A Twitter account was matched to a voter record if their full name exactly matched and they were the only person with that name in either the city- or state-level geographic area specified in both datasets. While the reliance on full names and disclosed locations eliminates many fake, automated (bot), and organizational accounts, it does raise concerns about potential selection bias. However, rigorous comparison of this panel with a gold-standard survey conducted by Pew Research Center showed that only small demographic and ideological differences exist between the two samples of registered U.S. voters (Hughes et al., 2021). Importantly, this matched dataset provides comprehensive data on individuals' social media behavior through Twitter, as well as the basic socio-demographic information. Age and gender are retrieved directly from public voter registration records, while race/ethnicity and party affiliation are based on TargetSmart inferences (see validation in Shugars et al. 2021, Appendix B).

The primary dataset used in this work is a set of 606,112 panel members for whom we have at least one indication of activity on Twitter during the 2020 presidential election (August to November, inclusive). This set includes all individuals who posted or liked at least one tweet during the four months of the study period, which is important for capturing not only highly-active users but also users who rarely or never post on Twitter but still consume content. Our target population is therefore restricted to registered U.S. voters who were minimally active on Twitter during the 2020 presidential election, and we make no claims about the important, yet

omitted, populations of eligible non-registered voters or inactive Twitter users. Appendix A provides socio-demographic information about these active panel members. By focusing on a period of a presidential election, we examine potential political exposure at its peak (Grinberg et al., 2019; Peterson et al., 2021), when it matters, and where most politically-relevant actors are likely to be active.

To model potential exposure, we follow the approach used in prior work to approximate individuals' social feed using the content available from the accounts they follow (Eady et al., 2019; Grinberg et al., 2019). Due to Twitter rate limits, we could not collect all tweets posted by the 51 million users followed by our sample during the observation period, and base our estimates on the 10% random sample of the Twitter Decahose, similar to the approach used in Grinberg et al. (2019). It is important to note that this approach only captures *potential* exposure and not exposure per se, partially due to the sampling of followees' content, but more importantly because exposure requires knowledge about user activity and the outcomes of personalized ranking algorithms – two ingredients that are unavailable to the research community. Nevertheless, in lieu of more precise information from social media platforms about exposure, this approach reflects the most accurate and reproducible estimate currently available to the public about the composition of people's social feeds from their ego-network.

To identify political tweets, we train a Machine Learning classifier and validate its accuracy against human coders, similar to the approach used in prior work (Bakshy et al., 2015; Eady et al., 2019; Grinberg et al., 2019). The classifier resulted in a precision of 88.8%, a recall of 80.0% for tweets about U.S. politics, and a recall of 96.4% in the subcategory of election-related tweets. More details about the classifier and its validation are in Appendix B.

Identifying Different Actors in Political Exposure

Following the curated flows framework, we examine different types of actors that curate political content for individuals in our panel. We focus on four types of actors directly mentioned in Thorson and Wells' (2016) framework – media organizations, journalists, politicians, and social peers – and include a fifth category of "opinion leaders" who have been identified as important in recent research. Specifically, opinion leaders have large followship on social media, nonpolitical even more than political opinion leaders (Mukerjee et al., 2022), and a demonstrated ability to influence public opinion (Alatas et al., 2019). To date, however, the share of political content originating from opinion leaders' accounts in day-to-day political exposure has not been directly quantified.

To identify accounts of different actors, we rely on manually curated lists of accounts by recent academic works, develop methods to identify additional accounts, and validate the accuracy of our inferences and the robustness of results. We identify media organizations by using the list of media organizations in McCabe et al. (2022), which started with a seed list of known media organizations and used snowball sampling to expand it iteratively. We supplement this list with the media organizations listed in Wojcieszak et al. (2022a). We also rely on Wojcieszak et al.'s (2022a) extensive list to identify 1,951 journalists' accounts.

Politician accounts are identified through an original list we compiled by linking an official list of the 116th MoC names to a list of MoC accounts on Twitter (Wrubel and Kerchner, 2020). Our list of 927 accounts includes both the accounts of MoC and their election campaigning accounts, which is important for capturing all messages originating from politicians during an election cycle. We supplemented this list with 51 additional politician accounts found

in Wojcieszak et al. (2022a). Appendix C details our identification strategy for politician accounts.

To identify opinion leaders, we rely on the manually labeled list of accounts of nonpolitical opinion leaders (e.g. public figures, popular brands, celebrities) by Mukerjee et al. (2022), and extend it using Bail et al.'s (2018) approach of considering as an opinion leader any account followed by 15 or more active MoC. Since the accounts followed by multiple MoC may themselves belong to media organizations, journalists, or politicians, we use a combination of automatic and manual annotation of accounts using profile information to distinguish opinion leaders from other actor types. Validating this approach using a held-out random sample of accounts showed an accuracy of 80.0%, which is considerably higher than random assignment with four categories. Appendix D details our identification strategy for opinion leader accounts.

We consider as a social peer any account that does not appear on any of the aforementioned lists of media organizations, journalists, politicians, or opinion leaders. Importantly, however, the same content can be attributed to multiple people due to the complex nature of retweets, quotes, replies or mentions. To support different attributions of content and interpretations of the results, we distinguish between *direct* and *indirect* exposure. Direct exposure comes from directly following the accounts of media organizations, journalists, politicians, or opinion leaders. Indirect exposure is mediated through social peers who retweet, quote, mention or reply to a tweet by these actors. To complete our documentation of the analytical work we conducted to identify distinctive actors in political exposure, Appendix E validates account inferences and robustness, and Appendix F provides summary information about all curating actors analyzed in this study.

Measuring Political Alignment

Modeling the ideological leaning of news content and politicians is fundamental to assessing people's online media diet, and different methods have been proposed for this purpose. Our analysis of political alignment focuses on three aspects of citizens' political exposure: first, exposure to left- and right-leaning MoC; second, exposure to left- and right-leaning opinion leaders; and third, exposure to political news sites based on the ideological alignment of people who share links to this site.

For Members of Congress, we consider their party affiliation to be representative of their political leaning, excluding four Independents and one Libertarian. For opinion leaders, we infer their political leaning based on the composition of MoC that follow them. For news sites, we follow Bakshy et al.'s (2015) approach by estimating the political leaning based on the political alignment of people sharing links to the website. Appendix G provides further detail on our estimation of the political alignment of news sites.

Clustering Methodology to Infer Prototypical Types of Political Exposure

We use state-of-the-art clustering methods to identify prototypical types of political exposure on Twitter. To this end, we take into account three categories of information about panel members' political potential exposure, namely: (i) the overall magnitude of political exposure and its share out of all content from peers, (ii) the curating sources (partitioned by direct and indirect exposure), and (iii) ideological leaning of news sites in the feed. The first two categories directly speak to RQ1, while the third category enables the clusters to capture political alignment of content, due to the importance of this dimension. Appendix H provides the full list and

description of the 15 features that we analyze that provide measures of these three key categories of political exposure.

To meaningfully identify clusters in this high-dimensional data, we follow the standard practice in machine learning of reducing the dimensionality of the data first (Allaoui et al., 2020; Grootendorst, 2022), and only then apply the clustering algorithm. Specifically, we use Uniform Manifold Approximation and Projection (UMAP) to reduce dimensionality (McInnes et al., 2020), and then apply the clustering algorithm of HDBSCAN (McInnes et al., 2017). HDBSCAN is well-suited for our task because it can identify clusters of any geometry (in contrast to algorithms like K-means that assume a Gaussian distribution), and because clusters are formed based on density, starting with the densest areas first and then split or prune sparser sub-areas (see McInnes et al., 2017 for more details). This approach results in clusters that capture the more common types of political exposure in our sample and is robust to outliers.

Results

In this section, we report results regarding our two research questions: What are the prototypical types of political exposure on Twitter, and how does the distribution of these exposure types vary across distinctive socio-demographic groups?

In order to identify robust patterns of political exposure, users that did not meet a minimum threshold of political exposure were assigned to a separate cluster of ‘nonpolitical’ users. Consistent with prior research, we set this threshold at one observed political tweet a day on average in the decahose, i.e., a total of 122 observations throughout the entire election period (Grinberg et al., 2019). Based on this criterion, 8.9% of the population was directly assigned to the nonpolitical cluster.

Clustering the political exposure of users with a minimal amount of exposure to politics resulted in seven clusters that cover 99.1% of the population, 0.4% of accounts that the algorithm identified as outliers, and three small clusters with several hundred people that together amount to 0.5% of the population. Hereafter, we omit these outlier and small-cluster accounts from further analysis and focus on the core exposure patterns identified in 99.1% of the population.

Figure 1 presents the prototypical types identified by clustering the political exposure of panel members. Each point in Figure 1A represents an individual and their political consumption at the reduced two-dimensional space computed by the UMAP algorithm with its color designating its cluster assignment. Points that are closer together represent individuals with similar properties of political exposure. Figure 1B shows the median amount of political exposure available in people's ego-networks¹ and its share out of all content available to people on Twitter, for each cluster separately. For example, the cluster referred to as ‘media superconsumers’ consists of 4.7% of the population, and this cluster’s median user has nearly 6,000 political tweets available to them each day, which comprises 52% of their total daily tweet exposure. Cluster labels reflect our assessment of the distinctive features of each cluster in terms of size in the population, the amount of political exposure it represents, and the composition of political curators and sources in the cluster. In particular, we labeled the clusters as follows: one cluster as nonpolitical due to low level of exposure to politics (not meeting our minimal threshold), one cluster as opinion leaders (OL) oriented due to an elevated level of exposure to opinion leaders, one cluster as average consumers based on its large share in the population (50.5%), two clusters as partisan due to the political alignment of content, and three clusters based on elevated levels of media consumption of increasing degrees.

¹ Absolute number estimates are based on the multiplication of observed amounts in the 10% random sample ‘decahose’ by 10.

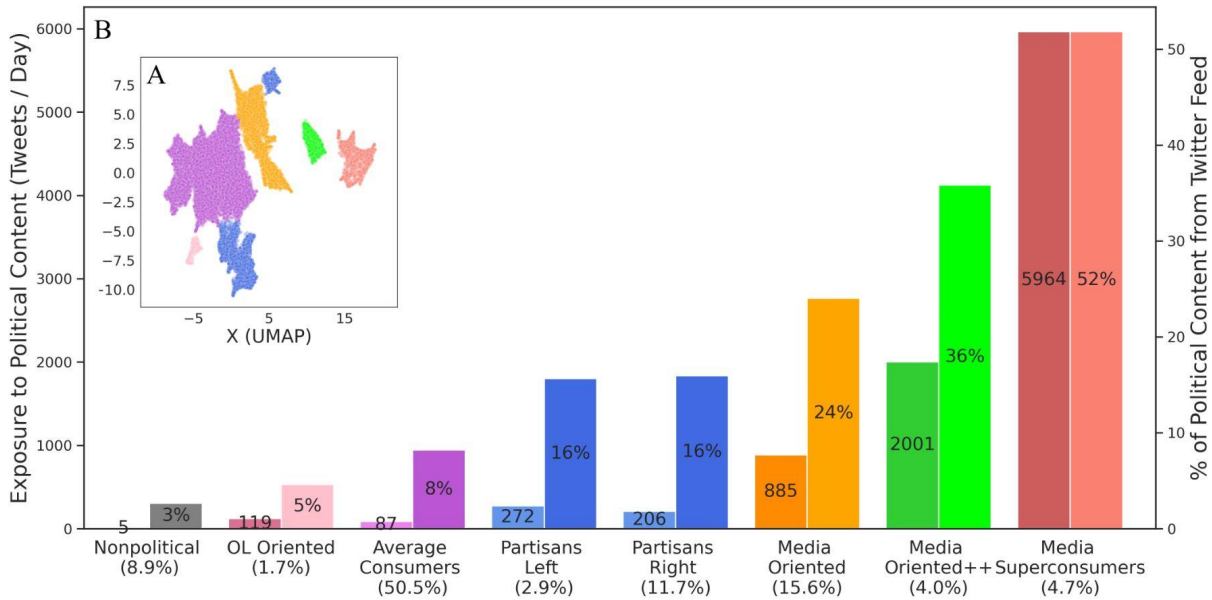


Figure 1: Prototypical types of individual political exposure.

Each point in panel (A) represents the political exposure of a single panel member, reduced to two dimensions using the UMAP algorithm, and colored by the cluster assignment obtained from HDBSCAN. Panel (B) shows the median number of political tweets available to individuals per day (left bars), and their percentage out of all tweets available to them on Twitter (right bars). Cluster labels and their share in the population are specified on the x-axis. Colors are consistent between the two figure panels. 95% bootstrapped CIs are omitted from the figure due to their small magnitude, which are upper bounded by 27 exposures to tweets and 0.28%, respectively.

Figure 1 provides two key observations. First, we observe that the bulk of the population has a meaningful share of politics in their Twitter feeds. This finding is consistent with prior work that shows that Twitter users are above average in their political engagement (McClain et al., 2021). Except for the two clusters with the lowest share of politics (Nonpolitical and OL Oriented), all other clusters, which account for nearly 90% of the population, have 8% or more of politics in the feed. Even if this finding only applies to registered U.S. voters on Twitter and to a lesser extent to other social media platforms, it presents a picture of an engaged public during an election cycle. Second, we observe that the Partisan Left and Partisan Right clusters exhibit very similar levels of political consumption to one another, and that the media-oriented clusters, which combine to 15-24% of the population, have a larger share of politics in the content

available from social peers than the partisan clusters. Next-step research can use these findings to investigate the causal relationship between the overall level of direct and indirect media exposure and subsequent attitudes and behaviors such as ideological polarization.

In addition to the overall level of exposure to politics, the clusters we identified vary in the composition of political exposure from distinctive curating actors. Figure 2 shows the breakdown of political exposure by different actor types including media organizations, journalists, politicians, opinion leaders, and social peers. Lighter-colored bars indicate indirect exposure, where the focal user received content from a peer that referred to a media organization, journalist, politician, or opinion leader (i.e., through a “retweet”). For example, the group of Average Consumers receives more than four times indirect exposure (22.8%) than direct exposure (5.2%) to politicians. In stark contrast, the Media Superconsumers group receives nearly 90% of their political exposure directly from media organizations with hardly any indirect exposure. A similar pattern appears for the OL Oriented cluster that gets more than 70% of its political exposure directly from opinion leaders. Figure E2 in Appendix E provides alternative ordering of the bars, first by direct and indirect exposure, then by actor type.

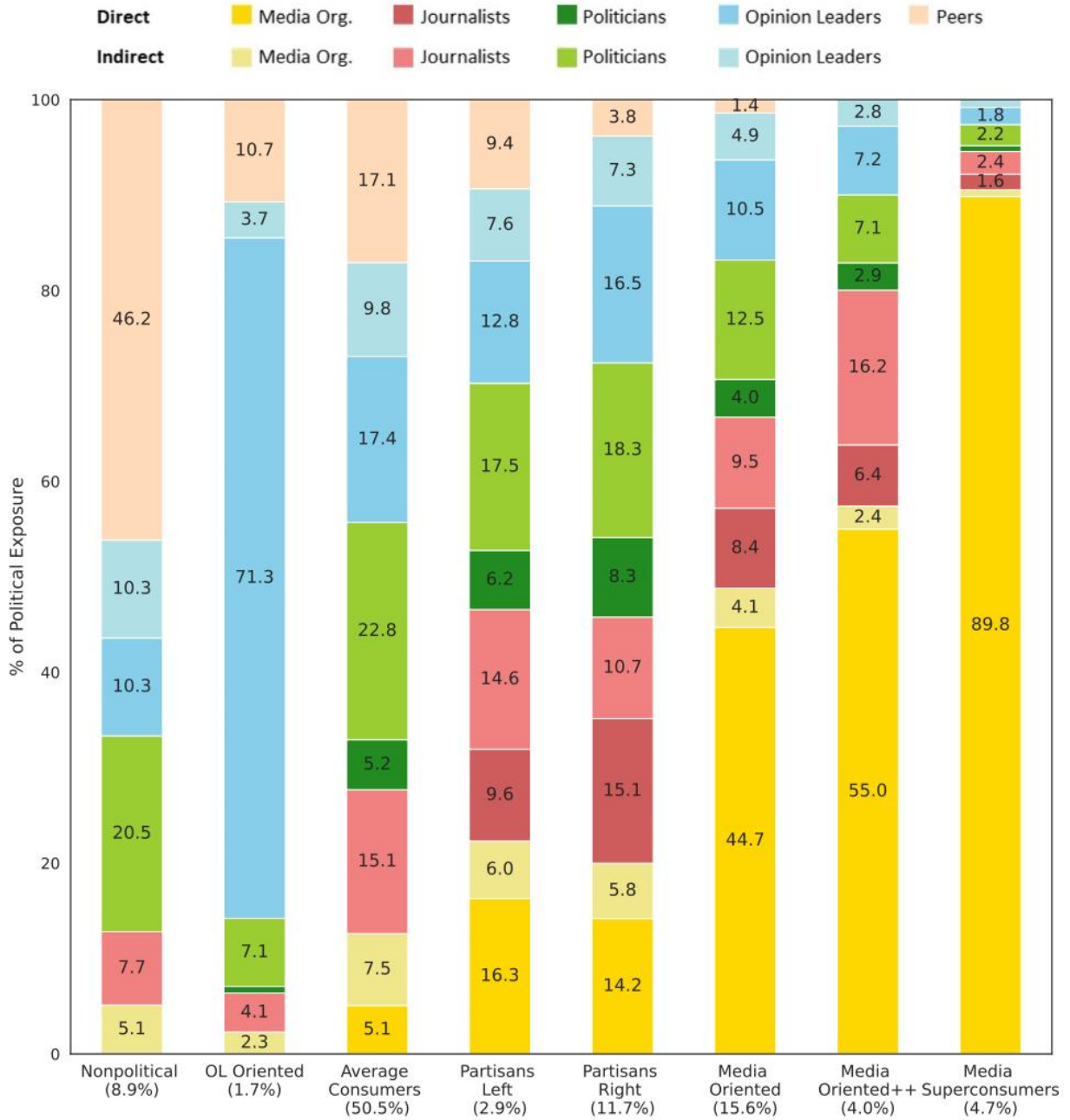


Figure 2: The composition of political exposure across clusters.

The share of politics curated by different actor types (y-axis) across clusters (x-axis). Darker-colored bars represent direct exposure to media organizations, journalists, politicians, opinion leaders, and social peers. Lighter-colored bars represent indirect exposure to media organizations, journalists, politicians, or opinion leaders through social peers.

Focusing again on the six clusters representing the bulk of the population (average, partisans and media consumers; nearly 90%), Figure 2 provides three key observations. First, it shows that more than half of political exposure for these clusters comes from traditional sources of political information – media organizations, journalists, and politicians – and that share increases with the increased share of politics in the feed (reflected in cluster ordering from left to right as shown in Fig. 1). Second, the clusters also vary considerably in terms of direct and indirect exposure. Nonpolitical consumers are only indirectly exposed to traditional sources, while Average Consumers get most political exposure indirectly, through social peers and not directly from traditional sources. Partisans have the largest share of political exposure directly from politicians and journalists, and relatively little direct exposure to media organizations compared to the media oriented clusters. Finally, leaving aside the more extreme superconsumers, we see that the Media Oriented and Media Oriented++ clusters, which combine to nearly 20% of the sample population, get about half of their political exposure directly from media organizations. Taken together, these findings highlight the importance of considering both direct and indirect exposure to traditional sources as well as opinion leaders, particularly for people who receive a smaller share of political content.

We now turn to our second research question, which focuses on how different socio-demographic groups engage with different types of political consumption. Figure 3 shows how age, gender, race/ethnicity, and party affiliation (y-axes) are distributed across the different exposure types (x-axis; following the same order of increasing share of political exposure from left to right). Specifically, the figure shows the average age estimate for each cluster, and the percentage of women, Caucasians, and registered Democrats in each cluster (See Appendix A for

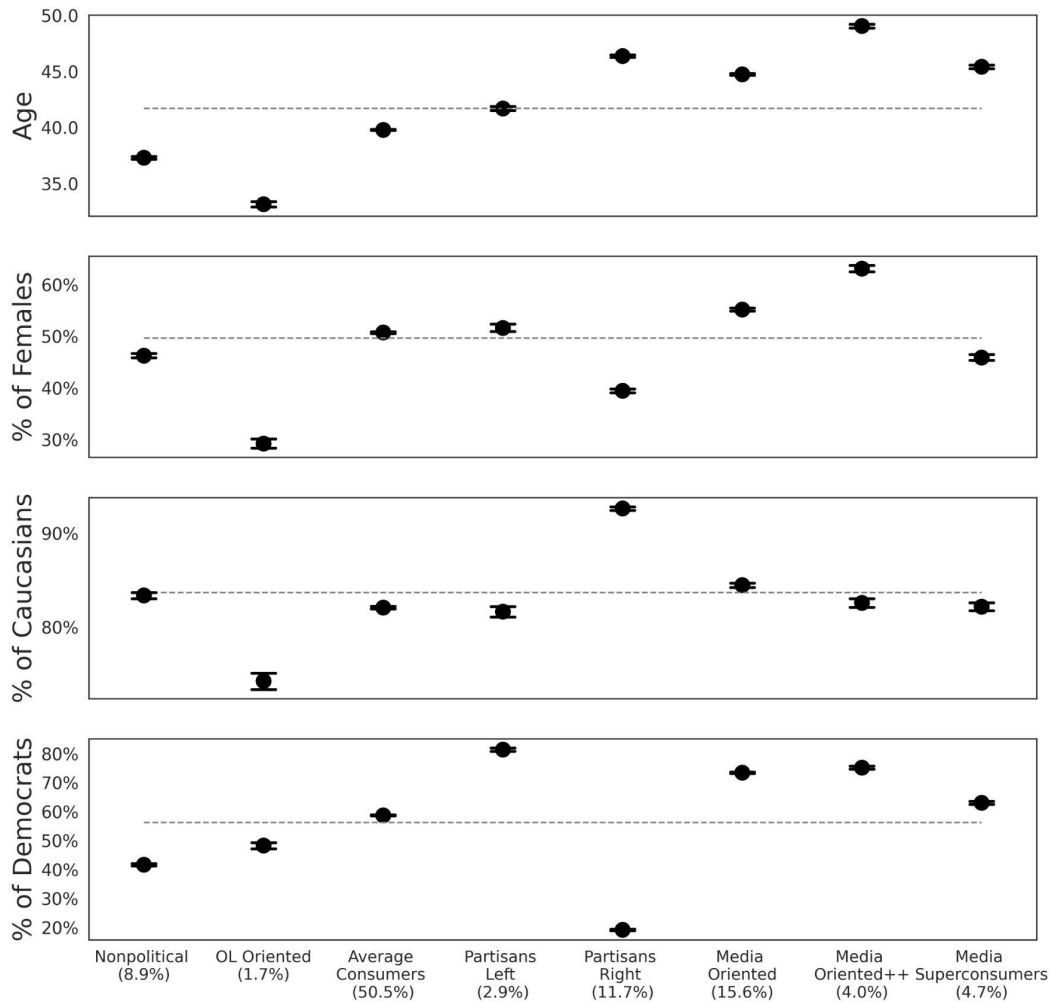


Figure 3: Socio-demographic characteristics among different political exposure types.
 Note: Sample averages are marked in a gray dashed line. 95% bootstrapped CIs are shown (mostly occluded due to their small size).

further detail on socio-demographic characteristics). The dashed horizontal line in each panel designates the sample average as a baseline for comparison.

Figure 3 provides several key observations. First, there is a clear positive association between the average age and the share of political exposure of the clusters, as the literature generally predicts at the individual-level (e.g., Verba et al., 1995). Figure I1 in Appendix I further supports this using the full age distributions. If our cluster-level findings persist at the individual-level, it suggests that age is linked with different overall amounts of political exposure

and composition of curating actors. It should be noted that cluster- and individual-level results may not be aligned due to within-cluster heterogeneity. Second, the figure shows meaningful gender and race/ethnicity differences between Partisan Right and Partisan Left. The fact that the two partisan clusters have different demographic characteristics, yet a similar breakdown of actors in their political feeds, suggests that there may be some commonalities in the polarization processes across political ideology. Third, the OL Oriented cluster is distinctively young, male, and non-Caucasian. Together with its small size in the sample (1.7%) and overrepresentation of opinion leaders, this seems like a niche cluster that gets exposed to politics incidentally through nonpolitical opinion leaders.

Finally, we find that the Media Oriented and Media Oriented++ clusters, which together combine to nearly 20% of the sample population, have significantly higher percentages of women, registered Democrats, and older adults. Prior work has documented a partisan gender gap in American politics (Doherty et al., 2018), with women more likely to identify as Democrats. However, to the best of our knowledge, no prior work has shown such large gender differences in political consumption directly from media organizations.

Discussion

Much of the discussion about societal factors that may be contributing to democratic backsliding in advanced democracies – including rising populism, decreasing trust in media and political establishment, increased polarization, and misinformation – has been linked to the increased prevalence of digital media, and in particular, to social media. Social platforms are, indeed, widely adopted as a source of political information, and a primary source for many young adults. These trends in political content exposure call for a better theoretical understanding of political

exposure on these platforms, including next-step causal examination of the impact of different types of political exposure on subsequent political attitudes and political behaviors. Robust analysis of these phenomena requires new computational methods for making valid inferences based on digital trace data that complement traditional methods.

Grounded in the curated flows theoretical framework, this work contributes to the conceptualization and measurement of actors responsible for this curation. The empirical findings describe the types of actors that are responsible for political content distribution to registered U.S. voters on Twitter, and the demographic characteristics of distinctive types of political consumers. We found that the bulk of the population on the platform was exposed to non-negligible amounts of political content during the 2020 U.S. Presidential election, ranging on an average day from 87 political tweets (8% of the overall feed) to a few thousand political tweets (52% of the overall feed). Notably, more than half of political tweets originated from traditional sources of political information – media organizations, journalists, and politicians. The observational findings of the current study pave the way to investigate the causal impact of political content curation by distinctive actors on people’s subsequent attitudes and behaviors, such as left-right ideological polarization and affective polarization.

Another key finding is that media organizations are an important source of political information for a large proportion of the sample, with much of this exposure taking place directly and without any mediation by peers. These findings contribute to the debate about the erosion of traditional gatekeepers, as most media organizations on our lists have, fundamentally, the same editorial processes that Kurt Lewin (1943) wrote about when he first introduced Gatekeeping theory. Our results show that a substantial amount of modern consumers of political content on Twitter *choose* to replicate traditional gatekeeping in new media. Future research could

investigate the curation roles and impacts that these media-oriented individuals have on their local network and examine the role of media organizations in influencing subsequent political attitudes and behaviors of specific socio-demographic groups.

Along with these contributions, this research has several important limitations previewed earlier in the study. First, while the findings are likely to capture political exposure of American adults on Twitter in 2020, which were about a fifth of American adults (Odabaş, 2022), without direct measurement on other platforms and populations it is unclear how these findings will generalize. On the one hand, previous research had found some similar media effects to Twitter and the more widely-used Facebook (e.g., Valenzuela et al. 2018). On the other hand, numerous studies have emphasized the importance of considering specific contextual features in the relationship between social media use and political behavior (e.g., Vaccari and Valeriani, 2021). Additional comparative research is needed to fully contextualize these findings. A second key limitation is the empirical focus of our analysis on potential political exposure, meaning content that is available to people and not necessarily the content that is actually seen by them. Although this is a limitation that affects all scholarship on these topics, it is important to note that the difference between these two populations may be systematically biased by factors such as the time when individuals visit their feeds, the duration of their visits, and the algorithmic content ranking conducted by social platforms. A third key limitation is that since we relied on manually curated lists and verification for identifying distinctive curation actors (e.g., media organizations, opinion leaders), we cannot guarantee the comprehensiveness of the lists. For example, the list of politicians does not include state and local politicians, which may have different levels of exposure and audiences.

There are also several avenues for future work to expand this research. In terms of theory, the curated flows framework puts much of its emphasis on the actor who is doing the curation. Our study shows that there is room to expand the theory to consider the producer of the content in addition to the person who curates it as it propagates through the network. Content attribution is also a major challenge that calls for methodological contributions. Furthermore, future research can examine how the different types of political content exposure are related to pro-democratic attitudinal measures known to be crucial for robust democratic functioning, such as political knowledge, and political efficacy. Future research can explore how exposure varies along other socio-demographic dimensions such as educational attainment or socio-economic status. In addition, the current study paves the way for next-step experimental research that makes clear causal identification of how different types of political consumers engage in and mobilize to political action both online and offline.

Acknowledgments

The research was approved by BGU's Departmental Ethics Review Board (Approval #SISE-2022-32).

Funding: Funded by the European Union (to Oser: ERC, PRD, project number 101077659); Ben-Gurion University Data Science Research Center's Young Faculty Interdisciplinary Grant 2021-2022 (to Grinberg and Oser as co-PI's); by Ben-Gurion University's Interdisciplinary Fellowship for Outstanding Graduate Students 2021-2023 (to Shamir); by the Israel Science Foundation (to Oser, grant number 1246/20). We would like to acknowledge the helpful comments received from Diyi Liu, Eran Amsalem, Patrícia Rossini and participants of the 8th IJPP conference, the 2022 Zurich Digital Democracy workshop, and the 2022 American Political Science Association, Politicla Communication pre-conference. Other valuable colleagues

contributed insightful comments including Alon Zoizner and Oren Tsur. Finally, we would like to thank the anonymous reviewers who helped crystalize and articulate more clearly the core contributions of this work.

Data and materials availability: All data and code necessary to evaluate the final conclusions in the paper are available in a public GitHub repository at <https://github.com/Socially-Embedded-Lab/pol-exp-curators> and in Harvard Dataverse at <https://doi.org/10.7910/DVN/OP9JQT>. Although the Twitter panel was originally constructed by linking two public resources (public voter record and public Twitter activity), we do not release any personally identifying information as this may expose individuals to additional unanticipated risks, violate people's privacy expectations (Williams et al., 2017), and breach Twitter's Terms of Service.

References

- Alatas, V., Chandrasekhar, A.G., Mobius, M., Olken, B.A., Paladines, C., 2019. When celebrities speak: A nationwide twitter experiment promoting vaccination in Indonesia. National Bureau of Economic Research.
- Allaoui, M., Kherfi, M.L., Cheriet, A., 2020. Considerably Improving Clustering Algorithms Using UMAP Dimensionality Reduction Technique: A Comparative Study. pp. 317–325.
- Allen, D., 2015. Reconceiving Public Spheres, in: From Voice to Influence: Understanding Citizenship in a Digital Age. University of Chicago Press, p. 179.
- Arababa'h, A., Marble, W., Mousa, S., Siegel, A.A., 2021. Can Exposure to Celebrities Reduce Prejudice? The Effect of Mohamed Salah on Islamophobic Behaviors and Attitudes. *Am Polit Sci Rev* 115, 1111–1128. <https://doi.org/10.1017/S0003055421000423>

- Amsalem, E., Zoizner, A., 2022. Do people learn about politics on social media? A meta-analysis of 76 studies. *Journal of Communication*. <https://doi.org/10.1093/joc/jqac034>
- Anderson, M., 2023. After Musk's takeover, big shifts in how Republican and Democratic Twitter users view the platform. Pew Research Center.
- Bail, C.A., Argyle, L.P., Brown, T.W., Bumpus, J.P., Chen, H., Hunzaker, M.B.F., Lee, J., Mann, M., Merhout, F., Volfovsky, A., 2018. Exposure to opposing views on social media can increase political polarization. *Proc Natl Acad Sci USA* 115, 9216–9221. <https://doi.org/10.1073/pnas.1804840115>
- Bakshy, E., Messing, S., Adamic, L.A., 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science* 348, 1130–1132.
- Barberá, P., 2015. How Social Media Reduces Mass Political Polarization. Evidence from Germany, Spain, and the U.S. 46.
- Barzilai-Nahon, K., 2008. Toward a theory of network gatekeeping: A framework for exploring information control. *J. Am. Soc. Inf. Sci.* 59, 1493–1512. <https://doi.org/10.1002/asi.20857>
- Berinsky, A.J., 2017. Measuring Public Opinion with Surveys. *Annual Review of Political Science* 20, 309–329. <https://doi.org/10.1146/annurev-polisci-101513-113724>
- Bode, L., 2016. Political News in the News Feed: Learning Politics from Social Media. *Mass Communication and Society* 19, 24–48. <https://doi.org/10.1080/15205436.2015.1045149>
- Davis, J.L., Jurgenson, N., 2014. Context collapse: theorizing context collusions and collisions. *Journal of Computer-Mediated Communication* 17, 476–485. <https://doi.org/10.1080/1369118X.2014.888458>
- de Vreese, C.H., Boomgaarden, H., 2006. News, Political Knowledge and Participation: The Differential Effects of News Media Exposure on Political Knowledge and Participation.

- Acta Polit 41, 317–341. <https://doi.org/10.1057/palgrave.ap.5500164>
- Doherty, C., Kiley, J., O’Hea, O., 2018. Wide gender gap, growing educational divide in voters’ party identification. Pew Research Center.
- Eady, G., Nagler, J., Guess, A., Zilinsky, J., Tucker, J.A., 2019. How Many People Live in Political Bubbles on Social Media? Evidence From Linked Survey and Twitter Data. SAGE Open 9, 215824401983270. <https://doi.org/10.1177/2158244019832705>
- Flaxman, S., Goel, S., Rao, J.M., 2016. Filter Bubbles, Echo Chambers, and Online News Consumption. Public Opinion Quarterly 80, 298–320.
- Fletcher, R., Nielsen, R.K., 2018. Are people incidentally exposed to news on social media? A comparative analysis. New media & society 20, 2450–2468.
- Garrett, R.K., Gvirsman, S.D., Johnson, B.K., Tsfati, Y., Neo, R., Dal, A., 2014. Implications of Pro- and Counterattitudinal Information Exposure for Affective Polarization: Partisan Media Exposure and Affective Polarization. Hum Commun Res 40, 309–332. <https://doi.org/10.1111/hcre.12028>
- Graham, T., Jackson, D., Wright, S., 2015. From everyday conversation to political action: Talking austerity in online ‘third spaces.’ European Journal of Communication 30, 648–665. <https://doi.org/10.1177/0267323115595529>
- Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., Lazer, D., 2019. Fake news on Twitter during the 2016 U.S. presidential election. Science 363, 374. <https://doi.org/10.1126/science.aau2706>
- Grootendorst, M., 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure.
- Guess, A., Munger, K., Nagler, J., Tucker, J., 2019. How accurate are survey responses on social media and politics? Political Communication 36, 241–258.

- Guess, A.M., 2021. (Almost) Everything in Moderation: New Evidence on Americans' Online Media Diets. *American Journal of Political Science* *ajps*.12589.
<https://doi.org/10.1111/ajps.12589>
- Guess, A.M., 2015. Measure for Measure: An Experimental Test of Online Political Media Exposure. *Political Analysis* *23*, 59–75. <https://doi.org/10.1093/pan/mpu010>
- Habermas, J., 1984. *The theory of communicative action*. Beacon press.
- Henderson, M., Jiang, K., Johnson, M., Porter, L., 2021. Measuring Twitter use: validating survey-based measures. *Social Science Computer Review* *39*, 1121–1141.
- Hindman, M., 2009. *The myth of digital democracy*. Princeton university press, Princeton (N.J.).
- Hughes, A.G., McCabe, S.D., Hobbs, W.R., Remy, E., Shah, S., Lazer, D.M.J., 2021. Using Administrative Records and Survey Data to Construct Samples of Tweeters and Tweets. *Public Opinion Quarterly* *85*, 323–346. <https://doi.org/10.1093/poq/nfab020>
- Jackson, S.J., Foucault Welles, B., 2015. Hijacking# myNYPD: Social media dissent and networked counterpublics. *Journal of communication* *65*, 932–952.
- Jürgens, P., Stark, B., 2022. Mapping Exposure Diversity: The Divergent Effects of Algorithmic Curation on News Consumption. *Journal of Communication* *72*, 322–344.
<https://doi.org/10.1093/joc/jqac009>
- Katz, E., Blumler, J.G., 1974. *The uses of mass communications: Current perspectives on gratifications research*.
- King, G., Persily, N., 2020. A new model for industry–academic partnerships. *PS: Political Science & Politics* *53*, 703–709.
- Lewin, K., 1943. Forces behind food habits and methods of change. *Bulletin of the national Research Council* *108*, 35–65.

- McCabe, S., Green, J., Wan, A., Lazer, D., 2022. New Tweetscores. OSF, Available at: osf.io/794va.
- McClain, C., Widjaya, R., Rivero, G., Smith, A., 2021. The behaviors and attitudes of US adults on Twitter. Pew Research Center.
- McInnes, L., Healy, J., Astels, S., 2017. hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software* 2. <https://doi.org/10.21105/joss.00205>
- McInnes, L., Healy, J., Melville, J., 2020. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.
- Mellon, J., Prosser, C., 2017. Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Research & Politics* 4. <https://doi.org/10.1177/2053168017720008>
- Merten, L., 2021. Block, Hide or Follow—Personal News Curation Practices on Social Media. *Digital Journalism* 9, 1018–1039. <https://doi.org/10.1080/21670811.2020.1829978>
- Messing, S., Westwood, S.J., 2014. Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online. *Communication research* 41, 1042–1063.
- Mitchell, A., Shearer, E., Stocking, G., 2021. News on Twitter: Consumed by Most Users and Trusted by Many. Pew Research Center.
- Morrow, G., Swire-Thompson, B., Polny, J.M., Kopec, M., Wihbey, J.P., 2021. The emerging science of content labeling: Contextualizing social media content moderation. *Journal of the Association for Information Science and Technology*.
- Mukerjee, S., Jaidka, K., Lelkes, Y., 2022. The Political Landscape of the U.S. Twittersverse. *Political Communication* 39, 565–588. <https://doi.org/10.1080/10584609.2022.2075061>

- Nanz, A., Matthes, J., 2022. Democratic consequences of incidental exposure to political information: A meta-analysis. *Journal of Communication* 72, 345–373.
- Odabaş, M., 2022. 10 facts about Americans and Twitter. Pew Research Center.
- Oser, J., Boulianne, S., 2020. Reinforcement Effects between Digital Media Use and Political Participation: A Meta-Analysis of Repeated-Wave Panel Data. *Public Opinion Quarterly*.
- Oser, J., Hooghe, M., Marien, S., 2013. Is Online Participation Distinct from Offline Participation? A Latent Class Analysis of Participation Types and Their Stratification. *Political Research Quarterly* 66, 91–101. <https://doi.org/10.1177/1065912912436695>
- Peterson, E., Goel, S., Iyengar, S., 2021. Partisan selective exposure in online news consumption: evidence from the 2016 presidential campaign. *Political Science Research and Methods* 9, 242–258. <https://doi.org/10.1017/psrm.2019.55>
- Prior, M., 2013. Media and Political Polarization. *Annu. Rev. Polit. Sci.* 16, 101–127. <https://doi.org/10.1146/annurev-polisci-100711-135242>
- Prior, M., 2009. The Immensely Inflated News Audience: Assessing Bias in Self-Reported News Exposure. *Public Opinion Quarterly* 73, 130–143. <https://doi.org/10.1093/poq/nfp002>
- Robertson, R.E., Green, J., Ruck, D.J., Ognyanova, K., Wilson, C., Lazer, D., 2023. Users choose to engage with more partisan news than they are exposed to on Google Search. *Nature*. <https://doi.org/10.1038/s41586-023-06078-5>
- Schlozman, K.L., Brady, H.E., Verba, S., 2018. Unequal and Unrepresented, in: *Unequal and Unrepresented*. Princeton University Press.
- Sears, D.O., Freedman, J.L., 1967. Selective exposure to information: A critical review. *Public Opinion Quarterly* 31, 194–213.
- Shearer, E., 2018. Social media outpaces print newspapers in the US as a news source. Pew

research center 10.

Shmargad, Y., Klar, S., 2020. Sorting the News: How Ranking by Popularity Polarizes Our Politics. *Political Communication* 37, 423–446.

<https://doi.org/10.1080/10584609.2020.1713267>

Shugars, S., Gitomer, A., McCabe, S., Gallagher, R.J., Joseph, K., Grinberg, N., Doroshenko, L., Foucault Welles, B., Lazer, D., 2021. Pandemics, Protests, and Publics: Demographic

Activity and Engagement on Twitter in 2020. *Journal of Quantitative Description: Digital Media* 1. <https://doi.org/10.51685/jqd.2021.002>

Suuronen, A., Reinikainen, H., Borchers, N.S., Strandberg, K., 2021. When Social Media Influencers go Political: An Exploratory Analysis on the Emergence of Political Topics Among Finnish Influencers. *Javnost - The Public* 1–17.

<https://doi.org/10.1080/13183222.2021.1983367>

Taylor, S.J., Muchnik, L., Kumar, M., Aral, S., 2022. Identity effects in social media. *Nature Human Behaviour*. <https://doi.org/10.1038/s41562-022-01459-8>

Thorson, K., Wells, C., 2016. Curated flows: A framework for mapping media exposure in the digital age. *Communication Theory* 26, 309–328.

Vaccari, C., Valeriani, A., 2021. *Outside the bubble: Social media and political participation in western democracies*. Oxford University Press.

Vaccari, C., Valeriani, A., Barberá, P., Bonneau, R., Jost, J.T., Nagler, J., Tucker, J.A., 2015. Political expression and action on social media: Exploring the relationship between lower-and higher-threshold political activities among Twitter users in Italy. *Journal of Computer-Mediated Communication* 20, 221–239.

Valeriani, A., Vaccari, C., 2016. Accidental exposure to politics on social media as online

- participation equalizer in Germany, Italy, and the United Kingdom. *New media & society* 18, 1857–1874.
- Verba, S., Schlozman, K.L., Brady, H.E., 1995. *Voice and equality: Civic voluntarism in American politics*. Harvard University Press.
- Weeks, B.E., Lane, D.S., Kim, D.H., Lee, S.S., Kwak, N., 2017. Incidental exposure, selective exposure, and political information sharing: Integrating online exposure patterns and expression on social media. *Journal of computer-mediated communication* 22, 363–379.
- Wells, C., Thorson, K., 2017. Combining Big Data and Survey Techniques to Model Effects of Political Content Flows in Facebook. *Social Science Computer Review* 35, 33–52.
<https://doi.org/10.1177/0894439315609528>
- Williams, M.L., Burnap, P., Sloan, L., Jessop, C., Lepps, H., 2017. Users' views of ethics in social media research: Informed consent, anonymity, and harm, in: *The Ethics of Online Research*. Emerald Publishing Limited.
- Wojcieszak, M., Casas, A., Yu, X., Nagler, J., Tucker, J.A., 2022a. Most users do not follow political elites on Twitter; those who do show overwhelming preferences for ideological congruity. *Science Advances* 8, eabn9418. <https://doi.org/10.1126/sciadv.abn9418>
- Wojcieszak, M., Menchen-Trevino, E., Goncalves, J.F., Weeks, B., 2022b. Avenues to news and diverse news exposure online: Comparing direct navigation, social media, news aggregators, search queries, and article hyperlinks. *The International Journal of Press/Politics* 27, 860–886.
- Wojcik, S., Hughes, A., 2019. *Sizing Up Twitter Users*. Pew Research Center.
- Wrubel, L., Kerchner, D., 2020. 116th U.S. Congress Tweet Ids.
<https://doi.org/10.7910/DVN/MBOJNS>

Zhuravskaya, E., Petrova, M., Enikolopov, R., 2020. Political Effects of the Internet and Social Media. *Annual Review of Economics* 12, 415–438.

<https://doi.org/10.1146/annurev-economics-081919-050239>

Zúñiga, H.G. de, Diehl, T., 2019. News finds me perception and democracy: Effects on political knowledge, political interest, and voting. *New Media & Society* 21, 1253–1271.

<https://doi.org/10.1177/1461444818817548>