The Agenda-Setting Function of Social Media

Anonymous Author(s)

Abstract

As people increasingly use social media as a primary news source, it becomes critical to understand how online platforms affect peoples' experience of the news. Through the media effects of agenda-setting and framing, different news sources can vary in their influence on public opinion regarding which issues people consider important and how particular aspects of these issues should be interpreted. However, little is known about how issues and frames shift and segregate across partisan lines as traditional news on social media gets filtered by the selective exposure effects of social media. In this study, we investigate the issues and frames invoked in news article shares across Reddit over 16 years and measure their traditional media and social media partisanship. We measure the change between *production* (news articles posted on Reddit) and *consumption* (news articles posted on Reddit, weighted by their score). We find that issues are shared in a co-partisan manner across traditional media and social media lines. Issues are also more polarized in social media than traditional media and more polarized in consumption than production. We find that frames across several issues are also subject to co-partisan sharing behavior. In contrast to the significant polarization of news outlets on Reddit in 2016, issues and frames do not polarize more over time. Finally, looking at case studies of frames within specific issues, we disaggregate the shift from production to consumption by distinguishing between issues where the frames polarize and issues that simply receive less exposure on one side of the political spectrum. Our results give insight into broader phenomena like political polarization by highlighting the dimensions of precisely what polarizes and how polarization occurs. Overall, our study showcases the importance of understanding how social media distorts the perception of the news via its agenda-setting and framing functions.

CCS Concepts

• **Applied computing → Law, social and behavioral sciences**.

Keywords

agenda-setting, framing, news on social media, political polarization

1 Introduction

In 1972, McCombs and Shaw published their seminal article "The Agenda-Setting Function of Mass Media", in which they demonstrated that mass media outlets such as newspapers, by virtue of their collective choices of which news to cover, play a pivotal function in setting the national political agenda [27]. They observed that "the information in the mass media becomes the only contact many have with politics." In our day, a similar sentiment could be expressed for social media.

Social media has been a rapidly growing part of people's information ecosystems. Indeed, in 2024, approximately 72% of Americans used social media at least to some extent as a news source [45]. While social media platforms can encourage exposure to a wider diversity of news sources [15, 31], there is also widespread concern surrounding the potential for social media to cause and exacerbate societal issues like political polarization [10, 16] and echo chambers [3, 14]. Consequently, there is an ever-increasing body of work studying how self-selection and algorithmic curation on online platforms impacts the content that people are exposed to [2, 17, 35, 41].

Media sources employ both agenda-setting, placing emphasis on certain issues over others through increased coverage or other means, [27, 44], and framing, "selecting some aspects of a perceived reality and making them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described" [11]. These media effects influence public opinion by prescribing *what* issues people should think about and *how* they should think about them [26, 42]. Shifts in public opinion can have downstream impacts on policy.

Past work surrounding potential exposure, actual exposure, and user engagement on news articles on a social media platform has primarily analyzed news articles at the domain-level [2, 17] or the news story-level using article URLs [17]. Meanwhile, previous studies on agenda-setting and framing have been limited in scope to a single issue [10, 13, 29], subset of issues [29, 42], or a single news source [24]. Thus, there is a gap in the literature: little work explores media effects in news shared on social media at a comprehensive scale. How does the online social media environment influence user perception of news via agenda-setting and framing effects?

In our work, we aim to fill this gap by studying the different issues (corresponding to agenda-setting) and frames (corresponding to framing) present in news articles shared on Reddit. We choose Reddit since it is a major social media platform, ranking as the 6th most visited website globally [28]. We study *production*—the news articles that are shared on Reddit—as well as *consumption*—the news articles that are shared on Reddit, weighted by their score (the difference between the number of upvotes and downvotes). A higher score means that more users have signalled support for the article, and the article appears higher on a Reddit user's feed [28]. We analyze the partisan nature of agenda-setting and framing by looking at traditional media partisanship (the partisan bias of the news outlet an article was published in) as measured by Allsides

ratings[1] and social media partisanship (the partisan bias of the community that the news article was posted in) as measured through community embeddings [48]. We address the following research questions:

- **RQ1.** How do issues shift from production to consumption along traditional and social partisanship lines?
- **RQ2.** How do frames across issues shift from production to consumption on traditional and social partisanship lines?
- **RQ3.** How do the shifts in issues and frames change over time?
- **RQ4.** How do frames within specific issues change from production to consumption on traditional and social partisanship lines?

The Present Work. We analyze 3.8M hard news articles (news discussing issues like economics, international affairs, and politics as opposed to arts, entertainment, and sports) shared over a 16-year time period as a submission—that is, a top-level post—on Reddit. For each article, we extract the main issue discussed and the primary frame invoked.

With regards to agenda-setting, we find that issues are shared in a co-partisan manner across traditional media and social media lines. When comparing production to consumption, we find that first, issues are more polarized along social media lines than traditional media lines, and second, issues polarize from production to consumption.

With regards to framing patterns across several issues, we find that frames are also subject to co-partisan sharing behaviour. Considering the shift from production to consumption, we find that while frames polarize along traditional media lines, they de-polarize along social media lines.

Over time, the level of polarization in issues along traditional and social media lines and in production and consumption does not change. For frames across issues, the level of polarization changes even less than issues. These observations are in stark contrast to the level of social media polarization of news sources, which experienced a large polarization event in 2016. Put together, these results indicate that while the news sources that left- and right-wing communities consume polarizes, the actual issues discussed and the frames invoked in the *text* of the news articles stays more similar across partisan lines.

Finally, we disaggregate the overall shift in issues from production to consumption by looking at frames within specific issues. We distinguish between issues where the frames polarize from production to consumption from issues that simply receive less engagement on one side of the political spectrum.

Overall, our results give insight into how user selection behaviours on social media platforms can affect news articles' shift from production to consumption. Our study adds a media effects dimension to the growing body of work on selective exposure, political polarization, and echo chambers. More broadly, our work illustrates the importance of studying the *textual* content of news on social media, and how these analyses can complement existing knowledge on the news shared and read on social media.

---

[1]https://www.allsides.com/media-bias

2 Related Work

Our work builds on a growing body of research studying discussions about political issues and news sharing on social media. This section highlights previous research on agenda-setting, framing, and news sharing on social media.

Agenda-Setting. Past work has looked at agenda-setting in issues like the Persian Gulf Crisis [21], coverage of disasters [23], and coverage of the US in Russian media [13]. The advent of social media has ushered work studying how social media affects agenda-setting. Russell Neuman et al. [42] compared the agenda-setting functions of social media content and traditional media content and found that social media places greater emphasis on social issues and less on economic policy. Similar to our work, Feezell studied the effect of social media in mainstream media dissemination [12]. Using an experimental design, they found that news exposure through Facebook increased people's perceived importance of certain policy issues [12].

Related to our work looking at the partisan leanings of different issues, issue ownership theory states that different political parties are perceived to be better at handling different issues [37]. Puglisi [39] found that the New York Times, which leans left-wing, gives more coverage to issues that Democrats "own" during presidential campaign periods. Larcinese et al. [25] also find partisan differences in coverage of economic issues. Across different news outlets, topics like healthcare tend to receive more left-wing coverage, while topics like troops and veterans receive more right-wing coverage [34]. Agenda-setting is also closely related to selection and coverage bias: which issues are covered by the news and how much space is given to each issue [43]. Previous research has found that coverage bias in news articles shared on Twitter depends more on geography than political leaning [43].

Framing. Computational approaches to frame detection have been used to study many different forms of media, including congressional speeches [6], news articles [13], and tweets [10, 29, 30, 36]. Much of past work has analyzed the use of framing within a specific political issue such as gun control and shootings [1, 10], immigration [6, 29], policing [50], and the Russia-Ukraine war [36]. These studies have found that framing differs along partisan lines [10, 29] and that the use of frames has polarized over time [6]. Demszky et al. [10] find that in gun control discussions, *how* topics are discussed (that is, framing) polarizes more than *what* topics are discussed. Many of these studies on framing [13, 19, 22, 24, 29] use the Policy Frames Codebook [4], which contains 15 issue-general frames that can be applied to any policy issue.

Some work has also grouped frames in various issues together to explore framing more broadly as a communicative device. Kwak et al. [24] perform a systemic analysis of frames in the New York Times over 17 years. Mendelsohn et al. [30] analyze how frames in immigration, gun control, and LGBTQ rights-related tweets vary depending on the role of the author (for example, whether they are a journalist) and the type of tweet (for example, original tweets versus retweets).

News Sharing on Social Media. People prefer news sources that align with their pre-existing views [20], and social media platforms can exacerbate these partisan differences in news consumption [9].

Accordingly, there has been a growing body of literature on the effects of news sharing and consumption on online platforms. On Reddit, one study found that news articles are shared in co-partisan contexts, but right-leaning news is shared disproportionately more in right-wing communities [32]. On Facebook, ideological segregation increases when considering news articles' potential exposure, actual exposure, and engagement [17]. This segregation was observed at the news source and news story levels. Moreover, like-minded sources on Facebook constitute most of what users are exposed to [35]. Despite this evidence that social media encourages out-sized exposure to co-partisan content, other work [2, 41] has found that exposure to partisan news is driven more by user choice rather than algorithmic curation. Additionally, different social media platforms vary in their influence on the spread of information. For example, news consumption is less segregated on Reddit than Facebook [8], and Reddit users' curation and amplification behaviours can reduce exposure to highly biased content [49].

3 Data and Methodology

We begin with a dataset of all Reddit submissions from June 2005 until June 2021. To identify news sharing from these submissions, we retrieve a list of news sources from Allsides, an organization that promotes balanced news consumption by estimating the perceived political bias of news outlets[2]. Similar to prior work, we match the Allsides news sources' web domains to Reddit submissions that link directly to external websites [32].

We perform the following steps to further filter our dataset of news article shares. First, we discard submissions with fewer than two comments or with vote scores lower than two. This selects for higher-quality content and ignores users who comment on and upvote their own posts. Second, to leverage Reddit community embeddings (described in further detail in 3.1), we filter out news article shares that do not occur in the largest 10,006 subreddits. Third, we remove articles with an empty title or body. Fourth, we look at the top 25 most commonly occurring titles. We find many titles, such as "Are you a robot?," that appear to be paywalled. For these titles, we check the body to see whether the article is indeed paywalled; if it is, we remove the article from the dataset.

Hard News Filtering. For this study, we follow previous work by restricting our attention to "hard news" (such as economics, international affairs, and politics) rather than "soft news" (such as arts, entertainment, and sports). To filter out soft news articles, we largely follow the methodology from Bakshy et al. [2]; we identify sets of hard and soft news articles using URL regular expressions and train a linear Support Vector Machine (SVM) on the title and body of these articles. The 10-fold accuracy, precision, and recall are 93.5%, 95.3% , and 91.6%, respectively. Filtering out soft news articles by applying the SVM on the entire dataset, we obtain our final dataset of 4.57M news articles. More details on the hard news filtering can be found in Section A.1.

3.1 Traditional and Social Media Partisanship

Traditional Media Partisanship. Each news source on Allsides (and our dataset) is rated with one of five political leanings: −2, −1,

---

0, 1, 2, corresponding to left-wing, center-left, center, center-right, and right-wing, respectively. Following previous work, we use these ratings as a measure of traditional media partisan skew [32, 40, 46].

Social Media Partisanship. To gain a behavioural understanding of where the news article shares occur on Reddit, we use community embeddings, an established technique used to study online social media through a behavioural lens [47]. Community embeddings have three advantages for our purposes: first, they are behavioral, being calculated purely from the collective actions of millions of users, instead of relying on text, annotators, or survey that may be used; second, they are granular, allowing us to classify not just *whether* a subreddit is left- or right-wing but *how* left- or right-wing it is; and third, they allow us to determine the partisan leaning of every subreddit, not just explicitly political ones.

The community embeddings are created using word2vec, taking subreddits as "words" and commenters as "contexts." Thus, two subreddits are closer together in the community embeddings if they share more users. The top 10,006 subreddits (which account for 95.4% of the comments on Reddit) are embedded into a 150-dimensional space. Hyperparameters are then tuned with community analogies, such as sports teams to cities and universities to cities. We use the social partisanship scores of every subreddit from previous work (more details can be found in Appendix A.2) [48].

3.2 Issue Identification

To study agenda-setting on Reddit, we follow previous computational work [10, 13] and use topic modelling to identify issues. Specifically, we apply a method similar to BERTopic [18], and that of Movva et al. [33]: using semantic text embeddings, performing dimensionality reduction on the embeddings, then clustering them. For the clustering algorithm, we perform soft HDBSCAN clustering, which assigns to each document a vector of probabilities, where each component represents the probability that the document belongs to a cluster. We tune the hyperparameters through manual inspection of the articles in the clusters so that each cluster contains articles about one issue rather than multiple clusters on the same issue (which would result in smaller but more clusters) or one cluster containing multiple issues (which would result in larger but fewer clusters). We perform the clustering on 25% of the dataset and use these clusters to assign the remaining 75% of articles to a cluster. We used this method because HDBSCAN did not scale to the entirety of our dataset, and alternative clustering methods that did (such as k-means) produced less interpretable clusters. HDBSCAN is a conservative clustering algorithm that does not necessarily assign all points to a cluster. In our topic model, 16% of documents are classified as outliers. Our topic model contains 35 topics; we remove one of topics, Sports, from our analyses because we are only interested in hard news articles.

We manually validate the quality of our topic model by first examining a random sample of 25 documents per cluster and counting the number of documents that belong to each cluster. On average, 87.7% of the documents belong to their corresponding clusters. Additionally, we take a sample of 25 documents from the entire dataset that are not classified as outliers. We manually assigned each of these documents to the cluster that we felt the article should belong

---

to and compared them with the assignments our topic model produced. The topic model's top-1 accuracy was 64.0%. More details on the issue identification and manual validation can be found in A.3. Table 1 in Appendix B.1 also contains an example of an article headline for each topic.

3.3 Frame Detection

Following previous computational research on framing [13, 19, 22, 24, 29], we use the 15 frames outlined in the Policy Frames Handbook [4] to label each news article. We label each article with its primary frame. We use a subset of Media Frames Corpus (MFC) [5] containing 5,347 articles to train, validate, and test a classification model. We used a subset of the MFC rather than the entire set because some articles are no longer accessible in Nexis Uni, the database from which the MFC pulls article bodies; for example, the Washington Post is no longer available. We also restrict the MFC to articles that are labelled with a primary frame that is agreed upon by both annotators. Using this subset of the MFC, we finetune a `roberta-large` model, and achieve an accuracy of 72.3% and a macro-F1 of 64.2% on a test set of 535 articles. To our knowledge, the highest accuracy achieved in primary frame classification using the MFC was 71.3% by Kwak et al. [24]; our accuracy surpasses this number. We discard articles labelled with Other frames, and consider the remaining 14 frames. Table 2 in Appendix B.1 contains an example of an article headline for each frame. After removing outlier articles from our topic model, articles belonging to the Sports topic, and articles labelled with Other frames, we are left with a final dataset of 3.82M news articles for our analyses.

4 Results

4.1 Agenda-Setting

We first analyze the issues present in news articles shared on Reddit. Then, we explore how the consumption of issues differs from their production in traditional media and social partisanship leaning, volume, and time.
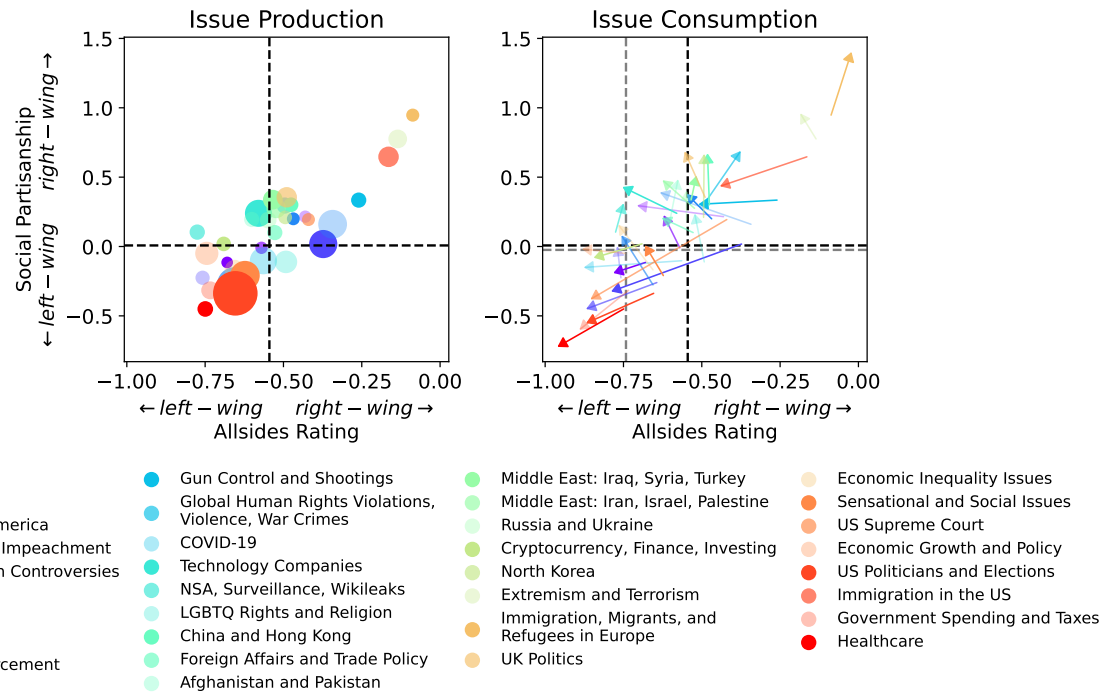
Agenda Production. We begin by examining the issues that are discussed in news articles posted on Reddit, which we refer to as *production*. Each news article shared in our dataset is labelled with the Allsides rating of the news outlet publishing the article, the social partisanship score of the subreddit it was posted in, and the issue it discusses. For each issue, we look at all the articles on that issue and take the average Allsides rating, which represents partisanship on traditional media, and its average social partisanship score, which represents partisanship on social media. In Figure 1, the plot to the left shows a positive correlation between the average Allsides rating of each issue and the social partisanship score. Issues that are left-wing on both Allsides rating and social partisanship score tend to center around US politics, the economy, and healthcare. Issues that are right-wing on both Allsides rating and social partisanship tend to center around immigration, extremism, and terrorism. Our findings on the political leanings of issues roughly correlate with previous studies on the political leanings of issues in news articles containing at least one quote from a presidential speech [34].

Agenda Consumption: Volume. Not all the news articles shared on Reddit are seen by many users. We approximate exposure to news articles by using the score of the news article, which is the difference between the number of upvotes and the number of downvotes a post receives. While this is a proxy for actual exposure, we use the score since the score is used as a component in ranking the submissions on a user's Reddit feed [7, 28]: submissions with a higher score appear higher on a user's feed. Furthermore, other studies [49] using Reddit data have used imperfect proxies to measure exposure to content since submission view counts are not publicly available. We weigh each post by its score and call this the *consumption* of a news article.

We begin our analyses on consumption by comparing issues that are disproportionately consumed more than they are produced and the ones that are disproportionately produced more than they are consumed. We measure disproportionate consumption and production for each issue by taking the fraction of total consumption that belongs to the issue, and dividing it by the fraction of total production that belongs to the issue. Thus, a ratio greater than 1 means that an issue makes up a larger part of total consumption than production, meaning that it is consumed more than it is produced. In Figure 2, we find that overall, issues that pertain primarily to US politics are disproportionately consumed, while global issues are disproportionately produced. This result indicates that the Reddit ecosystem brings US news to the forefront while suppressing world news. Reddit users are thus likely to see a disproportionate amount of US news relative to what exists on the platform.

Agenda Consumption: Traditional and Social Media Partisanship. How do issues move from production to consumption? This shift enables us to understand how behaviour on Reddit skews issues in news articles, both along traditional partisan lines and social media partisan lines. For each issue, we measure the traditional and social media partisanship of the consumption of each issue by taking the weighted average Allsides rating and social partisanship score, weighted by the score (the difference between the number of upvotes and downvotes) of each article share. From the plot to the right in Figure 1, we find that most issues shift leftward on both Allsides rating and social partisanship or rightward on social partisanship. Issues in the former group include US Politicians and Elections, Trump Administration Controversies, and Healthcare, which are all centred around US politics. Issues in the latter group include China and Hong Kong, Middle East: Iraq, Syria, Turkey, and North Korea, which are all centred around global issues.

Following previous work [32], we measure the polarization of a given set of issues by taking the interquartile range (IQR) of the issues' Allsides ratings (for traditional media polarization) or social partisanship scores (for social media polarization). To compare polarization in Allsides ratings to polarization in social partisanship scores, we convert the Allsides ratings into z-scores taken across the entire dataset. In news article production, we find that the IQR of the issues' Allsides ratings is 0.162, and the IQR of the social partisanship scores is 0.380. When considering consumption, these values become 0.227 and 0.533, respectively. This result has two key implications: first, issues are more polarized along social partisanship lines than in traditional media, and second, issues in news articles on Reddit are more polarized in consumption

Figure 1: Distribution of issues by traditional and social media. The left plot shows where the issues are produced across this distribution, and the right plot shows how the issues have shifted from production to consumption. The size of the dots is proportional to the number of news articles belonging to the issue. The black dashed line represents the mean Allsides rating and social partisanship score for production; the grey dashed line represents these values for consumption.

than production. These results are consistent with previous work on domain-level and news story-level shifts from production to consumption on social media [17].

Agenda Consumption: Temporal Shifts. Previously, we found that issue polarization increases from traditional to social media and from production to consumption. How does issue polarization change over time in news article production and consumption across traditional and social media? In Figure 3, we find that issues are almost always more polarized along social partisanship lines than traditional media and in consumption than production. The level of polarization across the four measures remains relatively static over time. In contrast, news source production and consumption experience a large polarization event in 2016 along social partisanship lines, a phenomenon also observed by Mok et al. [32]. These observations indicate that even though left- and right-wing communities have become increasingly segregated in the news article sources that are produced and consumed, the news articles are not becoming as segregated in the issues that they discuss.
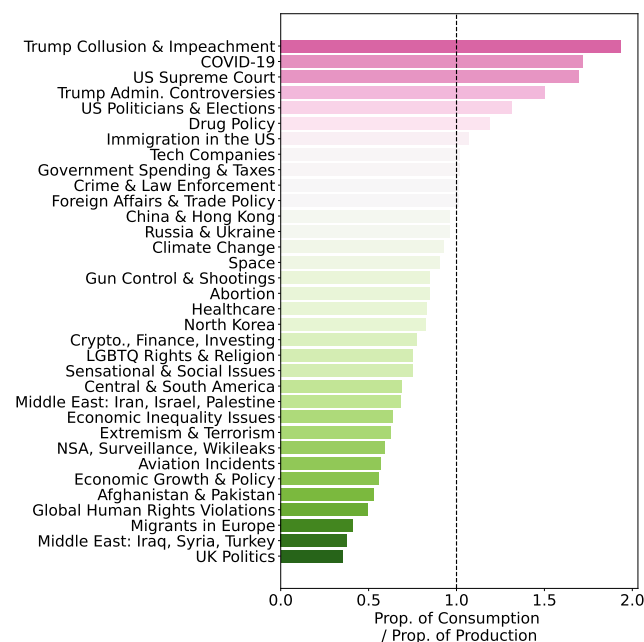
4.2 Framing Across Issues

In the previous section, we explored the production and consumption of the various issues present in news articles on Reddit. Now, we turn to framing, which occurs when specific aspects of an issue are highlighted to promote particular interpretations of the issue at
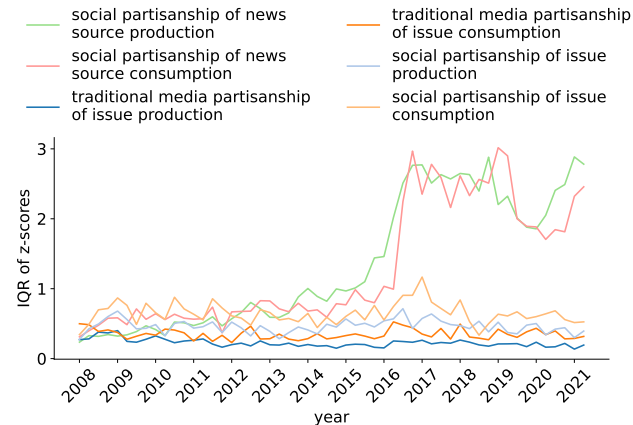
hand [11]. In other words, while 4.1 is about *what* is being discussed, framing is more about *how* the issues in 4.1 are discussed. Issue frames experience regularities across different policy debates [4], which enables us to examine the same frame across all the different issues explored in the previous section. Similar to the issues, we begin by examining the production of frames. Then, we explore how the consumption of frames differs from their production in traditional media and social partisanship leaning, volume, and temporally.

Frame Production. How are frames divided across different partisanship contexts? Figure 4 highlights the frames that are produced on Reddit along Allsides ratings and average social partisanship scores. Security and defense frames, Capacity and resources frames, and Law and order, crime and justice frames are more frequently used by right-wing traditional media and shared in right-wing social media communities, whereas Quality of life frames and Economic frames are more frequently used by left-wing traditional and social media contexts. Our analysis shows partisanship trends exist in the production of frames across a wide variety of issues, adding to the body of literature surrounding the relationship between partisanship leaning and framing in specific policy debates.

Frame Consumption: Volume. Which frames are disproportionately consumed and which are disproportionately produced? We find that the three most disproportionately consumed frames are Public opinion frames, Political frames, and Constitutionality and
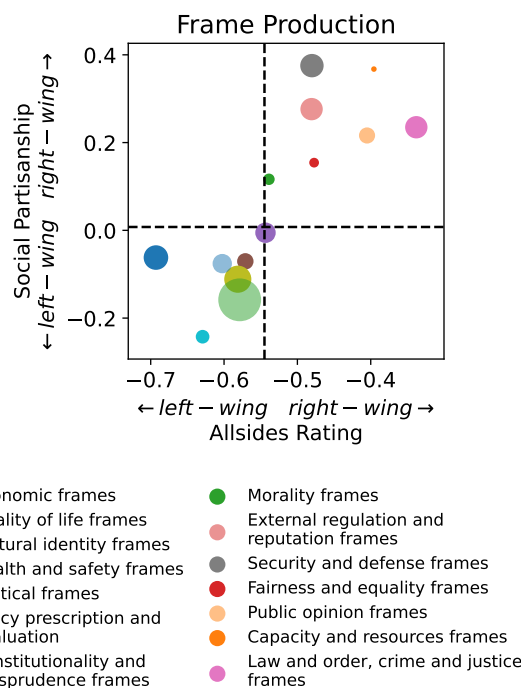
Figure 2: Ratio between the consumption and the production for each issue. The consumption and production of each issue are measured as a proportion of the total consumption and production, respectively.

Figure 3: Issue and news outlet polarization from 2008 to mid-2021, by three-month periods.
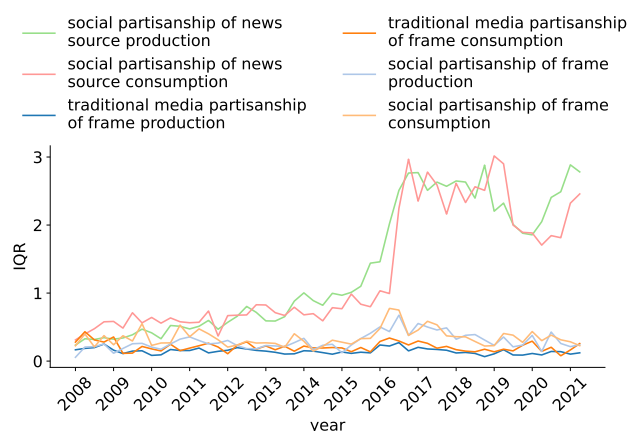
jurisprudence frames. The three most disproportionately produced frames are Capacity and resources frames, External regulation and reputation frames, and Cultural identity frames. Overall, frames like Public opinion frames and Political frames, which are more relevant to societal attitudes, seem to receive a disproportionate amount of exposure. In contrast, frames like Capacity and resources frames and External regulation and reputation frames, which are more focused on the structural factors of a nation than the general public, receive less exposure on Reddit.

Figure 4: Distribution of frame production by traditional and social media. The black dashed line represents the mean Allsides rating and social partisanship score for production.

Frame Consumption: Traditional and Social Media Partisanship. How do frames move from production to consumption? To capture the movement from production to consumption at an aggregate level, we again measure the IQR of the frames in traditional and social media and in production and consumption. We find that the IQR of the issues' Allsides ratings is 0.090, and the IQR of the social partisanship scores is 0.305. When considering consumption, these values become 0.118 and 0.280, respectively. Like issues, frames are more polarized along social partisanship lines than in traditional media. Unlike issues, frames in news articles on Reddit are *less* polarized in consumption than production along social media lines. These results indicate that while frames invoked by left-wing sources become more left-wing and frames invoked by right-wing sources become more right-wing, the frames are shared on aggregate in a more cross-cutting manner across Reddit communities.

Frame Consumption: Temporal Shifts. How does frame polarization change over time in news article production and consumption across traditional and social media? Similar to issue polarization, for a given three-month time period in Figure 5, we measure the polarization of a given set of news sources by taking the IQR of the news sources' mean social partisanship scores during that time period. We find that, like issues, frames are almost always more polarized along social partisanship lines than traditional media and in consumption than production. The level of polarization increases less than the topics (Figure 3). These observations indicate that even

Figure 5: Frame and news outlet polarization from 2008 to mid-2021, by three-month periods.

though left- and right-wing communities have become increasingly segregated in the news article sources that are produced and consumed, the news articles are not becoming as segregated in the frames that they invoke.

4.3 Case Studies Involving Issues and Frames

In the previous sections, we looked at the production and consumption of issues and frames on an aggregate level. While these analyses allowed us to see the broad picture of where issues and frames are on traditional and social media partisanship lines and how they shift from production to consumption, it does not allow us to see the details regarding frames within issues. In this section, we choose two issues, Immigration, Migrants, and Refugees in Europe and LGBTQ Rights and Religion, as case studies and conduct a fine-grained analysis by looking at frames *within* each of these issues.

Immigration, Migrants, and Refugees in Europe. In 4.1, we saw that Immigration, Migrants, and Refugees in Europe as a whole shifts right along both traditional and social media lines. How do the individual frames within this issue shift? We inspect the frames of the articles surrounding this issue that have been shared over 100 times, and we find polarization of the frames along traditional media lines (Figure 6). Security and defense frames, Quality of life frames, and Cultural identity frames move left. By qualitatively examining the top consumed articles for these topics, these articles discuss shootings and other violent events and their connection with immigrants (Security and defense frames), immigrants adjusting to European society (Quality of life frames), and demographics of immigrants, including age, gender, and race (Cultural identity frames).

In contrast, Economic frames, Policy prescription and evaluation frames, and External regulation and reputation frames move right along both traditional and social media lines. Examining the top consumed articles that invoke these frames, they concern immigration's effect on jobs and government spending (Economic frames), discussions on integration v.s. repatriation of migrants

(Policy prescription and evaluation), and comparisons between immigration policies in Europe and the US, especially with regards to border security (External regulation and reputation). These insights tell us that while news on Immigration, Migrants, and Refugees in Europe as a whole is consumed more in right-wing contexts, articles from left-wing outlets that frame immigration to highlight the immigrants' quality of life and culture appeal more to Reddit users.

LGBTQ Rights and Religion. LGBTQ Rights and Religion as a whole shifts slightly along traditional media lines and right along social media lines. Inspecting the frames within LGBTQ Rights and Religion that have appeared in 100 or more news articles (Figure 6), we see that all frames surrounding LGBTQ Rights and Religion move rightward along social partisanship lines. This observation indicates that on Reddit, most aspects surrounding the issue of LGBTQ Rights and Religion receive more attention from right-wing contexts.

These observations can also be compared to the previous case study surrounding Immigration, Migrants, and Refugees in Europe, where the frames polarized along both traditional media lines. Overall, these two case studies show that issues can shift from production to consumption in many ways: for instance, both sides may differ in the aspects of the issue they consume (in the case of Immigration, Migrants, and Refugees in Europe) or one side of the political spectrum may consume all aspects of the issue more (in the case of LGBTQ Rights and Religion).
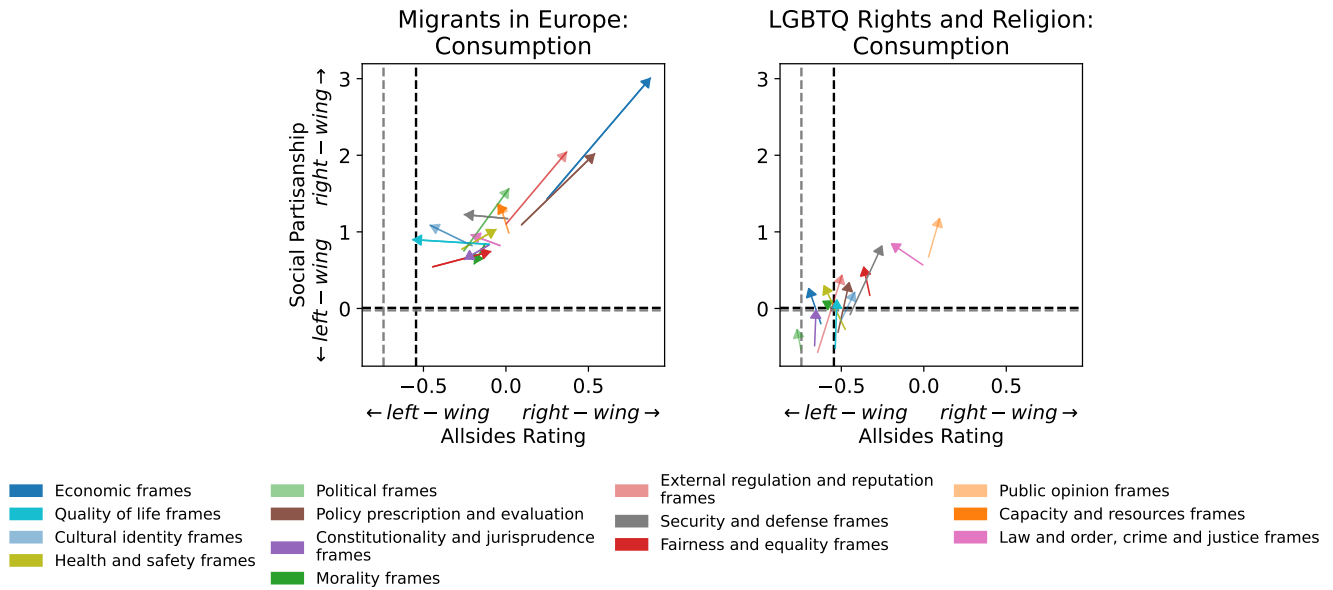
5 Discussion

In this paper, we expand upon the growing body of work on social media's role in shaping public discourse by analyzing the agenda-setting and framing effects of the Reddit ecosystem. Our findings demonstrate that issues and frames shared on social media platforms are subject to co-partisan sharing patterns, but the dynamics of polarization are more complex than previously understood.

A key contribution of our work lies in bridging a significant gap in the literature. Prior studies have either compared production and consumption of news at the domain-level, which ignores topics and therefore agenda-setting and framing effects, or they examine agenda-setting and framing of traditional media as opposed to social media. Our work studies the agenda-setting and framing functions of *social* media, complementing theories of these media effects in *mass* media [11, 27] in a world where social media is playing an ever-increasing role in people's daily lives. Additionally, our analysis of all 3.8 million hard news articles shared over 16 years of Reddit history provides an unprecedented comprehensive scale for understanding these shifts across partisan lines.

In terms of agenda-setting, we find that issues are shared in a co-partisan manner when comparing traditional and social media partisanship. Furthermore, issues are more polarized in social media than in traditional media, with polarization increasing from production to consumption. This result underscores the significant role that social media platforms play in amplifying certain issues over others, reinforcing existing partisan divides. However, frames also exhibit similar co-partisan sharing behaviours, but their polarization on social media decreases from production to consumption.

Figure 6: Distribution of frame production and consumption by traditional and social media for Immigration, Migrants, and Refugees in Europe (left) and LGBTQ Rights and Religion (right). The black dashed line represents the mean Allsides rating and social partisanship score for production; the grey dashed line represents these values for consumption.

For both issues and frames, the described polarization patterns change little over time. This forms a stark contrast with news outlets, which experienced a significant polarization event along social media partisanship lines in 2016. This result suggests that although Reddit communities polarize in the sources that they consume, the actual issues discussed and frames invoked in the news articles consumed on Reddit remain similar between the left- and right-wing. Finally, through case studies of frames within two different issues, we demonstrate that an issue's shift from production to consumption could either result in its frames polarizing or moving in the same direction along partisan lines. Our case studies highlight the importance of understanding the fine-grained nature of how issues shift from potential to actual exposure.

This work opens future avenues for studying the textual content of news articles and its relation to production and consumption on social media platforms. For example, future work could explore why news sources polarize in 2016, but issues and frames do not. One possibility is that linguistic properties of news articles—for example, stance or tone—differ between different news sources. Another possibility is that other aspects of the news article, such as the photos or videos it includes, vary between news sources. Alternatively, users might simply have biases regarding different news sources. For example, regardless of the content of a news article, a user may choose not to post it on a specific subreddit or upvote it solely based on the article's source.

Limitations. There are a couple of limitations with our analyses. First, using a Reddit submission's score as a measure proportional to its actual exposure to Reddit users is an imperfect proxy. The actual algorithm also depends on how much time has passed since the submission was posted [28] and is closed-source [7]. Second,

our analyses on framing focus exclusively on the issue-general frames outlined in the Policy Frames Handbook [4]. Many other types of issue-general frames, such as diagnostic, prognostic, and motivational frames [30], have not been explored in this study. Furthermore, incorporating issue-specific frames would allow for richer analyses in our case studies.

References

[1] Kevin Aslett, Nora Webb Williams, Andreu Casas, Wesley Zuidema, and John Wilkerson. 2022. What was the problem in parkland? Using social media to measure the effectiveness of issue frames. *Policy studies journal* 50, 1 (2022), 266–289.

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

[3] Andrei Boutyline and Robb Willer. 2017. The social structure of political echo chambers: Variation in ideological homophily in online networks. *Political psychology* 38, 3 (2017), 551–569.

[4] Amber E Boydstun, Dallas Card, Justin H Gross, Philip Resnik, and Noah A Smith. 2014. Tracking the development of media frames within and across policy issues. In *APSA 2014 annual meeting paper*.

[5] Dallas Card, Amber Boydstun, Justin H Gross, Philip Resnik, and Noah A Smith. 2015. The media frames corpus: Annotations of frames across issues. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. 438–444.

[6] Dallas Card, Serina Chang, Chris Becker, Julia Mendelsohn, Rob Voigt, Leah Boustan, Ran Abramitzky, and Dan Jurafsky. 2022. Computational analysis of 140 years of US political speeches reveals more positive but increasingly polarized framing of immigration. *Proceedings of the National Academy of Sciences* 119, 31 (2022), e2120510119.

[7] Mark Carman, Mark Koerber, Jiuyong Li, Kim-Kwang Raymond Choo, and Helen Ashman. 2018. Manipulating visibility of political and apolitical threads on reddit via score boosting. In *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*. IEEE, 184–190.

[8] Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. 2021. The echo chamber effect on social

media. *Proceedings of the National Academy of Sciences* 118, 9 (2021), e2023301118.

[9] Sylvain Dejean, Marianne Lumeau, and Stéphanie Peltier. 2022. Partisan selective exposure in news consumption. *Information Economics and Policy* 60 (2022), 100992.

[10] Dorottya Demszky, Nikhil Garg, Rob Voigt, James Zou, Matthew Gentzkow, Jesse Shapiro, and Dan Jurafsky. 2019. Analyzing polarization in social media: Method and application to tweets on 21 mass shootings. *arXiv preprint arXiv:1904.01596* (2019).

[11] Robert M Entman. 1993. Framing: Toward clarification of a fractured paradigm. *Journal of communication* 43, 4 (1993), 51–58.

[12] Jessica T Feezell. 2018. Agenda setting through social media: The importance of incidental news exposure and social filtering in the digital era. *Political research quarterly* 71, 2 (2018), 482–494.

[13] Anjalie Field, Doron Kliger, Shuly Wintner, Jennifer Pan, Dan Jurafsky, and Yulia Tsvetkov. 2018. Framing and agenda-setting in Russian news: a computational analysis of intricate political strategies. *arXiv preprint arXiv:1808.09386* (2018).

[14] Seth Flaxman, Sharad Goel, and Justin M Rao. 2016. Filter bubbles, echo chambers, and online news consumption. *Public opinion quarterly* 80, S1 (2016), 298–320.

[15] Richard Fletcher and Rasmus Kleis Nielsen. 2018. Are people incidentally exposed to news on social media? A comparative analysis. *New media & society* 20, 7 (2018), 2450–2468.

[16] Kiran Garimella, Tim Smith, Rebecca Weiss, and Robert West. 2021. Political polarization in online news consumption. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 15. 152–162.

[17] Sandra González-Bailón, David Lazer, Pablo Barberá, Meiqing Zhang, Hunt All-cott, Taylor Brown, Adriana Crespo-Tenorio, Deen Freelon, Matthew Gentzkow, Andrew M Guess, et al. 2023. Asymmetric ideological segregation in exposure to political news on Facebook. *Science* 381, 6656 (2023), 392–398.

[18] Maarten Grootendorst. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794* (2022).

[19] Mareike Hartmann, Tallulah Jansen, Isabelle Augenstein, and Anders Søgaard. 2019. Issue framing in online discussion fora. *arXiv preprint arXiv:1904.03969* (2019).

[20] Shanto Iyengar and Kyu S Hahn. 2009. Red media, blue media: Evidence of ideological selectivity in media use. *Journal of communication* 59, 1 (2009), 19–39.

[21] Shanto Iyengar and Adam Simon. 1993. News coverage of the Gulf crisis and public opinion: A study of agenda-setting, priming, and framing. *Communication research* 20, 3 (1993), 365–383.

[22] Kristen Johnson, I-Ta Lee, and Dan Goldwasser. 2017. Ideological phrase indicators for classification of political discourse framing on Twitter. In *Proceedings of the Second Workshop on NLP and Computational Social Science*. 90–99.

[23] Haewoon Kwak and Jisun An. 2014. A first look at global news coverage of disasters by using the gdelt dataset. In *Social Informatics: 6th International Conference, SocInfo 2014, Barcelona, Spain, November 11-13, 2014. Proceedings 6*. Springer, 300–308.

[24] Haewoon Kwak, Jisun An, and Yong-Yeol Ahn. 2020. A systematic media frame analysis of 1.5 million new york times articles from 2000 to 2017. In *Proceedings of the 12th ACM Conference on Web Science*. 305–314.

[25] Valentino Larcinese, Riccardo Puglisi, and James M Snyder Jr. 2011. Partisan bias in economic news: Evidence on the agenda-setting behavior of US newspapers. *Journal of public Economics* 95, 9-10 (2011), 1178–1189.

[26] Maxwell McCombs. 2002. The agenda-setting role of the mass media in the shaping of public opinion. In *Mass Media Economics 2002 Conference, London School of Economics: http://sticerd. lse. ac. uk/dps/extra/McCombs. pdf*.

[27] Maxwell E McCombs and Donald L Shaw. 1972. The agenda-setting function of mass media. *Public opinion quarterly* 36, 2 (1972), 176–187.

[28] Alexey N Medvedev, Renaud Lambiotte, and Jean-Charles Delvenne. 2019. The anatomy of Reddit: An overview of academic research. *Dynamics on and of Complex Networks III: Machine Learning and Statistical Physics Approaches 10* (2019), 183–204.

[29] Julia Mendelsohn, Ceren Budak, and David Jurgens. 2021. Modeling framing in immigration discourse on social media. *arXiv preprint arXiv:2104.06443* (2021).

[30] Julia Mendelsohn, Maya Vijan, Dallas Card, and Ceren Budak. 2024. Framing Social Movements on Social Media: Unpacking Diagnostic, Prognostic, and Motivational Strategies. *Journal of Quantitative Description: Digital Media* 4 (2024).

[31] Solomon Messing and Sean J Westwood. 2014. Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online. *Communication research* 41, 8 (2014), 1042–1063.

[32] Lillio Mok, Michael Inzlicht, and Ashton Anderson. 2023. Echo Tunnels: Polarized News Sharing Online Runs Narrow but Deep. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 17. 662–673.

[33] Rajiv Movva, Sidhika Balachandar, Kenny Peng, Gabriel Agostini, Nikhil Garg, and Emma Pierson. 2024. Topics, Authors, and Institutions in Large Language Model Research: Trends from 17K arXiv Papers. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*. 1223–1243.

[34] Vlad Niculae, Caroline Suen, Justine Zhang, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2015. Quotus: The structure of political media coverage as revealed by quoting patterns. In *Proceedings of the 24th International Conference on World Wide Web*. 798–808.

[35] Brendan Nyhan, Jaime Settle, Emily Thorson, Magdalena Wojcieszak, Pablo Barberá, Annie Y Chen, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, et al. 2023. Like-minded sources on Facebook are prevalent but not polarizing. *Nature* 620, 7972 (2023), 137–144.

[36] Chan Young Park, Julia Mendelsohn, Anjalie Field, and Yulia Tsvetkov. 2022. Challenges and opportunities in information manipulation detection: An examination of wartime Russian media. *arXiv preprint arXiv:2205.12382* (2022).

[37] John R Petrocik. 1996. Issue ownership in presidential elections, with a 1980 case study. *American journal of political science* (1996), 825–850.

[38] Horst Pöttker. 2003. News and its communicative quality: the inverted pyramid—when and why did it appear? *Journalism Studies* 4, 4 (2003), 501–511.

[39] Riccardo Puglisi. 2011. Being the New York Times: the political behaviour of a newspaper. *The BE journal of economic analysis & policy* 11, 1 (2011).

[40] Steve Rathje, Jay J Van Bavel, and Sander Van Der Linden. 2021. Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences* 118, 26 (2021), e2024292118.

[41] Ronald E Robertson, Jon Green, Damian J Ruck, Katherine Ognyanova, Christo Wilson, and David Lazer. 2023. Users choose to engage with more partisan news than they are exposed to on Google Search. *Nature* 618, 7964 (2023), 342–348.

[42] W Russell Neuman, Lauren Guggenheim, Set al Mo Jang, and Soo Young Bae. 2014. The dynamics of public attention: Agenda-setting theory meets big data. *Journal of communication* 64, 2 (2014), 193–214.

[43] Diego Saez-Trumper, Carlos Castillo, and Mounia Lalmas. 2013. Social media news communities: gatekeeping, coverage, and statement bias. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*. 1679–1684.

[44] Dietram A Scheufele and David Tewksbury. 2007. Framing, agenda setting, and priming: The evolution of three media effects models. *Journal of communication* 57, 1 (2007), 9–20.

[45] Christopher St. Aubin and Jacob Liedke. 2024. Social Media and News Fact Sheet. https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/

[46] Jennifer A Stark and Nicholas Diakopoulos. 2017. Using Baselines for Algorithm Audits. In *European Conference on Data and Computational Journalism*, Vol. 3.

[47] Isaac Waller and Ashton Anderson. 2019. Generalists and specialists: Using community embeddings to quantify activity diversity in online platforms. In *The World Wide Web Conference*. 1954–1964.

[48] Isaac Waller and Ashton Anderson. 2021. Quantifying social organization and political polarization in online platforms. *Nature* 600, 7888 (2021), 264–268.

[49] Galen Weld, Maria Glenski, and Tim Althoff. 2021. Political bias and factualness in news sharing across more than 100,000 online communities. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 15. 796–807.

[50] Caleb Ziems and Diyi Yang. 2021. To protect and to serve? analyzing entity-centric framing of police violence. *arXiv preprint arXiv:2109.05325* (2021).

A Supplementary Methodology

A.1 Hard News Filtering

For this study, we are only interested in "hard news" (such as economics, international affairs, and politics) rather than "soft news" (such as arts, entertainment, and sports). To filter out soft news articles, we largely follow the methodology from Bakshy et al. [2]. First, we count vectorize all the titles and the bodies of the articles. Second, we identify hard and soft news articles using the same URL regular expressions as Bakshy et al. [2]. We remove the 13.1K articles that match the hard news regex and the soft news regex. There are 9.56 times as many hard news articles as soft news articles. We randomly sample the hard news articles to have an equal number of hard and soft news articles. Third, we train a linear Support Vector Machine (SVM) using the balanced sample of news articles. The 10-fold accuracy, precision, and recall are 93.5%, 95.3% , and 91.6% , respectively. Fourth, we use this SVM to classify the entire dataset. This process results in our final dataset of 4.57M news articles.

A.2 Social Media Partisanship

In our analyses, we use Waller and Anderson's [48] social partisanship scores from community embeddings. The social partisanship scores are obtained as follows. Waller and Anderson begin with a pair of communities, r/democrats and r/Conservative, which differ in political partisanship but are similar in other regards. The initial partisanship dimension is created by taking the vector difference between r/democrats and r/Conservative. Additional steps, such as the augmentation of seed pairs using nine other similar pairs of communities, are performed to make the partisanship dimension more robust. Then, each subreddit's 150-dimensional representation in the community embeddings is projected onto the partisanship vector. Finally, we take the z-scores across the projected subreddits to get a social partisanship score for each subreddit, and use these scores as a measure of social media partisan bias.

A.3 Issue Identification

To identify the issues in our dataset, we apply a method similar to BERTopic [18], and that of Movva et al. [33]: using semantic text embeddings, performing dimensionality reduction on the embeddings, then clustering them. In particular, first, we use the INSTRUCTOR-LARGE model to embed the headline and the first three sentences of each news article in a 768-dimensional space. We opt to use the first three sentences rather than the entire text of the news articles because of their "inverted pyramid" structure: most of the important information related to a news article is contained in its beginning [38]. Second, we reduce the dimensionality of these embeddings by applying PCA to reduce the embeddings to 274 dimensions (which explain 90% of the variance present in the original embeddings) and then UMAP to 5 dimensions. Third, we take a random sample of 25% of the dataset and cluster them using the HDBSCAN algorithm. Using these clusters, we assign the remaining 75% to a cluster. We used this method because HDBSCAN did not scale to the entirety of our dataset, and alternative clustering methods that did (such as k-means) produced less interpretable clusters. We tune the hyperparameter min_cluster_size to achieve our desired granularity. We opt for min_cluster_size = 7000 because setting this number to 6500 produces two Climate Change clusters while setting this number to 7500 produces a broad cluster, including both articles that mention plant extinction and repatriation in Australia. This process creates 35 clusters and classifies 25.7% of documents as outliers. We assigned an initial name to each cluster by looking at a random sample of five articles per cluster.

We wanted to check if we could reduce the number of outlier documents while potentially expanding the To reduce the number of outliers, we use the soft clustering version of HDBSCAN, which assigns each document a vector of probabilities, where each component represents the probability that the document belongs to a cluster. We then experiment with various probability thresholds, classifying documents as outliers if none of the probabilities exceed the threshold. We manually examined five documents per cluster with different thresholds and chose the lowest threshold where the sample of documents in each cluster was coherent and thematically related. After this process, 15.7% of documents were classified as outliers. Our topic model contains 35 topics; we remove one of topics, Sports, from our analyses because we are only interested in hard news articles. Table 1 contains an example of an article headline for each topic.

Manual Validation. We manually validate the quality of our topic model in three ways. First, we examine a random sample of 25 documents per cluster (using a different seed from the seed used to label the documents) and count the number of documents that belong to each cluster. On average, 87.7% of the documents belong to their corresponding clusters.

Second, we took a sample of 25 documents from the entire dataset that are not classified as outliers. We manually assigned each of these documents to the cluster that we felt the article should belong to and compared them with the assignments our topic model produced. The topic model's top-1 accuracy was 64.0%.

Third, since many of our results involve analyzing the relationship between news article production and consumption, we look at the news article with the highest score for each topic and check whether it belongs to that topic. We find that 80% of the top-consumed articles belong to their corresponding cluster.

B Supplementary Results

B.1 Example Issues and Frames

Table 1: Topic names and example headlines.

Topic Name	Example Article Headline
Abortion	Alabama governor signs near-total abortion ban into law
Space	30 Years After Explosion, Challenger Engineer Still Blames Himself
Central and South America	Fidel Castro, Cuba's leader of revolution, dies at 90
Trump Collusion and Impeachment	Impeachment Results: How Democrats and Republicans Voted
Trump Administration Controversies	Churchill's grandson slams Trump for skipping cemetery visit because of weather
Drug Policy	NYPD officers can no longer search a vehicle due to the smell of marijuana alone, new memo says
Aviation Incidents	Man filmed being dragged off United flight causes outrage in China
Climate Change	Corbyn ballot challenge ruling & Boris Johnson in Paris talks
Crime and Law Enforcement	Protests about police brutality are met with wave of police brutality across US
Gun Control and Shootings	Court documents ask Parkland parents to prove mental anguish
Global Human Rights Violations, Violence, War Crimes	Sudan scraps apostasy law and alcohol ban for non-Muslims
COVID-19	President Trump, first lady Melania test positive for coronavirus
Technology Companies	F.C.C. Repeals Net Neutrality Rules
NSA, Surveillance, Wikileaks	Julian Assange: Wikileaks co-founder arrested in London
LGBTQ Rights and Religion	Supreme Court rules workers can't be fired for being gay or transgender
China and Hong Kong	The Hong Kong migrants fleeing to start new lives in the UK
Foreign Affairs and Trade Policy	Trump signs order withdrawing U.S. from Trans-Pacific trade deal
Afghanistan and Pakistan	Pakistani troops killed by Indian shelling in Kashmir, says military
Middle East: Iraq, Syria, Turkey	Distressing footage emerges of children with 'chemical burns' in Syria
Middle East: Iran, Israel, Palestine	Trump: We must all work together to make a deal with Iran
Russia and Ukraine	Doctor in charge of treating Alexei Navalny in Russia after Novichok poisoning dies suddenly
Cryptocurrency, Finance, Investing	Chinese tech billionaire Jack Ma has not been seen in public for TWO MONTHS
North Korea	Koreas to march under single 'united' flag in Olympic Games
Extremism and Terrorism	Auschwitz museum: Important to remember Holocaust 'did not start from gas chambers'
Immigration, Migrants, and Refugees in Europe	Merkel rules out migrant policy reversal after attacks
UK Politics	Jeremy Corbyn tables no-confidence motion after May defeat – as it happened
Economic Inequality Issues	World's witnessing a new Gilded Age as billionaires' wealth swells to $6tn
Sensational and Social Issues	Millennials earn 20% less than Boomers did at same stage of life
US Senate and Supreme Court	Democrats prepare bill limiting U.S. Supreme Court justice terms to 18 years
Economic Growth and Policy	Losses on short positions in U.S. firms top $70 billion - Ortex data
US Politicians and Elections	Bernie Sanders drops out of the presidential race
Immigration in the US	Trump orders wall to be built on Mexico border
Government Spending and Taxes	50 years of tax cuts for the rich failed to trickle down, economics study says
Healthcare	Hospitals Sued to Keep Prices Secret. They Lost.

Table 2: Primary frames and example headlines. The topic is also included to illustrate how frames are invoked across different issues.

Primary Frame	Topic	Example Article Headline
Economic frames	US Politicians and Elections	NY Times: Trump paid $750 in US income taxes in 2016, 2017
Capacity and resources frames	Healthcare	ObamaCare signups surge in early days to set new record
Morality frames	COVID-19	Russia's top doctor quits over 'gross violations' of ethics that rushed through Covid-19 'vaccine'
Fairness and equality frames	Technology Companies	Facebook Fired An Employee Who Collected Evidence Of Right-Wing Pages Getting Preferential Treatment
Constitutionality and jurisprudence frames	LGBTQ Rights and Religion	Supreme Court rules workers can't be fired for being gay or transgender
Policy prescription and evaluation	Technology Companies	F.C.C. Repeals Net Neutrality Rules
Law and order, crime and justice frames	Technology Companies	Facebook failed to remove sexualised images of children
Security and defense frames	Trump Collusion and Impeachment	Trump revealed highly classified information to Russian foreign minister and ambassador
Health and safety frames	COVID-19	Coronavirus declared global health emergency by WHO
Quality of life frames	China and Hong Kong	The Hong Kong migrants fleeing to start new lives in the UK
Cultural identity frames	Immigration, Migrants, and Refugees in Europe	Muslim Representative: It's Up to White British to Integrate More
Public Opinion frames	Crime and Law Enforcement	Protests about police brutality are met with wave of police brutality across US
Political frames	Healthcare	US healthcare: Senate 'skinny repeal' bill fails
External regulation and reputation frames	Climate Change	19 of 20 World Leaders Just Pledged to Fight Climate Change. Trump Was the Lone Holdout.