

3-9-2022

Ideological variation in preferred content and source credibility on Reddit during the COVID-19 pandemic

Tendai Gwanzura

Department of Epidemiology, Florida International University

Wallace Chipidza

Center for Information Systems and Technology, Claremont Graduate University

Christopher Krewson

Political Science Department, Brigham Young University

Nicole Gatto

School of Community and Global Health, Claremont Graduate University

Follow this and additional works at: https://digitalcommons.fiu.edu/covid-19_research


Recommended Citation

Gwanzura, Tendai; Chipidza, Wallace; Krewson, Christopher; and Gatto, Nicole, "Ideological variation in preferred content and source credibility on Reddit during the COVID-19 pandemic" (2022). *Coronavirus Research at FIU*. 96.

https://digitalcommons.fiu.edu/covid-19_research/96

This work is brought to you for free and open access by FIU Digital Commons. It has been accepted for inclusion in Coronavirus Research at FIU by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

Ideological variation in preferred content and source credibility on Reddit during the COVID-19 pandemic

Big Data & Society
 January–June: 1–14
 © The Author(s) 2022
 Article reuse guidelines:
sagepub.com/journals-permissions
 DOI: 10.1177/20539517221076486
journals.sagepub.com/home/bds


Wallace Chipidza¹ , Christopher Krewson² , Nicole Gatto³,
 Elmira Akbarpourdibazar¹ and Tendai Gwanzura⁴ 

Abstract

In this exploratory study, we examine political polarization regarding the online discussion of the COVID-19 pandemic. We use data from Reddit to explore the differences in the topics emphasized by different subreddits according to political ideology. We also examine whether there are systematic differences in the credibility of sources shared by the subscribers of subreddits that vary by ideology, and in the tendency to share information from sources implicated in spreading COVID-19 misinformation. Our results show polarization in topics of discussion: the Trump, White House, and economic relief topics are statistically more prominent in liberal subreddits, and China and deaths topics are more prominent in conservative subreddits. There are also significant differences between liberal and conservative subreddits in their preferences for news sources. Liberal subreddits share and discuss articles from more credible news sources than conservative subreddits, and conservative subreddits are more likely than liberal subreddits to share articles from sites flagged for publishing COVID-19 misinformation.

Keywords

COVID-19, Reddit, misinformation, topic analysis, ideology, health, politics

Introduction

Background

Increasingly, Americans get their news from new media such as Reddit, Facebook, and Twitter, which have been implicated in intensifying political polarization (Pew Research Center, 2019). While it is well-understood that such new media democratize access to information thus mollifying the gatekeeping effect of legacy media (Andersen et al., 2016), these media make it easy for users to customize their media diets, potentially increasing political polarization (Young and Anderson, 2017). Reddit is the fifth most visited website in the USA with over 400 million unique monthly visitors¹. Underscoring Reddit's importance in academic research, a recent review found more than 700 studies leveraging data sets from Reddit over the 2010–2020 period (i.e. Proferes et al., 2021). The website is a collection of independently moderated sub-communities—known as subreddits—organized around shared interests by subscribers. There are political subreddits devoted to ideologies such as conservatism, libertarianism, liberalism, and socialism (Centivany and Glushko,

2016). Given that there is no editorial process and individuals are free to post news articles that conform to their ideological biases on Reddit, it is worthwhile to explore whether information sharing regarding COVID-19 on Reddit reflects political polarization (i.e. are conservatives and liberals getting their COVID-19-related news from the same sources?).

Concerns with spread of misinformation, especially through social media platforms, have accompanied the COVID-19 pandemic. The problem of misinformation is

¹Center for Information Systems and Technology, Claremont Graduate University, California, USA

²Political Science Department, Brigham Young University, Provo, Utah, USA

³School of Community and Global Health, Claremont Graduate University, California, USA

⁴Department of Epidemiology, Florida International University, Miami, USA

Corresponding author:

Wallace Chipidza, Center for Information Systems and Technology, Claremont Graduate University, Claremont, CA 91711, USA.
 Email: wallace.chipidza@cgu.edu

serious enough to influence the outcomes of national elections (Allcott and Gentzkow, 2017), but when it affects public health, misinformation can have life and death implications (Baum, 2011). Thus, the value of credible information concerning public health measures for COVID-19 prevention cannot be emphasized enough. Reddit has been implicated in spreading misinformation in the past (Au et al., 2020; Tasnim et al., 2020). The phenomenon of health misinformation has been dubbed the infodemic and has been highlighted by the World Health Organization as posing serious threats to public health (Zarocostas, 2020). Given the amplifying effect of social media on infodemics, it is important to examine the credibility of information shared on Reddit when users discuss COVID-19.

Our study is based on Reddit data; thus, it is useful to point out some attendant limitations of Reddit that have a bearing on the interpretation of our findings. The Reddit user base is mostly young (58% in the 18–34 years range), male (57%), and resides in the USA (plurality of 49%) (Proferes et al., 2021; Statista, 2021). Our findings should be understood incorporating that context. As we note earlier, because of its popularity—that is, 52 million daily active users (Proferes et al., 2021)—and impact on real world events, Reddit is worth studying in its own right. Reddit’s affordances make it possible to answer questions on whether aspects of the COVID-19 pandemic are ideologically politicized in ways that might not be possible otherwise. The demographic limitations notwithstanding, our findings are still revelatory given how many people visit Reddit daily and the platform’s affordances that allow people to subscribe to subreddits based on shared interests.

To get a complete picture of the COVID-19 pandemic as it unfolds on Reddit, it is imperative to understand (a) the topics that comprise its discussion, (b) whether ideology influences the topics that are emphasized in political subreddits, (c) whether ideology influences preferred information sources, and (d) whether ideology influences preference for sharing COVID-19-related misinformation. The study’s research questions are summarized as follows:

1. What are the topics that comprise the discussion of the COVID-19 pandemic on Reddit?
2. How does a subreddit’s expressed or implied political ideology/bias influence the topics preferred for discussion by its subscribers?
3. How does a subreddit’s expressed or implied political ideology/bias influence the COVID-19-related information sources preferred by its subscribers?
4. How does a subreddit’s expressed or implied political ideology/bias influence the COVID-19-related misinformation shared by its subscribers?

Social media platforms like Reddit play an increasingly important role in disseminating information on health and pandemics (Baum, 2011). Examining these questions on Reddit potentially explains how COVID-19 is being discussed online, while also highlighting how political polarization affects public receptiveness to pandemic mitigation messaging. We use structural topic modeling (STM)—a machine learning technique to automate topic discovery within large corpora of data (Roberts et al., 2014)—to uncover the topics (clusters of co-occurring words) associated with COVID-19 on Reddit. We then assign qualitative labels to the topics based on the most prevalent words in each topic and the most representative text loading on each topic. In addition to topic discovery, STM allows for examining the effect of covariates (in our case political ideology) on preferred topics (i.e. which topics do conservative subreddits tend to prefer relative to liberal subreddits?). The *r/coronavirus* and *r/science* subreddits provide useful comparisons for how non-political subreddits discuss COVID-19 relative to their political counterparts.

Why Reddit?

Answering our research questions helps inform the degree to which Reddit and other social media platforms foster ideological polarization regarding COVID-19. Our study also builds knowledge on the effect of ideology on tendency to share COVID-19 misinformation potentially helping broader efforts to mitigate the pandemic. Our findings regarding ideological polarization on Reddit may reflect underlying political polarization in the USA in the context of COVID-19.

Reddit has multiple affordances that make it amenable to answering our research questions. First, it has an API that allows for all sorts of queries regarding content submitted and discussed, the timing of that content, as well as other important metadata (e.g. number of upvotes and downvotes, upvote ratio, number of comments, etc.). The Reddit API is uniquely generous among social media platforms in the level of transparency that it offers to researchers, possibly because Reddit users are typically anonymous and the privacy concerns that matter for Facebook for example may not be as salient.

Even more important, the organization of Reddit by subreddits makes it possible to examine how users subscribing to certain interests come to view certain phenomena. Our case of COVID-19 is telling; the numerous ideological subreddits surely weigh in on contentious issues such as possible cures for the disease, vaccine safety and efficacy, as well as the effectiveness of mitigation measures such as masks and lockdowns.

In the next section, we describe related theory and research. We proceed to describe our data collection and analyses processes. After, we describe the results of our

study, discuss them as they relate to existing research, and conclude.

Related research and theory

Reddit features and political polarization

Existing communication theories regarding health information sharing largely draw from studies on traditional media, such as newspapers, television, or other broadcasting services (Dalmer, 2017). However, social media platforms like Reddit are emerging as new media from which the increasing majorities of the population obtain their news (Shearer and Matsa, 2018).

As Reddit is organized as a collection of common interest subreddits, with individuals free to subscribe to subreddits of their choice, it enables “individuals to select outlets that conform to their prior ideologies as in an ‘echo chamber’” (Campante and Hojman, 2013: 80). In the context of Reddit, extant research suggests that individuals may become dependent on specific subreddits for their political news. In fact, Reddit’s subscription feature reinforces this dependency effect to the extent that it facilitates default exposure to content from subreddits to which a given user has subscribed (Massanari, 2017). Thus, we are likely to find political polarization in the discussion of COVID-19 wherein subscribers to political subreddits discuss those issues which are helpful in justifying their ideological projects. Furthermore, we expect political subreddits to select those information sources that are congruent with their worldviews, such that subscribers to conservative subreddits may prefer right-leaning information sources and subscribers to liberal subreddits may prefer left-leaning information sources. Consequently, we expect limited overlap in the information sources preferred across conservative and liberal subreddits.

The concept of a “gatekeeper” (a person who determines what makes the news and what does not) is less relevant today as nearly anyone with modest resources can report news via the internet (Young and Anderson, 2017). On Reddit, users organize themselves into subcommunities known as subreddits structured along shared interests. Politics form a dominant means of organizing these subreddits, bringing together users subscribing to similar ideologies. Each user may then subscribe to a set of subreddits of their choice, the most popular and current content of which subreddits constitutes the user’s default front page when logged in. One of the most common activities on Reddit is the sharing of links to external news sources (Gaudette et al., 2021). These links are important sources of news which may unintentionally reveal important biases in news exposure and consumption.

For our study, we expect political subreddits to cover COVID-19 because of the magnitude of the importance of the event globally. While we expect liberal, conservative,

and neutral subreddits to focus on a similar agenda (COVID-19) during our period of study, we also expect that the various subreddit communities will frame this event differently. Furthermore, we also expect that community members will be more likely to share news sources that are biased toward their community’s preferences.

Metzger et al. (2020) show that people perceive attitude-consistent news sources as more credible than attitude-challenging news sources, all else equal. Accordingly, we might expect individual evaluations of a news source’s credibility to be correlated with the ideological leanings of its content rather than on the actual quality or factuality of the news being reported. This may lead to the selective sharing of information to Reddit based on partisanship. On political subreddits, selective sharing is likely amplified because moderation rules, norms, and groupthink are likely to shape media diets to reflect their organizing ideologies, leading to variations in information source credibility and propensity to share COVID-19 misinformation according to ideology.

Earlier research on Reddit suggests various mechanisms by which polarization emerges on the platform. For example, Mills (2018) found that the subreddits supporting Donald Trump (i.e. r/the_donald) and Bernie Sanders (i.e. r/SFP) during the 2016 US presidential elections sourced their news from different outlets and that it was relatively easy for these communities to obtain news conforming to their preferred narratives. Content that conforms to a subreddit’s dominant narrative is more likely to be upvoted than content that challenges the narrative (Mills, 2018; Soliman et al., 2019). Rules in political subreddits that prohibit certain content and limit participation to supporters of a political cause also encourage polarization (Mills, 2018; Robards, 2018). Thus, multiple features on Reddit including organization by subreddits, default page, moderation policies, link sharing, and upvoting (or downvoting) may contribute toward polarization.

Data and methods

We used PRAW—Reddit’s Python API—to collect titles of all posts containing the terms “coronavirus” or “Covid” from 21 January 2020 to 9 July 2020. A variety of subreddits dedicated to politics exist. On these subreddits, subscribers share information/news through posts of their own content or links to news articles which other users can also comment and vote on. (Votes signaling approval are called upvotes, and those signaling disapproval are called downvotes.) We followed the IRB guidelines at our institution that determine that publicly available data is exempt from review. Furthermore, data is reported only in the aggregate, with no means to link it to the real identities of Reddit users.

We collected posts linking to external information sources from 12 subreddits with noted political biases

(conservative and liberal), and from 2 that are politically neutral but with relevant coverage of the COVID-19 pandemic (r/coronavirus and r/science)². The number of posts was 6268 but ultimately 5188 posts were used in the analysis after filtering out self-posts (i.e. those not linking to external news sources). For our purposes, the r/science and r/coronavirus subreddits are neutral based on their stated missions, that is, “Increase Knowledge” (sic) and “... monitor the spread of the disease COVID-19 ...,” respectively. In fact, one of the rules for participating in r/coronavirus is to “(a)void politics.” Thus, the stated foci of these subreddits are not to promote ideological goals and are thus considered politically neutral. We relied on the descriptions displayed on the subreddits’ front pages to determine their ideological biases, and in the case of r/politics and r/esist, we confirmed the ideological bias to be liberal as noted in recent literature (Soliman et al., 2019). The data for each post title included the source subreddit, the source URL for the post, and the time of posting. Although Reddit allows posting of original content (referred to as self-posts) by users, we only considered posts that linked to other information sources. Table 1 shows the details of these subreddits and counts of titles fulfilling our search criteria.

We employ STM to discover hidden topics within the titles of COVID-19-related news articles posted in the 12 political subreddits (six left-leaning and six right-leaning) and the two neutral subreddits (r/science focused on science in general and r/coronavirus on COVID-19 in particular).

STM is a generative algorithm wherein data are generated until parameters are found that maximize the probability of the observed input corpus (Roberts et al., 2014). Much like other topic modeling algorithms (e.g. Latent Dirichlet Allocation (LDA)), the goal of STM is to find distributions of words over topics, and of topics over documents, that is, each topic is a mixture of words, and each document is a mixture of topics (Roberts et al., 2014). Thus, STM is similar to a soft clustering scheme in that a word can belong to multiple topics but with varying probabilities. Topics can be understood as sets of words occurring together with greater than expected chance (Chandelier et al., 2018). Unlike LDA, however, in STM topics can be correlated, the topic proportion for a given document can be influenced by a set of covariates (i.e. using ideological bias as an example covariate, certain topics become much more prevalent in a document for liberals compared to conservatives). The use of a given word in a topic can also be influenced by a set of covariates that is, conservatives may emphasize certain words within a given topic much more than their liberal counterparts.

The incorporation of covariates in topic modeling thus injects useful information into the inference process. As with all topic modeling approaches, it is still up to the researcher to interpret discovered topics within a corpus

of data. The major advantage of using an automated approach lies in uncovering the dominant themes within large corpora of text, which would require orders of magnitude more effort if left to human labor.

Here we make no a priori hypotheses as to the topics we might find; rather, we adopt an inductive approach to topic discovery. We employed STM to understand (a) the topics of COVID-19-related discussions on Reddit (RQ1) and (b) how those topics varied by ideology (RQ2) in the corpus since 21 January 2020. We use the stm package in R to carry out our analyses (R Development Core Team, 2010; Roberts et al., 2014). After varying K —the number of topics contained within the corpus—we settled on $K = 9$ because that is the number of topics that maximizes topic coherence and held-out log-likelihood of generating the observed corpus (Mimno et al., 2011). We ran the STM algorithm over the Reddit corpus with spectral initialization to ensure that results were consistent across different runs of the algorithm (Roberts et al., 2014), with maximum number of iterations equaling 1000. We used the function findThoughts in the stm package to help derive the qualitative labels of the generated topics. The function retrieves the most representative posts loading on a given topic.

Figure 1 summarizes the process of data collection and analysis for the topics that comprise discussion of the COVID-19 pandemic on Reddit.

We use source credibility ratings from MediaBiasFactCheck.com to examine differences in credibility of information sources shared by Reddit users across ideological lines with linear regression (RQ3). The site MediaBiasFactCheck classifies over 3300 media sources according to factual rating and ideological bias and has been cited in academic research for the same purpose as our study (e.g. Mensio and Alani, 2019; Yu et al., 2019). We assigned values between 1 and 6 corresponding to the following factual accuracy ratings by MediaBiasFactCheck: very low (1), low (2), mixed (3), mostly factual (4), high (5), and very high (6). For robustness checks, we tested whether the credibility differences across conservative versus liberal subreddits would still be observed using another media credibility rating service. We employed NewsGuard³, a service that “tell(s) you if a site is reliable as you browse online news.”

To understand how a subreddit’s expressed or implied political ideology/bias influence the COVID-19-related misinformation shared by its subscribers (RQ4), we used the NewsGuardTech’s database of sites flagged for COVID-19 misinformation⁴. To date, 336 sites have been flagged as sources of COVID-19 misinformation, with the flagged content including myths such as that Sars-Cov-2 was stolen by Chinese spies from a Canadian lab, that 5G technology is linked to the spread of COVID-19, and that the National Institute of Allergy and Infectious Diseases Director Dr Anthony Fauci would personally profit from a COVID-19 vaccine⁵. We assigned a flag of 1 to sites

Table 1. Subreddits, self-descriptions, biases, and COVID-19 post counts.

Subreddit	Subreddit self-description	Subreddit bias	Number of COVID-19 posts
AskThe_Donald	We are a PRO Trump, PRO Administration, and MAGA zone. This sub is for people to learn and talk about Trump and Conservative subjects	Conservative	372
Conservative	The place for Conservatives on Reddit	Conservative	561
Coronavirus	In December 2019, a novel coronavirus strain (SARS-CoV-2) emerged in the city of Wuhan, China. This subreddit seeks to monitor the spread of COVID-19, declared a pandemic by the WHO	Neutral	738
DrainTheSwamp	Our country is now taking so steady a course as to show by what road it will pass to destruction, to wit: by consolidation of power first, and then corruption, its necessary consequence—Thomas Jefferson	Conservative	548
EnoughTrumpSpam	Because the amount of Trump spam is *too damn high!* Enough Трампа Spam	Liberal	430
Esist	Welcome to esist	Liberal	362
Libertarian	A place to discuss libertarianism, politics, related topics, and to share things that would be of interest to libertarians	Conservative	575
Political_Revolution	This subreddit is a part of the political revolution as envisioned by Senator Bernie Sanders. We represent a movement promoting activism, raising support for progressive candidates, and spreading awareness for the issues focused on by the progressive cause	Liberal	495
politics	/r/Politics is for news and discussion about US politics	Liberal	525
redacted	Welcome to redacted	Conservative	154
SandersForPresident	Bernie Sanders 2020!	Liberal	396
Science	This community is a place to share and discuss new scientific research. Read about the latest advances in astronomy, biology, medicine, physics, social science, and more. Find and submit new publications and popular science coverage of current research	Neutral	441
Socialism	Welcome to r/socialism! This is a community for socialists to discuss current events in our world from an anti-capitalist perspectives	Liberal	518
tucker_carlson	Tucker Carlson is the sworn enemy of lying, pomposity, smugness and groupthink. His goal is to pierce pomposity, translate double-speak, mock smugness and barbecue nonsense as he debates people from all across the political spectrum every weeknight on Tucker Carlson Tonight!	Conservative	153
Total			6268

flagged for misinformation, and 0 otherwise. We regressed the flag variable on subreddit ideology to determine the effect of subreddit ideology on sharing COVID-19 news from sites that spread misinformation.

Results

Generated topics

Setting $K = 9$ as suggested by the topic coherence measure resulted in topics ordered by proportion in the corpus which we labeled as follows: deaths (14.3%), Trump (13.5%), scientific research (12.9%), economic relief (11.6%), China (9.8%), fighting the pandemic (9.7%), protests (9.7%), White House (9.6%), and reopening debate (8.9%). It is worth noting that the scientific research topic references many ideas that have been deemed misinformation by the WHO, for example, the idea that hydroxychloroquine is an effective treatment for COVID-19 (World Health Organization, 2021). Table 2 summarizes the

generated topics; it includes the labels we assigned to each topic, the most frequent words appearing in each topic, its description, and example posts loading highly on each topic—one each from neutral, conservative, and liberal subreddits.

Effect of ideological bias on topic distribution

The analysis of the Reddit corpus using STM shows that there were statistically significant differences in preferences for seven of the nine topics. Conservative subreddits were more likely than liberal subreddits to post about scientific research ($\beta = 0.06, p = .00$) (the beta corresponds to a probability, meaning that the probability of a post loading on the scientific research topic was 6% higher if posted in a conservative subreddit relative to a liberal subreddit), deaths ($\beta = 0.05, p = .00$), and China ($\beta = 0.01, p = .04$) topics. On the other hand, liberal subreddits were more likely than conservative subreddits to post about the

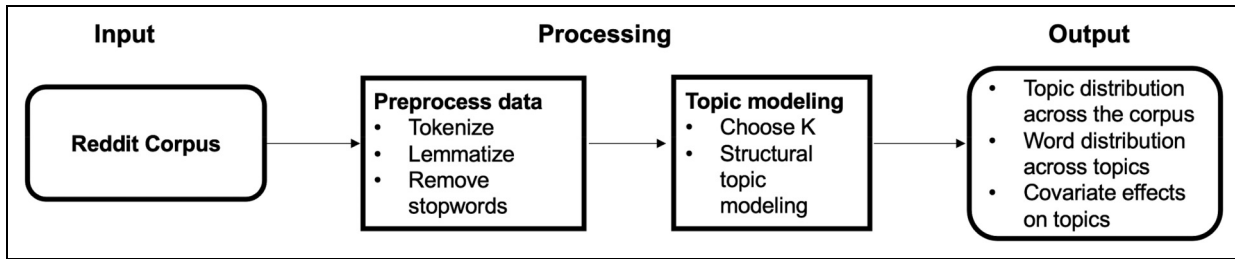


Figure 1. Structural topic modeling process.

Trump topic ($\beta = 0.05$, $p = .00$), economic relief ($\beta = 0.03$, $p = .00$), White House ($\beta = 0.02$, $p = .00$), and reopening debate ($\beta = 0.02$, $p = .00$). There were no significant differences in preference for the Black Lives Matter protests and fighting the pandemic topics. Figure 2 graphically summarizes these differences.

It is also worth noting that the neutral subreddits (*r/coronavirus* and *r/science*) were more likely than biased Reddit subreddits to post about scientific research ($\beta = 0.15$, $p = .00$), deaths ($\beta = 0.11$, $p = .00$), and fighting the pandemic ($\beta = 0.08$, $p = .00$) topics. They were less likely to post about Trump ($\beta = -0.16$, $p = .00$), economic relief ($\beta = -0.09$, $p = .00$), White House ($\beta = -0.06$, $p = .00$), China ($\beta = -0.02$, $p = .00$) and Black Lives Matter protests ($\beta = -0.02$, $p = .00$) topics. The neutral subreddits were just as likely as biased subreddits to post about the reopening debate. Figure 3 summarizes this information.

Effect of subreddit moderation on topic distribution

We considered two kinds of moderation as covariates potentially influencing topic distribution in discussing COVID-19 news on Reddit. The first kind—editing restriction—relates to permission to submit edited titles; this kind of moderation was not significant in all topics. The second kind—domain restriction—which relates to allowing submissions from all domains, was not significant save for two topics, that is, the White House topic, which was more popular in subreddits enforcing this moderation, and the fighting the pandemic topic which was less popular. The effect of moderation on topic distribution is thus limited to only two topics, and is dwarfed by the effect of ideology.

Effect of ideology on COVID-19 news source accuracy

Of the 312 news sources shared by Reddit users with a rating from MediaBiasFactCheck, 10% had a very low factual rating (e.g. Breitbart News and Gateway Pundit), 1% low (e.g. Children’s Health Defense and Global Research Canada), 19% mixed (e.g. CNN and Fox News), 11% mostly factual (e.g. Bloomberg News and

National Review), 55% high (e.g. ABC News and Chicago Tribune), and 3% very high factual rating (e.g. Associated Press and Reuters). Figure 4 shows the top 20 news sources shared by subscribers of conservative and liberal subreddits, respectively. We assign low credibility to news sources with very low, low, and mixed factual ratings from MediaBiasFactCheck, and high credibility to sources with mostly factual, high, and very high factual ratings. Of the top 20 news sites posted in conservative subreddits, 60% had mixed or lower factual ratings. In contrast, the equivalent number for liberal subreddits is 20%.

On average, liberal subreddit subscribers shared news from sources with higher factual ratings (mean = 4.46/6) than their conservative counterparts (mean = 3.07), a difference that is statistically significant ($\beta = 1.00$, $p = .00$). Furthermore, the neutral subreddits shared from sources with higher credibility (mean = 4.61) than the ideologically biased subreddits (mean = 4.07) and the difference was also significant ($\beta = 0.55$, $p = .00$).

Using NewsGuard, five (25%) of the top 20 sites shared by conservatives “severely violate basic journalistic standards.” The equivalent number for liberal subreddits is zero. On average, using the NewsGuard trustscore, neutral and liberal sources had average credibility ratings of 93.3% and 91.5%, respectively, which were higher than the conservative average of 75.2%. The difference between the liberal and conservative means was significant ($\beta = 16.4$, $p = .00$), as was the difference between neutral and biased sources ($\beta = 7.88$, $p = .00$). Thus, we can be reasonably confident that conservative subreddits on Reddit consume COVID-19 news from sites with lower credibility than their liberal and neutral counterparts and that the biased subreddits consume news from less credible sites than their neutral counterparts.

Effect of ideology on preference for COVID-19 misinformation

Of the 5258 posts from news sources, 260 (5%) were from sites flagged for COVID-19 misinformation. We found that conservative subreddits share COVID-19 news from sites flagged for COVID-19-related misinformation at higher rates than liberal subreddits ($\beta = -0.20$, $p = .00$) and

Table 2. Structural topic modeling (STM)-generated topics from Reddit corpus.

Topic	Keywords	Description	Representative text from neutral subreddits	Representative text from conservative subreddits	Representative text from liberal subreddits
Deaths	Test, death, case, die, positive, hospital, report, first, mask, protest, home, week, nurse, cdc, number, govern, patient, Florida, York, even	Lamenting the deaths caused by COVID-19	New York, New Jersey, and Michigan saw 20,000 more deaths than average in March and April. These were on top of the deaths already accounted for by Covid death counts. This leads researchers to believe the Covid death count is actually much higher than is being reported.	The USA Is Dramatically Overcounting Covid Deaths. If you were in hospice and had already been given a few weeks to live, and then you also were found to have COVID that would be counted as a COVID death. Even if you died of clear alternative cause it still listed as a COVID death.	Corona death toll: Americans are almost certainly dying of Covid but being left out of the official count
Trump	Trump, response, crisis, claim, use, America, brief, administration, blame, treatment, govern, rally, show, campaign, global, fail, press, video, danger, media	Criticism and defense of President Trump's response to the pandemic	Trump suggests "injection" of disinfectant to beat corona and "clean" the lungs	Trump is Right: WHO Failed the World With the corona Pandemic	Trump Touts "Game-Changing" Drug Cocktail For corona Linked To Fatal Arrhythmia—The president, who is not a doctor, recommends a potentially dangerous drug combo to his 74 million Twitter followers. "What do we have to lose?" he asked. The US government gave hydroxychloroquine to 1300 veterans infected with Covid, despite evidence that the drug is ineffective and could increase the risk of death
Scientific research	Infect, patient, study, research, rate, find, disease, show, use, virus, symptom, severe, drug, suggest, can, sars-cov-2, human, may, spread, found	Results of scientific studies that characterize the nature of the virus and possible treatments	Researchers Identify Multiple Molecules that Shut Down SARS-Cov-2 Polymerase Reaction: A library of molecules with unique structural and chemical features inhibit the novel corona polymerase, a key drug target for Covid	Hydroxychloroquine rated "most effective therapy" by doctors for corona: Global survey	
Economic relief	Sander, bill, amid, relief, million, senate, help, need, fund, vote, stock, support, month, economic, Pelosi, report, can, stimulus, gop, aid	Legislative efforts to mitigate against the economic damage wrought by the pandemic	Mitt Romney: Every American adult should immediately receive \$1000 to help ensure families and workers can meet their short-term obligations and increase spending in the economy.	Speaker Nancy Pelosi Caught Trying to Include Abortion Funding in Bill to Combat corona	AOC Takes Brave, Lonely Stand Against "Unconscionable" Covid Relief Package That Doesn't Sufficiently Help Those Hurt the Most
China	Health, China, spread, just, stop, news, work, order, public, Chinese, doctor, confirm, nation, south, polit, see, Korea, issue, Wuhan, fox	Criticisms of China's role in the spread of the virus	The Chinese doctors squad just arrived in Italy to help to fight the Covid with their experience	China "intentionally concealed the severity" of corona outbreak to hoard supplies: DHS report	Referring to the corona, Trump says he was told by China's President Xi. By April, during the month of April, the heat generally kills this kind of virus, so that would be a good thing.

(continued)

Table 2. Continued

Topic	Keywords	Description	Representative text from neutral subreddits	Representative text from conservative subreddits	Representative text from liberal subreddits
Fighting the pandemic	Virus, worker, time, social, country, fight, lockdown, media, organ, black, know, way, emergency, lead, attack, still, last, better, near, nation	Efforts to fight against the spread of COVID-19, including lockdowns, social distancing, and declared states of emergency.	Rihanna Donates \$4.2 Million to Domestic Violence Victims Impacted by Covid Lockdowns. The pop star teams up with Twitter CEO Jack Dorsey to make the sizable donation	President Trump made a commitment to donate his salary while in office. Honoring that promise and to further protect the American people, he is donating his 2019 Q4 salary to HHS to support the efforts being undertaken to confront, contain, and combat corona.	Still No Widespread Covid Testing, But the Fed Has a \$500 Billion Bank Stimulus—The government is forking over a grand total of \$1.5 trillion to...nudge the stock market
Black Lives Matter protests	American, world, democrat, right, lie, try, live, face, man, want, kill, force, threat, save, capital, keep, give, get, think, tell	Link of Black Lives Matter protests to spread of COVID-19.	Police Tactics Could Turn Protests Into Covid Hot Spots —Sure, large crowds already carry a risk of transmission. It's just worse when you tear gas people, make them cough on each other, and bus them to jail.	Democrats cheering "Black Lives Matter" protests now say Trump rallies pose corona risk	DOJ Wants to Suspend Constitutional Rights During Corona Emergency
White House	House, call, white, republican, take, make, Biden, warn, plan, governor, expert, push, concern, don't, sick, top, hold, leave, joe, hoax	Criticism and defense of White House response to the pandemic.	Bipartisan group pitches the White House on a \$46.5 billion Covid plan: to hire an army of 180,000 contact-tracers, book blocks of vacant hotel rooms so Americans sick with Covid can self-isolate, and pay sick individuals to stay away from work until they recover	Only 632 People Watch Sleepy Joe Biden's corona Town Hall on YouTube	The Trump Administration Paid Millions for Test Tubes—and Got Unusable Mini Soda Bottles; The plastic tubes supplied for Corona testing by Fillakit, a first-time federal contractor with a sketchy owner; don't even fit the racks used to analyze samples.
Reopening debate	Pandemic, state, outbreak, office, care, due, vaccine, medic, demand, year, medicare, company, system, expose, data, poll, healthcare, refuse, Texas, reopen	Debates on whether states should reopen or not.	Texas Has Over 1700 New Covid Cases Since Abbott Announced State's Reopening	Elon Musk says he plans to move Tesla out of California and sue county after corona restrictions	Texas Began Opening Businesses May 1. Now, They're Averaging 1000 New Cases of Covid a Day

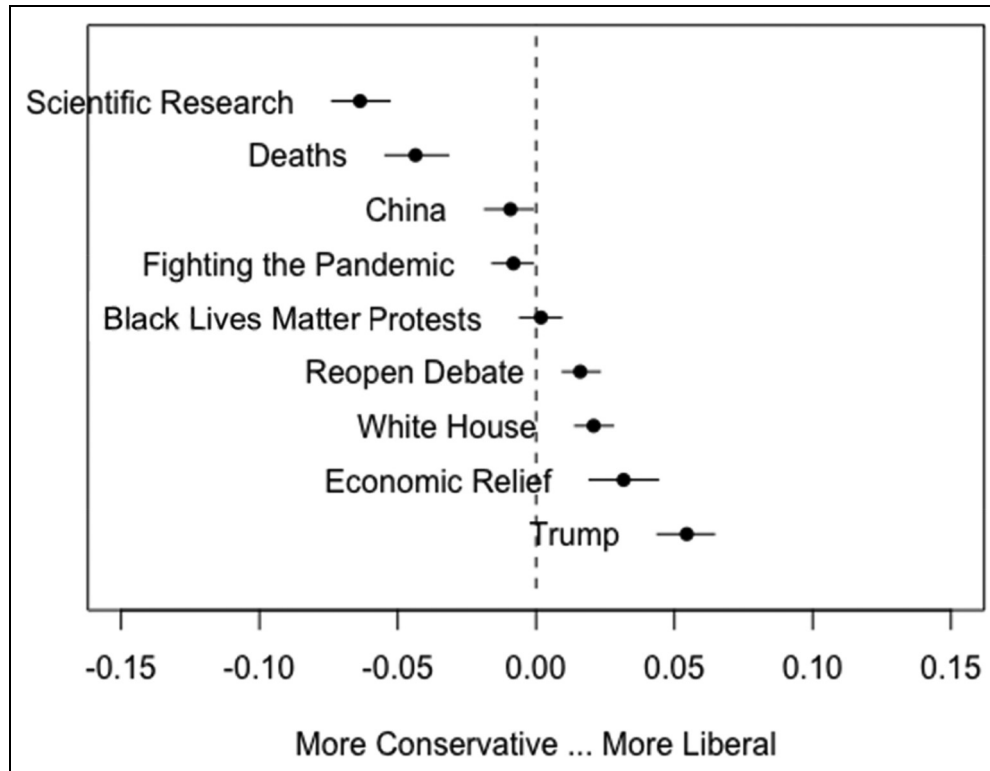


Figure 2. Topic preferences by ideological bias.

neutral subreddits ($\beta = -0.21, p = .00$). In fact, 15% of all posts in conservative subreddits originated from sites flagged for misinformation; the equivalent number for both liberal and neutral subreddits was 0%. There was no significant difference in sharing COVID-19 articles from sites flagged for misinformation between liberal and neutral sources. Subreddit moderation policies—that is, domain and title editing restrictions—had no significant effect on the tendency to share news sources from sites flagged for misinformation.

Discussion

We include in our sample some of the most influential subreddits on Reddit; *r/science*, *r/coronavirus*, *r/politics*, and *r/Conservative* for example, boast 27 million, 2.4 million, 7.6 million, and 800,000 subscribers⁶, respectively. The median subscriber count across all the subreddits is 385,000. Submissions had 44 comments on average and were upvoted 238 times, with an upvote ratio of 83% (meaning that 17% of votes were downvotes). These metrics show significant engagement of COVID-19-related content in political subreddits as well as in *r/science* and *r/coronavirus*.

Our study finds that discussion of the COVID-19 pandemic is polarized across political ideology on Reddit. Reddit affords adherents to political ideologies the ability

to coalesce around common interests, further bolstering the echo chamber effect whereby people primarily expose themselves to those beliefs that reinforce their world views. Thus, whereas discussion in liberal subreddits framed the COVID-19 phenomenon as an indictment of Trump’s perceived mishandling of the crisis than did conservative subreddits, the latter were much more likely to frame the resulting fallout as the fault of China. These findings suggest that conservatives and liberals on Reddit consume COVID-19 information from different sources, and they emphasize those aspects of the pandemic that might be politically helpful for their ideological projects. At the same time, information that helps with COVID-19 prevention was little discussed relative to other topics, suggesting that political polarization might be undercutting the public health message.

Why do we observe political polarization on the healthcare phenomenon that is COVID-19 as discussed on Reddit? Part of the answer might lie in certain Reddit features, which inadvertently foster political polarization. Features such as upvoting, default sorting filter, post aggregation across subreddits, and limited tool support for moderators have been implicated in fostering “toxic technocultures” and ideological extremism on Reddit (Gaudette et al., 2021; Massanari, 2017). Our study implicates Reddit’s subreddit subscription feature in making COVID-19 a politically polarizing phenomenon. Reddit

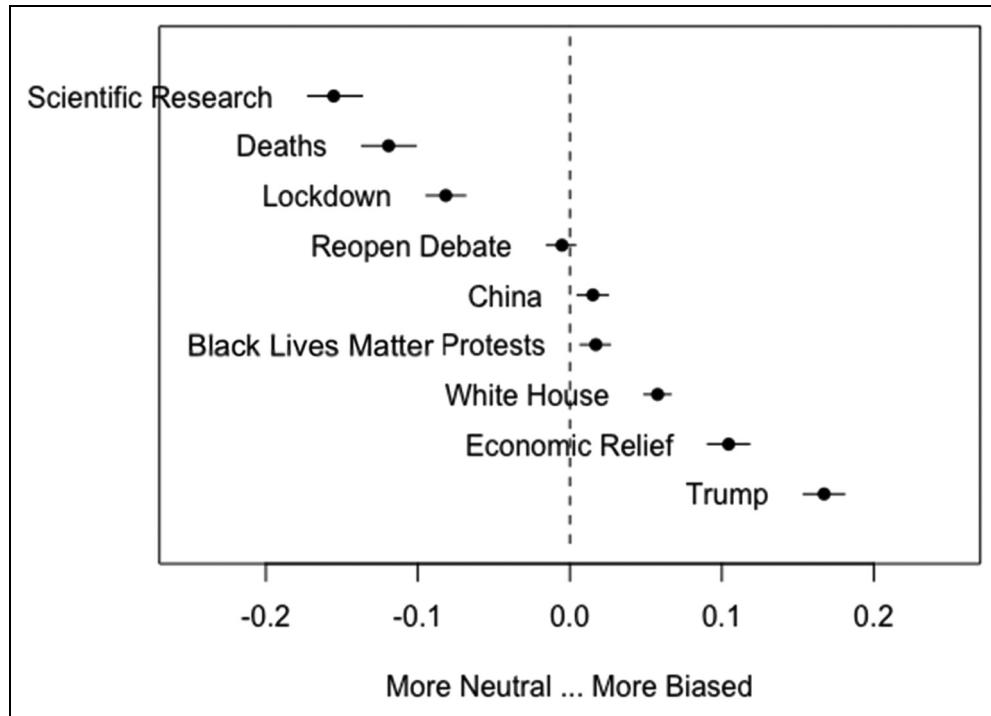


Figure 3. Topic preferences: biased versus neutral.

users subscribe to those subreddits which interest them, and these form the default front page for the user when accessing reddit.com (Massanari, 2017). If a user is subscribed to subreddits that align only with their ideology, they will primarily see content conforming to the same. Users are also likely to share information from sources that reinforce their existing viewpoints, making it much harder to change minds in ways that are beneficial for public health. Thus, users may only be exposed to the narrative of COVID-19 as “not as deadly as the flu” thus downplaying its risk. Furthermore, users may decide to spend most of their time on Reddit interacting with others within select subreddits, where narratives may crystallize into echo chambers (Gaudette et al., 2021). People upvote content and by voting for certain topics, it might appear that certain COVID-19 viewpoints are valid and mainstream when they are not. For example, a post titled “The US Is Dramatically Overcounting COVID-19 Deaths...” was upvoted about 1000 times, even if the scientific consensus is clear that these ideas are wrong. Reddit’s voting algorithm thus facilitates mistrust in scientific reporting and skepticism of COVID-19 mitigation measures.

Our findings demonstrate a link between political polarization and misinformation. On a platform like Reddit with relatively few editorial guardrails, one may submit information from other sources whether vetted or not; here the argument is that Reddit is a content aggregator rather than content host and thus should not be responsible for the content submitted and discussed by users (see Massanari,

2017). Given that the goal of political subreddits is to advocate for viewpoints that are beneficial for their particular ideological projects, it follows that misinformation will likely be prevalent in political subreddits. We do find that conservative subreddits share articles from sites flagged for COVID-19 misinformation at much higher rates than liberal and neutral subreddits.

The content of the misinformation is instructive in terms of its perceived political benefits for proponents of conservative ideology. The ideas that COVID-19 death tolls in the USA are inflated and that hydroxychloroquine is effective at curing the disease are meant to suggest that the USA is overreacting to the pandemic and that lockdowns which are seen as economically and politically costly to the Republican party are therefore unnecessary. Our findings are interesting in light of recent research that examines the spread of misinformation promoting hydroxychloroquine as a “cure” of COVID-19 on Twitter, especially given former US President Trump’s promotion of the drug. (Haupt et al., 2021). Similar to our findings, claims of hydroxychloroquine’s efficacy in treating COVID-19 leverage scientific authority (i.e. the fringe group of physicians “America’s Frontline Doctors”), but in truth most healthcare experts and extant scientific studies find hydroxychloroquine an ineffective, even damaging, as a cure for COVID-19. The feature of Reddit as a content aggregator even from sites flagged for COVID-19 misinformation facilitates the dissemination of unverified information.

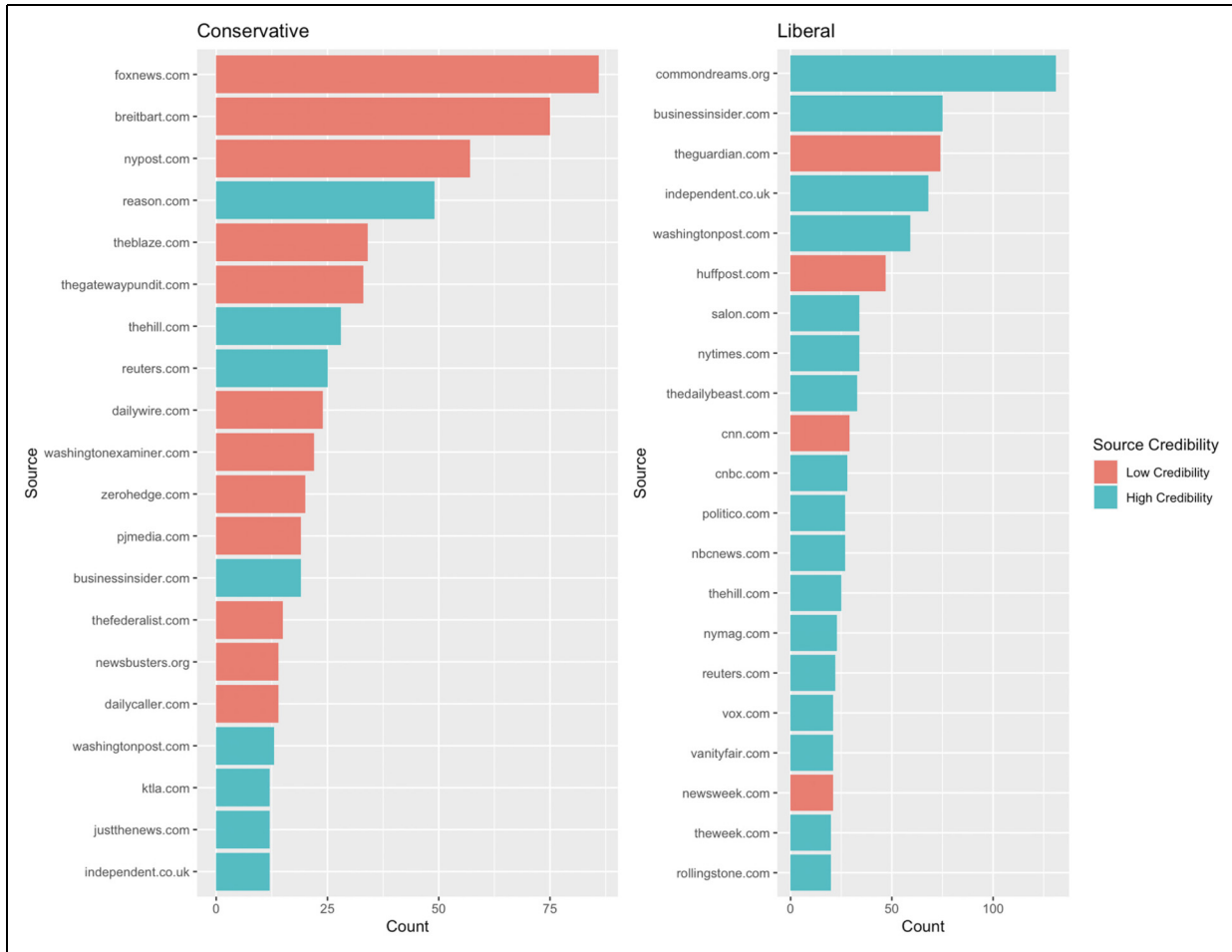


Figure 4. Top 20 news sources posted by subscribers of conservative and liberal subreddits.

Public health and health care interventions that are designed with an understanding of the political systems and institutions in which they are to be implemented may be more successful in generating a sustainable health impact (Greer et al., 2017; Trachtman, 2019). We recommend that public health professionals become active on Reddit to help with messaging on COVID-19 prevention and mitigation. Although Reddit aims to be laissez-faire as possible in terms of moderating content, the problem of misinformation could be partially alleviated if infectious disease specialists, epidemiologists, medical doctors, lab scientists, and other public health professionals were to contribute to subreddits to push factual information about COVID-19. Organizations representing these health professionals such as the American Medical Association or the American Public Health Association may also consider adjusting their advocacy activities to combat misinformation. This could be accomplished, for example, by dedicating efforts to monitoring news sites for health-related misinformation and use Reddit features such as flairs to display the flag on

posts linking to flagged sites. Over time, the flags would likely nudge users away from posting information from sites with low credibility. In fact, the WHO advocates that trusted sources such as the WHO and national health authorities are used for accurate COVID-19 information through their “Stop the Spread” campaign and encourages everyone to help stop the spread of misinformation. See <https://www.who.int/campaigns/connecting-the-world-to-combat-coronavirus/how-to-report-misinformation-online>

We also find that the neutral subreddits shared articles from more credible sites compared to the ideological subreddits. Subreddit moderation potentially explains why the neutral subreddits share more credible information; by allowing submissions only from peer reviewed research, r/science prevents the spread of unverified COVID-19 information and likely limits the politicization of the pandemic as well in that subreddit. Disseminating accurate information about infection is a key component of managing a pandemic. Reddit may help push credible information to their users by algorithmically promoting

information from science-focused subreddits. Such promotion would educate users on how to avoid infection based on best available peer-reviewed material, while also limiting user exposure to potentially harmful misinformation. In essence, for users of Reddit, credible information about the pandemic is best found in the neutral subreddits, and conservative subreddits are much less credible than liberal subreddits.

Moderation of content also has a role to play. If users are found to repeatedly post unverified health-related information to a given subreddit, moderators may impose Reddit penalties such as suspensions or bans on offenders. Awarding flairs of expertise to verified users is another mechanism that may help promote credible information on Reddit. These verified users would have flair indicating expertise in epidemiology, medicine, and/or public health, for example. The flair would nudge other users to trust their responses more than from unverified users.

Study limitations and future research

Our study has several limitations. First, our findings show that conservative and liberal subreddits consume COVID-19 information from disparate sources, but it does not show whether Reddit causes this polarization as opposed to just reflecting it. Future research could investigate the extent to which Reddit itself serves as a polarizing or misinformation source. Furthermore, the demographics of Reddit users skew young, white, and male (Massanari, 2017), meaning that the platform is not representative of the US or global population. Methodologically, there are limitations with Reddit's PRAW API, which is not guaranteed to return all the items requested from a specified search⁷. Another limitation is that we relied on MediaBiasFactCheck.com ratings of media credibility to understand the effect of ideology on information source credibility. Although we confirmed the robustness of our findings using ratings from NewsGuard, there is an opportunity for future research to make use of alternative rating systems, for example, crowdsourced ratings.

There are other political and COVID-19 themed subreddits that we did not include in our study at this exploratory phase, because they did not meet our inclusion criteria. We did not include COVID-19 themed subreddits such as *r/COVID-19*, *r/China_Flu*, and *r/COVID19_support* because they were created much later than the beginning of the phenomenon. We consider two politically neutral subreddits—*r/science* and *r/coronavirus*—in our analyses because they offer relevant coverage of COVID-19 related issues. Future research might examine COVID-19 topics on Reddit as a whole, to understand how other non-political but highly subscribed subreddits like *r/sports*, *r/news*, and *r/videos* discuss the pandemic. Results from such studies may demonstrate how issues become

politicized over time. Last, subreddit moderation policies beyond the domain and editing restrictions that we identified in this study may also influence discussion of COVID-19 on Reddit. Notably, we examined the effect of these moderation policies, not their actual enforcement. Future research may examine the effect of active moderation activities, for example, the extent to which the manual flagging of posts impacts the popularity of submitted posts and the extent to which deletion of posts that violate subreddit policies impacts topic discussion and spread of misinformation.

Conclusions

Social media has previously been implicated in fueling political polarization. In this paper, we investigated whether the discussion of COVID-19-related news reflects political polarization on Reddit, a popular online discussion platform. The topics of discussion in political subreddits are polarized on ideological lines: Trump, White House, and economic relief topics are statistically more prominent in liberal subreddits, and China and deaths topics are more prominent in conservative subreddits. We found significant differences between liberal and conservative subreddits in their preferences for news sources. Liberal subreddits share and discuss more credible news sources than conservative subreddits. Conservative subreddits are more likely than liberal subreddits to share articles from sites flagged for publishing COVID-19 misinformation.

COVID-19 misinformation has been proliferating at alarming rates on social media (Allington et al., 2021). Our study shows that the misinformation is largely shared in conservative subreddits on Reddit. Social media sites should intervene to algorithmically promote credible news sources in relation to health information.




Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Wallace Chipidza  <https://orcid.org/0000-0002-3980-4861>
 Christopher Krewson  <https://orcid.org/0000-0002-1286-1778>
 Tendai Gwanzura  <https://orcid.org/0000-0002-2110-3530>

Notes

- <https://www.redditinc.com/>
- The list of political and neutral subreddits chosen in this study is not exhaustive. Political subreddits were included by the

following criteria: if they were in the top six by both number of subscribers and number of coronavirus or COVID-19 posts. We applied these criteria so as to equalize the conservative and liberal subreddits and to exclude subreddits with low counts of COVID-19 related posts. Although other coronavirus themed subreddits exist (e.g., r/COVID19 and r/COVID19_support), we included r/coronavirus as representative as it is the oldest (created in May 2013) and the largest subscribed of subreddits dedicated to coronaviruses and the COVID-19 pandemic.

3. <https://www.newsguardtech.com/>
4. <https://www.newsguardtech.com/coronavirus-misinformation-tracking-center/>, retrieved 10/4/20. (accessed 4 October 2020)>
5. <https://www.newsguardtech.com/covid-19-myths/>, retrieved 10/4/20. (accessed 4 October 2020)
6. <http://redditlist.com/>
7. https://praw.readthedocs.io/en/v3.6.2/pages/getting_started.html

References

- Allcott H and Gentzkow M (2017) Social media and fake news in the 2016 election. *Journal of Economic Perspectives* 31(2): 211–236.
- Allington D, Duffy B, Wessely S, et al. (2021) Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency. *Psychological Medicine* 51(10): 1763–1769. DOI: 10.1017/S003329172000224X.
- Andersen K, de Vreese CH and Albæk E (2016) Measuring media diet in a high-choice environment-testing the list-frequency technique. *Communication Methods and Measures* 10(2–3): 81–98.
- Au H, Bright J and Howard PN (2020) *Social media misinformation about the WHO*.
- Baum MA (2011) Red state, blue state, flu state: Media self-selection and partisan gaps in swine flu vaccinations. *Journal of Health Politics, Policy and Law* 36(6): 1021–1059.
- Campante FR and Hojman DA (2013) Media and polarization. *Journal of Public Economics* 100: 79–92.
- Centivany A and Glushko B (2016) ‘Popcorn tastes good’: Participatory policymaking and Reddit’s Amageddon. In: *Proceedings of the SIGCHI conference on human factors in computing systems* New York, NY, USA, 7 May 2016, pp. 1126–1137. CHI ’16. Association for Computing Machinery. DOI: .
- Chandelier M, Steuckardt A, Mathevet R, et al. (2018) Content analysis of newspaper coverage of wolf recolonization in France using structural topic modeling. *Biological Conservation* 220: 254–261.
- Dalmer NK (2017) Questioning reliability assessments of health information on social media. *Journal of the Medical Library Association* 105(1): 61–68.
- Gaudette T, Scrivens R, Davies G, et al. (2021) Upvoting extremism: Collective identity formation and the extreme right on Reddit. *New Media and Society* 23(12): 3491–3508. DOI: 10.1177/1461444820958123
- Greer SL, Bekker M, De Leeuw E, et al. (2017) Policy, politics and public health. *European Journal of Public Health* 27-(suppl_4): 40–43.
- Haupt MR, Li J and Mackey TK (2021) Identifying and characterizing scientific authority-related misinformation discourse about hydroxychloroquine on twitter using unsupervised machine learning. *Big Data and Society* 8(1): 20539517211013844.
- Massanari A (2017) #Gamergate and The Fappening: How Reddit’s algorithm, governance, and culture support toxic technocultures. *New Media and Society* 19(3): 329–346.
- Mensio M and Alani H (2019) News source credibility in the eyes of different assessors.
- Metzger MJ, Hartsell EH and Flanagin AJ (2020) Cognitive dissonance or credibility? A comparison of two theoretical explanations for selective exposure to partisan news. *Communication Research* 47(1): 3–28.
- Mills RA (2018) Pop-up political advocacy communities on reddit.com: SandersForPresident and The Donald. *AI & Society* 33(1): 39–54.
- Mimno D, Wallach HM, Talley E, et al. (2011) Optimizing semantic coherence in topic models. In: *Proceedings of the conference on empirical methods in natural language processing*, USA, 27 July 2011, pp. 262–272. EMNLP ’11. Association for Computational Linguistics.
- Pew Research Center (2019) Key findings about the online news landscape in America. Available at: <https://www.pewresearch.org/fact-tank/2019/09/11/key-findings-about-the-online-news-landscape-in-america/> (accessed 27 February 2020).
- Proferes N, Jones N, Gilbert S, et al. (2021) Studying reddit: A systematic overview of disciplines, approaches, methods, and ethics. *Social Media + Society* 7(2): 205630512110190.
- R Development Core Team RC (2010) R: A language and environment for statistical computing.
- Robards B (2018) Belonging and Neo-Tribalism on Social Media Site Reddit. In: Hardy A, Bennett A, and Robards B (eds) *Neo-Tribes: Consumption, Leisure and Tourism*. Cham: Springer International Publishing, pp. 187–206. DOI: 10.1007/978-3-319-68207-5_12
- Roberts ME, Stewart BM and Tingley D (2014) Stm: R package for structural topic models. *Journal of Statistical Software* 10(2): 1–40.
- Shearer E and Matsa KE (2018) News use across social media platforms 2018. Available at: <https://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/> (accessed 19 October 2020).
- Soliman A, Hafer J and Lemmerich F (2019) A characterization of political communities on Reddit. In: *Proceedings of the 30th ACM Conference on Hypertext and Social Media*, New York, NY, USA, 12 September 2019, pp. 259–263. HT ’19. Association for Computing Machinery. DOI: 10.1145/3342220.3343662.
- Statista (2021) Reddit: traffic by country. Available at: <https://www.statista.com/statistics/325144/reddit-global-active-user-distribution/> (accessed 24 August 2021).
- Tasnim S, Hossain MM and Mazumder H (2020) Impact of rumors and misinformation on COVID-19 in social media. *Journal of Preventive Medicine and Public Health* 53(3): 171–174.
- Trachtman S (2019) Polarization, participation, and premiums: How political behavior helps explain where the ACA works, and where It doesn’t. *Journal of Health Politics, Policy and Law* 44(6): 855–884.
- World Health Organization (2021) Drugs to prevent COVID-19: A WHO living guideline. Available at: <https://app>

- magicapp.org/#/guideline/L6RxYL/section/jM1N5j (accessed 21 July 2021).
- Young DG and Anderson K (2017) Media diet homogeneity in a fragmented media landscape. *Atlantic Journal of Communication* 25(1): 33–47.
- Yu S, Martino GDS and Nakov P (2019) Experiments in detecting persuasion techniques in the news. *arXiv preprint arXiv:1911.06815*.
- Zarocostas J (2020) How to fight an infodemic. *The Lancet* 395-(10225): 676.