
Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

International Journal of Information Management 61 (2021) 102397

Contents lists available at ScienceDirect

International Journal of Information Management

journal homepage: www.elsevier.com/locate/ijinfomgt

Research Article

# The effect of toxicity on COVID-19 news network formation in political subcommunities on Reddit: An affiliation network approach

Wallace Chipidza

*Center for Information Systems and Technology, Claremont Graduate University, Claremont, CA 91711, USA*

ARTICLE INFO

*Keywords:*
Social media
Affiliation networks
Political polarization
Exponential random graph modeling
Toxicity
News sharing

ABSTRACT

Political polarization remains perhaps the "greatest barrier" to effective COVID-19 pandemic mitigation measures in the United States. Social media has been implicated in fueling this polarization. In this paper, we uncover the network of COVID-19 related news sources shared to 30 politically biased and 2 neutral subcommunities on Reddit. We find, using exponential random graph modeling, that news sources associated with highly toxic – "rude, disrespectful" – content are more likely to be shared across political subreddits. We also find homophily according to toxicity levels in the network of online news sources. Our findings suggest that news sources associated with high toxicity are rewarded with prominent positions in the resultant network. The toxicity in COVID-19 discussions may fuel political polarization by denigrating ideological opponents and politicizing responses to the COVID-19 pandemic, all to the detriment of mitigation measures. Public health practitioners should monitor toxicity in public online discussions to familiarize themselves with emerging political arguments that threaten adherence to public health crises management. We also recommend, based on our findings, that social media platforms algorithmically promote neutral and scientific news sources to reduce toxic discussion in subcommunities and encourage compliance with public health recommendations in the fight against COVID-19.

## 1. Introduction

The COVID-19 pandemic has wreaked much havoc on human health and well-being through loss of lives, enforced isolation, and devastated economies (Baker, Bloom, Davis, & Terry, 2020). Humans have no natural immunity to SARS-CoV-2, the coronavirus that causes COVID-19, and as yet, vaccines are in limited supply both in the United States and elsewhere (McClung et al., 2020). There are also no known effective pharmaceutical therapies for the disease (Grossman, Kim, Rexer, & Thirumurthy, 2020), meaning that preventative measures are paramount in curtailing loss of human life and other negative effects accompanying the pandemic. Although populations across the world have complied with public health recommendations to slow the disease's spread, there has also been significant resistance, especially in the US (Barrios & Hochberg, 2020). Recently, evidence is accumulating that intention to receive COVID-19 vaccines is politically polarized, with Democrats much more willing to receive it than Republicans (Fridman, Gershon, & Gneezy, 2021). This resistance has been fueled by political polarization and is consistent with past responses to the H1N1 pandemic (Lee & Basnyat, 2013).

The traditional media has been implicated in fueling political polarization, and social media intensifies it further (Halberstam & Knight, 2014). Increasingly, Americans report getting more of their news from social media rather than traditional media (Pew Research Center, 2019). Social media has surpassed search engines as the main disseminator of news on the internet (Boxer, 2015). The characteristics of news content influence its sharing; content that invokes strong emotions is more likely to be shared than content that invokes weaker emotions (Berger & Milkman, 2012). It is feasible that news sources which create emotionally provoking content are likely to be shared more often on social media platforms and thus occupy influential positions of the resulting network of news sources.

Understanding news sharing on social media is important, because users on the same social media platform may choose different information sources and are thus exposed to and ultimately consume different content (Halberstam & Knight, 2014). COVID-19 is a serious threat to societal well-being and warrants shared understanding of the nature of the threat and concerted action to prevent its spread. Ideally, qualified public health professionals should be conveying COVID-19 related information to the public (Regidor et al., 2007); but currently a significant proportion of society obtains health news from social media (Allington, Duffy, Wessely, Dhavan, & Rubin, 2020). This study

*E-mail address:* wallace.chipidza@cgu.edu.

https://doi.org/10.1016/j.ijinfomgt.2021.102397
Received 21 October 2020; Received in revised form 26 July 2021; Accepted 29 July 2021
Available online 9 August 2021

uncovers the social network of COVID-19 news sources as shared on the popular online community Reddit and investigates the factors that affect its formation. Our main research question is:

> RQ: What are the factors that affect tie formation in the network of news sources as it relates to COVID-19 related information sharing on Reddit?
> The main research question is further disambiguated into the following more specific research questions:
> RQ1: How does toxicity influence the number of ties formed by a news source in the network of COVID-19 news sources on Reddit?
> RQ2: How does similarity in toxicity influence tie formation between two news sources in the network of COVID-19 news sources on Reddit?

We build an affiliation network (also known as a bipartite or two-mode network) comprised of two entities: *news sources* shared to different *subreddits*. Analysis of the news sources network using exponential random graph modeling shows that news sources associated with highly toxic content are popular within the network, even after controlling for structural effects in the model. The results also show that news sources associated with similar levels of toxicity are more likely to be shared together within the same subreddits than those with dissimilar levels of toxicity. These findings suggest that toxic news titles – further editorialized by Reddit users or not – are a mechanism by which users build community in their subreddits. At the same time, the toxicity in discussions may also fuel political polarization by denigrating ideological opponents and politicizing responses to the COVID-19 pandemic, all to the detriment of mitigation measures. Based on our findings, we recommend that public health professionals create messaging strategies that take into account the political context, and in particular monitor toxicity of online discussions to understand when public health crises are becoming politicized.

This paper is organized as follows. We first explore relevant literature and derive hypotheses regarding the formation of the network of news sources on Reddit. Next, we outline our methods of data collection and analyses. We then report the results the study. Last, we discuss the results and theoretical and practical implications of our study before concluding.

## 2. Literature review

In this section, we first summarize the COVID-19 literature that appears in information systems journals. We proceed to a general overview of the literature on news sharing, pandemics, and political polarization. We conclude this section with a brief summary of the literature on network formation.

### 2.1. COVID-19 and IS research

Already, much conceptual work has been published on the COVID-19 pandemic in the information systems (IS) discipline. Researchers have identified opportunities for IS research to solve pandemic-related problems; for example, facilitating data sharing across nations to help with global management of COVID-19, building more resilient supply chain systems (Barnes, 2020), and digitalizing healthcare systems (Carroll & Conboy, 2020). Another emerging theme in the literature is the role played by information technologies in causing or exacerbating the problems faced by societies during the pandemic. Such problems include technostress (Venkatesh, 2020), loss of privacy (Leclercq-Vandelannoitte & Aroles, 2020), loss of control (Rowe, 2020; Rowe, Ngwenyama, & Richet, 2020), and reinforcing of unequal access to healthcare and education (Dwivedi et al., 2020). A third and related stream of research are the changes in human behavior caused by the pandemic and how they implicate information technologies. For example, citing earlier research, Venkatesh (2020) reasons that because technological changes have usually wrought negative impacts such as increased job stress and decreased job satisfaction, the shift to teleworking will likely negatively impact workers and motivate them to alter coping strategies. Table 1 summarizes the COVID-19 related IS literature.

**Table 1**
Summary of COVID-19 related literature in IS.

| Category of Literature | Description | Examples |
|---|---|---|
| Opportunities for design science | Information technology-oriented solutions to COVID-19 problems | Using IS to build resilient supply chain information systems (Barnes, 2020), designing "a balance between moderating and freewheeling discussion on social media" (Sein, 2020, p. 2), modeling dynamics of COVID-19 outbreaks (Pee, Pan, Wang, & Wu, 2021). |
| Problems of the pandemic | How information technologies contribute to the problems of the pandemic | Spread of COVID-19 misinformation on social media ( Laato, Islam, Islam, & Whelan, 2020), risks of contract tracing apps and other COVID-19 mitigation technologies to privacy (Leclercq-Vandelannoitte & Aroles, 2020; Rowe et al., 2020; Rowe, 2020), reinforcement of the digital divide in education (Iivari, Sharma, & Venta-Olkkonen, 2020) |
| Behavior changes after pandemic | How information technologies are implicated in behavioral changes during the pandemic | The impact of teleworking on job outcomes (Richter, 2020; Venkatesh, 2020), changes in sensitive information disclosure behavior on social media ( Nabity-Grover, Cheung, & Thatcher, 2020), the use of web conferencing tools to create "virtual togetherness" (Hacker, vom Brocke, Handali, Otto, & Schneider, 2020) |

Much of the early COVID-19 related work in IS has been conceptual. Comparatively little empirical work has been conducted in the realm of IS and COVID-19, e.g., using design science to create a COVID-19 outbreak analytics system (Pee et al., 2021), conducting experiments to determine the impact of convenience and privacy concerns on the acceptance of COVID-19 tracing apps (Trang, Trenz, Weiger, Tarafdar, & Cheung, 2020), and analyzing tweets to investigate the affordances and constraints of web conferencing tools (Hacker et al., 2020). This study is most related to the latter study to the extent that our study is empirical and employs user generated content to investigate the factors that contribute to the way COVID-19 information is disseminated and discussed across online communities.

### 2.2. News sharing, pandemics, and political polarization on social media

Previous research has investigated the factors motivating news sharing on social media. Individuals share news for status-building, socialization, self-expression, and entertainment purposes (Valenzuela, Piña, & Ramírez, 2017). Other individuals are motivated by emotions to share news articles, i.e. articles that induce strong emotions like anger, happiness, or awe are shared more often than average (Berger & Milkman, 2012; Stieglitz & Dang-Xuan, 2013). Social media users filter out some news content from direct news sources by their sharing choices, amplifying or excluding certain content (Bright, 2016). An important difference between our study and past research lies in our utilizing an anonymous platform to understand information sharing on the COVID-19 pandemic. We surmise that on such anonymous platforms, factors such as status- and influence-building weaken as motivators for information sharing. Rather, the characteristics of the content becomes

much more important in determining how it is shared and its ultimate position in the resulting network of news sources.

Preliminary research implicates social media in the spread of COVID-19 misinformation. Individuals that frequently check social media for COVID-19 news are more likely to hold conspiracy beliefs about the disease than those who do not; and individuals who believe conspiracies about COVID-19 are less likely to engage in health-protective behaviors (Allington et al., 2020). Social media has been castigated for facilitating the spread of COVID-19 related panic, fearmongering, and racism (Depoux et al., 2020). In China, social media exposure was found to exacerbate depression and anxiety during COVID-19 (Gao et al., 2020). Our study employs social affiliation network analysis to uncover the network of news sources shared on Reddit and in the process derive implications for public health.

User generated content, defined as "media content created or produced by the general public rather than by paid professionals and primarily distributed on the Internet" (Daugherty, Eastin, & Bright, 2008, p. 16), plays an important role in influencing political outcomes. Online subcommunities segregate along ideological lines, with political blogs linking preferentially to blogs sharing similar ideologies (Adamic & Glance, 2005). On Twitter, in the US, users are more likely to share posts from users with similar political ideology (Conover et al., 2011), and in Canada, users cluster according to political ideology (Gruzd & Roy, 2014). Social media also incentivizes political extremism; members of the US House of Representatives that express extreme ideological views attract more followers than their moderate counterparts (Hong & Kim, 2016). Political candidates employ social media in multifaceted and different ways when conducting their campaigns, to differential outcomes; for example, former US president Trump's use of Twitter may have helped him foster a more favorable image compared to his opponent in the 2016 US presidential election (Grover, Kar, Dwivedi, & Janssen, 2019). On the other hand, social media facilitates discussion among individuals with differing ideological viewpoints, although such conversations generally do not persuade ideological opponents, but may even devolve into hostility (Gruzd & Roy, 2014). These studies are helpful to the extent that they show the intersection of social media and politics; they also show that social media discourse affects real-world events. Our study intersects social media, politics, and public health. We investigate the link between a news source's toxicity and its influence in the news source's social network with the goal of understanding how toxicity in social media discussions may contribute to politicization of a health crisis.

### 2.3. Formation of social networks

Relative to studies that investigate the impact of social network position on various outcomes, social network formation is understudied (Bhattacharya, Phan, Bai, & Airoldi, 2019; Rose Kim, Howard, Cox Pahnke, & Boeker, 2016). Once the structure of a network is revealed, it is instructive to investigate the tendencies that the network exhibits. These tendencies can be structural, meaning that the configurations of the network's relationships exhibit certain patterns e.g. transitivity, reciprocity, and preferential attachment (Harris, 2013). The tendencies can also be demographic in nature, i.e., they originate outside the network and arise from the influence of node characteristics e.g., the average toxicity associated with a news source or its ideological bias. Our study utilizes affiliation networks to understand how toxicity influences tie formation in the network of COVID-19 news sources on Reddit.

## 3. Theoretical background and hypotheses development

In the case of COVID-19, the choice of which news outlet to get information from is likely influenced by its accuracy; after all, news coverage of a health or medical issue influences the specific actions taken by readers in response to the issue (Chang, 2015). The more accurate the health-related information, the more useful it is in preventing the spread of disease. But in the case of COVID-19, there is mounting evidence that the response to the pandemic is becoming polarized along ideological lines in the US (Adolph, Amano, Bang-Jensen, Fullman, & Wilkerson, 2021; Grossman et al., 2020). Much of the spread of COVID-19 news is facilitated by social media through sharing posts on individual and subcommunity pages. Because many subcommunities form with the intention to attract people of like-minded political persuasion (Bishop, 2014; Norris, 2002), such communities likely adopt homogeneous attitudes regarding the response to COVID-19. The content posted in these communities is meant for an audience of like-minded individuals; thus, content that riles up community members is likely to elicit more approval and engagement than mild content. One way to mobilize community members in political subcommunities is by posting toxic content, either verbatim from original sources or by editing the original content so that it becomes more toxic and invoke strong emotions in users. It is important to note that our hypotheses reference "the toxicity associated with a news source" rather than "the toxicity of a news source" because as we noted earlier, some links posted to Reddit are titled by Reddit users rather than taken verbatim from the original source.

### 3.1. The effect of news source title toxicity on tie formation

Toxicity of online content is defined as "the usage of rude, disrespectful, or unreasonable language that will likely provoke or make another user leave a discussion" (Jigsaw & Google, n.d.; Obadimu, Mead, Hussain, & Agarwal, 2019, p. 3). Much extant work on toxicity has focused on the harm it causes insofar as it targets other users with abuse, insults, and harassment within a given online discussion (Bishop, 2014; Hosseini, Kannan, Zhang, & Poovendran, 2017). The toxicity literature has also largely assumed that toxic content is aimed at specific or broad adversarial audiences (e.g. Almerekhi, Kwak, Jansen, & Salminen, 2019; Guberman, Schmitz, & Hemphill, 2016; Märtens, Shen, Iosup, & Kuipers, 2015). However, we theorize that in the case of specific subcommunities such as on Reddit where posts submitted by users are meant to initiate discussion (Soliman, Hafer, & Lemmerich, 2019), toxic content may not be directly aimed at participating adversaries in discussions but may aim to elicit collective outrage by like-minded individuals within the subcommunity. This is especially true when the toxic content is contained within a post's title, as compared to comments and is posted to a political subreddit. For example, a post titled "*What a pathetic little man: Trump says he knew coronavirus was a pandemic 'long before' it was declared*" and shared in a liberal subreddit is not meant to aggravate specific subscribers of that subreddit, but is likely meant to mobilize readers against the former US President's response to the pandemic. Toxic titles usually identify a target for derision (e.g. a political opponent), making them effective in mobilizing discussion around a political issue. A non-toxic title, such as "*Positive RT-PCR Test Results in Patients Recovered From Covid*" shared to any political subreddit is less likely to excite the emotions of readers, given that no target is explicitly identified.

Social capital theory, which postulates that collectivities with relatively large numbers of internal ties derive benefits from access to shared resources and cognition (Adler & Kwon, 2002), also adds further justification for our hypothesis. For contentious phenomena like pandemics that are politically implicated, individuals' ideological persuasions influence their news sources preferences. On social media platforms like Reddit, individuals seek to reinforce their world views (Massanari, 2017). Neutral headlines may not comport with the combative inclinations sought by subscribers to highly partisan subreddits. In contrast to traditional media, users are able to shape what they are exposed to on social media regarding political information and viewpoints, contributing to the "echo chamber" effect (Kim & Ellison, 2021). Toxic comments are therefore likely to attract new subscribers to subreddits to the extent that "content that can be degrading, abusive, or

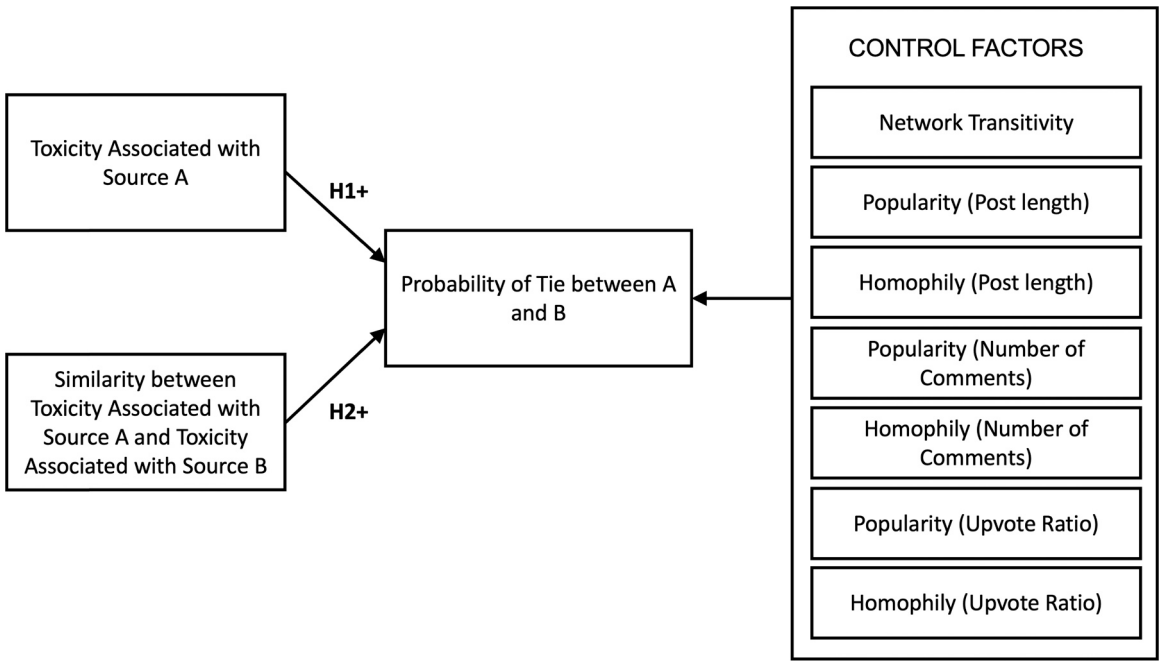


**Fig. 1.** Research model.

otherwise harmful [to political opponents]" (Aroyo, Dixon, Thain, Redfield, & Rosen, 2019, p. 1102) is useful for building social capital against opposing political ideologies.

Another important consideration is the anonymous nature of Reddit participation; anonymity is a key driver of toxicity on online platforms (Chen, Muddiman, Wilner, Pariser, & Stroud, 2019). Toxicity in the form of profanity draws attention to polarizing viewpoints and encourages political participation; both are components of community building as happens on political subreddits (Brooks & Geer, 2007). Previous research has found that incivility against political opponents to be an effective mobilization tool (Chen et al., 2019). Certainly, the purpose of toxicity in political messaging is multi-pronged; it is used to belittle or provoke opponents but is also used to signal to potential supporters that they hold morally or logically superior views to their opponents. News sources with toxic headlines are thus likely to be shared more within specific political subcommunities to the extent that they help mobilize support for political causes, leading to the following hypothesis:

**H1**. In Reddit's political subcommunities, a news source associated with high toxicity is shared more often with other news sources than a news source associated with low toxicity.

The principle of homophily, often described with the proverb "birds of a feather flock together," is the tendency for entities with similar characteristics to form ties at greater rates than random (McPherson & Smith-Lovin, 1987). Thus, in marriage networks, individuals of similar race, ethnicity, income, and educational status are more likely to marry than individuals of dissimilar statuses (McPherson, Smith-Lovin, & Cook, 2001). For our current context, the principle of homophily suggests that news sources associated with similar levels of toxicity are likely to be shared together more often than sources associated with different levels of toxicity. There are two theoretical explanations for why homophily will be expected in tie formation according to toxicity. The first explanation is social selection. According to social selection theory, individuals join online subcommunities to form connections based on shared interests, values, and attitudes (Goel, Mason, & Watts, 2010; Soliman et al., 2019). Online platforms bring together "like--minded communities that serve as echo chambers for their existing opinions" (Goel et al., 2010, p. 1); thus, the individuals subscribing to a given subreddit are likely to enjoy similar content. A dimension on which users might choose to select a given subreddit is therefore likely to be the toxicity of the content that users share. For example, a left-leaning user might intend to share her COVID-19 related content to one political subreddit but would need to choose among many candidate subreddits. Given a choice between one that has more neutral language and one that characterizes conservatives as logically or morally inferior to left-leaning voters, an individual is likely to be drawn to the more ideologically combative subreddit, especially in highly politically polarized contexts (Chen et al., 2019).

The second theoretical explanation is social influence, i.e., when over time the toxicity of submitted posts become similar to each other (Luke, 2015). The content shared in political subcommunities elicit engagement which, when sustained, maintains group cohesion and enables groups to attract new members. In certain subcommunities, civility is enforced by moderation (Lampe & Resnick, 2004). In others, there is much leeway as to the civility or lack thereof allowed. Across levels of content moderation, group norms (e.g., how to title posts, lengths of posts, and etiquette on welcoming new members) emerge over time that regulate the type of content that best engages the subcommunity (Wise, Hamman, & Thorson, 2006). As issues get heated over time, so might polarization, which suggests that over time people would need to ratchet up their levels of toxicity to galvanize the community. In addition to toxicity at the subreddit level, toxicity at the news source level might also implicate tie formation. Thus, sources that publish highly toxic content might be more likely to be shared in politically partisan subreddits than sources that do not publish toxic content. Therefore, individuals that enjoy content with high toxicity will tend to subscribe to subreddits that publish such content, and the opposite is true. It is likely that news sources associated with similar levels of toxicity will be shared together more often than news sources associated with different levels of toxicity. The second hypothesis:

H2: On Reddit, news sources associated with similar levels of toxicity will be shared together more often than news sources associated with different levels of toxicity.

Fig. 1 summarizes our research model.

## 4. Methodology

Reddit has emerged as an important social media platform over the

**Table 2**
Examples of Reddit post titles, their sources, and toxicity scores.

| Reddit post title | Editorialized by user? | Original Title if editorialized | Source | Toxicity Score |
|---|---|---|---|---|
| Georgia's Idiot Republican Governor Says He Didn't Know People Could Spread Covid Without Symptoms | No | | Gizmodo | 0.90 |
| Trump ignorant son-in-law is running the corona response. That unacceptable. - Trump Admin is Nothing but a Ship of Criminal Fools. | Yes | Trump's ignorant son-in-law is running the coronavirus response. That's unacceptable. | Washington Post | 0.84 |
| Shithole Company YouTube Takes Down Bio-Tech Firm's Video Explaining Potential UV Light Treatment for corona Patients | Yes | YouTube Takes Down Bio-Tech Firm's Video Explaining Potential UV Light Treatment for Coronavirus Patients | Gateway Pundit | 0.71 |
| Democrats cheering 'Black Lives Matter' protests now say Trump rallies pose corona risk | No | | Fox News | 0.52 |
| The majority of Covid infections may be attributable to silent transmission from presymptomatic and asymptomatic cases. Symptom-based isolation must be supplemented by rapid contact tracing and testing that identifies asymptomatic and presymptomatic cases, to safely lift current restrictions. | Yes | The implications of silent transmission for the control of COVID-19 outbreaks | Proceedings of the National Academy of Sciences | 0.01 |

past decade. It is heavily trafficked, with over 430 monthly active users making it the 7th most trafficked website in the US (even above Twitter) (Backlinko, 2021). Users participate anonymously by subscribing to subreddits – subcommunities devoted to specific interests –, submitting links, writing posts, commenting on posts, and expressing approval (by upvoting) and disapproval (by downvoting) on submitted posts (Soliman et al., 2019). Reddit also allows for searching posts by a given keyword e.g., "COVID". These affordances of Reddit make it useful for investigating this study's research questions; to the extent that our study develops a priori hypotheses and tests them while considering the (i) characteristics of the technology platform and (ii) the sociopolitical context of COVID-19, it fulfills the central tenets articulated in Kar and Dwivedi (2020).

We investigate our research questions by examining overlaps in shared news sources across a collection of 32 subreddits, 11 conservative (a combined 1.1 million subscribers), 19 liberal (9.2 million subscribers), and 2 neutral (26.7 million subscribers). We used the categorization from Soliman et al. (2019) and subreddit descriptions to determine the ideological bias of a given subreddit: "conservative" for right-leaning subreddits, "liberal" for left-leaning sites, and "neutral" for the r/coronavirus and r/science subreddits. The aggregated distribution of the COVID-19 related posts was 2601 (33%) conservative, 4028 (52%) liberal, and 1179 (15%) neutral from a total 1180 unique news sources. We collected the data using the Reddit PRAW API (Boe, 2016), with search terms "covid," "corona," and "coronavirus." The earliest returned post was on January 21, 2020 and the latest post was on July 10, 2020. The PRAW API has been used for the same purposes in this study of collecting Reddit posts, comments, upvote and downvote counts, and other metadata in other highly cited studies (e.g., Buntain & Golbeck, 2014; Olson & DeFrain, 2003; Park & Conway, 2018).

### 4.1. Building the affiliation network

The affiliation network had two types of entities: news source and subreddit. Thus, there was a tie between a news source *ns* and subreddit *sr* for every post shared from *ns* to *sr*. The affiliation network was weighted, meaning that the final tie had weight *w* if source *ns* was posted *w* times to subreddit *sr*. For example, the weight of the Fox News -> r/Conservative tie was 30, while the equivalent for the Business Insider to r/politics tie was 43. We created two one-mode networks by projecting the affiliation network to the composite one-mode networks (the news sources network and the subreddit network).

### 4.2. Measuring the effect of news source title toxicity on tie formation

To investigate hypotheses H1 and H2, we employed exponential random graph modeling (ERGM) on the one-mode projection of the news sources network. ERGM is a statistical inference method for predicting tie formation in social networks. ERGM estimates parameters for structural (e.g., transitivity, reciprocity, and degree distribution) and node attributes (e.g., variations in tie formation, attribute similarity, and attribute mixing) effects on network formation and returns conditional log-odds that can be converted to the probability of tie formation given node attributes (see Hunter, Handcock, Butts, Goodreau, and Morris (2008) for more details). Networks violate the independence of observation assumption that should be met for techniques such as logistic and linear regression. ERGM does not rely on that assumption, and in fact can control for the dependence that nodes have on each other through their relationships. Mathematically, an ERGM is specified as follows.

$$P(Y = y) = \frac{exp\{\theta^{T} z(y)\}}{\kappa(\theta, \quad Y)}, \quad y \in Y \tag{1}$$

The left side of the equation is the probability that a given network *y* is observed given all the possible networks *Y* of the same size or number of nodes. On the right side of the equation, the numerator is an exponentiation of the sum of the products of each parameter $\theta$ and each model effect z. The denominator is the sum of the numerator across all possible networks of the same size. The goal of ERGM is to calculate the vector of parameters $\theta^{T}$ that maximizes the probability of the observed network. Parameter estimation uses MCMC sampling to estimate a distribution of all possible networks. The parameters that ERGM calculates make it possible to estimate the excess effect of a given structural or node attribute on the probability of a tie between two nodes.

For each article posted containing the term "corona" or "covid" in the political subreddits, we retrieved the title. The titles we used were either verbatim from the source websites or edited by users. We calculated the toxicity score of each title using Google's Perspective API; the API uses machine learning to derive a toxicity score for a piece of text content (Jigsaw & Google, n.d.). The algorithm used to assign a toxicity score on a given piece of text relies on binary text classification based on the features (i.e., the character n-grams) of the text (Wulczyn, Thain, & Dixon, 2017). The API creators initially leveraged crowdsourcing by human evaluators to score text in the corpus of discussion comments on English Wikipedia; they proceeded to build the binary text classifier based on the toxicity scores assigned by human evaluators. Thus, the toxicity score for each post in our corpus is a value assigned by querying the Perspective API. The score varies from 0 to 1, with higher scores indicating higher toxicity. This API is used by organizations such as Wikipedia and The New York Times to enforce civility and prevent abuse in online discussions (Google, 2018; Hosseini et al., 2017). The API has also been exposed to adversarial training (Hosseini et al., 2017) and has been used before for the purpose of assigning toxicity scores to online content in academic research as we do in this study (e.g., Obadimu et al., 2019). Table 2 shows examples of titles posted to Reddit and the toxicity

**Table 3**
Mathematical and statnet specifications of tested network formation models.

| Model | Mathematical specification | Statnet specification |
|---|---|---|
| Null | $Pr(X = sources\_net\|\theta) = \frac{1}{\kappa(\theta)}(\theta_0 * edges)$ | $M_0 \leftarrow ergm(sources\_net \sim edges)$ |
| Node Attributes Only | $Pr(X = sources\_net\|\theta) = \frac{1}{\kappa(\theta)}(\theta_0 * edges + \theta_1 * absdiff('Toxicity') + \theta_2 * nodecov('Toxicity') + \theta_3 * absdiff('Upvote\_Ratio') + \theta_4 * nodecov('Upvote\_Ratio') + \theta_5 * absdiff('Number\_of\_Comments') + \theta_6 * nodecov('Number\_of\_Comments') + \theta_7 * absdiff('Word\_Count') + \theta_8 * nodecov('Word\_Count'))$ | $M_1 \leftarrow ergm(sources\_net \sim edges + absdiff('Toxicity') + nodecov('Toxicity') + absdiff('Upvote\_Ratio') + nodecov('Upvote\_Ratio') + absdiff('Number\_of\_Comments') + nodecov('Number\_of\_Comments') + absdiff('Word\_Count') + nodecov('Word\_Count'))$ |
| Structural and Node Attributes Model | $Pr(X = sources\_net\|\theta) = \frac{1}{\kappa(\theta)}(\theta_0 * edges + \theta_1 * absdiff('Toxicity') + \theta_2 * nodecov('Toxicity') + \theta_3 * absdiff('Upvote\_Ratio') + \theta_4 * nodecov('Upvote\_Ratio') + \theta_5 * absdiff('Number\_of\_Comments') + \theta_6 * nodecov('Number\_of\_Comments') + \theta_7 * absdiff('Word\_Count') + \theta_8 * nodecov('Word\_Count') + \theta_9 * gwesp(0.2, T)))$ | $M_2 \leftarrow ergm(sources\_net.filt \sim edges + gwesp(0.2, T) + absdiff('Toxicity') + nodecov('Toxicity') + absdiff('Upvote\_Ratio') + nodecov('Upvote\_Ratio') + absdiff('Number\_of\_Comments') + nodecov('Number\_of\_Comments'))$ |

**Table 4**
Descriptive statistics for node covariates in tie formation model.

| | Mean | Standard deviation | Minimum | Maximum |
|---|---|---|---|---|
| **Toxicity** | 0.17 | 0.11 | 0.00 | 0.93 |
| **Upvote_Ratio** | 0.25 | 0.40 | 0.00 | 1.00 |
| **Number_of_Comments** | 284.63 | 645.96 | 0.00 | 6143.00 |
| **Word_Count** | 17.96 | 9.61 | 2.00 | 56.33 |

scores calculated by Perspective API.

We calculated an average toxicity score for each news source in our data set. To calculate the homophily effect of toxicity on the probability of a tie between two nodes, we used the *absdiff* ERGM term which captures how, for a given attribute, the absolute difference between the nodes' values varies with tie formation. When *absdiff* is added to a model, the effect is to add one network statistic that equals the sum of the absolute differences between the average toxicity scores of every pair of directly connected news sources (Handcock, Hunter, Butts, Goodreau, & Morris, 2008); lower values of the absolute difference between the toxicity scores of two news sources indicate homophily. The popularity effect of the toxicity term was calculated using the *nodecov* term. The intercept was measured using the *edges* term; this term returns a parameter that is the conditional log-odds of the density of the network. We also control for (i) transitivity in the network (a structural term) using the geometrically weighted edgewise shared partners (GWESP) term, (ii) the approval of a news source's articles, as measured by ratio of upvotes (iii) the engagement of a news source's articles, as measured by the number of comments and (iv) the average length of the source's articles' titles. For the ERGM analysis, we dichotomized the ties between any two news sources into 1 and 0 based on whether the tie weight exceeded 4 in the network; we chose the number 4 in a data-driven manner to ensure model convergence as recommended in the network formation literature (see Snijders, Van de Bunt, & Steglich, 2010).

To aid with reproducibility of the research, we include in Table 3 the specifications of the tested models as expressed mathematically and in *R*.

Table 4 summarizes the statistics of the node covariates in the model.

## 5. Results

### 5.1. Description of the affiliation network

Fig. 2 is a plot of the affiliation network of the top 90 news sources and 13 subreddits (5 conservative, 6 liberal, and 2 neutral). The plot shows conservative and liberal subreddits occupying different regions of the network and generally sharing different news sources. While subscribers to liberal subreddits seem to share COVID-19 news from mainstream outlets like The Washington Post, Reuters, and NBC, conservative subreddits seem to prefer news from right-wing sites such as RedState, National Review, and Breitbart. Notably, whereas two conservative subreddits – r/Libertarian and r/redacted seem to span the boundary and share news from mainstream and/or left-leaning news sources, no liberal subreddits seem to do the same. The neutral r/coronavirus subreddit occupies a similar region of the network with liberal subreddits, sharing news from mainstream sites such as BBC, The Washington Post, and Reuters. The other neutral subreddit, r/science, shares news from scientific journals and institutions like the CDC and NIH.

On average, toxicity of submitted COVID-19 posts was higher in conservative than liberal (+0.03 points on the toxicity scale, $p = .000$) and neutral (+0.09, $p = .0000$) subreddits. Fig. 3 summarizes how toxicity of COVID-19 discussions varied across subreddits.

### 5.2. Characteristics of news sources network

The network has 1180 unique news sources, a density of 0.22, and a clustering coefficient of 0.65.

#### 5.2.1. Clusters of the news sources network

Using the fast-greedy algorithm to detect clusters within the network of news sources, we found five clusters. A cluster, in network vocabulary often termed a community, is a subgroup of the network with a relatively large number of internal ties and relatively few ties to other subgroups (Luke, 2015). A measure for the goodness of a clustering is *modularity* – the extent to which nodes within a given cluster have dense ties within the subgroup and few ties to outside groups (Newman, 2006). The modularity of the clustering was reasonably high at 0.36. We assigned descriptive labels to the clusters as follows: *mainstream, scientific, international, right-wing, and left-wing*. The right-wing cluster of news sources exhibited the highest toxicity in its coverage of COVID-19 ($\beta = 0.03$, $p = 0.03$), followed by the mainstream, international, and left-wing sources (no significant differences among them). The scientific news cluster exhibited the lowest average toxicity ($\beta = -0.07$, $p = 0.00$). Table 5 summarizes the statistics of the clusters.

#### 5.2.2. The relationship between source news title toxicity and tie formation

We first built a baseline model with just the *edges* term to capture the propensity of tie formation in the network. The baseline model is the equivalent of the null hypothesis, i.e. the hypothesis that the network arises as a result of ties being formed randomly. The negative parameter for the edges term shown in Table 6, column 2 means that tie formation in the network is less likely than would be expected from chance, contingent on the number of nodes in the network. The parameter can be transformed to a baseline probability of tie formation similar to how parameters for logistic regression are calculated. In this case, the baseline probability for a tie between any two news sources in the network is

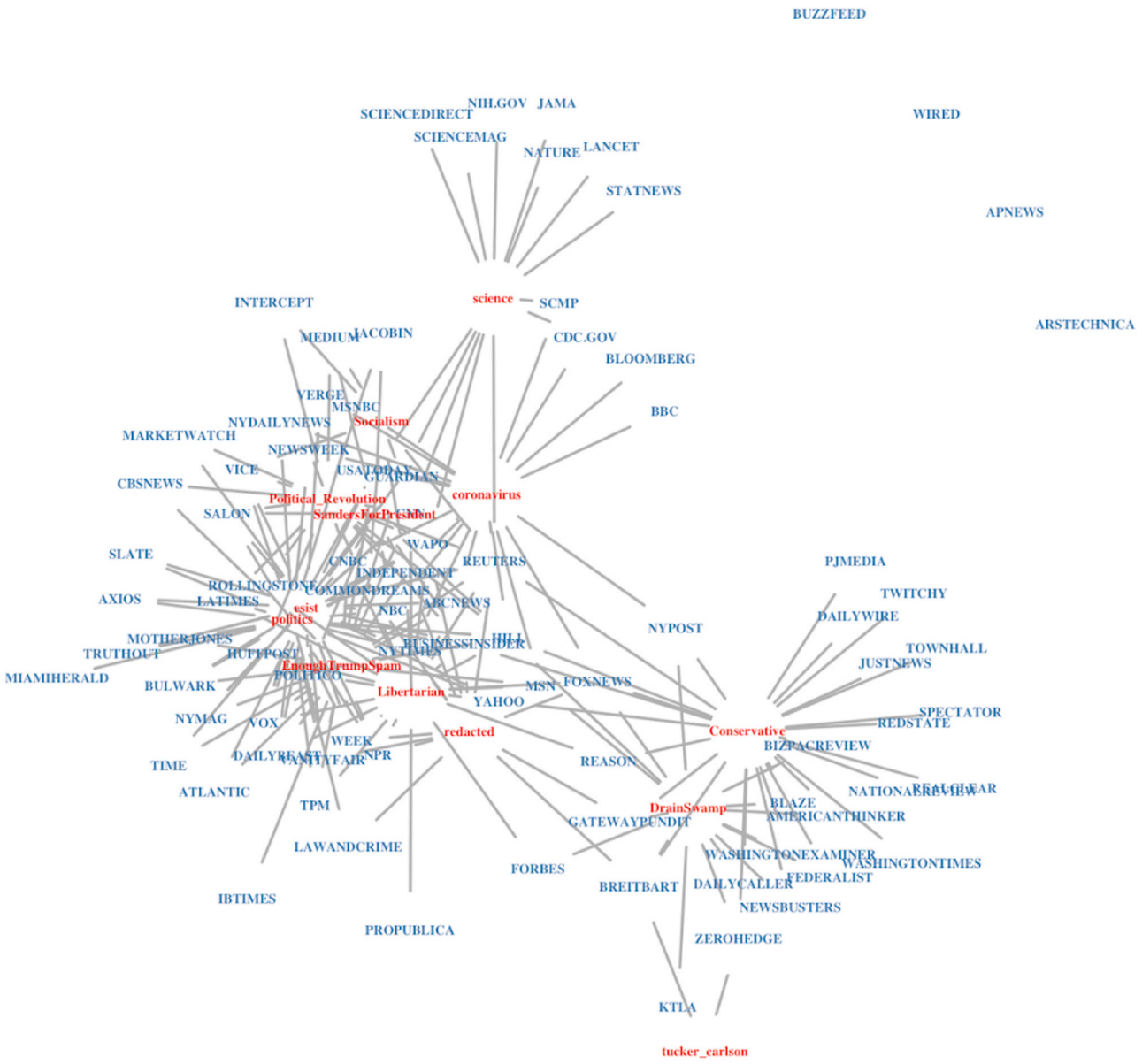


**Fig. 2.** Affiliation network of the top 90 news sources and 13 subreddits.

0.35.

We hypothesized that tie formation between any pair of news sources in the social network is a result of the individual sources' toxicity, approval (measured by upvotes), and engagement (measured by comments). We also control for the average title length for each news source (measured by word count). We add the homophily and covariate effects of these four factors to the baseline model to form the node attributes only model (Table 6, column 3). This model is superior to the baseline model as shown by the decrease in the Akaike Information Criterion (AIC) from 9582.49 to 7789.50 at the inclusion of the new terms. The third model (Table 6, column 4), which also includes the transitivity effect as measured by the geometrically weighted edgewise shared partners (GWESP) outperforms the null and node attributes-only models, further decreasing the AIC to 7767.06. Further, the structural and node attributes model has the highest log likelihood among the three models. Results show homophily effects according to toxicity, approval, engagement, and title length, supporting hypothesis H1. Thus, the change expected in the log-odds of a tie between two news sources decreases by 11.08 with a unit increase in the *difference* between their toxicity levels, by 1.58 with a unit increase in the difference between the upvote ratios, by 0.002 with a unit increase in the difference between the number of comments, and by 0.07 with a unit increase in the difference between the average title lengths of the news sources. The change in the log-odds that a news source will form a tie with another news source in the network increase by 5.03 with a unit increase in its toxicity, supporting hypothesis H2, and by 0.67, 0.001, and 0.04 with unit increases in the upvote ratio, number of comments, and title length respectively (i.e., news sources with longer titles are more likely to form ties than sources with short titles). The positive parameter for GWESP shows significant transitivity within the network. The results are detailed in Table 6.

There are two ways in which we demonstrate robustness for the parameter estimates of the homophily and covariate effects of toxicity on tie formation in the network of news sources. First, the parameters for the homophily and covariate effects of toxicity on tie formation are relatively constant under the node attributes only model and the structural and node attributes model, indicating robust effects. In fact, controlling for transitivity significantly reduces the standard errors of the effects. Second, we considered the model under a lower threshold for tie formation i.e., at a threshold of 3, meaning that the network

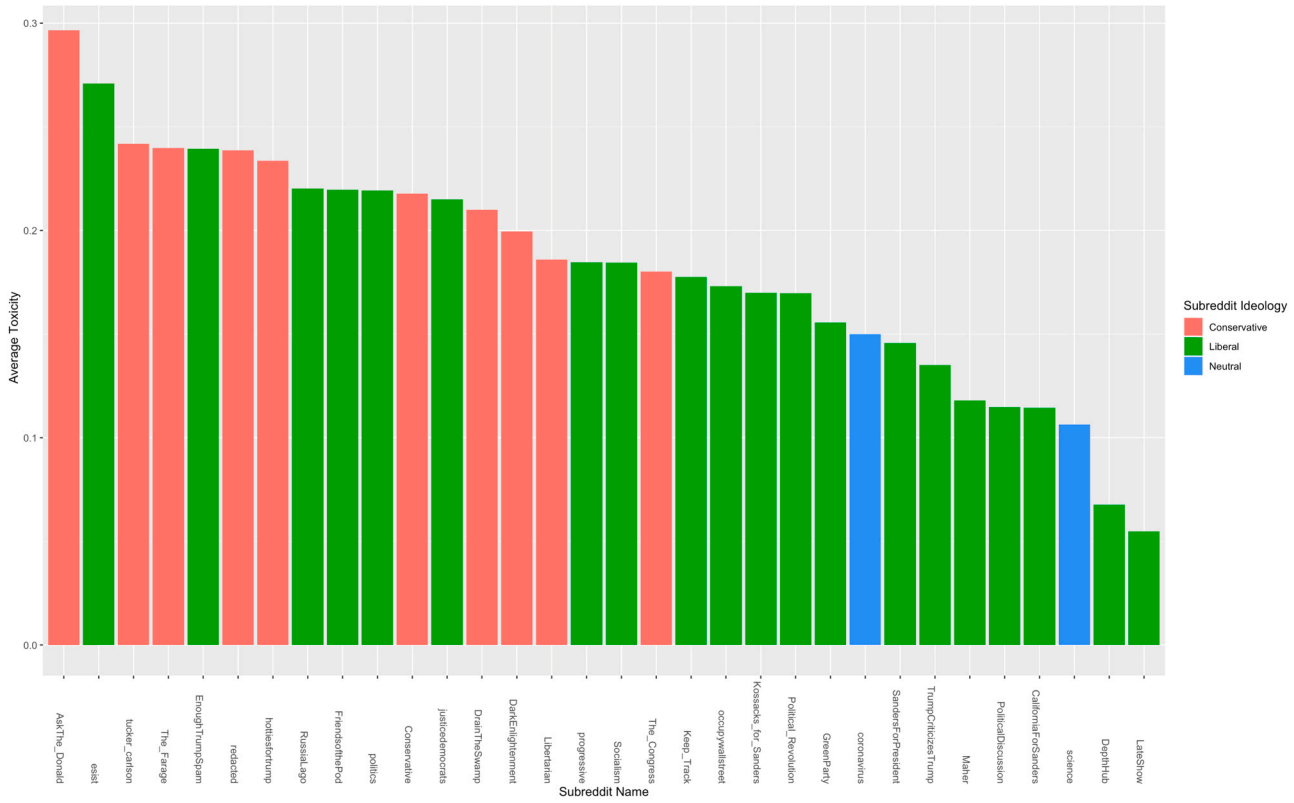


Figure 3: Average toxicity per post submitted to a given subreddit

**Fig. 3.** Average toxicity per post submitted to a given subreddit.

**Table 5**
Labels, descriptive statistics, and examples of clusters in network of news sources.

| Cluster Label | Number of News Sources | Density | Average Count (sd) | Average Toxicity (sd) | Examples of News Sources in Cluster |
|---|---|---|---|---|---|
| Mainstream | 429 | 0.38 | 14.24 (94.05) | 0.18 (0.11) | nytimes.com, bbc.com, reuters.com |
| International | 264 | 1.00 | 1.40 (0.97) | 0.15 (0.10) | timesofindia.com, arabnews.com, newsinenglish.no |
| Right-wing | 230 | 0.96 | 3.35 (8.09) | 0.20 (0.12) | foxnews.com, gatewaypundit.com, townhall.com |
| Scientific | 144 | 0.50 | 2.54 (4.18) | 0.10 (0.06) | jamanetwork.com, thelancet.com, nature.com |
| Left-wing | 113 | 0.65 | 1.49 (1.10) | 0.17 (0.09) | greenparty.ca, change.org, blackagendareport.com |

contained some relatively weak ties. Although the homophily and covariate effects of toxicity were weaker in this network, they were still significant and positive (Table 6, column 5). These findings held with a threshold of 5 as well i.e., the network contained relatively strong ties (Table 6, column 6). These findings, combined, indicate that the homophily and covariate effects of toxicity are robust under different assumptions of tie strength.

### 5.3. Goodness of fit

We employed the mcmc.diagnostics function in R to assess our models' goodness of fit. The MCMC process converged correctly for the parameters estimated by ERGM in modeling the formation of the network of news sources, as shown by the diagnostic plots in Fig. A1 in the Appendix. Further, the simulated networks' statistics are not different from the observed network's statistics indicating that the models have good fit for the observed networks (Table A1 in the appendix).

## 6. Discussion

This study provides contributions to three areas of social networks: affiliation networks, formation of networks, and the nature of discussion of COVID-19 on social media. We employ techniques from network formation modeling to uncover a hidden but consequential network of COVID-19 information sources as revealed by news sharing patterns on Reddit.

Our main contribution is identifying homophily and covariate effects of toxicity on tie formation in the network of COVID-19 related news sources on Reddit. These findings support our hypotheses. Our findings are remarkable to the extent that they show that the underlying news sources' network does not arise randomly, but that influences from human activity help structure the network. In particular, although Reddit users may not be aware, by deciding to submit content from a given news source, they help build the prominence of the news source in determining COVID-19 discussion on the platform. Further, by editorializing the titles of news post, they may also help increase or decrease the prominence of the news source in the attendant COVID-19 discussions. Our findings suggest that editorializing to make the titles more toxic would benefit the associated news source, and the reverse is true. In the next section, we summarize the theoretical contributions from this study and its attendant implications.

**Table 6**
Results of fitting structural and node attribute model to observed news sources social network.

| ERGM Term | Null Model | Node Attributes Only Model (Threshold = 4) | Structural and Node Attributes Model (Threshold = 4) | Node Attributes Only Model (Threshold = 3) | Node Attributes Only Model (Threshold = 5) |
|---|---|---|---|---|---|
| | Parameter (Standard Error) | | | | |
| edges | -0.61 (.02)*** | -4.23 (0.34)*** | -10.14 (0.12)*** | -3.32 (0.17)*** | -2.62 (0.55)*** |
| gwesp | | | 4.98 (0.14)*** | | |
| absdiff.Toxicity | | -11.08 (0.59)*** | -11.12 (0.03)*** | -7.25 (0.36)*** | -14.10 (0.91)*** |
| nodecov. Toxicity | | 5.03 (0.35)*** | 5.06 (0.05)*** | 3.21 (0.22)*** | 4.37 (0.52)*** |
| absdiff.Upvotes | | -1.58 (0.18)*** | -1.83 (0.16)*** | -1.09 (0.09)*** | -2.82 (0.34)*** |
| nodecov. Upvotes | | 0.67 (0.14)*** | 0.58 (0.12)*** | 0.53 (0.06)*** | -0.13 (0.26) |
| absdiff. Comments | | -0.00 (0.00)*** | -0.00 (0.00)*** | -0.00 (0.00)*** | -0.00 (0.00)*** |
| nodecov. Comments | | 0.00 (0.00)*** | 0.00 (0.00)*** | 0.00 (0.00)*** | 0.00 (0.00)*** |
| absdiff. Word_Count | | -0.07 (0.01)*** | -0.07 (0.01)*** | -0.07 (0.00)*** | -0.04 (0.01)*** |
| nodecov. Word_Count | | 0.04 (0.01)*** | 0.04 (0.01)*** | 0.04 (0.00)*** | 0.03 (0.01)*** |
| AIC | 9582.49 | 7789.5 | 7767.06 | 17,905.25 (AIC for null model = 20,939) | 4525.86 (AIC for null model = 20,939) |
| Log Likelihood | -4790.25 | -3885.75 | -3873.53 | 8943.63 (null model = −10468.5) | -2253.93 (null model = −2785.73) |

* indicates $p < 0.05$, ** indicates $p < 0.01$, *** indicates $p < 0.001$

### 6.1. Theoretical contributions and implications

#### 6.1.1. Affiliation networks and tie formation

We contribute by demonstrating the utility of affiliation networks in revealing hidden but consequential networks on social media (see Borgatti & Halgin, 2011; Kane & Alavi, 2008). Much work has explored networks of users or networks of subcommunities on Reddit; our work combined these aspects to the extent that an affiliation network is composed of two modes of networks: the *news sources network* which is a result of users submitting external content to a *network of specific subreddits*. To that end, affiliation networks simplify the study of a complex phenomenon that intersects users, information sources, a characteristic of submitted content i.e., toxicity, a selected topic (COVID-19), and political polarization. Here the affordances of Reddit were helpful. The political subreddits for the most part have defined ideological leanings in their descriptions, and Reddit allows subscribers to submit news articles *and* change the article headlines, facilitating the injection of user bias into coverage of an issue. Our technique can be applied to other social media platforms where users can create and join subcommunities, e.g., Facebook, LinkedIn, and Voat. Affiliation networks are also useful beyond groups; social media platforms with any kind of event management capability (e.g., Strava, Meetup, Facebook, etc.) can also be analyzed using an affiliation network approach to the extent that the event would be one type of entity and individuals attending that event would be another type of entity. Other entities like hashtags on Twitter and Facebook, shows on Netflix and Hulu, and book purchases on Amazon and eBay are also amenable to affiliation network analyses depending on the research question. Our work adds even to the area of affiliation network analysis by considering a little-studied attribute of user generated content i.e., toxicity and how it influences tie formation in a one-mode projection of the network.

#### 6.1.2. The utility of toxicity in online discussions

Past research has highlighted the controversy as to whether toxic content in politics is a positive or negative phenomenon. On one hand, toxicity is negative to the extent that it attacks and belittles political opponents, thus distracting from substantive issues (Almerekhi et al., 2019; Guberman et al., 2016). On the other hand, toxic content can be a useful mobilizing tool for issues where the public has limited awareness (Chen et al., 2019). Our study illustrates both of these aspects relating to toxic content. News sources associated with highly toxic content are rewarded with high centrality in the resultant network, supporting the notion that toxicity may be a mobilizing tool for contentious political positions. At the same time, toxicity of content plays an instrumental role in determining the structure of news sources suggesting that it is a driver of the observed ideological polarization. Such polarization for an issue with life and death implications such as COVID-19 does not bode well for the effectiveness of public health messaging.

Extant research suggests that individuals are more likely to share positive news (e.g., technology and social welfare stories) than negative news (e.g., political and crime stories). The filtering is ostensibly meant to present "less of a 'mean world'" to social media users (Bright, 2016 p. 344), making positive news more likely to be shared than negative news (Valenzuela et al., 2017). Our study suggests that in a politicized context, the meaning of positive news perhaps translates to "news that fits one's viewpoint." The more toxic the news, the more influential its source in the network. Our findings add to the literature on the effect of news content characteristics on sharing by users on social media in politically polarized and anonymous platform contexts. In these contexts, toxicity is rewarded with influence in the resultant network.

Typically, toxic content is meant to abuse some opponent to ensure they capitulate in an online argument (Lampe & Resnick, 2004; Wise et al., 2006). Here, it is evident that toxic content shared to a political subreddit is meant to elicit approval and engagement from fellow subscribers, and indeed we find evidence that news outlets associated with high toxicity are rewarded with influential positions in the resulting network meaning that they are shared more often across political subreddits. Overall, subscribers to conservative subreddits post the most toxic submissions, which reflects dedicated politicization of COVID-19 on these subreddits relative to liberal and neutral subreddits. Health information presented in toxic form is likely to further politicize the issue, complicating effective public health intervention measures.

#### 6.1.3. The affordances of Reddit and the sociopolitical context of COVID-19

In the public health communication literature, best practices were generally created pre-social media and for newspapers and broadcast media (i.e., radio and television). Recently, the need to adapt to modern forms of communication has been recognized by public health professionals and researchers. Even then, social media platforms are treated

as a monolith in this regard. Our study focuses on Reddit in particular, and especially on the characteristics unique to the platform such as user anonymity and organization by subreddits; these characteristics might incentivize high toxicity in discussing COVID-19 and other politically fraught issues. Thus, we are not treating the platform as a black box, in congruency with Orlikowski and Iacono (2001) admonishment to highlight relevant characteristics of the artifact in IS research. The features of a given artifact coupled with the broader sociopolitical context – in our case political polarization – have bearing on the resulting network of information sources. Thus, we answer the "why" concerning COVID-19 news network formation on Reddit; such a holistic treatment of an IS phenomenon is consistent with recent theorizing guidelines suggested in the realm of big data (Kar & Dwivedi, 2020). Researchers of public health and technology may learn from our approach to further explore factors that influence health news exposure on social media.

In the realm of social media research, attention has been placed on changes in information disclosure behavior (Nabity-Grover et al., 2020). During the COVID-19 pandemic, social media use increased significantly, and disclosure behavior changed such that previously sensitive personal health information was more likely to be posted by individuals, and information that might show one to be in contravention of public health recommendations such as travel was less likely to be posted. The arguments laid out in Nabity-Grover et al. (2020) are salient for understanding how subscribers to political subreddits may also change their posting behavior. The lockdowns and other social distancing measures would have decreased social interactions for which social media may fill the void. The shift to interacting via social media platforms like Reddit would have intensified activity, leading to stronger affinities for political subreddits and strengthening of echo chambers.

#### 6.1.4. Toxicity and communities of news sources

From our clustering of news sources using community detection, we found that scientific sources were the least toxic by some distance. This finding adds to research in health communication which has largely focused on journalistic sources (see Duncan, 2009; Gholami, Hosseini, Ashoorkhani, & Majdzadeh, 2011; Lee & Basnyat, 2013). New media platforms such as Reddit offer avenues for scientific information to be disseminated to the public in palatable ways. Traditional media could learn from new media in this regard. Our findings suggest that scientific sources are careful to use non-incendiary language to inform readers. Even mainstream sources at times posted toxic headlines expressing frustration with people unwilling to wear masks, politicians reluctant to institute stricter lockdowns to curb the spread of the virus, and the Trump administration's response to the pandemic. Research from the psychological reactance literature suggests that such toxic headlines may even be counterproductive in persuading people to comply with public health recommendations if they view compliance as threatening to individual freedom (Brehm, 1989). Social media platforms may algorithmically promote scientific sources regarding pandemic-related information to militate against misinformation and politicization of the issue using non-combative rhetoric.

### 6.2. Implications for practice

#### 6.2.1. Recommendations for Reddit and other social media platforms

First, social media platforms can make decisions that foster spread of helpful information during public health crises and other disasters. Perhaps the best silver lining in Reddit coverage of COVID-19 is reflected in the growth of the r/coronavirus subreddit. The subreddit seems to bridge the political subreddits with the r/science subreddit, suggesting a role in translating scientific research to more accessible language. By the eigenvector centrality measure (Bonacich, 2007), r/coronavirus is the second most influential subreddit in the network of subreddits. The influential role of r/coronavirus underscores the importance of organic mobilization by users in formulating solutions to emergent communication needs during a health crisis. By highlighting the fast growth of the subreddit on its front page, Reddit helped facilitate this influence which helps in disseminating credible COVID-19 information. Other social media platforms could learn from this example.

Second, to the extent that toxicity may encourage politicization of a public health crisis, it is important to curtail toxic content as much as possible. One strategy to reduce toxicity would be to employ algorithmic monitors that alert social media users to rethink posts that are deemed too incendiary leading to breakdown of civil discussion.

Third, human moderation of content may be useful to encourage the dissemination of information from credible sources during times of crises. Platforms such as Reddit in their missions tout the values of free expression, and anonymity of participation may be one such mechanism encouraging free exchange of ideas. Even then, subreddits vary on the levels of human moderation that they allow; some subreddits are fairly strict and have stringent vetting processes for the news sources they allow, while others are more laissez-faire and allow all kinds of content, which has led to the subreddits being restricted or banned (see Gaudette, Scrivens, Davies, & Frank, 2020; Massanari, 2017). We recommend that public health practitioners join key subreddits and offer their expertize, and they might even create their own subreddits so that they reach large swaths of users.

#### 6.2.2. Recommendations for public health messaging

The spread of COVID-19 misinformation has been recognized by the World Health Organization (World Health Organization, 2020) – which has dubbed it an infodemic – as a risk factor against its mitigation. The problem with infodemics is that they are initially birthed and spread in the fringes before they catch on (Nielsen, Fletcher, Newman, Brennen, & Howard, 2020). Although our study did not specifically study infodemics, there are ways in which our techniques of data collection can help stem their spread. Public health professionals may monitor the evolution of the network of news sources on public platforms such as Reddit to identify influential news sources in specific subcommunities. By so doing, influential sources that spread misinformation may also be revealed and thus inform responses from a public health messaging perspective.

Related to the above recommendation, the focus on misinformation may be necessary but may lead public health professionals to ignore the effect of political polarization on the public's receptiveness to public health messaging. While relatively few people may believe conspiracy theories about SARS-CoV-2 – the virus causing COVID-19 – more might nevertheless believe that requiring mask-wearing in public to be an infringement on individual liberty, which is a political position and not necessarily an instance of misinformation. The challenge for public health messaging would be to convince people not only of the benefits of mask-wearing, but to persuade them that such an action would not infringe on their liberties but also help the collective. Public health practitioners can learn about emerging political arguments by monitoring public online discussions about the pandemic as they happen on Reddit, Twitter, Facebook, and other platforms.

Our findings show that toxicity of news titles influence news source popularity in the resultant network. Given that toxicity in discussions might signal polarization in a given discussion, public health practitioners might want to monitor toxicity in public online discussions of pandemics and other diseases to determine the extent to which public health issues are becoming politicized. With that information, public health messaging can be massaged to reduce politicization and ensure maximum compliance regardless of an individual's political leanings.

### 6.3. Limitations and future research direction

Our study has various limitations. First, there are posts related to COVID-19, especially image posts in the form of memes that would not be returned by searches using the Reddit API. Second, political posts may appear in ostensibly non-political subreddits, and would not appear in our search results. Third, another limitation stems from Reddit's

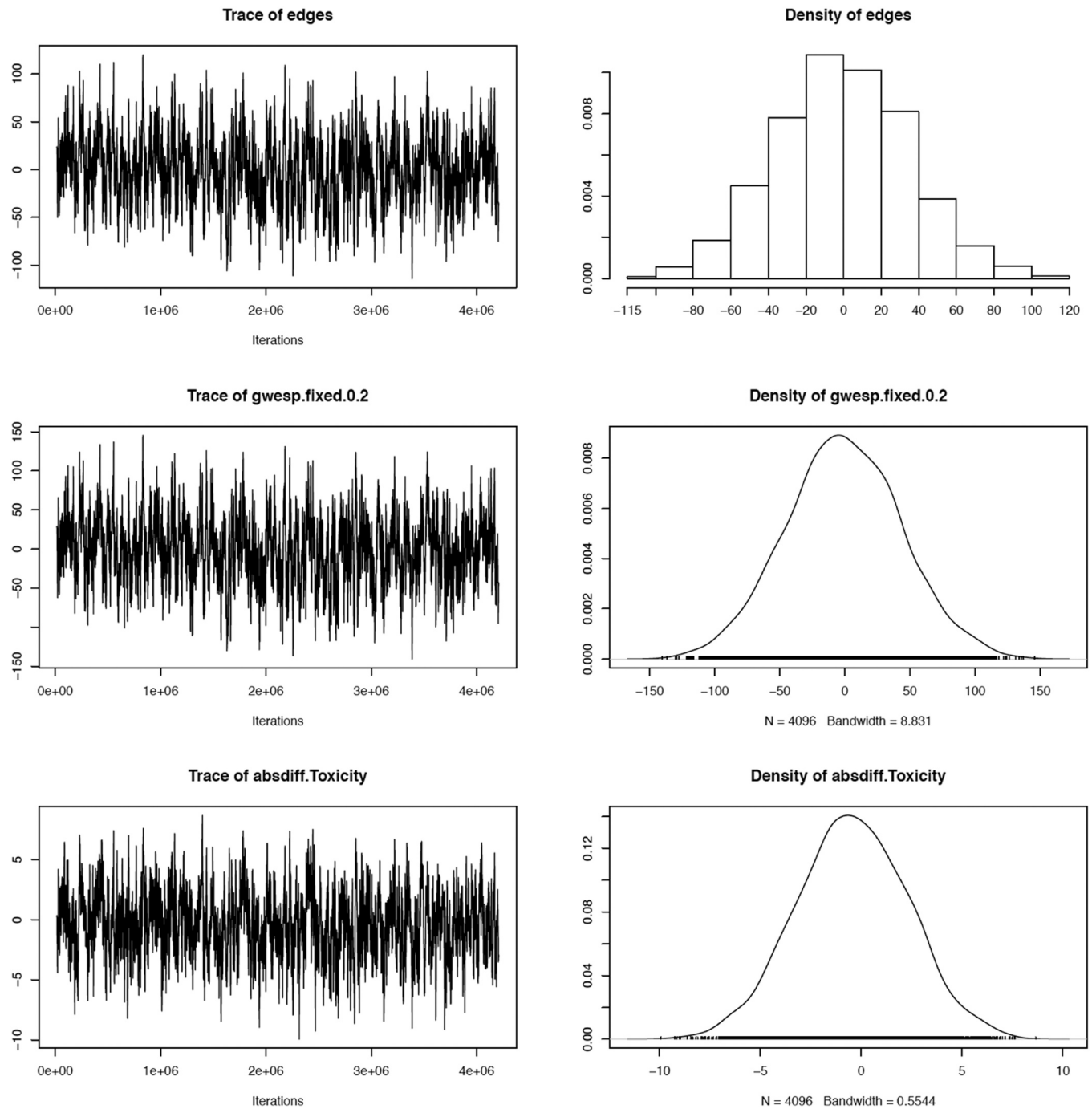


**Fig. A1.** Partial MCMC diagnostics for structural + node attribute model of news sources network tie formation.

PRAW API which in certain cases limits the number of returned results for a given keyword search (Boe, 2016). Fourth, there are also limitations arising from the demographics of Reddit users who are overwhelmingly young, white, and male and thus do not reflect the general US or global population (Massanari, 2017). Fifth, while certain subreddits allow for users to change the titles of their linked articles, others do not. It is not yet clear the extent to which users choose to further editorialize their submissions, and it is nontrivial to determine which titles have been editorialized and which ones have not. Any exercise to make such determinations would be complicated by the fact that webpages are frequently updated.

Our study contributes to the understanding of the network of news sources, but future work can explore whether other platforms (e.g., Facebook and Twitter) exhibit similar patterns of news source sharing and the degree to which toxicity influences tie formation in the resultant networks. Certainly, news source networks created on Facebook may be different from the networks arising from anonymous participation as in Reddit, and future research examining these questions may help inform public health messaging strategies on social media with characteristics of the technology platform in mind. Longitudinal research to confirm which of social selection and social influence mechanisms of the homophily effect of toxicity may also prove a fruitful tangent of future research.

**Table A1**
Observed vs. simulated network statistics from ERGM-estimated model parameters for network of news sources.

| | observed | min | mean | max | p-value |
|---|---|---|---|---|---|
| edges | 2602.00 | 2500.00 | 2593.58 | 2668.00 | 0.84 |
| gwesp.fixed.0.2 | 3176.37 | 3052.28 | 3166.07 | 3257.72 | 0.80 |
| absdiff.Toxicity | 140.56 | 133.00 | 140.11 | 146.09 | 0.94 |
| nodecov.Toxicity | 1079.26 | 1037.66 | 1075.76 | 1104.76 | 0.84 |
| absdiff.Upvotes | 294.92 | 263.11 | 292.01 | 324.24 | 0.66 |
| nodecov.Upvotes | 4440.94 | 4303.15 | 4430.79 | 4555.33 | 0.88 |
| absdiff.Comments | 811,234.51 | 770,231.59 | 807,108.25 | 836,589.68 | 0.78 |
| nodecov.Comments | 1,991,882.69 | 1,929,435.64 | 1,987,138.97 | 2,036,364.92 | 0.86 |
| absdiff.Word_Count | 10,673.51 | 10,022.16 | 10,639.53 | 11,222.82 | 0.94 |
| nodecov.Word_Count | 95,105.43 | 91,179.03 | 94,837.28 | 97,248.69 | 0.90 |

## 7. Conclusion

Toxicity has been characterized mainly as a malignant force of abuse and cyberbullying in online discussions. Our study found it beneficial for news sources associated with it, to the extent that such news sources are rewarded with high network centrality. In that regard, toxicity is a tool not only for defeating opponents in conflict-laden discussions but also for building social capital for one's political faction.

The news sources shared to political subreddits are clustered into five communities: mainstream, international, right-wing, left-wing, and science; the latter exhibits the least toxicity. Although the right-wing cluster exhibits the highest toxicity, coverage in mainstream sources is also oftentimes highly toxic in its frustration at noncompliance of pandemic mitigation measures by primarily right-leaning people. The toxic coverage in mainstream sources may invoke psychological reactance, where individuals that perceive compliance with social distancing, mask-wearing and other measures even more strongly view the measures as threats to individual freedom. We recommend that scientific and other neutral sources be algorithmically elevated in online communities to reduce the toxicity of COVID-19 coverage on social media.

Practically, we note that political polarization may be a potent force in discouraging compliance with pandemic mitigation measures. Public health practitioners should monitor toxicity of online discussions of public health crises to detect when they are becoming politicized. By understanding the political arguments made regarding responses to health crises, public health practitioners may be better equipped to create effective messaging strategies.

## CRediT authorship contribution statement

**Wallace Chipidza**: Conceptualization, Methodology, Data curation, Visualization, Formal analysis, Validation, Writing – Original draft preparation, Writing – Reviewing and Editing.

## Appendix A

See Fig. A1 and Table A1 Here.

## References


Adamic, L. A., & Glance, N. (2005). The political blogosphere and the 2004 US election: Divided they blog. Proceedings of the 3rd International Workshop on Link Discovery, 36–43.

Adler, P., & Kwon, S.-W. (2002). Social capital: Prospects for a new concept. *Academy of Management Review, 27*(1), 17–40. https://doi.org/10.2307/4134367.

Adolph, C., Amano, K., Bang-Jensen, B., Fullman, N., & Wilkerson, J. (2021). Pandemic politics: Timing state-level social distancing responses to COVID-19. *Journal of Health Politics, Policy and Law, 46*(2), 211–233.

Allington, D., Duffy, B., Wessely, S., Dhavan, N., & Rubin, J. (2020). Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency. *Psychological Medicine*, 1–7. https://doi.org/10.1017/S003329172000224X.

Almerekhi, H., Kwak, H., Jansen, B. J., & Salminen, J. (2019). Detecting toxicity triggers in online discussions. Proceedings of the 30th ACM Conference on Hypertext and Social Media, 291–292.

Aroyo, L., Dixon, L., Thain, N., Redfield, O., & Rosen, R. (2019). Crowdsourcing subjective tasks: The case study of understanding toxicity in online discussions. Companion Proceedings of the 2019 World Wide Web Conference, 1100–1105.

Backlinko. (2021). Reddit Usage and Growth Statistics: How Many People Use Reddit in 2021? Backlinko. ⟨https://backlinko.com/reddit-users⟩.

Baker, S. R., Bloom, N., Davis, S. J., & Terry, S. J. (2020). *COVID-induced economic uncertainty (No. w26983)*. National Bureau of Economic Research. https://doi.org/10.3386/w26983.

Barnes, S. J. (2020). Information management research and practice in the post-COVID-19 world. *International Journal of Information Management, 55*, Article 102175. https://doi.org/10.1016/j.ijinfomgt.2020.102175.

Barrios, J. M., & Hochberg, Y. (2020). *Risk perception through the lens of politics in the time of the COVID-19 pandemic (No. w27008)*. National Bureau of Economic Research. https://doi.org/10.3386/w27008.

Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research, 49*(2), 192–205.

Bhattacharya, P., Phan, T. Q., Bai, X., & Airoldi, E. M. (2019). A coevolution model of network structure and user behavior: The case of content generation in online social networks. *Information Systems Research, 30*(1), 117–132. https://doi.org/10.1287/isre.2018.0790.

Bishop, J. (2014). Dealing with internet trolling in political online communities: Towards the this is why we can't have nice things scale. *International Journal of E-Politics, 5*(4), 1–20.

Boe, B. (2016). Python Reddit API Wrapper (PRAW).

Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social Networks, 29*(4), 555–564.

Borgatti, S. P., & Halgin, D. S. (2011). On network theory. *Organization Science, 22*(5), 1168–1181.

Boxer, B. (2015). Facebook is changing the news game—Are publishers ready? - Google Search. ⟨https://www.forbes.com/sites/benjaminboxer/2015/12/02/facebook-is-changing-the-game-are-publishers-ready⟩.

Brehm, J. W. (1989). Psychological reactance: Theory and applications. ACR North American Advances, NA-16. ⟨https://www.acrwebsite.org/volumes/6883/volumes/v16/NA-16/full⟩.

Bright, J. (2016). The social news gap: How news reading and news sharing diverge. *Journal of Communication, 66*(3), 343–365.

Brooks, D. J., & Geer, J. G. (2007). Beyond negativity: The effects of incivility on the electorate. *American Journal of Political Science, 51*(1), 1–16.

Buntain, C., & Golbeck, J. (2014). Identifying social roles in reddit using network structure. Proceedings of the 23rd International Conference on World Wide Web, 615–620.

Carroll, N., & Conboy, K. (2020). Normalising the "new normal": Changing tech-driven work practices under pandemic time pressure. *International Journal of Information Management, 55*, Article 102186. https://doi.org/10.1016/j.ijinfomgt.2020.102186.

Chang, C. (2015). Inaccuracy in health research news: A typology and predictions of scientists' perceptions of the accuracy of research news. *Journal of Health Communication, 20*(2), 177–186.

Chen, G. M., Muddiman, A., Wilner, T., Pariser, E., & Stroud, N. J. (2019). We should not get rid of incivility online. *Social Media + Society, 5*(3). https://doi.org/10.1177/2056305119862641, 2056305119862641.

Conover, M. D., Ratkiewicz, J., Francisco, M. R., Gonçalves, B., Menczer, F., & Flammini, A. (2011). Political polarization on twitter. Icwsm, 133(2011), 89–96.

Daugherty, T., Eastin, M. S., & Bright, L. (2008). Exploring consumer motivations for creating user-generated content. *Journal of Interactive Advertising, 8*(2), 16–25.

Depoux, A., Martin, S., Karafillakis, E., Preet, R., Wilder-Smith, A., & Larson, H. (2020). *The pandemic of social media panic travels faster than the COVID-19 outbreak*. Oxford University Press.

Duncan, B. (2009). How the media reported the first days of the pandemic (H1N1) 2009: Results of EU-wide media analysis. *Euro Surveillance: Bulletin Europeen sur les maladies transmissibles = European Communicable Disease Bulletin, 14*(30), 19286.

Dwivedi, Y. K., Hughes, D. L., Coombs, C., Constantiou, I., Duan, Y., Edwards, J. S., … Upadhyay, N. (2020). Impact of COVID-19 pandemic on information management research and practice: Transforming education, work and life. *International Journal of Information Management, 55*, Article 102211. https://doi.org/10.1016/j.ijinfomgt.2020.102211.

Fridman, A., Gershon, R., & Gneezy, A. (2021). COVID-19 and vaccine hesitancy: A longitudinal study. *PLoS One, 16*(4), Article 0250123. https://doi.org/10.1371/journal.pone.0250123.

Gao, J., Zheng, P., Jia, Y., Chen, H., Mao, Y., Chen, S., … Dai, J. (2020). Mental health problems and social media exposure during COVID-19 outbreak. *PLoS One, 15*(4), Article 0231924.

Gaudette, T., Scrivens, R., Davies, G., & Frank, R. (2020). Upvoting extremism: Collective identity formation and the extreme right on Reddit. *New Media & Society*. https://doi.org/10.1177/1461444820958123, 1461444820958123.

Gholami, J., Hosseini, S. H., Ashoorkhani, M., & Majdzadeh, R. (2011). Lessons learned from H1N1 epidemic: The role of mass media in informing physicians. *International Journal of Preventive Medicine, 2*(1), 32–37.

Goel, S., Mason, W., & Watts, D. J. (2010). Real and perceived attitude agreement in social networks. *Journal of Personality and Social Psychology, 99*(4), 611–621.

Google. (2018, May 23). New York Times: Using AI to host better conversations. Google ⟨https://blog.google/technology/ai/new-york-times-using-ai-host-better-conversations/⟩.

Grossman, G., Kim, S., Rexer, J. M., & Thirumurthy, H. (2020). Political partisanship influences behavioral responses to governors' recommendations for COVID-19 prevention in the United States. *Proceedings of the National Academy of Sciencesof the United States of America, 117*(39), 24144–24153. https://doi.org/10.1073/pnas.2007835117.

Grover, P., Kar, A. K., Dwivedi, Y. K., & Janssen, M. (2019). Polarization and acculturation in US Election 2016 outcomes – Can twitter analytics predict changes in voting preferences. *Technological Forecasting and Social Change, 145*, 438–460. https://doi.org/10.1016/j.techfore.2018.09.009.

Gruzd, A., & Roy, J. (2014). Investigating political polarization on Twitter: A Canadian perspective. *Policy & Internet, 6*(1), 28–45.

Guberman, J., Schmitz, C., & Hemphill, L. (2016). Quantifying toxicity and verbal violence on Twitter. Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion, 277–280.

Hacker, J., vom Brocke, J., Handali, J., Otto, M., & Schneider, J. (2020). Virtually in this together—How web-conferencing systems enabled a new virtual togetherness during the COVID-19 crisis. *European Journal of Information Systems, 29*(5), 563–584. https://doi.org/10.1080/0960085X.2020.1814680.

Halberstam, Y., & Knight, B. (2014). *Homophily, group size, and the diffusion of political information in social networks: Evidence from twitter (No. w20681)*. National Bureau of Economic Research. https://doi.org/10.3386/w20681.

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of Statistical Software, 24*(1), 1548–7660.

Harris, J. K. (2013). *An introduction to exponential random graph modeling* (Vol. 173). Sage Publications. ⟨https://books.google.com/books?hl=en&lr=&id=lkYXBAAAQBAJ&oi=fnd&pg=PR1&ots=It97xWY6FV&sig=r87iXvmDy4hz7LP5H-Wf2KcSmN0⟩.

Hong, S., & Kim, S. H. (2016). Political polarization on twitter: Implications for the use of social media in digital governments. *Government Information Quarterly, 33*(4), 777–782.

Hosseini, H., Kannan, S., Zhang, B., & Poovendran, R. (2017). Deceiving Google's Perspective API Built for Detecting Toxic Comments. ArXiv:1702.08138 [Cs]. ⟨http://arxiv.org/abs/1702.08138⟩.

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software, 24*(3), 54860.

Iivari, N., Sharma, S., & Venta-Olkkonen, L. (2020). Digital transformation of everyday life—How COVID-19 pandemic transformed the basic education of the young generation and why information management research should care? *International Journal of Information Management, 55*, Article 102183. https://doi.org/10.1016/j.ijinfomgt.2020.102183.

Jigsaw, & Google. (n.d.). Perspective API. Retrieved September 3, 2020, from ⟨https://www.perspectiveapi.com/#/home⟩.

Kane, G. C., & Alavi, M. (2008). Casting the net: A multimodal network perspective on user-system interactions. *Information Systems Research, 19*(3), 253–272.

Kar, A. K., & Dwivedi, Y. K. (2020). Theory building with big data-driven research – Moving away from the "What" towards the "Why". *International Journal of Information Management, 54*, Article 102205. https://doi.org/10.1016/j.ijinfomgt.2020.102205.

Kim, D. H., & Ellison, N. B. (2021). From observation on social media to offline political participation: The social media affordances approach, 1461444821998346 *New Media & Society*. https://doi.org/10.1177/1461444821998346.

Laato, S., Islam, A. K. M. N., Islam, M. N., & Whelan, E. (2020). What drives unverified information sharing and cyberchondria during the COVID-19 pandemic? *European Journal of Information Systems, 29*(3), 288–305. https://doi.org/10.1080/0960085X.2020.1770632.

Lampe, C., & Resnick, P. (2004). Slash (dot) and burn: Distributed moderation in a large online conversation space. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 543–550.

Leclercq-Vandelannoitte, A., & Aroles, J. (2020). Does the end justify the means? Information systems and control society in the age of pandemics. *European Journal of Information Systems, 29*(6), 746–761. https://doi.org/10.1080/0960085X.2020.1820912.

Lee, S. T., & Basnyat, I. (2013). From press release to news: Mapping the framing of the 2009 H1N1 A influenza pandemic. *Health Communication, 28*(2), 119–132. https://doi.org/10.1080/10410236.2012.658550.

Luke, D. A. (2015). *A user's guide to network analysis in R*. Springer.

Märtens, M., Shen, S., Iosup, A., & Kuipers, F. (2015). Toxicity detection in multiplayer online games. 2015 International Workshop on Network and Systems Support for Games (NetGames), 1–6.

Massanari, A. (2017). # Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures. *New Media & Society, 19*(3), 329–346.

McClung, N., Chamberland, M., Kinlaw, K., Bowen Matthew, D., Wallace, M., Bell, B. P., … Dooling, K. (2020). The advisory committee on immunization practices' ethical principles for allocating initial supplies of COVID-19 vaccine—United States, 2020. *Morbidity and Mortality Weekly Report, 69*(47), 1782–1786. https://doi.org/10.15585/mmwr.mm6947e3.

McPherson, J. M., & Smith-Lovin, L. (1987). Homophily in voluntary organizations: Status distance and the composition of face-to-face groups. *American Sociological Review, 52*(3), 370–379. https://doi.org/10.2307/2095356.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*(1), 415–444. https://doi.org/10.1146/annurev.soc.27.1.415.

Nabity-Grover, T., Cheung, C. M. K., & Thatcher, J. B. (2020). Inside out and outside in: How the COVID-19 pandemic affects self-disclosure on social media. *International Journal of Information Management, 55*, Article 102188. https://doi.org/10.1016/j.ijinfomgt.2020.102188.

Nielsen, R. K., Fletcher, R., Newman, N., Brennen, J. S., & Howard, P. N. (2020). *Navigating the 'infodemic': How people in six countries access and rate news and information about coronavirus*. Reuters Institute.

Norris, P. (2002). *The bridging and bonding role of online communities*. Thousand Oaks, CA: Sage Publications Sage CA.

Obadimu, A., Mead, E., Hussain, M. N., & Agarwal, N. (2019). Identifying toxicity within youtube video comment. International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation, 214–223.

Olson, D. H., & DeFrain, J. (2003). *Marriage and families* (pp. 100–105). Boston: McGrow Hill.

Orlikowski, W. J., & Iacono, C. S. (2001). Research commentary: Desperately seeking the "IT" in IT research—A call to theorizing the IT artifact. *Information Systems Research, 12*(2), 121–134.

Park, A., & Conway, M. (2018). Tracking Health Related Discussions on Reddit for Public Health Applications. AMIA Annual Symposium Proceedings, 2017, 1362–1371.

Pee, L. G., Pan, S. L., Wang, J., & Wu, J. (2021). Designing for the future in the age of pandemics: A future-ready design research (FRDR) process. *European Journal of Information Systems, 30*(2), 157–175. https://doi.org/10.1080/0960085X.2020.1863751.

Pew Research Center. (2019). *Key findings about the online news landscape in America*. Pew Research Center. ⟨https://www.pewresearch.org/fact-tank/2019/09/11/key-findings-about-the-online-news-landscape-in-america/⟩.

Regidor, E., de la Fuente, L., Gutiérrez-Fisac, J. L., de Mateo, S., Pascual, C., Sánchez-Payá, J., & Ronda, E. (2007). The role of the public health official in communicating public health information. *American Journal of Public Health, 97*(Suppl 1), S93–S97. https://doi.org/10.2105/AJPH.2006.094623.

Richter, A. (2020). Locked-down digital work. *International Journal of Information Management, 55*, Article 102157. https://doi.org/10.1016/j.ijinfomgt.2020.102157.

Rose Kim, J. Y., Howard, M., Cox Pahnke, E., & Boeker, W. (2016). Understanding network formation in strategy research: Exponential random graph models: Understanding Network Formation in Strategy Research: ERGMs. *Strategic Management Journal, 37*(1), 22–44. https://doi.org/10.1002/smj.2454.

Rowe, F. (2020). Contact tracing apps and values dilemmas: A privacy paradox in a neoliberal world. *International Journal of Information Management, 55*, Article 102178. https://doi.org/10.1016/j.ijinfomgt.2020.102178.

Rowe, F., Ngwenyama, O., & Richet, J.-L. (2020). Contact-tracing apps and alienation in the age of COVID-19. *European Journal of Information Systems, 29*(5), 545–562. https://doi.org/10.1080/0960085X.2020.1803155.

Sein, M. K. (2020). The serendipitous impact of COVID-19 pandemic: A rare opportunity for research and practice. *International Journal of Information Management, 55*, Article 102164. https://doi.org/10.1016/j.ijinfomgt.2020.102164.

Snijders, T. A., Van de Bunt, G. G., & Steglich, C. E. (2010). Introduction to stochastic actor-based models for network dynamics. *Social Networks, 32*(1), 44–60.

Soliman, A., Hafer, J., & Lemmerich, F. (2019). A Characterization of Political Communities on Reddit. Proceedings of the 30th ACM Conference on Hypertext and Social Media, 259–263. ⟨https://doi.org/10.1145/3342220.3343662⟩.

Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media—Sentiment of microblogs and sharing behavior. *Journal of Management Information Systems, 29*(4), 217–248.

Trang, S., Trenz, M., Weiger, W. H., Tarafdar, M., & Cheung, C. M. K. (2020). One app to trace them all? Examining app specifications for mass acceptance of contact-tracing apps. *European Journal of Information Systems, 29*(4), 415–428. https://doi.org/10.1080/0960085X.2020.1784046.

Valenzuela, S., Piña, M., & Ramírez, J. (2017). Behavioral effects of framing on social media users: How conflict, economic, human interest, and morality frames drive news sharing. *Journal of Communication, 67*(5), 803–826.

Venkatesh, V. (2020). Impacts of COVID-19: A research agenda to support people in their fight. *International Journal of Information Management, 55*, Article 102197. https://doi.org/10.1016/j.ijinfomgt.2020.102197.

Wise, K., Hamman, B., & Thorson, K. (2006). Moderation, response rate, and message interactivity: Features of online communities and their effects on intent to participate. *Journal of Computer-Mediated Communication, 12*(1), 24–41.

World Health Organization. (2020). Myth busters. ⟨https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters⟩.

Wulczyn, E., Thain, N., & Dixon, L. (2017). Ex machina: Personal attacks seen at scale. Proceedings of the 26th International Conference on World Wide Web, 1391–1399.

Wallace Chipidza (PhD, Baylor University, 2018) is an assistant professor in the Center for Information Systems and Technology at Claremont Graduate University. He researches social networks and their evolution, applications of big data and machine learning, as well as the use of information technology for development.