RESEARCH ARTICLE

Understanding high- and low-quality URL Sharing on COVID-19 Twitter streams

Lisa Singh[1] · Leticia Bode[1] · Ceren Budak[2] · Kornraphop Kawintiranon[1] · Colton Padden[1] · Emily Vraga[3]

Abstract

This article investigates the prevalence of high and low quality URLs shared on Twitter when users discuss COVID-19. We distinguish between high quality health sources, traditional news sources, and low quality misinformation sources. We find that misinformation, in terms of tweets containing URLs from low quality misinformation websites, is shared at a higher rate than tweets containing URLs on high quality health information websites. However, both are a relatively small proportion of the overall conversation. In contrast, news sources are shared at a much higher rate. These findings lead us to analyze the network created by the URLs referenced on the webpages shared by Twitter users. When looking at the combined network formed by all three of the source types, we find that the high quality health information network, the low quality misinformation network, and the news information network are all well connected with a clear community structure. While high and low quality sites do have connections to each other, the connections to and from news sources are more common, highlighting the central brokerage role news sources play in this information ecosystem. Our findings suggest that while low quality URLs are not extensively shared in the COVID-19 Twitter conversation, a well connected community of low quality COVID-19 related information has emerged on the web, and both health and news sources are connecting to this community.

Keywords COVID-19 · Coronavirus · Twitter conversation · Misinformation · Low quality domains

✉ Lisa Singh
    lisa.singh@georgetown.edu

1   Georgetown University, Washington, DC, USA

2   University of Michigan, Ann Arbor, USA

3   University of Minnesota, Minneapolis, USA

Introduction

Social media are a significant conduit for news and information in the modern media environment, with one in three people in the world engaging in social media, and two thirds of those on the Internet using it [44]. The popularity is higher in the United States with 68% of American adults reporting that they get their news on social media [37]. This is particularly true for health and science information, with a third of people reporting that social media are an "important" source of science news [29]. Twitter, in particular, is known for sharing and consuming news: 59% of its users describing it as "good" or "extremely good" for sharing preventive health information [63].

Of course, there is a great deal of research that examines the existence and spread of misinformation on Twitter [3, 14, 53], including that spread by bots [21, 50]. Most notably, several researchers took interest in this phenomenon following the 2016 US Presidential Election [8, 14, 25]. Clearly, misinformation abounds on Twitter, and the problem may be growing relative to other platforms [3]. Given its prevalence on Twitter, we would expect to see it proliferate during a pandemic as well.

More specifically, social media are also rife with health misinformation. Health misinformation—often defined as information that counters best available evidence from medical experts at the time ([62]; see also [23, 43, 56])—has been documented across almost all social media platforms, including Facebook, Twitter, YouTube, Pinterest, and Instagram [12, 13, 20, 27, 45, 52]. Moreover, health misinformation is not limited to any one issue, and may be of special concern for global health crises like the Ebola outbreak in 2014 and the spread of Zika in 2016, where research documented high prevalence and popularity of health misinformation topics [20, 45, 52].

One illustrative example of misinformation on social media relates to the emergence of online communities around anti-vaccination attitudes and beliefs. Although so-called "anti-vaxxers" are a minority of the population, they are a vocal and growing community on social media platforms like Twitter [28]. The communities that form around these sentiments also tend to be "highly clustered" [66], engaging with one another, but not with other networks of users [28]. They are also vulnerable to misinformation, which spreads easily on social media [6], and is most rampant among the overconfident – that is, those who think they know more than experts [39].

Unfortunately, there is reason to be even more concerned about the quality of such information in today's news ecosystem compared to that of earlier epidemics. As recent research shows, trust in institutions is eroding [58] and this is accompanied by renewed concern about the spread of misinformation online. In response, the World Health Organization raised alarms about an "infodemic" regarding the novel coronavirus that causes COVID-19, which they defined as "overabundance of information— some accurate and some not—that occurs during an epidemic. It can lead to confusion and ultimately mistrust in governments and public health response" [65]. Citing social media as a key driver in the

infodemic, the WHO called upon researchers to better define and understand the scope of high and low quality information spread on social media.

This article attempts to answer this call. We investigate webpages (information sources) being shared on Twitter when users discuss COVID-19, distinguishing between high quality health sources, traditional news sources, and low quality information sources. We then investigate the networks that exist among the information sources. This gives us insight into whether or not communities are emerging around sources of high quality, low quality, and news information, or between them, and more generally what these ecosystems look like. We find that (1) misinformation, in terms of URL links to low quality information sites, is shared at a higher rate than links to high quality health information, but remains a relatively small proportion of the COVID-19 Twitter conversation; (2) news sources are shared at a higher volume than either low or high quality sources; (3) the networks of each group of information sources are well connected, with clear community structure, indicating an emergence of both a high quality information subnetwork and a low quality information subnetwork; and (4) while high and low quality sites do have connections to each other, the number of connections to and from news sources is larger, highlighting the central role news sources play in the sharing of both high and low quality information. These findings suggest that even though low quality misinformation sources related to COVID-19 are not shared extensively on Twitter, the community structure that connects these sources to credible sources provides pathways for individuals to be exposed to low quality content related to COVID-19, or vice versa.

Research questions

During a pandemic or other emerging crisis, public interest in news and information tends to be quite high [5], and the COVID-19 pandemic is no exception [4]. However, the nature of these crises–especially in terms of the uncertainty surrounding a rapidly emerging infectious disease like COVID-19 - may lead people to share suboptimal information.

Types and frequency of shared URL content

While we might want the public to be relying on health information shared by reputable health organizations, we do not know if this is the case. Likewise, people are expected to depend heavily on the news media during crisis situations to orient themselves to new information and build community [5]. Finally, low quality information and misinformation may also be prevalent online. While the 2016 election brought attention to this issue [7], it is nothing new-especially in the health domain. For instance, a 2010 study by [47] examined 1000 randomly selected tweets mentioning antibiotics and found that 700 of them contained medical misinformation or malpractice.

Therefore, it is not surprising that similar trends appear to be emerging when considering the prevalence of misinformation on social media surrounding COVID-19

[1, 33, 51, 55]. While the type of misinformation being studied varies in these studies, all the studies show that misinformation sharing is occurring. According to a European Union External Action Committee Report, "substantive amount of both misinformation and disinformation are spreading on- and offline" [22]. Kousy and colleagues [33] used a random sample of tweets containing different coronavirus keywords and hashtags, and found that misinformation and unverifiable content within the tweets was being shared at a high rate, particularly by individual and informal group accounts (33.8%). Researchers have identified a number of conspiracy theories being shared [1, 51, 55], e.g., linking 5G to COVID-19, but the levels of sharing, the information cascades related to some of these conspiracies, and the belief in the conspiracy vary depending upon the user group studied and the specific conspiracy. [30] analyzed a set of misinformation claims identified by Google Fact Check Explorer and found that 88% of these claims were posted on social media sites, and that most of the information was recontextualization or 'spinning' of factual information. The focus of these mentioned studies has been on studying content of a tweet and identifying specific pieces of misinformation in that content. We take a different yet complementary approach by focusing on the URLs being shared and categorizing them according to their web-domains. This allows us to focus on the original producers of content being shared.

Our first descriptive analysis thus focuses on understanding the original producers of content that users are sharing with each other when using a COVID-19 hashtag. Twitter is generally a low trust environment (for example, in 2020 54% of those who had heard of Twitter said they distrusted it [32]), which might suggest that much of the content is low-quality. How does that apply when it comes to COVID-19? Are individuals who share URLs linking to sources that generally share misinformation, traditional news sources, or health organizations'? Which are shared most frequently, and how does this change over time?

COVID-19 information ecosystem

Earlier work on Twitter networks focused on how to model message diffusion and propagation [15, 31]. Later, scholars focused on characterizing specific networks of misinformation using Twitter data [19, 57, 67]. More recently, researchers moved beyond case studies and investigated the diffusion of true and false information, and found that lies spread faster than truths on Twitter [15]. This previous work leads us to consider relationships among the sources/domains of the shared URL content.

The second descriptive analysis, therefore, focuses on understanding the ecosystem of the information sources shared by Twitter users. Specifically, our goal is to understand the connectivity (URL link structure) of the domains of the shared URLs. This allows us to investigate the following questions: Do different categories of sources link to each other? Which categories link to one another most often? Understanding these dynamics begins to gives us insight into whether or not networks of subgroups in the ecosystem have formed, and the pathways that exist between high and low quality information.

Methodology

This section begins by describing our Twitter data set. We then explain the methodology used for each analysis.

Description of data set

Using the Twitter Streaming API, we began collecting tweets related to COVID-19 on January 16, 2020. Data collection continues, but the data we present in this study is from January 16, 2020 to April 15, 2020. Table 6 in Appendix A shows the English hashtags we used to collect data and the date we began collecting data for the hashtag. Most of the data collection began in January, and additional hashtags were added in mid-March to reflect the changing nature of the conversation around COVID-19 online. [1] During the study period of January 16 through April 15, 11.2 million tweets, 1.5 million quotes, and 54.5 million retweets were shared.

URL shares methodology

We begin by identifying all of the URLs that are shared in each tweet, retweet, and quote in our data set. Our tweets are not truncated and the data from the API is a JSON record. We extract the URLs from the JSON record, they are a separate field in the JSON record. If we have a retweet or a quote, we extract the URLs in the parent tweet and remove URLs to the original tweets.

We then reduce each URL to the web domain and count the frequency of each domain to determine the most popular domains. Considering only the domain, rather than the content itself, is a relatively blunt measure of misinformation, but one that is commonly used for similar research purposes [8, 9, 25, 50]. As one article described it, "the attribution of "fakeness" is thus not at the level of the story but at that of the publisher" [25].

To differentiate the quality of the information shared, we categorize the sources or webpage domains of the shared content, and focus on three relevant groups for understanding information quality: high-quality health sources, news sources, and low-quality/questionable content providers. While there are other groups that may be relevant, we focus on traditional information sources that the general public are more likely to find and share on social media.

High-Quality Health Sources (HQHS): In April 2020, we identified the set of reputable web domains that publish health information as follows. We first determined all the countries identified by the CDC as a Level 3 travel health notice country (that is, with the recommendation to "avoid all non-essential travel"). For each of these countries, we identify the web domain of each country's equivalent to a Center for Disease Control. Next, we augmented this list by including

---

[1] Note that due to a data collection glitch with the Twitter Streaming API, some of the hashtags were unavailable between March 13, 2020 and March 15, 2020.

top medical journals and hospitals, and by identifying additional US government agencies that had official COVID-19 related recommendations (for example, while not a public health organization, the EPA released information about effective disinfectants). After the White House announcement regarding the America's Health Insurance Plan's collaboration with the White House Coronavirus Task Force, the AHIP Statement page clarifying the free testing plan was also included on this list. In total, there are 39 sources included in this list.

Traditional News Sources (TNS) To identify reputable news sources, we adopt the definition and list of traditional news sites shared by MediaBias FactCheck—an independent online media outlet maintained by a small team of researchers and journalists [38]. This list has over one thousand three hundred web domains listed as reliable news sources. We distinguish these from HQHS because we want to understand the relationship between links in articles they post and HQHS and LQMS sites.

Low-quality/Misinformation Sources (LQMS) We identify the set of low-quality/questionable sources in two ways. First, we aggregate information using a list curated by NewsGuard [42]. NewsGuard is a journalistic organization that generally rates websites on their tendency to spread true or false information. Since the COVID-19 outbreak, they have kept a separate list of websites identified as propagating misinformation specifically related to the virus. Second, we rely on lists of low-quality and fake news producers aggregated by various scholars and fact-checking organizations [11]. A short summary of the lists are described below:

1. *ZIMDARS*: Zimdars et al. [68] tag websites with at most 3 of the following subcategories: *fake, satire, bias, conspiracy, rumor, state, junksci, hate, clickbait, political, reliable, unidentified*, and *unreliable*.
2. *MBFC*: Media Bias/Fact Check is an independent online media outlet maintained by a small team of researchers and journalists [61]. Similar to [68], the list they create assigns domains to subcategories. Their list uses the following three subcategories: *fake, conspiracy, satire*.
3. *POLITIFACT*: The staff of PolitiFact, in collaboration with Facebook, created a list of the most-shared fake news sites leading up to the 2016 U.S. Presidential election on Facebook [46]. This list labels sites using the following categories: *fake, imposter, some fake*, or *parody*.
4. *DAILYDOT*: *the Daily Dot*—a mainstream online news site created a list by referencing other pre-existing fake news lists.
5. *ALLCOTT*: Allcott et al. [2] aggregated the following five lists shared by: Politifact, Grinberg et al. [24], Silverman [54], Schaedel [48], and Guess et al. [26]. The subcategorization process is as follows: Politifact subcategories were ignored and all the domains were relabeled as fake. The subcategories *black, red, orange* (black: completely false, red/orange: has unreliable claims) of [24] were maintained. All domains from other referenced lists were labeled as fake.

Because these lists have different subcategories of low quality information, for consistency, we focus on using the *fake* category across the different sources. We

perform robustness checks to understand how consistent our results are across the various lists since past work shows results can vary depending on the list used [11]. In total, our list contains 1249 low quality sources.

To determine the set of HQHS, TNS, and LQMS domains observed in our Twitter data set, we first identify all of the URLs that are shared in our data set. We resolve redirects to determine the base web address. We then download the referenced webpage. We initially downloaded webpages in March 2020 and then added the remaining webpages in late April 2020.[2] Finally, we identify all of the URLs that are embedded on the downloaded webpage, e.g. a link to the CDC site in a particular New York Times article. We then aggregate this information to determine how often different sources are shared. In total, 92% of the URL content was downloaded and analyzed using this process. The remaining content resulted in URLs that were not resolved or resolved to different file types such as gifs or pdfs. The largest portion of the unlabeled data was bit.ly addresses that did not resolve to a valid URL.

Information sources network methodology

For each webpage we downloaded, we extracted all the URLs that were embedded in the webpage. We then reduced them to their domains. We construct a directed, weighted network by identifying domain source and destination pairs. For example, if three CNN news articles reference a CDC webpage, we would construct a directed edge from the CNN domain node to the CDC domain node. The edge weight of the directed edge would be three. We built an overall network of all the different types of information sources to determine whether or not there was connectivity across all of them. We then built separate networks for each of the information groups: HQHS, TNS, and LQMS. We compute standard network statistics, including degree, betweenness, eigenvector, and clustering coefficients. We also use the modularity clustering algorithm [41] to better understand the community structure for each information group. We used the NetworkX implementation of the algorithm and conducted an extensive sensitivity analysis to determine the final number of clusters.

Findings and analysis

Social media users commonly rely on external information to convey ideas, support claims, and serve information needs. Social media use around COVID-19 is no exception. Our analysis of tweets related to the disease shows that 40.04% of original tweet content, 4.85% of the retweeted content, and 10.76% of the overall content includes a URL. While we do not have a random sample of tweets to compare this to, studies of different user groups have shown high levels of information sharing among the group, e.g. computer scientists [49]. While those levels are higher than the levels in our data set, it is clear that URL information sharing is occurring and

---

[2] Note that some of the content may have changed between the initial sharing of the webpage on Twitter and the moment we downloaded the page.

Table 1 Most frequent domains shared by at least 100 user accounts

CBCEFB CBCEFB Domain	Tweet frequency	User frequency
youtube.com	357,717	145,929
instagram.com	214,001	93,605
twitter.com	129,031	29,867
nytimes.com	67,314	30,034
theguardian.com	63,744	23,950
bbc.co.uk	38,871	14,338
facebook.com	38,297	20,596
cnn.com	33,262	13,960
linkedin.com	32,613	15,066
change.org	32,456	21,012

may reflect the incredible need for information in this uncertain time [4, 5]. Uncertainty is strongly related to information seeking, and this has been shown specifically in the realm of health information seeking and sharing online: "When there is a lack of sufficient information from traditional medical professionals, uncertainties arise and online media provide individuals with an opportunity for further information seeking and sharing so as to evaluate, verify, or even challenge the prescriptions" [35]. We also acknowledge that not all the URLs being shared are motivated by information seeking behavior. As we will show, some of the top domains shared are social media sources.

It is also interesting to note the large difference in tweets and retweets containing a URL. Specifically, Twitter users appear less likely to retweet content containing a URL. This is a rather unexpected finding. Social media users turn to social media for a variety of needs, including emotional, informational, and instrumental support [16]. This observed pattern may result because different types of needs are driving tweeting versus retweeting behavior. This is only speculation, but informational needs may be driving COVID-19 tweeting behavior, while emotional needs may be driving COVID-19 retweeting behavior [17].

Popular domains

We begin by examining these shared links to determine the most popular domains. In this data set, there are over two hundred and thirty seven thousand unique domains that people share in their tweets. Table 1 presents the top-10 domains with respect to their tweet frequency. We focus on those domains that are not only frequently tweeted, but also have more than 100 user accounts tweeting the domain. By setting a threshold of 100 users sharing a URL of a specific domain, we focus on content that a substantial number of users chose to share, as opposed to content that was frequently shared by only a handful of users.

Inspecting these top-10 domains reveal some interesting patterns. First, people are linking to other social media platforms from Twitter. Indeed, the top two domains linked to in these URLs are both competing social media platforms (YouTube and

Table 2 Frequency of HQHS, TNS, and LQMS URLs shared in tweets, retweets, and quotes

Data set	Tweet Count	Tweets with URLs	HQHS URL Count	TNS URL Count	LQMS URL Count
Original Tweets	11,276,874	4,515,469 (40.0%)	62,539 (0.6%)	857,601 (7.6%)	93,894 (0.8%)
Quotes	1,500,099	101,240 (6.7%)	1,853 (0.1%)	7764 (0.5%)	760 (0.05%)
Retweets	54,825,063	2,658,933 (4.8%)	64,229 (0.1%)	654,513 (1.2%)	104,680 (0.2%)
Combined	67,602,036	7,275,642 (10.8%)	128,621 (0.2%)	1,519,878 (2.2%)	199,334 (0.3%)

Instagram). Other social media platforms in the top-10 domains include Facebook and LinkedIn. This is consistent with research showing that disinformation campaigns link across social media platforms in their efforts [64]. Second, news media sites tend to round out the top-10 in terms of web domains shared in tweets, with the New York Times, the Guardian, the BBC, and CNN. This is an indication that these news organizations are important for informing Twitter users about the pandemic. The tenth most popular domain shared is change.org. As of mid-July, over 4500 petitions related to the COVID-19 epidemic have been shared [18]. Petition topics range from giving every American extra money during this crisis to mandating schools say closed within a county in the United States until no new cases are identified in the county.

We also took a closer look at the most shared URL (not including retweets or quotes) within the Twitter domain to see the types of information people shared when posting a Twitter URL (Table 1 shows that Twitter is in 3rd place). We found that this is a reflection of the information Twitter is sharing about COVID-19: the most shared Twitter URL links to the official announcements about COVID-19 from Twitter [60].

In general, Table 1 highlights the lack of diversity in external sources shared with respect to COVID-19. Instead of linking to different types of sources, they link predominantly to social media and news media. Moreover, the dominance of cross-platform sharing reinforces the important role that social media needs to play in information sharing on Twitter.

Sharing of high quality health, traditional news, and low-quality sources

Table 2 provides a high level description of the overall prevalence of HQHS, TNS, and LQMS domains mentioned during Twitter conversation about COVID-19. We see that the tweets containing a link to reputable health sources (HQHS) account for 0.55% of tweets and 0.12% of retweets. Traditional news (TNS) accounts for 7.6% of tweets and 1.2% of retweets. Finally, low-quality/fake news sources (LQMS) account for 0.83% of original tweets and 0.19% of retweets.

A few important patterns can be seen. First, we see that reliance on HQHS links is minimal, despite limiting our analysis to tweets that explicitly relate to the

Table 3 Frequency of low-quality URLs from different blacklists shared in tweets, retweets, and quotes

CBCEFBData set	NewsGuard	ALLCOTT	DAILYDOT	MBFC	POLITIFACT	ZIMDARS
Original Tweets	31,999	5654	2446	64,677	893	1455
Quotes	345	23	8	446	10	17
Retweets	28,855	1244	5298	84,009	338	1000
Combined	61,199	6921	7752	149,132	1241	2472

Table 4 Pearson correlation among daily tweet counts of low quality information identified on different blacklists

Pearson Correlation Coefficient	NEWSGUARD	ALLCOTT	DAILYDOT	MBFC	POLITIFACT	ZIMDARS
NEWSGUARD	–	0.7588	0.6696	0.8215	0.6120	0.7398
ALLCOTT	0.7588	–	0.7271	0.9317	0.5399	0.7765
DAILYDOT	0.6696	0.7271	–	0.7851	0.4318	0.6225
MBFC	0.8215	0.9317	0.7851	–	0.5238	0.8082
POLITIFACT	0.6120	0.5399	0.4318	0.7851	–	0.7318
ZIMDARS	0.7398	0.7765	0.6225	0.9317	0.7318	–

coronavirus—a health topic. Indeed, the reliance on LQMS is comparable to HQHS, with LQMS having an edge in terms of both original tweets and retweets. But while there are more tweets and retweets with a URL link to LQMS than HQHS domains, together, LQMS and HSHS account for less than one percent of the combined overall shared content and less than 5% of the overall shared URL content. We also see that traditional news source content is shared at much larger rates compared to either HQHS or LQMS, representing 7.6% of tweets and 1.2% of retweets. While the news media still are not a large part of the overall conversation, these sources fare better when compared against only those tweets that contain a link. Here, links to TNS comprise over 20% of links in our dataset, while LQMS represent 2.7% and HQHS account for 1.8%. One possible explanation for this discrepancy is that users may follow online news sources as part of their regular information gathering process and have confidence sharing articles from those news sources they trust.

We next inspect the breakdown of the LQMS shares in Table 3 and their Pearson correlation among each of the low quality lists in Table 4. We see that the prevalence of low quality content depends significantly on which list is chosen to make that judgement. This finding is in line with past work [11]. Unsurprisingly, lists such as Politifact and DailyDot that have a strong emphasis on political fake news do not identify a large number of shares. Most low quality sources are identified when considering lists shared by MediaBias/FactCheck and NewsGuard. One important finding is that even though the frequency of tweets associated with each list varies considerably, all of the lists' frequency are highly correlated to the NewsGuard frequency (between 0.61 and 0.82) and the MediaBias/FactCheck frequency (between
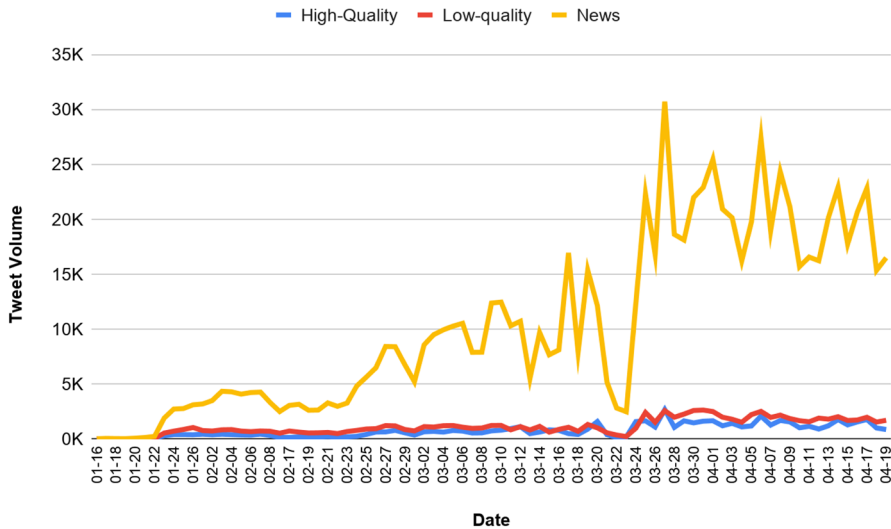
Fig. 1 Daily Tweet Volume of High Quality (HQHS—blue), Low Quality (LQMS—red), and News Sources (TNS—yellow). The *x* axis is the date and the *y* axis is the volume

0.78 and 0.93). Focusing in on only the Fake News parts of the lists, some correlations go up and others go down. DailyDoT has a high correlation to ZDARS (0.97), as does Alcott and Politifact (0.98). However, all the other correlations drop to between 0.14 and 0.57. In general, when the volume of tweets containing misinformation sources listed on these lists increases, so do the volume of tweets containing misinformation from low quality sources identified on the other lists. This may be an indication that webpages associated with low quality information sources have URLs embedded to other types of low quality information sources, e.g. a health fake news site may reference a political fake news site.

Finally, we want to understand how the volume of high quality, low quality, and news sources change over time. Figure 1 shows the daily volume of high quality (HQHS), low quality (LQMS) and news sources (TNS). The *x* axis is the date and the *y* axis is the volume. We see that the volume of all the sources are increasing, but the sharing of news sources is increasing at a greater rate than the sharing of high and low quality information sources.[3] The general increase is not surprising since the overall volume of conversation about COVID-19 increased during this time period. If we instead look at the share of overall conversation of these sources, we see a different trend. Figure 2 shows the daily proportion of the overall conversation of each of these sources. We see that the shares of HQHS, LQMS, and TNS all decreased as the crisis continued and plateaued in April. This decreasing trend is also similar for URL sharing as a whole. This may result because there is a larger need for information at the onset of the pandemic and it decreases as more information is available

---

[3] As a reminder, there was a data glitch between March 13, 2020 and March 15, 2020
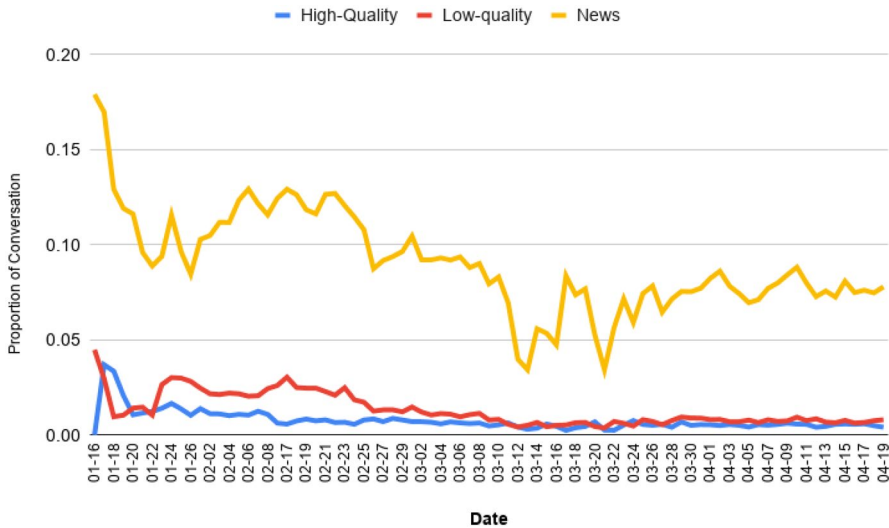
Fig. 2 Daily proportion of high quality (HQHS—blue), Low Quality (LQMS—red), and News Sources (TNS—yellow) in Tweets

from other sources. Still, it is important to remember that even though the share has decreased, the increase in volume means that more people are viewing/sharing more low quality URLs in April than in February or March. The impact could be significant given the magnitude of the pandemic.

COVID-19 information ecosystem

In an effort to characterize the COVID-19 information ecosystem, we examine the content of the shared URLs to determine which sources are being linked to within the webpages themselves. We focus our discussion on the high, low, and news domains described in the previous section. However, we pause to mention that the webpages shared by Twitter users contained links to over 70 million webpages across 1.1 million domains. Future work will investigate ways to include more of these sources in new analyses.

Traditional news source content When checking the content of the shared webpages of traditional news sources for URLs to high and low quality sources, we find that over 112,447 of the tweets have links to webpages from high quality sources, over 110,390 have links to webpages from low quality sources, and over 526,069 have links to webpages from news or other sources. This indicates that a comparable amount of news content links to both high and low quality sources, although we cannot speak to the nature of the links. For example, it is possible that some of these news sites are debunking the information posted on the LQMS.

Focusing on more frequently shared news sources (news domains), we find that 478 news domains were mentioned in at least 100 tweets. Of those, 407 news

domains contain at least one article that links to at least one HQHS site or LQMS. We want to classify these news sources based on the proportion of high quality and low quality information they share. We say a news site has *high reliability* if it references high quality information sources at least 80% of the time. The news site has *mixed reliability* if it references between 50 and 80% high quality content, it has *low reliability* if it references less than 50% high quality content, and it has *no reliability* if it links to only low quality content. We evaluated this subset and found that 272 out of 407 (66.83%) of these sources have high reliability, 31 (7.62%) have mixed reliability, 38 (9.34%) have low reliability and 66 (16.22%) have no reliability. In other words, those news domains that are mentioned most frequently in tweets generally link to high quality domains, with over half being highly reliable news sources, but a reasonable fraction link to low quality misinformation domains, with over 16% only linking to low quality misinformation.[4]

For the long tail of less popular domains (mentioned in less than 100 tweets), the results are somewhat comparable. There are 850 news domains with at least one article containing at least one link to a HQHS site or LQMS site. Of those, 567 out of 850 (66.71%) of these sources have high reliability, 26 (3.06%) have mixed reliability, 15 (1.76%) have low reliability and 242 (28.47%) have no reliability.

Even though the majority of sources have high reliability, a significant proportion have no reliability, an indication that the quality of news sources varies considerably with regards to COVID-19 information.

The last part of this analysis focuses on the connectivity structure of each of our information sources, i.e., HQHS, LQMS, and TNS, in order to understand if communities have begun forming. When sources conveying similar information about the pandemic reference each other, it may give more legitimacy to the information if the information is not being debunked.

Table 5 shows the network properties of each of our information source networks. All of these statistics are computed after removing self-edges. The number of nodes in each information group varies considerably, with few nodes in HQHS and many nodes in TNS and LQMS. The overall density for all three networks is considered high relative to random networks, with HQHS being the highest of our group. In other words, webpages of high quality health sources connect to or reference each other most frequently. This is confirmed when looking at the average clustering coefficient. For HQHS, it is 0.61, well above random connectivity. TNS and LQMS are 0.43 and 0.26, respectively. While not as high as HQHS, they are still high. This tells us that many nodes have neighbors connected to each other, indicating a substantial community structure within each of these information sharing groups.

All three networks are disassortative, meaning they exhibit a hub-and-spoke pattern. This finding is in line with analyses of the Web, but all three three networks have more disassortativity than most technological and biological networks studied

---

[4] Note that the *reliability* labels used here solely describe the reliability of content from those sources shared on Twitter. It is entirely possible, for instance, for a *low-reliability* site to link to high quality domains more often than 50% through other articles not shared on Twitter. But Twitter users show a preference for their articles with more connections to low-quality sites.

Table 5 Network statistics for each group of information sources

Network	Nbr nodes (connected nodes)	Nbr edges	Avg degree	Avg Eigen vector	Avg clusters coefficient	Graph density	Avg betweenness	Avg assortativity	Gini coefficient of node degree
HQHS	39 (38)	172	9.05	0.12	0.61	0.12	0.023	− 0.5476	0.5219
TNS	1198 (1003)	31,489	62.79	0.018	0.43	0.031	0.0011	− 0.2922	0.6516
LQMS	1249 (351)	1,845	10.51	0.025	0.26	0.015	0.0030	− 0.2306	0.6560
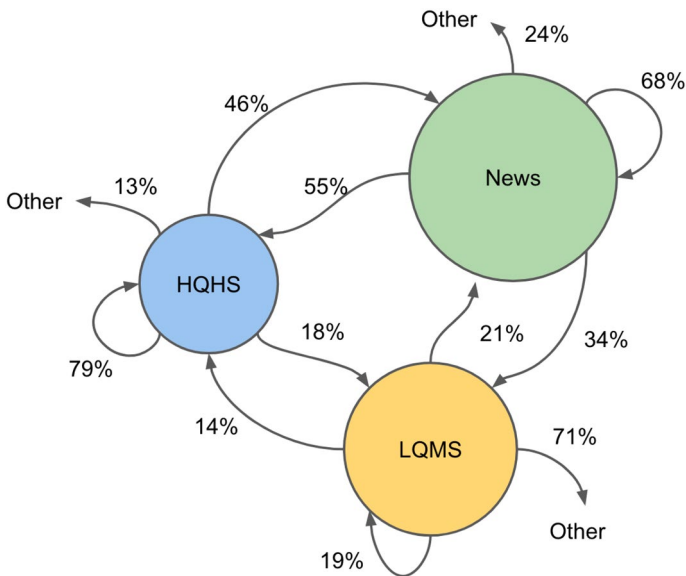
Fig. 3 Directed connectivity among different source Types. The percentage of information sources that connect to different types of information sources. A singe source can connect to multiple source types

in the past [40]. Assortativity has a direct relationship with network robustness. Disassortative graphs are less robust to targeted vertex removal [40]. In assortative graphs, a failure of a central node would be less detrimental to the overall connectedness of the overall ecosystem since high degree nodes are connected to one another, creating plentiful paths to allow dissemination of information and/or pathways for web users to explore the information space. In disassortative networks, however, high degree nodes are less connected to one another. As such, failure of a high degree node in a disassortative network, e.g dailymail.co.uk or rt.com in our data set, would have a larger impact on the connectedness of the network. Recent work has focused on strategies to take down misinformation sites [10]. In disassortative networks, such a strategy can have a broader impact on the overall misinformation ecosystem. It is interesting to note that the HQHS is the most disassortative network, suggesting that they are less robust to vertex removal [41]. In other words, if certain central nodes do not continue to link to the smaller health agencies and journals, the information shared by those organizations will not be disseminated as broadly since redundant pathways do not exist.

We also examine the centralization of the networks to determine the degree to which centrality is evenly distributed in these networks. Our analysis shows that the networks are highly skewed, more similar in structure to power law networks than to random networks. This high heterogeneity of the nodes in the network can also be seen by examining the values of the gini index, where a gini index of zero indicates more homogeneous connectivity structure. All of our networks have a similar gini index with LQMS having the highest (0.6560), i.e. having the most heterogeneity of the nodes.
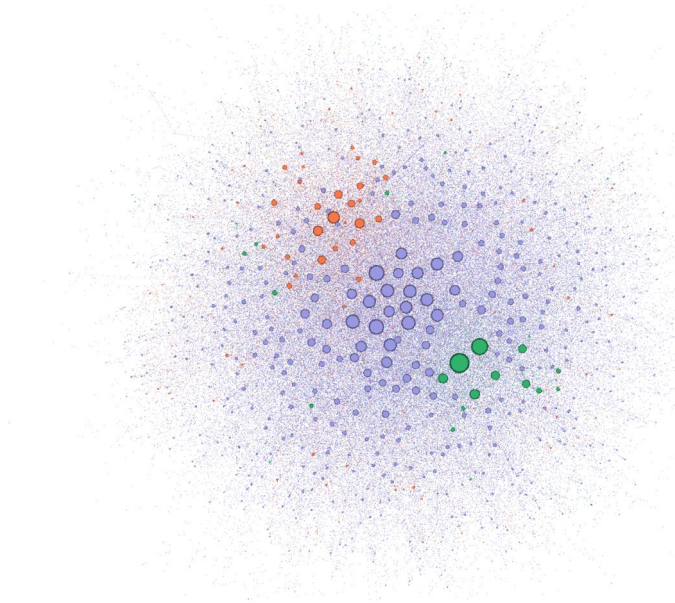
Fig. 4 TNS-HQHS-LQMS Network. The size of the node is based on the number of connections (node degree), and the edge color is based on the source node: TNS (purple), HQHS (green), and LQMS (orange)

Looking at the overall landscape, Fig. 3 shows the percentage of information sources that connect to different types of sources. For example, articles from 55% of TNS sources connect to one or more articles from HQHS, 34% connect to one or more articles from LQMS, 68% connect to one or more articles from other TNS, and 24% do not connect to HQHS, LQMS, or TNS.[5] At a high level, we see that news sources are connected to most from HQHS (46%), and 21% of LQMS. LQMS and HQHS do connect to each other, but at lower rates than the other group connects. LQMS tend to contain URLs to sources that are not HQHS or TNS (71%). This could be an indication that part of the LQMS network is not being shared through Twitter, so it is not visible in this study. It could mean that links to information not connected to COVID-19 are also shared on these LQMS. While we are uncertain about the other sources that are linked to, it is clear that HQHS and TNS are more connected to each other than to LQMS.

To get a better sense of how the different sources connect to each other individually, Fig. 4 shows the network based on webpage URL links across all the sources. The purple nodes are TNS, the orange nodes are the LQMS, and the green nodes are the HQHS. The edges are the color of the source and the node size is based on

---

[5] Because a single article can connect to multiple source types, the percentages will total more than 100%.

Fig. 5 Network between HQHS (green) and LQMS (orange). The size of the node is based on the number of connections (node degree), and the edge color is based on the source node
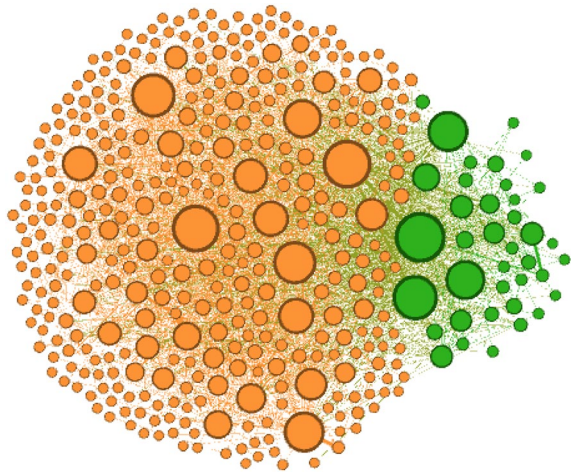
overall degree. The figure highlights two things. Not surprisingly, the news sources are dominant in the network. With regards to HQHS, this makes sense since there are many more news sources. However, the number of news sources and misinformation sources are comparable, thereby reinforcing our previous finding that high and low quality information sources have less connectivity to each other than to news sources. Second, the subnetworks corresponding to LQMS, HQHS, and TNS are not well separated. There are many individual sources across source types that are connecting to each other. While most LQMS sites are being connected to the broader network through their connections with TNS, some LQMS domains on the periphery are surprisingly more closely clustered with HQHS. Perhaps connections from TNS and HQHS are a result of refuting claims, but even if that is the case, clear pathways exist between the two types of sources. Overall, the figure reinforces the important role news organizations play for those sharing and seeking information on social media. Previous work has suggested that when traditional news covers fake news, it gives the fake news oxygen, even if they are trying to refute the content [37]

Figure 5 focuses on the connectivity between the HQHS (green nodes) and the LQMS (orange nodes). While there is some connectivity between them, the majority of nodes in the network do not have edges across source types, recall that there are 39 HQHS and 1249 LQMS. What is interesting is that while there are some LQMS connecting to HQHS, most of the nodes have very few connections and the few they have are to the more prominent health sources, i.e. the large nodes. Most of the connections from the high degree HQHS are to other HQHS. A small number are to LQMS. This again highlights that the weakest pathways across the different source types are between HQHS and LQMS.

Fig. 6 HQHS network clustering using modularity, three clusters are each shown in a different color. The node size is based on node degree
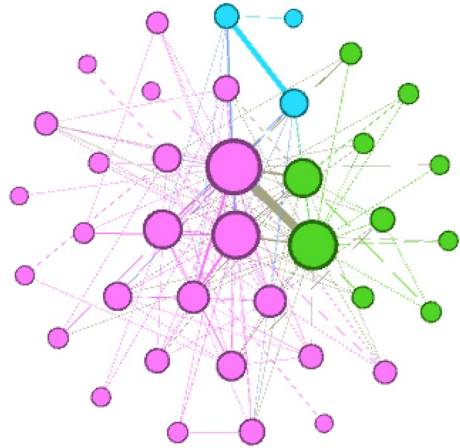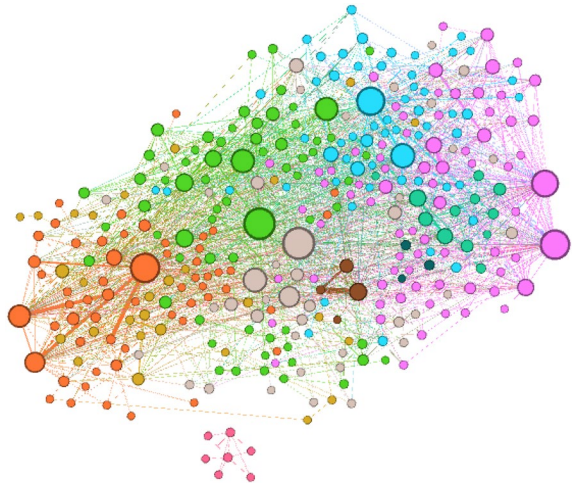

Fig. 7 LQMS network clustering using modularity. the network contains 10 clusters, each shown in a different color. The node size is based on node degree

Figures 6 and 7 show the clusters generated when using modularity clustering for the HQHS, and the LQMS, respectively. Figure 6 has three clusters. Each cluster is shown as a different color, and node size is based on node degree. The pink cluster mainly contains the WHO, CDC, and other health/government organizations. The green cluster contains predominantly reliable research journals, including Nature and the Lancet, as well as some international agencies. The third small, blue cluster contains medical journals and groups, including JAMA and AMA. We do see strong ties across clusters, with the strongest being between the WHO and Lancet. While the high clustering coefficient was an indication that high levels of connectivity existed in this network, the modularity clustering

algorithm showcases that there are two larger subgroups and one small one that have higher than expected connectivity structures.

Figure 7 contains ten clusters. Again, each cluster is shown as a different color, and node size is based on node degree. Focusing on the larger clusters, the orange and yellow clusters contain low quality health websites, while the pink and green contain fake news about COVID-19 identified by MediaBias/FactCheck. Each community contains stronger connectivity and more pathways within the cluster than outside the cluster. This may result because of prevalent themes in the information shared or reciprocal link agreements. Future research will investigate the similarities and differences in content across the clusters.

Limitations

This study has a number of limitations worth noting. First, our results are purely based on domain level analysis. In other words, all URLs from the same domain are classified under the same category. However, high-quality sites can sometimes, albeit rarely, share misinformation. For example, during the early stages of the pandemic, the CDC recommended not wearing masks [30]. Similarly, sites that aim to misinform can at times share reliable information. Despite these rare cases, in this paper, we opt for this simplification for two main reasons: (i) a domain level analysis allows us to focus on producers and therefore intent to deceive [35] and (ii) identifying factualness/low quality information is a notoriously difficult task that is hard to scale to the size of our corpus [9].

Our next limitation has to do with the identified sources. The lists used to identify the set of health, traditional news, and misinformation sites are heavily US and English based. As such, our analysis has that particular bias.

Finally, we collected our Twitter data using a set of COVID-19 related hashtags. While this provides a rich dataset to examine information sharing behavior related to the pandemic, hashtag focused data collection has its limitations. For instance, tweets including different hashtags can differ in important cultural and socio-political dimensions [60]. To address this point, we rely on multiple relevant hashtags instead of only one. However, this list will still not identify all COVID-19 related conversations. Some of the hashtags were also added after the beginning of the study period. If misinformation levels vary according to hashtag, we may miss that. Hashtag usage also has important temporal implications. For instance, [60] examined the Gezi Uprising and found that movement related hashtag usage was dropping while the protests were intensifying. The interviews revealed that this was at least partially due to hashtags becoming less useful, and thus a wasteful use of characters allocated to a tweet, once everyone knew the topic. A similar pattern could be observed for a phenomena as widely spread and impactful as COVID-19. Encouragingly, we do not see a drop-off in activity in our analysis other than the days we had data collection issues.

Conclusions and future directions

This article attempts to understand the types of URLs that are being shared on Twitter within the COVID-19 conversation. Our analysis focuses on health related domains, news domains, and domains containing misinformation. We find that while domains containing misinformation are shared at a higher rate than domains containing high quality health information, neither is prevalent in the COVID-19 conversation. Even though they are not tweeted at a high rate, a network analysis of links between webpages shared by misinformation sources shows that the network is dense, well connected, and disassortative. This means that even though community exists, the network is not robust to the removal of nodes and can be fragmented with the right interventions. While the networks we created based on the web content shared involving the HQHS and TNS are connected to nodes/sources in one of our source groups, the majority of links in the LQMS webpages are linking to other sources outside of our source groups. This may represent a part of the misinformation ecosystem that was not captured by following links from Twitter. This highlights that understanding the entire connectivity structure of the COVID-19 information ecosystem requires following all these links to other sources and identifying other links shared on other social media sites. Without question, this is a larger ecosystem than the 2000+ sources we focused on.

There are also future directions with regards to misinformation about COVID-19 on Twitter. For example, while this paper investigates the link structure of the webpages, we do not analyze the content of each webpage shared. Future work will build topic models to understand how the topics relate to the types of information being shared more broadly using the COVID-19 hashtags and the network structure of the different categories of web domains. Another direction of research would investigate misinformation shared in the content of the tweet and understand its prevalence. Finally, not only is the misinformation being shared, but correcting information is also being shared. We need to find ways to measure how much correcting information is being shared and whether or not it propagates at the same rate as the misinformation.

Compliance with ethical standards

Conflicts of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Appendix A

See Table 6.

Table 6 Hashtags and start date of collection

Hashtag	Date of first collection
#2019nCoV	1/16/2020
#ChinaPneumonia	1/16/2020
#ChinesePneumonia	1/16/2020
#Corona	1/16/2020
#WuhanCoronavirus	1/16/2020
#WuhanPneumonia	1/16/2020
#CoronavirusOutbreak	1/20/2020
#VirusChina	1/21/2020
#Coronavirus*	1/23/2020
#CoronaOutbreak	3/02/2020
#COVID	3/02/2020
#CoronavirusUpdate	3/11/2020
#COVID_19	3/11/2020
#COVID19**	3/11/2020

References

1. Ahmed, W., Vidal-Alaball, J., Downing, J., & Seguí, F. L. (2020). Covid-19 and the 5g conspiracy theory: social network analysis of twitter data. *Journal of Medical Internet Research, 22*(5), e19458.
2. Allcott, H., Gentzkow, M., & Yu, C. (2019). Trends in the diffusion of misinformation on social media. *Research & Politics, 6*(2), 2053168019848554.
3. Amy Mitchell, J.B.O. (2020). Americans immersed in covid-19 news; most think media are doing fairly well covering it (03/18/2020). https://www.journalism.org/2020/03/18/americans-immersed-in-covid-19-news-most-think-media-are-doing-fairly-well-covering-it/
4. Ball-Rokeach, S. J., & DeFleur, M. L. (1976). A dependency model of mass-media effects. *Communication Research, 3*(1), 3–21.
5. Bandari, R., Zhou, Z., Qian, H., Tangherlini, T. R., & Roychowdhury, V. P. (2017). Trends in the diffusion of misinformation on social media. *Computer, 50*(11), 60–67.
6. Bode, L., Budak, C., Ladd, J. M., Newport, F., Pasek, J., Singh, L. O., et al. (2020). *Words that matter: How the news and social media shaped the 2016 Presidential campaign*. Washington, DC: Brookings Institution Press.
7. Bovet, A., & Makse, H. A. (2019). Influence of fake news in twitter during the 2016 us presidential election. *Nature Communications, 10*(1), 1–14.
8. Bozarth, L., & Budak, C. (2020). Toward a better performance evaluation framework for fake news classification. *Proceedings of the International AAAI Conference on Web and Social Media*, Atlanta, Georgia, U.S. *14*, 60–71.
9. Bozarth, L., & Budak, C. (2001). Market forces: Quantifying the role of top credible ad servers in the fake news ecosystem. In: Proceedings of the International AAAI Conference on Web and Social Media (2021, Forthcoming)

10. Bozarth, L., Saraf, A., & Budak, C. (2020). Higher ground? How ground truth labeling impacts our understanding of fake news about the 2016 us presidential nominees. *Proceedings of the International AAAI Conference on Web and Social Media,* Atlanta, Georgia, U.S. *14*, 48–59.

11. Briones, R., Nan, X., Madden, K., & Waks, L. (2012). When vaccines go viral: An analysis of hpv vaccine coverage on youtube. *Health Communication, 27*(5), 478–485.

12. Broniatowski, D. A., Jamison, A. M., Qi, S., AlKulaib, L., Chen, T., Benton, A., et al. (2018). Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. *American Journal of Public Health, 108*(10), 1378–1384.

13. Budak, C. (2019). What happened? the spread of fake news publisher content during the 2016 us presidential election. In: Proceedings of the International on the World Wide Web, pp. 139–150.

14. Budak, C., Agrawal, D., & El Abbadi, A. (2011). Limiting the spread of misinformation in social networks. In: Proceedings of the International Conference on the World Wide Web, pp. 665–674.

15. Budak, C., & Agrawal, R. (2013) On participation in group chats on twitter. In: Proceedings of the International Conference on World Wide Web, pp. 165–176.

16. Chen, G. M. (2011). Tweet this: A uses and gratifications perspective on how active twitter use gratifies a need to connect with others. *Computers in Human Behavior, 27*(2), 755–762.

17. Coronavirus epidemic. (2007). https://www.change.org/t/coronavirus-epidemic-en-us.

18. De Domenico, M., Lima, A., Mougel, P., & Musolesi, M. (2013). The anatomy of a scientific rumor. *Scientific Reports, 3*, 2980.

19. Dredze, M., Broniatowski, D. A., & Hilyard, K. M. (2016). Zika vaccine misconceptions: A social media analysis. *Vaccine, 34*(30), 3441–3442.

20. Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM, 59*(7), 96–104.

21. Eeas special report update: Short assessment of narratives and disinformation around the covid-19 pandemic (update 23 april–18 may). (2020). https://euvsdisinfo.eu/eeas-special-report-update-short-assessment-of-narratives-and-disinformation-around-the-covid19-pandemic-updated-23-april-18-may/.

22. Garrett, R. K., Weeks, B. E., & Neo, R. L. (2016). Driving a wedge between evidence and beliefs: How online ideological news exposure promotes political misperceptions. *Journal of Computer-Mediated Communication, 21*(5), 331–348.

23. Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2018). Fake news on twitter during the 2016 us presidential election. Tech. rep.

24. Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). The rise of social bots. *Science, 363*(6425), 374–378.

25. Guess, A., Nyhan, B., & Reifler, J. (2018). Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 us presidential campaign. *European Research Council, 9*, 4.

26. Guidry, J. P. D., Carlyle, K., Messner, M., & Jin, Y. (2015). On pins and needles: How vaccines are portrayed on pinterest. *Vaccine, 33*(39), 5051–5056.

27. Gunaratne, K., Coomes, E. A., & Haghbayan, H. (2019). Temporal trends in anti-vaccine discourse on twitter. *Vaccine, 37*(35), 4867–4871.

28. Hitlin, P., & Olmstead, K. (2019). The science people see on social media. Pew Research Center (12/30/2019). https://www.pewresearch.org/science/2018/03/21/the-science-people-see-on-social-media. Accessed 16 July 2020.

29. Hollowood, E., Mostrous, A., Cummings, B., Newell, C., Bothwell, O., Macintyre, B., Kessler, J., & Nyenwa, T. (2020). Fake news in the time of c-19 (2020). https://members.tortoisemedia.com/2020/03/23/the-infodemic-fake-news-coronavirus/content.html. Accessed 16 July 2020.

30. Jingnan H, Aubrey, A, Wroth, C. (2020) Should We All Be Wearing Masks In Public? Health Experts Revisit The Question. https://www.npr.org/sections/health-shots/2020/03/31/824560471/should-we-all-be-wearing-masks-in-public-health-experts-revisit-the-question. Accessed 31 Mar 2020.

31. Jin, F., Dougherty, E., Saraf, P., Cao, Y., & Ramakrishnan, N. (2013). Epidemiological modeling of news and rumors on twitter. In: Proceedings of the Workshop on Social Network Mining and Analysis, pp. 1–9.

32. Jurkowitz, M., Mitchell, A., Shearer, E., & Walker, M. (2020). *U.s. media polarization and the 2020 election: A nation divided*. Washington, DC: Pew Research Center.

33. Kouzy, R., Abi Jaoude, J., Kraitem, A., El Alam, M. B., Karam, B., Adib, E., et al. (2020). Coronavirus goes viral: quantifying the covid-19 misinformation epidemic on twitter. *Cureus*, *12*(3), e7255.

34. Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., et al. (2018). The science of fake news. *Science, 359*(6380), 1094–1096.

35. Lin, W. Y., Zhang, X., Song, H., & Omori, K. (2016). Health information seeking in the web 2.0 age: Trust in social media, uncertainty reduction, and self-disclosure. *Computers in Human Behavior, 56*, 289–294.
36. Marwick, A., Lewis, R. (2020). Media manipulation and disinformation online (2020). https://datasociety.net/library/media-manipulation-and-disinfo-online. Accessed 16 July 2020.
37. Matsa, K.E., Shearer, E. (2018). News use across social media platforms 2018. Pew Research Center (2018). https://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018. Accessed 16 July 2020.
38. Media Bias/Fact Check: Media bias/fact check: The most comprehensive media bias recourse. (2018). https://mediabiasfactcheck.com/. Accessed 16 July 2020.
39. Motta, M., Callaghan, T., & Sylvester, S. (2018). Knowing less but presuming more: Dunning-kruger effects and the endorsement of anti-vaccine policy attitudes. *Social Science & Medicine, 211*, 274–281.
40. Newman, M. E. (2002). Assortative mixing in networks. *Physical Review Letters, 89*(20), 208701.
41. Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences, 103*(23), 8577–8582.
42. Newsguard: Coronavirus misinformation tracking center. (2020). https://www.newsguardtech.com/coronavirus-misinformation-tracking-center. Accessed 16 July 2020.
43. Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior, 32*(2), 303–330.
44. Ortiz-Ospina, E. (2020). The rise of social media https://ourworldindata.org/rise-of-social-media. Accessed 16 July 2020.
45. Oyeyemi, S. O., Gabarron, E., & Wynn, R. (2014). Ebola, twitter, and misinformation: a dangerous combination? *BMJ (Clinical research ed.), 349*, g6178.
46. Politifact staff. (2018). Politifact guide to fake news websites and what they peddle
47. Scanfeld, D., Scanfeld, V., & Larson, E. L. (2010). Dissemination of health information through social networks: twitter and antibiotics. *American Journal of Infection Control, 38*(3), 182–188.
48. Schaedel, S (2017) Websites that post fake and satirical stories. FactCheck. Available at https://www.factcheck.org/2017/07/websites-post-fake-satirical-stories/. Accessed 3 Sept 2018.
49. Schmitt, M., & Jäschke, R. (2017). What do computer scientists tweet? Analyzing the link-sharing practice on twitter. *PLoS One, 12*(6), e0179630.
50. Shao, C., Ciampaglia, G. L., Varol, O., Yang, K. C., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature Communications, 9*(1), 1–9.
51. Sharma, K., Seo, S., Meng, C., Rambhatla, S., & Liu, Y. (2020). Covid-19 on social media: Analyzing misinformation in twitter conversations. arXiv preprint arXiv:2003.12309
52. Sharma, M., Yadav, K., Yadav, N., & Ferdinand, K. C. (2017). Zika virus pandemic-analysis of facebook as a social media health information platform. *American Journal of Infection Control, 45*(3), 301–302.
53. Shin, J., Jian, L., Driscoll, K., & Bar, F. (2018). The diffusion of misinformation on social media: Temporal pattern, message, and source. *Computers in Human Behavior, 83*, 278–287.
54. Silverman, C. (2016). Here are 50 of the biggest fake news hits on facebook from 2016.
55. Singh, L., Wahedi, L., Wang, Y., Wei, Y., Kirov, C., Martin, S., Donato, K., Liu, Y., & Kawintiranon, K. (2019). Blending noisy social media signals with traditional movement variables to predict forced migration. In: Proceedings of the ACM International Conference on Knowledge Discovery & Data Mining, pp. 1975–1983. ACM.
56. Southwell, B. G., Niederdeppe, J., Cappella, J. N., Gaysynsky, A., Kelley, D. E., Oh, A., et al. (2019). Misinformation as a misunderstood challenge to public health. *American Journal of Preventive Medicine, 57*(2), 282–285. https://doi.org/10.1016/j.amepre.2019.03.009.
57. Starbird, K., Maddock, J., Orand, M., Achterman, P., & Mason, R.M. (2014). Rumors, false flags, and digital vigilantes: Misinformation on twitter after the 2013 boston marathon bombing. In: Proceedings of IConference.
58. Swift, A. (2016). Americans' trust in mass media sinks to new low. Gallup. https://news.gallup.com/poll/195542/americans-trust-mass-media-sinks-to-new-low.aspx. Accessed 16 July 2020.
59. Tufekci, Z. (2014). Big questions for social media big data: Representativeness, validity and other methodological pitfalls. In: E. Adar, P. Resnick, M.D. Choudhury, B. Hogan, A.H. Oh (eds.) Proceedings of the Eighth International Conference on Weblogs and Social Media, ICWSM 2014, Ann Arbor, Michigan, USA, June 1-4, 2014. The AAAI Press. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8062. Accessed 16 July 2020.

60. Twitter. (2020). Covid-19: Latest news updates from around the world (2020). https://twitter.com/i/events/1219057585707315201. Accessed 16 July 2020.
61. Van Zandt, D. (2018). Media bias/fact check (mbfc news). https://mediabiasfactcheck.com/about/. Accessed 16 July 2020.
62. Vraga, E. K., & Bode, L. (2020). Defining misinformation and understanding its bounded nature: Using expertise and evidence for describing misinformation. *Political Communication, 37*(1), 136–144.
63. Wilford, J., Osann, K., & Wenzel, L. (2018). Social media use among parents of young childhood cancer survivors. *Journal of Oncology Navigation & Survivorship, 9*(1).
64. Wilson, T., & Starbird, K. (2020). Cross-platform disinformation campaigns: lessons learned and next steps. *Harvard Kennedy School Misinformation Review*, *1*(1), 399–405.
65. World Health Organization. (2020). Coronavirus disease 2019 (covid-19) situation report. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200415-sitrep-86-covid-19.pdf. Accessed 16 July 2020.
66. Yuan, X., Schuchard, R. J., & Crooks, A. T. (2019). Examining emergent communities and social bots within the polarized online vaccination debate in twitter. *Social+ Media Society, 5*(3), 2056305119865465.
67. Zhou, Z., Bandari, R., Kong, J., Qjan, H., & Roychowdhury, V. (2020). Information resonance on twitter: watching iran. In: Proceedings of the Workshop on Social Media Analytics.
68. Zimdars, M. (2016). *My âfake listâ went viral. But made-up stories are only part of the problem.* Washington, DC: The Washington Post.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.