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.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
FibVID: Comprehensive fake news diffusion dataset during the COVID-19 period

Jisu Kim a, b, Jihwan Aum a, SangEun Lee a, Yeonju Jang a, **, Eunil Park a, b, *, Daejin Choi c

a Sungkyunkwan University, Seoul 03063, Korea
b Raon Data, Seoul 03073, Korea
c Incheon National University, Incheon 22012, Korea

ARTICLE INFO

Keywords:
COVID-19
Diffusion
Fact checking
Fake news

ABSTRACT

As the SARS-CoV-2 (COVID-19) pandemic has run rampant worldwide, the dissemination of misinformation has sown confusion on a global scale. Thus, understanding the propagation of fake news and implementing countermeasures has become exceedingly important to the well-being of society. To assist this cause, we produce a valuable dataset called FibVID (Fake news information-broadcasting dataset of COVID-19), which addresses COVID-19 and non-COVID news from three key angles. First, we provide truth and falsehood (T/F) indicators of news items, as labeled and validated by several fact-checking platforms (e.g., Snopes and Politifact). Second, we collect spurious-claim-related tweets and retweets from Twitter, one of the world’s largest social networks. Third, we provide basic user information, including the terms and characteristics of “heavy fake news” user to present a better understanding of T/F claims in consideration of COVID-19. FibVID provides several significant contributions. It helps to uncover propagation patterns of news items and themes related to identifying their authenticity. It further helps catalog and identify the traits of users who engage in fake news diffusion. We also provide suggestions for future applications of FibVID with a few exploratory analyses to examine the effectiveness of the approaches used.

1. Introduction

The outbreak of SARS-CoV-2 (COVID-19) has posed major challenges to our society’s health-care systems. Compounding the problem is the effect of the global spread of misinformation since the onset of the pandemic (Depoux et al., 2020). The term, “info-demic” has been employed to warn against the growing phenomenon of fake news related to COVID-19 being disseminated (Roitero et al., 2020).

COVID-19 fake news, which centers on health issues, has significantly impacted people’s behaviors and reduced the efficacy of health-care measures during this pandemic (Apuke and Omar, 2020; Pennycook et al., 2020). For example, many misled people in Nigeria have overdosed on chloroquine, an anti-malarial drug, after being misinformed by the news that it was an effective treatment for COVID-19 (Busari and Adebayo, 2020).

* Corresponding author at: 90310 International Hall, 25-2 Sungkyunkwan-ro, Jongno-gu, Seoul 03063, Korea.
** Co-corresponding author at: 9B303 International Hall, 25-2 Sungkyunkwan-ro, Jongno-gu, Seoul 03063, Korea.
E-mail addresses: cherish@g.skku.edu (Y. Jang), eunilpark@skku.edu (E. Park).

https://doi.org/10.1016/j.tele.2021.101688
Received 28 May 2021; Received in revised form 3 July 2021; Accepted 19 July 2021
Available online 28 July 2021
0736-5853/© 2021 Elsevier Ltd. All rights reserved.
Fake news has become a pervasive issue in the 21st century, especially since the 2016 United States presidential election (Zhou and Zafarani, 2018; Bovet and Makse, 2019; Grinberg et al., 2019). Previous studies have shown that it is necessary to consider both the topics and domains of fake news in order to successfully handle and respond to it (Janicka et al., 2019). Hence, we have leveraged specialized approaches to characterize such news items (Pehlivanoglu et al., 2020) to effectively mitigate its effects (e.g., prediction and correction). Relatedly, Hessain et al. (2020) stated that rapid changes in information require more time and topic-specific datasets for fake-news treatment.

As a way to detect fake news, several prior studies indicated that propagation features are useful for identifying its propagation on social media (Monti et al., 2019; Jin et al., 2014). Notable differences between true-or-false (T/F) news items have been found with several diffusion aspects (e.g., volume, speed, and depth) (Zhou and Zafarani, 2018; Zhou et al., 2020).

Considering the findings of prior studies, our proposed dataset can contribute to propagation pathway detection not only for COVID-19 news, but also for other major topics. In this study, we collect and present a dataset of tweets and retweets from Twitter, as investigated by two well-known fact-checking platforms (i.e., Politifact and Snopes), including the identification of their features and propagation patterns. Moreover, we suggest several opportunities for presenting future research directions for fake-news research.

The remainder of this article is organized as follows. After a literature review, we provide data collection pipelines and descriptions of fake news related to COVID-19. Then, we present discussion points and future research directions, which are then highlighted by the results of exploratory analyses.

2. Related work

The development of a T/F news dataset is needed to help analyze, detect, and prevent the propagation of misinformation. Most datasets from prior studies focused on news content. For example, Wang (2017) compiled the Liar dataset from Politifact, which included 12.8K human-labeled short statements collected through the Politifact application program interface (API). It contained claims, their speakers, contexts, labels, and explanations. The FakeCOVID dataset contained 5,182 T/F claims for COVID-19, as addressed in Shabi and Nandini (2020). Although the capabilities of these datasets provided notable benefits, several limitations remain. Representatively, they do not consider social-network services, which are among the most influential and vigorous content propagation environments in our society.

To address this, several datasets that included social contexts (e.g., posts, user comments, and user data) were compiled. The FakeNewsNet dataset contains both T/F news content and social contexts on political issues, as addressed by Politifact and GossipCop (Shu et al., 2020). This dataset includes 23,196 fake news items, 690,732 tweets, and user information collected through the Twitter API. Other more specialized datasets of COVID-19 issues (i.e., CoAID) presented 4,251 news items, 296,000 related user engagements, 926 social-platform posts, and T/F labels (Cui and Lee, 2020). CMU-MisCOV19 annotated 4,573 tweets and classified them into informed, misinformed, and irrelevant groups (Memon and Carley, 2020). The COVID-19 Dataset provided data based on COVID-19-related keywords from March 23rd to May 13th, 2020 (Inuwa-Dutse and Korkontzelos, 2020).

Although these datasets from their social contexts are useful for fake news detection, several limitations remain. For example, they do not consider data propagation, which is needed to provide more valuable insights into the spread of fake news. User behavior and perspectives toward rumor propagation have long been explored. Most scholars in this field have shown that there is a notable relationship between user behavior and rumor propagation (Guess et al., 2018; Choi et al., 2020; Hui et al., 2018). However, their datasets have only partially been made available, posing limitations on further studies.

In this regard, as shown in Table 1, there are no datasets that provide sufficient features of news content, social context, and propagation. The ones that came close considered either COVID-19 topics or general topics, but not both. To provide valuable insights into the contextual cues and the aspects of social engagement, we collected COVID-19 and non-COVID T/F news datasets during the pandemic period with a variety of features, FibVID (Fake news information-broadcasting dataset of COVID-19). FibVID is a collection of falsified and verified data sources related to COVID-19 and non-COVID news. FibVID is publicly available.\(^1\)

3. Data collection pipeline

In this study, we provide T/F news-item tagging alongside propagation data from Twitter related to the COVID-19 pandemic. Our data collection pipeline, as described in Fig. 1, consists of three stages: news claim collection; claim propagation collection; and user information collection. In this section, we describe the collection process for each stage.

3.1. News claim collection

We began our data collection process by crawling news claims published by two popular fact-checking services, Politifact and Snopes, where news claims are released alongside states of veracity, as investigated and labeled by experts (Monti et al., 2019). We developed a claim crawler to collect this information from these services. The collected news claims were forwarded to a keyword extractor to select terms that can be used to assess their propagation. The extractor tokenized the given claims, removed stop-words from the tokens, and extracted the top-k tokens using importance scores computed by TextRank (Mihalcea and Tarau, 2004). We used

\(^1\) http://doi.org/10.5281/zenodo.4441377.
the maximum number of keywords for a given claim (10) but only considered those having more than three.

3.2. Claim propagation collection

Here, we collect the fully propagated set of news items, including initial tweets and their associated retweets on Twitter. Three modules are employed, a tweet retriever, an original tweet finder, and a retweet crawler. Using Tweepy and the set of extracted keywords from a given news item, we retrieve the relevant tweets and their metadata, including the tweet identifier (tweet ID), the user screen name, the timestamp, and the origin tweet ID. The original tweet IDs of the retweets are then fed into Tweepy to collect the initial tweets. Finally, we collected all retweet data stemming from the original tweets using our own retweet crawler.

3.3. User information collection

Finally, the user profile crawler receives the screen names of the users of the collected tweets/retweets. Using those screen names, the user-profile crawler gathers profile information using Tweepy.

4. Data preprocessing and description

In this section, we explain the preprocessing method of the news claims and their associated tweets and retweets to enhance the feasibility of the research objective and to provide a description of the dataset.

4.1. Claim labeling

Snopes and Politifact use label sets to offer measures of veracity for news claims. In particular, Politifact provides six labels (i.e., “pants-on-fire,” “false,” “barely-true,” “half-true,” “mostly-true,” and “true”), whereas Snopes offers fourteen (i.e., “true,” “mostly-true,” “mixture,” “mostly-false,” “false,” “unproven,” “outdated,” “mispersioned,” “correct-attribute,” “misattributed,” “scam,”

| Table 1 |

| Data Description | Data Features |
|------------------|--------------|
| Period | COVID or non-COVID | News Content | Social Context |
| | Claim | Tweet | User | Propagation Data |
| Liar (Wang, 2017) | 01/01/2007-12/31/2016 | non-COVID | ✓ | – | – | – |
| FakeNewsNet (Shu et al., 2020) | 01/13/2016-05/21/2017 | non-COVID | ✓ | ✓ | ✓ | – |
| FakeCOVID (Shahi and Nandini, 2020) | 04/01/2020-15/05/2020 | COVID | ✓ | – | – | – |
| CoAID (Cui and Lee, 2020) | 02/01/2020-05/17/2020 | COVID | ✓ | ✓ | – | – |
| CMU-MisCOVID (Memon and Carley, 2020) | 3 days in 2020 (03/29, 6/15, and 6/24) | COVID | ✓ | – | – | – |
| COVID-19 Dataset (Inuwa-Dutse and Korkontzelos, 2020) | 03/23/2020-05/13/2020 | COVID | ✓ | ✓ | – | – |
| FibVID | 02/01/2020-12/31/2020 | COVID & non-COVID | ✓ | ✓ | ✓ | ✓ |
“legend,” “labeled-satire,” and “lost-legend”). Because such inconsistencies make it difficult for research communities to use these dataset together, we reassigned labels according to the claims, based on the findings of Khan et al. (2019). Thus, the claims labeled as “pants-fire,” “false,” “barely true,” “mostly-false,” “misattributed,” “mispacedioned,” “scam,” and “labeled-satire” are classified as “fake,” whereas those with other labels are classified as “true.” We did not consider claims that had mixtures or unproven labels from Snopes.

Tags are provided for easier categorization after fact-checking. To label correct news-topic issues, we classified them as “COVID” if the claim was tagged as such, and we classified them as “non-COVID” in all other cases. Then, we provided claim numbers for easier identification of topics. Lastly, we divided news claims into four groups (i.e., 0: COVID true; 1: COVID fake; 2: non-COVID true; and 3: non-COVID fake; Table 2).

4.2. Constructing a propagation tree

As shown in Fig. 2 (Won et al., 2020), to indicate the parent tweets in reference to tweet ID, we denote the parent user and parent id. We mark them as the NaN value if the tweet is the original one. Additionally, we provide depth, which is information about the number of steps that the retweet has traveled from the original tweet.

4.3. Dataset description

From the data collection and preprocessing steps described in the previous sections, we gathered 1,353 news claims and 221,253 relevant tweets written by 144,741 users, as summarized in Tables 3 and 4. In addition to providing the collected information, we also computed the values for a few fields, including pseudonymized user IDs, depths of claim propagation, and text similarity (i.e., original tweet). Text similarity indicates how the texts of news claims are similar to those of the original tweets. The word2vec approach considers similar words as close words. It allows each word to maintain its own meaning and estimates the syntax similarity among words. Then, it calculates the word2vec vector of each claim and tweet text. With these values, we compared the cosine similarity of the vectors(Fig. 3). Fig. 4 shows the cumulative distribution of tweets in all groups over time.

5. Exploratory analysis

We conducted exploratory analyses to demonstrate dataset quality. We analyzed the dataset from the perspective of news claims, claim propagation, and user information about COVID T/F and non-COVID T/F news.

5.1. News claim

To confirm whether or not the differences between fake and true claims are related to topics, we compared them as COVID and non-COVID issues. We utilized the latent Dirichlet allocation (LDA) algorithm (Blei et al., 2003), which is a topic-clustering algorithm that groups co-occurring words into a single topic. We referenced the coherence score to determine the optimal number of topics. Then, we grouped the LDA outputs into topics based on their meanings. The topics, keywords, and representative claims are listed in Tables 5 and 6. Additionally, we compared the distributions of T/F claims in COVID and non-COVID issues (Fig. 5). We found that the topics addressed by fake claims and true claims were similar.

5.2. Claim propagation

This study examined volume, virality, and depth differences of the T/F claims of COVID and non-COVID issues. We established social networks of claims to calculate the properties for each group (Fig. 2). The origin tweet is the first for each claim. The root node indicates the original tweet of a claim. The other nodes indicate retweets that subsequently spread the claim. Edges indicate the relationships between the origin tweet and retweets.

The depth represents the number of steps in which the claims are retweeted within a network. The volume denotes the number of nodes in the network model. The virality of a network indicates the number of users that prompted others to join the conversation (Kim et al., 2020). In this study, the ‘directed network virality’ of a tweet tree was measured using the Wiener index (Dobrynin et al., 2001; Garas et al., 2017), which is calculated as the average distance between all pairwise tweets in a propagation using the following equation (Goel et al., 2016):

| Table 2 |
|---|---|---|
| Issues | COVID | non-COVID |
| Veracity | True | 0 | 2 |
|        | Fake | 1 | 3 |
Fig. 2. Network model (e.g. Volume: 8, Virality: 0.267; $N_0$ is an origin tweet and $N_1$ is a parent tweet of $N_4$).

Table 3
Description of the datasets.

| Part                  | Features                                                                 | Size  |
|-----------------------|---------------------------------------------------------------------------|-------|
| News claim            | claim number, group, text, source                                          | 1,353 |
| Claim propagation     | claim number, parent user, parent ID, tweet user, tweet ID, retweet count, post text, hashtag, like count, create date, group, depth | 221,253|
| Origin tweet          | claim number, tweet user, tweet ID, retweet count, post text, hashtag, like count, create date, group, similarity | 1,774 |
| User information      | user ID, create date, description, follower count, following count        | 144,741|

Table 4
Statistics of the collected datasets.

|                   | COVID (772 claims) | non-COVID (581 claims) |
|-------------------|--------------------|------------------------|
|                   | True               | Fake                   | True               | Fake               |
| Claims            | 203 (26.29%)       | 569 (73.70%)           | 150 (25.81%)       | 431 (74.18%)       |
| Claims on Twitter | 43                 | 141                    | 30                 | 81                 |
| Tweet count       | 27,296             | 133,374                | 12,793             | 47,790             |
| Avg. tweets per claim | 634.79          | 945.91                 | 426.43             | 590.00             |
| Avg. origin tweets per claim | 6.28         | 6.37                   | 4.10               | 5.96               |
| # of unique users | 23,209             | 95,599                 | 11,454             | 39,887             |

Fig. 3. Cosine similarity results of claim and origin tweet Word2Vec (COVID True ($M = 0.554, SD = 0.179, SE = 0.010$), COVID Fake ($0.520, 0.198, 0.006$), non-COVID True ($0.578, 0.203, 0.018$), non-COVID Fake ($0.574, 0.209, 0.009$)).
$v(T) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}, \quad (1)$

where $v(T)$ is defined as the average distance between all pairs of nodes (Goel et al., 2016), and $T$ is the diffusion tree. $d_{ij}$ denotes the length of the shortest path between node $i$ and $j$.

We conducted a multivariate analysis of variance to analyze the effects of veracity (fake vs. true) and issue groups (COVID vs. non-COVID) of claims on their diffusion. However, there were no significant differences between fake and true claims ($p > 0.05$). Moreover, there were no significant differences between COVID and non-COVID claims in terms of volume and virality ($p > 0.05$; Fig. 6, and Fig. 7). However, there was a significant difference in tree depth between COVID ($M = 2.19, SD = 1.59, SE = 0.11$) and non-COVID claims ($M = 1.69, SD = 1.34, SE = 0.12$); $F(1, 291) = 14.83, p < .05$; Fig. 8). No interaction effect between veracity and issue groups was observed ($p > .05$).

We visualized network models of representative claims from each group: COVID true #69 claim (Fig. 9), COVID fake #144 claim...
Representative claims were selected based on their volume, depth, and virality. For instance, the #69 claim was selected because its volume, depth, and virality were closest to the average values of COVID true claims.

Fig. 5. Distributions of topics for true and fake news in COVID and non-COVID issues.

Fig. 6. Virality results (COVID True ($M = 0.019$, $SD = 0.053$, $SE = 0.008$), COVID Fake (0.021, 0.049, 0.004), non-COVID True (0.030, 0.053, 0.009), non-COVID Fake (0.020, 0.051, 0.005)).

Fig. 7. Volume results (COVID True ($M = 634.79$, $SD = 828.99$, $SE = 127.91$), COVID Fake (945.91, 21,84.57, 184.63), non-COVID True (426.43, 896.36, 166.45), non-COVID Fake (590.00, 799.37, 89.37)).

(Fig. 10), non-COVID true #91 claim (Fig. 11), and non-COVID fake #228 claim (Fig. 12). Representative claims were selected based on their volume, depth, and virality. For instance, the #69 claim was selected because its volume, depth, and virality were closest to the average values of COVID true claims.
5.3. User information

5.3.1. Influence of users

Follower count is widely regarded as a measure of influence for a user (Bakshy et al., 2011). We conducted a multivariate analysis of variance to examine the differences in the influence of users between the veracity group (fake vs. true) and their issues (COVID vs. non-COVID). We found no main effect of either veracity and issue groups. However, a significant interaction effect was found between them: $F(1, 221,249)=7.12, p<.01$ (Fig. 13). Users who diffused COVID true claims ($M=197,506.13, SE=21,322.07$) were found to have a greater influence than those who propagated non-COVID true claims ($M=108,772.50, SE=21,490.46$). In contrast, users who participated in propagating COVID fake claims ($M=156,640.14, SE=8,859.87$) had a greater influence than those who diffused non-COVID fake claims ($M=171,961.42, SE=15,340.62$).

5.3.2. Heavy fake-news user

Among the 144,741 users, 31,768 spread tweets more than twice. Fig. 14 shows a distribution graph of the number of tweets per user. Because detecting users who heavily spread fake news is a critical issue, we briefly identified and analyzed heavy users of fake
news. In this study, heavy fake news users were defined as the top 1% of fake-news disseminators based on the number of tweets, regardless of COVID or non-COVID issues. We drew a graph of heavy fake news users to determine their issue orientation (Fig. 15). The x-axis indicates the ratio of the number of COVID fake news out of the total fake news. Zero indicates that users spread only non-COVID fake news rather than COVID fake news. 0.8 means that users spread 20% non-COVID fake news and 80% COVID fake news. Heavy fake-news users showed a tendency to spread more about COVID issues, which suggests an insight into the diffusion of COVID fake news.

Fig. 10. COVID Fake (Claim #144; volume = 359, virality = 0.007).

Fig. 11. Non-COVID True (Claim #91; volume = 123, virality = 0.015).
6. Suggestion

In this study, we aimed to provide a valuable dataset having several applications for future research. We provided T/F claims for both COVID and non-COVID news issues, their propagation features, and user information. In the final section, we provide potential future research directions using this dataset.

6.1. Explainable fake-news detection during the COVID-19 period

6.1.1. Textual context or linguistic context

Previous studies revealed several theoretical explanations for the characteristics of fake news. For instance, Hancock et al. (2007) pointed out the unique stylometric characteristics of fake news. Buller and Burgoon (1996) revealed that fake content tends to use evasive language and is less structured. With our claim dataset, we utilized stylometric features using several extraction techniques, such as linguistic inquiry and word count, part of speech, and term frequency-inverse document frequency, and implied that stylometric differences played a key role in addressing T/F news during the COVID-19 period.

6.1.2. Propagation

Apuke and Omar (2020) suggested that false content concerning COVID-19 has become more pronounced in social media. Because
our dataset contains several properties of network structure, such as user information and depth of propagation, our COVID-19 fake-news dataset can be employed to develop statistical and computational approaches for detecting specific fake-news types. We also found no differences in propagation patterns between non-topic-specific fake and true news, which contradicts a previous study (Zhao et al., 2020). Because the amount of fake news is relatively smaller than true news in online media, it may be difficult to identify whether a specific information source is fake (Shu et al., 2019). However, we observed a difference between the COVID and non-COVID claims in terms of propagation patterns (Basile et al., 2021; Dinh and Parulian, 2020). Thus, we expect that our dataset may be specifically applied to reveal the characteristics of topic-based fake-news diffusion appearing in distinct propagation patterns based on time and depth.

6.1.3. User traits
Prior studies have indicated that there is a correlation between user profiles and fake-news detection. For example, Shu et al. (2019) suggested that explicit and implicit features from social-media user profiles can help improve fake-news detection. With fake news (COVID-19 or non-COVID) and user traits in our dataset, user traits can be applied as features for detecting fake news.

6.1.4. Comprehensive relationships among texts, propagation, and users
Previous studies provided textual information and/or fragmentary information, such as original tweets and single-depth retweets (Shahi and Nandini, 2020; Wang, 2017; Cui and Lee, 2020). However, such datasets are not sufficiently thorough for deeper and comprehensive analyses. For example, the features of the Liar dataset might be sufficient for textual analysis, but they might not be sufficient to reveal the propagation patterns of fake news (Wang, 2017).
In contrast, our dataset contains textual content as well as tweet and retweet pairs, which contain all retweets related to a tweet during the period of data collection. Furthermore, we expect that our dataset will provide a multidimensional perspective on the issues, because it includes how users react to the information and which users engage in the diffusion. Because the content of information is found to affect user behaviors (Dwivedi et al., 2021), we expect that our dataset will provide opportunities to address complex relations between textual traits of fake news and tendencies of its propagation based on an understanding of human behaviors. Additionally, our dataset can be used to understand the relationship between user traits and fake news propagation at an in-depth level.

6.2. Seeking differences between crisis fake news and general fake news: new category of fake news

Prior to the COVID-19 pandemic, there was no key platform for a major efficient misinformation distribution scheme. Hence, rapid fake-news distribution was made possible by the emergence of social media (Vosoughi et al., 2018). To the best of our knowledge, few studies have been conducted on disease-scenario fake-news distribution. Thus, our dataset provides a pioneering basis for analyzing pandemic fake-news distribution. This dataset is particularly worthy, because it provides pandemic fake-news distribution trends alongside a comparison of general fake-news trends.

We expect that our dataset will provide new perspectives on fake news, especially if it is utilized to investigate COVID-19 and non-COVID fake news using contextual cues, propagation features, and social engagements.

6.3. Knowledge-based fake news detection during the COVID-19 period

Our knowledge-based perspective analyzes and detects fake news automatically using a fact-checking process comprising entity resolution, time recording, and knowledge fusion (Zhou and Zafarani, 2018). To effectively respond to fake news using machine learning, features from a knowledge-based perspective should be considered. Hence, our dataset provides fact-paired information that can be utilized for knowledge graphing and detection. It can also provide training features to provide future knowledge-based detection models.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ICAN (ICT Challenge and Advanced Network of HRD) program (IITP-2020-0-01816) supervised by the IITP (Institute of Information & Communications Technology Planning & Evaluation). This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2020R1C1C1004324).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.tele.2021.101688 and https://doi.org/10.5281/zenodo.4441376.

References

Apuke, O.D., Omar, B., 2020. Fake news and covid-19: modelling the predictors of fake news sharing among social media users. Telematics Inform. 56, 101475.
Bakshy, E., Hofman, J.M., Mason, W.A., Watts, D.J., 2011. Everyone’s an influencer: quantifying influence on twitter. In: Proceedings of the fourth ACM international conference on Web search and data mining, pp. 65–74.
Basile, V., Cauteruccio, F., Terracina, G., 2021. How dramatic events can affect emotionality in social posting: the impact of covid-19 on reddit. Future Internet 13, 29.
Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. J. Mach. Learn. Res. 3, 993–1022.
Bovet, A., Makse, H.A., 2019. Influence of fake news in twitter during the 2016 us presidential election. Nat. Commun. 10, 1–14.
Buller, D.B., Burgoon, J.K., 1996. Interpersonal deception theory. Commun. Theory 6, 203–242.
Busari, S., Adebayo, B., 2020. Nigeria records chloroquine poisoning after trump endorses it for coronavirus treatment. cnn.https://edition.cnn.com/2020/03/23/africa/chloroquine-trump-nigeria-intl/index.html.
Choi, D., Chun, S., Oh, H., Han, J., et al., 2020. Rumor propagation is amplified by echo chambers in social media. Scientific Rep. 10, 1–10.
Cui, L., Lee, D., 2020. Covid-19 healthcare misinformation dataset. https://arxiv.org/abs/2006.00885.
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. J. Travel Med. 27, taas031.
Dinh, L., Parulian, N., 2020. Covid-19 pandemic and information diffusion analysis on twitter. Proc. Assoc. Inf. Sci. Technol. 57, e252.
Dobrynin, A.A., Estrinches, R., Gutman, I., 2001. Wiener index of trees: theory and applications. Acta Applicandae Mathematica 66, 211–249.
Dwivedi, Y.K., Ismagilova, E., Hughes, D.L., Carlson, J., Filieri, R., Jacobson, J., Jain, V., Karjaluoto, H., Kefi, H., Krishen, A.S., et al., 2021. Setting the future of digital and social media marketing research: perspectives and research propositions. Int. J. Inf. Manage. 59, 102168.
Garas, G., Cingolani, I., Panzarasa, P., Darzi, A., Athanasiou, T., 2017. Network analysis of surgical innovation: measuring value and the virality of diffusion in robotic surgery. PLoS One 12, e0183332.
Goel, S., Anderson, A., Hofman, J., Watts, D.J., 2016. The structural virality of online diffusion. Manage. Sci. 62, 180–196.

Grinberg, N., Joseph, K., Friedland, I., Swire-Thompson, B., Lazer, D., 2019. Fake news on twitter during the 2016 us presidential election. Science 363, 374–378.

Guess, A., Nyhan, B., Reifler, J., 2018. Selective exposure to misinformation: evidence from the consumption of fake news during the 2016 us presidential campaign. Eur. Res. Council 4, 1–23.

Hancock, J.T., Curry, L.E., Goorha, S., Woodworth, M., 2007. On lying and being lied to: a linguistic analysis of deception in computer-mediated communication. Discourse Processes 45, 1–23.

Hossain, T., Logan IV, R.L., Ugarte, A., Matsubara, Y., Singh, S., Young, S., 2020. Detecting covid-19 misinformation on social media. https://openreview.net/pdf/ doi/6fa0b8f34afee6795940b2a3796b90a8ab8725.pdf.

Hui, P.-M., Shao, C., Fiammingi, A., Menczer, F., Ciampaglia, G.L., 2018. The hoaxy misinformation and fact-checking diffusion network. In: Proceedings of the International AAAI Conference on Web and Social Media, pp. 528–530.

Inuwa-Dutse, I., Korkontzelos, I., 2020. A curated collection of covid-19 online datasets. https://arxiv.org/abs/2007.09703.

Janicka, M., Pszona, M., Wawer, A., 2019. Cross-domain failures of fake news detection. Computacion y Sistemas 23, 1089–1097.

Jin, F., Wang, W., Zhao, L., Dougherty, E., Cao, Y., Lu, C.-T., Ramakrishnan, N., 2014. Misinformation propagation in the age of twitter. Computer 47, 90–94.

Khan, J.Y., Khondaker, M., Islam, T., Iqbal, A., Afroz, S., 2019. A benchmark study on machine learning methods for fake news detection. https://arxiv.org/abs/1905.04749.

Kim, J., Jeong, D., Choi, D., Park, E., 2020. Exploring public perceptions of renewable energy: Evidence from a word network model in social network services. Energy Strategy Reviews 32, 100552.

Kim, S.A., Carley, K.M., 2020. Characterizing covid-19 misinformation communities using a novel twitter dataset. https://arxiv.org/abs/2008.00791.

Mihalcea, R., Tarau, P., 2004. Textrank: Bringing order into text. In: Proceedings of the 2004 conference on empirical methods in natural language processing, pp. 404–411.

Monti, F., Frasca, F., Eynard, D., Mannion, D., Bronstein, M.M., 2019. Fake news detection on social media using geometric deep learning. https://arxiv.org/abs/1902.06673.

Pehlivanoglu, D., Lin, T., Chi, K., Perez, E., Polk, R., Cabhill, B., Lighthall, N., Ehser, N., 2020. News veracity detection among older adults during the covid-19 pandemic: The role of analytical reasoning, mood, news consumption, and news content. https://pyarxiv.com/3kgp9/.

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J.G., Rand, D.G., 2020. Fighting covid-19 misinformation on social media: experimental evidence for a scalable accuracy-nudge intervention. Psychol. Sci. 31, 770–780.

Roitero, K., Soprano, M., Portelli, B., Spina, D., Delia Mea, V., Serra, G., Mizzaro, S., Demartini, G., 2020. The covid-19 infodemic: can the crowd judge recent misinformation objectively?. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, pp. 1305–1314.

Shahy, G.K., Nandini, D., 2020. Fakecovid—a multilingual cross-domain fact check news dataset for covid-19. https://arxiv.org/abs/2006.11343.

Shu, K., Mahudeswaran, D., Wang, S., Lee, D., Liu, H., 2020. Fakenewsnet: a data repository with news content, social context, and spatiotemporal information for studying fake news on social media. Big Data 8, 171–188.

Shu, K., Wang, S., Liu, H., 2019. Beyond news content: the role of social context for fake news detection. In: Proceedings of the twelfth ACM international conference on web search and data mining, pp. 312–320.

Shu, K., Zhou, X., Wang, S., Zafarani, R., Liu, H., 2019. The role of user profiles for fake news detection. In: Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 436–439.

Yosoughi, S., Roy, D., Aral, S., 2018. The spread of true and false news online. Science 359, 1146–1151.

Wang, W., 2017. “liar, liar pants on fire”: a new benchmark for fake news detection. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 422–426.

Wen, Y., Kang, J., Choi, D., Park, E., Han, J., 2020. Who drives successful online conversations? Unveiling the role of first user response. Kybernetes 49, 876–886.

Zhao, Z., Zhao, J., Sano, Y., Levy, O., Takayasu, H., Takayasu, M., Li, D., Wu, J., Havlin, S., 2020. Fake news propagates differently from real news even at early stages of spreading. EPJ Data Sci. 9, 7.

Zhou, X., Jain, A., Ploha, V.V., Zafarani, R., 2020. Fake news early detection: A theory-driven model. Digital Threats: Research and Practice 1, 1–25.

Zhou, X., Zafarani, R., 2018. Fake news: a survey of research, detection methods, and opportunities. https://arxiv.org/abs/1812.00315.