



Research Article

Leveraging volunteer fact checking to identify misinformation about COVID-19 in social media

Identifying emerging health misinformation is a challenge because its manner and type are often unknown. However, many social media users correct misinformation when they encounter it. From this intuition, we implemented a strategy that detects emerging health misinformation by tracking replies that seem to provide accurate information. This strategy is more efficient than keyword-based search in identifying COVID-19 misinformation about antibiotics and a cure. It also reveals the extent to which misinformation has spread on social networks.

Authors: Hyunuk Kim (1), Dylan Walker (2)

Affiliations: (1,2) Questrom School of Business / Institute for Health System Innovation & Policy, Boston University

How to cite: Kim, H., & Walker, D. (2020). Leveraging volunteer fact checking to identify misinformation about COVID-19 in social media. *The Harvard Kennedy School (HKS) Misinformation Review*, Volume 1, Special Issue on COVID-19 and Misinformation

Received: April 3rd, 2020 Accepted: May 14th, 2020 Published: May 18th, 2020

Research questions

- How can we amplify the efforts of volunteer fact checkers to identify emerging health misinformation on social networks?

Essay summary

- We implemented a strategy that leverages the efforts of volunteer fact checkers and a social network to identify misinformation as it is emerging.
- The strategy starts by identifying replies that have similar context to official advice from health authorities. These replies act as seeds from which misinformation is identified. A state of the art natural language model was used to calculate similarity between replies and official advice.
- With the strategy, we identified COVID-19 misinformation about antibiotics and a cure on Twitter.
- This strategy is more efficient than keyword-based search in identifying tweets containing misinformation and does not require advanced knowledge of the type or manner of misinformation, nor a set of URLs or domains that have been associated with misinformation.
- We observed that misinformation is also present in the upstream (friends) and downstream (followers) peers of the accounts who created posts that were fact checked by others, suggesting that network-oriented strategies can uncover emerging misinformation.
- We suggest a collaborative system that amplifies and is aided by the efforts of volunteer fact checkers in identifying and correcting emerging misinformation on social networks.

¹ A publication of the Shorenstein Center on Media, Politics and Public Policy at Harvard University's John F. Kennedy School of Government.

Implications

The abundance of health-related information in social media facilitates communication between patients and practitioners and reduces information disparity in the public (Moorhead et al., 2013). Apart from the positive effects of social media on health literacy, social media can also be a venue of health misinformation that aggravates users' misperceptions on health issues (Sharma et al., 2017). Therefore, identifying health misinformation online is necessary to mitigate public health threats in the digital age (Chou et al., 2018). Recent research has identified multiple methods to detect misinformation in general (Conroy et al., 2015) and more specifically, health misinformation (Ghenai & Mejova, 2018; Dhoju et al., 2019). However, these methods are not well adapted for identifying emerging misinformation, which may be critical for rapid intervention and policy response that can affect health behaviors, because existing methods rely on intensive manual labeling or known misinformation sources (such as domains, URLs or accounts). The problem is that the manner and type of emerging health misinformation are often unknown. Moreover, misinformation can be buried amongst a large volume of accurate information. We are looking for a needle in a haystack.

How then can emerging health misinformation be identified in a timely manner? We suggest an approach that leverages the nature of social media, network structure, and the efforts of volunteer fact checkers. The approach uses semantic textual similarity with sources of accurate information (from verified sources) to identify replies that are likely intended as fact checks (posted by volunteer or casual fact checkers) in response to a parent post. If such replies are indeed accurate fact checks, then their parent posts are more likely to contain emerging misinformation. Furthermore, the local networks surrounding fact checked parent posts may also be more likely to contain misinformation (Figure 1). Misinformation may take many forms and can be challenging to detect when we lack concrete data on the manner and type of misinformation that is emerging. However, officially sanctioned sources of bona fide accurate information are often readily available. For example, in the case of COVID-19, officially sanctioned information sources include the World Health Organization (WHO) and the Centers for Disease Control and Prevention (CDC).

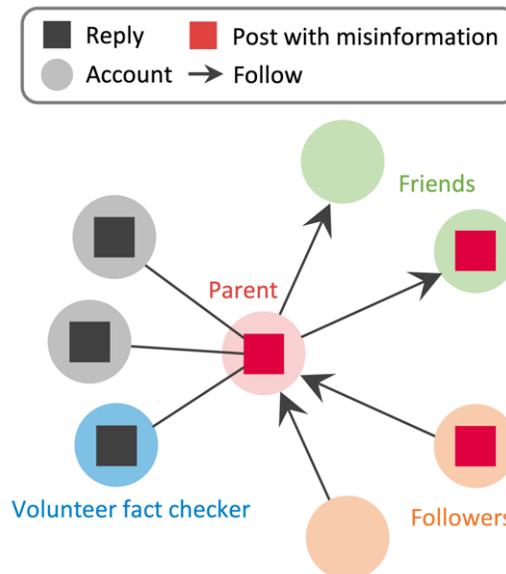


Figure 1. Schematic of our strategy that identifies misinformation online by tracking footprints of volunteer fact checkers. Starting from the volunteer fact checker (blue node) who provides accurate information, we can first identify misinformation in the parent post (red node) and then further detect misinformation in the upstream (friends; green nodes) and downstream (followers; orange nodes) peers of the parent.

The approach we describe here reduces the challenge of discovering arbitrary misinformation to that of identifying fact checked reply posts containing accurate information. Such posts act as seeds and indicate areas of a social network where misinformation is likely harbored. This approach substantially reduces the search space for identifying pertinent misinformation. To accomplish this, we collected public reply posts containing context-specific keywords (e.g., COVID-19) and calculated semantic textual similarity with official advice provided by a health authority. We then collected parent posts of replies of high similarity. A similar strategy was used by Vosoughi et al. (2018), who found rumors from replies linking to fact checking articles. However, our approach is not limited to specific fact check URLs or social media posts that were already labeled as misinformation by professional fact checkers. Furthermore, our strategy exploits the structure of the local networks surrounding fact checked parent posts while many existing approaches rely on linguistic features of misinformation (Conroy et al., 2015) or misinformation sharing behaviors (Shao et al., 2018).

We implemented this strategy on Twitter to identify misinformation about COVID-19. Two topics, antibiotics and a cure, were selected to check the applicability of our strategy in narrow and broad contexts. The query to collect relevant replies and the official advice that was used as seed of our strategy are provided in the Methods section.

1) Antibiotics: The claim “antibiotics are effective in preventing and treating the new coronavirus” is wrong because antibiotics do not work on viruses and furthermore antibiotic misuse has serious consequences for antimicrobial resistance, which contributes to the emergence of superbugs, as well as individual health. Unfortunately, this misinformation continues to spread on Twitter, despite public health education efforts.

2) A cure: The WHO mentioned that “While some western, traditional, or home remedies may provide comfort and alleviate symptoms of COVID-19, there is no evidence that current medicine can prevent or cure the disease.” We used the argument in the WHO’s webpage as official advice of our strategy to identify misinformation about COVID-19 and a cure. Note that we do not know what substances have been suggested as cures and do not presume the type of misinformation.

Our strategy uncovered tweets containing misinformation about COVID-19 for both topics with high signal-to-noise ratio (SNR) (i.e., the fraction of tweets containing misinformation) compared to naive keyword-based search that inspects misinformation in a set of tweets containing context-specific keywords. While our strategy seems to work well for the topics we studied, more research is needed to assess how it generalizes to other cases. We plan to test the performance of our strategy more comprehensively for diverse types of misinformation in future research.

Our strategy and findings have clear implications for organizations engaged in fact checking, for social media platform providers, and for academic researchers. Many organizations that practice professional fact checking do not systematically search social media for misinformation, perhaps because naive search methods yield a low signal-to-noise ratio. We suggest such organizations leverage our strategy and harness the wisdom of the crowds to enhance discovery of misinformation spreading on social media and lower search costs. This is important because the impact of fact-checking efforts may be hampered if they are unable to reach (potentially vulnerable) subpopulations or regions of social networks where misinformation is harbored. Notably, volunteer fact checkers have been found to be as effective as platform governed efforts in correcting health misinformation (Bode & Vraga, 2018).

Social network providers could also benefit from our findings. For example, platform providers may complement their Application Programming Interfaces (APIs) by making new tools or API endpoints available that enable discovery of and response to misinformation. We suggest that such tools or API endpoints might leverage content seeds to permit searching content within local network neighborhoods. Access to such tools could be governed by platform providers to prevent abuse. Alternatively, platforms could provide higher level offerings, such as targeted anti-misinformation campaigns (i.e., better targeted public service announcements).

Researchers could leverage our strategy to locate and better understand subpopulations where misinformation emerges. Once emerging misinformation has been identified, posting history and demographic information of the subpopulations can be used to infer personal traits (Qiu et al., 2012) associated with the spread of misinformation and to understand how misinformation reaches and negatively affects different groups. In this way our strategy and findings can aid practitioners and policy makers in designing targeted policies that reduce adverse effects of misinformation on society.

Findings

Finding 1: Misinformation about COVID-19 and antibiotics is classified into one of the following four categories: (1) antibiotics work against COVID-19, (2) antibiotics can treat viral pneumonia caused by COVID-19, (3) people can be resistant to antibiotics, and (4) other wrong claims including conspiracy theories.

We identified 58 out of 200 (SNR=0.29) tweets that incorrectly claimed that “antibiotics are effective against COVID-19 and viral infections caused by the new coronavirus.” To keep our estimates of the performance of our strategy conservative, we considered tweets containing general misconceptions of antibiotics or those asking medical advice containing incorrect presumptions as *non-misinformation* because they are out of the research scope. Example tweets are listed in Table 1.

Table 1. A list of identified tweets that contain misinformation about COVID-19 and antibiotics.

Category	Sample tweet
1	Scientists claim antibiotics already on the market can treat coronavirus
2	Corona virus causes Pneumonia. It’s the pneumonia and related conditions that kill you. Pneumonia is caused by a bacterial infection. It is treated with antibiotics. Is there enough antibiotics to treat the projected 7 million people who will need it?
3	#coronaviruss #CoronaVirusUpdates I’m stocking up on food, disposable gloves, N95 masks, and Purell. And I haven’t been on antibiotics in 35 years so they will be very effective.
4	If everyone took conventional antibiotics (to be safe from this #coronavirus) right now, I have this feeling that the stats would be much lower. First, it would boost the immune system and 2nd would pre kill the bacteria that the virus would feed on. That is the prevention & cure

The first piece of misinformation (Category 1) was created by the Daily Mail, which has been rated in the past as an untrustworthy source via crowdsourcing techniques (Pennycook & Rand, 2019). The associated article refers to and misinterprets scientific research with a manipulated figure, resulting in significant misconceptions about the use of antibiotics against COVID-19. The Daily Mail tweeted this piece of misinformation several times. Other tweets that we identified did not originate from a mass media source, but contain what appears to be individually-generated misinformation that reflects a poor understanding of antibiotics. The second piece of misinformation (Category 2) argued that pneumonia caused by the new coronavirus can be treated with antibiotics, which is incorrect as antibiotics are not effective against viral pneumonia. The third piece of misinformation (Category 3) implied that people can become resistant to antibiotics, which is not true. Rather, microbes can develop resistance to antibiotics. The fourth piece of misinformation (Category 4) argued that conventional antibiotics can be used as preventive drugs against COVID-19.

Finding 2: Misinformation about COVID-19 and a cure is classified into one of the following four categories: (1) alternative medicines or remedies are effective against COVID-19, (2) governors banned the prescription of hydroxychloroquine, (3) hydroxychloroquine definitively cures COVID-19, and (4) other wrong claims including conspiracy theories.

29 out of 200 (SNR=0.145) tweets that contain misinformation about a cure were identified. Similar to the previous case, we considered tweets asking questions about the effectiveness of medicines or substances against COVID-19 as *non-misinformation* because those tweets do not explicitly contain misinformation of our interest. This classification keeps our estimates of the performance of our strategy conservative. Example tweets are listed in Table 2.

Table 2. A list of identified tweets that contain misinformation about COVID-19 and a cure.

Category	Sample tweet
1	Sir, We can try some ayurvedic, Siddha, homeopathy medicine to recover the patients. Please sir discuss with some leading ayurvedic, Siddha, homeopathy doctors across our country and take the action soon. Let's fight together and lower our corona counts.
2	I am considering launching a lawsuit against Governor Sisolak of Nevada for overstepping his executive boundaries and banning the use of Hydroxychloroquine to treat COVID19 patients, even if doctors recommend it. Are there any lawyers that can help with this? Please DM.
3	I'M CALLING THIS OVER!!! Clinical trial results WILL prove combination therapy (H + A) eliminates COVID-19! Please continue preventative measures until #POTUS gives all clear. Studies all sorted by early April. Supplies secured and life back to normal by Easter
4	reminder that Cuba has an effective treatment for COVID-19 but the US is deluding the public into believing that they are actively searching for a cure/that there isnt one. there is an effective medicine but we have a blockade on Cuba.

The first piece of information (Category 1) advocates for alternative and traditional remedies against COVID-19, none of which has been proven to be effective. Other tweets in this category claimed that Vitamin C and green tea extract can treat COVID-19. The second piece of misinformation (Category 2) argued that the Governor of Nevada banned the prescription of hydroxychloroquine. This misinformation has already been debunked by [PolitiFact](#) which correctly points out that the Governor of Nevada (and New York) restricted unnecessary access to hydroxychloroquine to prevent stockpiling, but did not in any way restrict physicians' use of the drug. The third piece of misinformation (Category 3) showed a strong belief that hydroxychloroquine perfectly cures COVID-19, which is not consistent with evidence, and which also does not acknowledge its side effects, which can be fatal. The effect of hydroxychloroquine is still controversial. The fourth piece of misinformation (Category 4) raises a conspiracy theory that claims that Cuba has an effective medicine that is concealed by the U.S. government. Some other tweets in this last category claimed that a cure for COVID-19 was already developed but not distributed.

Finding 3: Our strategy is more efficient than keyword-based search in identifying COVID-19 misinformation about antibiotics and a cure. The strategy also shows that misinformation tends to be harbored in the local networks surrounding accounts that posted misinformation.

We compared the performance of our approach with that of keyword-based search which examines misinformation in a set of posts containing context-specific keywords. Keyword-based search was implemented by sampling 200 non-reply tweets having context-specific keywords to match the sample size of our strategy. We examined whether these tweets contained misinformation about COVID-19 and the two selected topics, antibiotics and a cure. From keyword-based search, we found 28 tweets containing misinformation about antibiotics (SNR=0.14) and 15 tweets containing misinformation about a cure (SNR=0.075). To evaluate the statistical significance of the SNR difference between our strategy and keyword-based search for both topics, we resampled 100 non-reply tweets from the 200 tweets 10,000 times and obtained the fraction of the resampled sets which return a higher SNR than our strategy (i.e., p-value). We found that our strategy is statistically significant and relatively efficient at discovering misinformation about the selected topics (SNR=0.29 for antibiotics, p-value=0.0001; SNR=0.145 for cure, p-value=0.0066).

Another advantage of our approach is that it identifies pockets in the social network where misinformation resides. It is well known that individuals tend to form social relationships with those who have characteristics similar to themselves -- i.e., "Birds of a feather flock together", a phenomenon that is commonly referred to as homophily (McPherson et al., 2001). Homophily can affect not only the structure of social networks, such as links between friends and friends of friends (Kossinets & Watts, 2009), but also how individuals are exposed to information generally (Bakshy et al., 2015), specifically for health (Centola & van de Rijt, 2015) and political information (Colleoni et al., 2014; Barberá et al., 2015; Halberstam & Knight, 2016), and can lead to echo chambers that reinforce existing beliefs (Garrett, 2009; Flaxman et al., 2016). If there is homophily in the tendency to spread or harbor misinformation, then upstream (friends) and downstream (followers) peers of accounts who posted tweets that were fact checked should be more likely to contain misinformation. To check whether and to what extent this is true, we investigated tweets about the selected topics in the up- and downstream account timelines and found moderate to high misinformation proportions (Figure 2), suggesting that users are homophilous in their tendency to harbor and spread misinformation, consistent with prior research (Del Vicario et al., 2016). By using fact checked posts as seeds on social networks, we expect to identify more misinformation candidate posts and reconstruct the network backbone of misinformation to reveal vulnerable subpopulations.

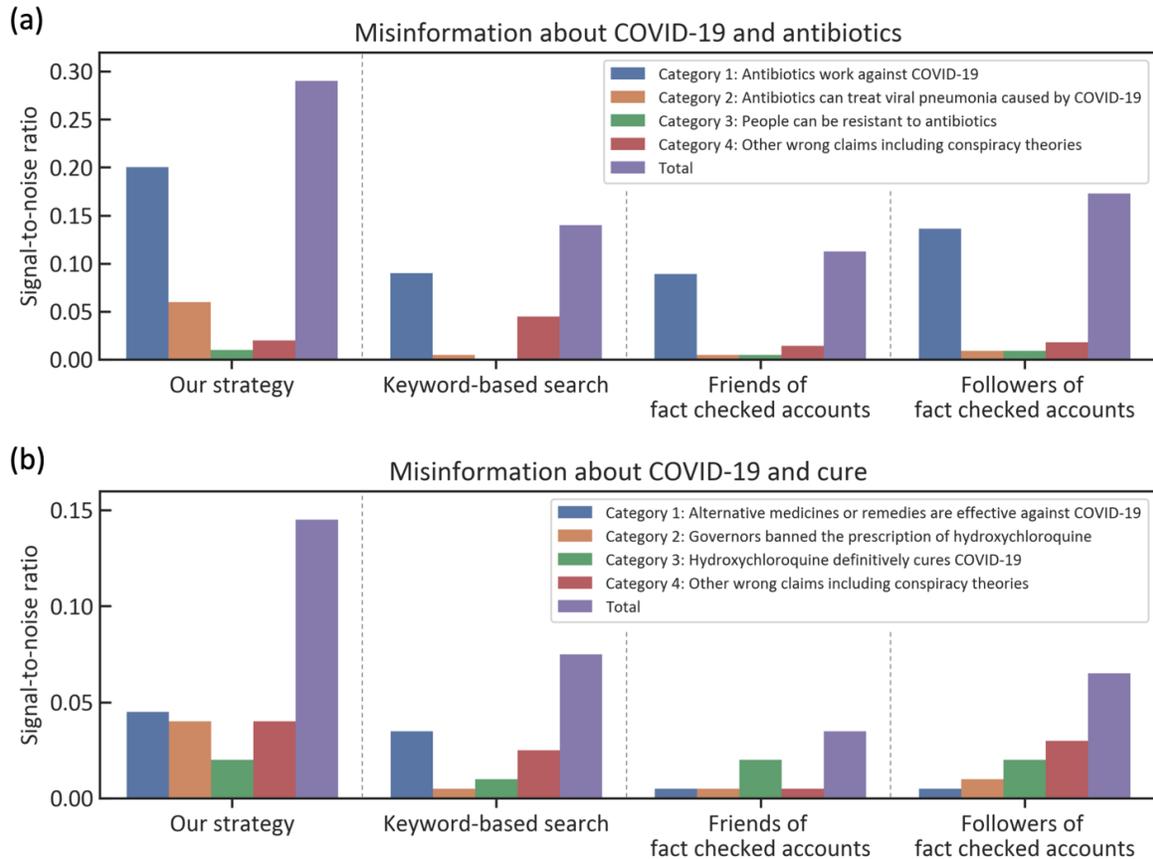


Figure 2. Signal-to-noise ratio (SNR), which is the fraction of misinformation posts, of our strategy and keyword-based search and in the local networks surrounding accounts that posted misinformation. Our strategy identified more tweets containing COVID-19 misinformation about (a) antibiotics (p -value=0.0001) and (b) cure (p -value=0.0066) than keyword-based search. From the timelines of friends and followers of the accounts who tweeted misinformation, we observed that misinformations was also present in the upstream (friends) and downstream (followers) peers of the accounts.

Methods

We collected 16,383 public tweet replies (in English) related to COVID-19 and antibiotics by querying “(corona OR virus OR coronavirus OR covid19 OR covid-19 OR 2019-ncov OR wuhanvirus OR (wuhan AND virus)) AND (antibiotic OR antibiotics)” in the Twitter search with the search period from January 1 to March 31, 2020. The reason we include the word “virus” in the query is that “virus” in tweet replies can represent the new coronavirus in parent tweets. COVID-19 related keywords in the query were chosen by the authors as they are terms related to the new coronavirus and were searched frequently after the COVID-19 outbreak according to Google Trends. The parent tweets of these tweet replies were collected if they were not self-replies, were written in English, and matched with the query “(corona OR coronavirus OR covid19 OR covid-19 OR 2019-ncov OR wuhanvirus OR (wuhan AND virus)) AND (antibiotic OR antibiotics)” (in order to restrict discourse to the topics of COVID-19 and antibiotics). Note that the word “virus” is not included in this query. This yielded 573 parent-reply tweet pairs (441 unique parents). The official advice of the WHO that we used as a proxy of accurate information is “No, antibiotics do not work against viruses, only bacteria. The new coronavirus (2019-nCoV) is a virus and, therefore, antibiotics

should not be used as a means of prevention or treatment. However, if you are hospitalized for the 2019-nCoV, you may receive antibiotics because bacterial co-infection is possible.” (retrieved from [the official WHO website](#)).

Similarly, to identify misinformation about COVID-19 and a cure, we collected 152,175 public tweet replies (in English) by querying “(corona OR virus OR coronavirus OR covid19 OR covid-19 OR 2019-ncov OR wuhanvirus OR (wuhan AND virus)) AND (medicine OR remedy OR cure OR treatment)” in the Twitter search with the search period from March 16 to March 31, 2020. The parents of these tweet replies were collected if they were not self-replies, were written in English, and satisfied the query “(corona OR coronavirus OR covid19 OR covid-19 OR 2019-ncov OR wuhanvirus OR (wuhan AND virus)).” The official advice of the WHO that we used in the strategy is “While some western, traditional or home remedies may provide comfort and alleviate symptoms of COVID-19, there is no evidence that current medicine can prevent or cure the disease.” (retrieved from [the official WHO website](#)).

To extract tweet replies likely providing accurate information about COVID-19 and the selected topics, we implemented sentence embedding, which can capture the context of text well relative to traditional approaches that rely only on keyword overlap (Camacho-Collados & Pilehvar, 2018). Sentence embeddings project text content onto a low-dimensional space and enable us to quantify the similarity between two arbitrary sentences. Among various sentence embedding models available, we chose “Sentence-BERT”, a state-of-the-art model that has been shown to achieve good performance in measuring semantic textual similarity (Reimers & Gurevych, 2019). We used the pretrained model “bert-base-nli-mean-tokens”, which is available through [a publicly accessible Python library](#). Mentions, emojis, and URLs in tweet replies were removed before feeding text into the model. We next calculated the cosine similarity between each reply vector and the vector of the official advice. High cosine similarity indicates that a reply would have similar context to the official advice. We manually examined parent-reply tweet pairs in descending order of cosine similarity until we inspected the first 200 unique parents to determine if they contained misinformation about COVID-19. To capture the extent to which misinformation is found in the local social network region surrounding fact checked parent posts, we fetched the timelines of friends and followers of the accounts who posted misinformation. Due to API constraints and to focus on personal users (who are less likely to be known misinformation spreaders), we selected only the accounts with less than 10,000 friends and less than 10,000 followers. For the topic about antibiotics, 164 out of 54,786 friends and 96 out of 57,712 followers created 214 and 110 topically relevant tweets respectively. For the topic about a cure, 1,163 out of 17,285 friends and 807 out of 26,547 followers created 2,699 and 1,436 topically relevant tweets respectively.

Keyword-based search, an alternative approach to identify misinformation online, was implemented for the selected topics by retrieving 8,769 non-reply tweets about antibiotics (in English; January 1 - March 31, 2020) that satisfy the query “(corona OR coronavirus OR covid19 OR covid-19 OR 2019-ncov OR wuhanvirus OR (wuhan AND virus)) AND (antibiotic OR antibiotics)” and 195,321 non-reply tweets about a cure (in English; March 16 - March 31, 2020) that satisfy the query “(corona OR coronavirus OR covid19 OR covid-19 OR 2019-ncov OR wuhanvirus OR (wuhan AND virus)) AND (medicine OR remedy OR cure OR treatment)”. For each topic, we analyzed a random sample of 200 non-reply tweets to determine if they contained misinformation. The sample size of 200 was chosen to ensure that this sample was directly comparable to the sample obtained by our proposed strategy.

Limitations

Our approach is not without its limitations. First, it requires well-defined keywords to obtain a sufficient number of relevant candidate posts and sufficient official advice to minimize the variance of similarity of replies. If official advice or correct topically relevant information is not available or sufficiently substantive, then our strategy could fail to reliably identify the fact checking posts it uses as seeds. Practitioners who

want to identify emerging health misinformation with our strategy are recommended to set proper context-specific keywords and official advice. We created [a public repository](#) with Python scripts to make our strategy available and accessible to readers who seek to identify misinformation for their topic and context of interest.

Second, our strategy does not comprehensively detect all topically relevant emerging misinformation. For example, misinformation that emerges beyond the local network regions where volunteer fact checkers have posted corrections will not directly be discovered. However, our strategy can be employed as part of a larger systematic approach, to amplify the efforts of volunteer fact checkers. In particular, when posts containing misinformation are identified, they may be incorporated into our strategy as new seeds, allowing for the discovery of misinformation in new regions of the social network.

Finally, we have not comprehensively evaluated the generalizability of our strategy to all contexts and topics of emerging misinformation. More research is warranted to determine when, under what circumstances, and to what extent it is more effective at discovering emerging misinformation than alternative approaches.

Bibliography

- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130-1132.
- Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber?. *Psychological Science*, 26(10), 1531-1542.
- Bode, L., & Vraga, E. K. (2018). See something, say something: Correction of global health misinformation on social media. *Health Communication*, 33(9), 1131-1140.
- Camacho-Collados, J., & Pilehvar, M. T. (2018). From word to sense embeddings: A survey on vector representations of meaning. *Journal of Artificial Intelligence Research*, 63, 743-788.
- Centola, D., & van de Rijt, A. (2015). Choosing your network: Social preferences in an online health community. *Social Science & Medicine*, 125, 19-31.
- Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of Communication*, 64(2), 317-332.
- Conroy, N. J., Rubin, V. L., & Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology*, 52(1), 1-4.
- Chou, W. Y. S., Oh, A., & Klein, W. M. (2018). Addressing health-related misinformation on social media. *JAMA*, 320(23), 2417-2418.
- Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3), 554-559.
- Dhoju, S., Main Uddin Rony, M., Ashad Kabir, M., & Hassan, N. (2019, May). Differences in health news from reliable and unreliable media. In *Companion Proceedings of The 2019 World Wide Web Conference* (pp. 981-987).
- Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, 80(S1), 298-320.
- Garrett, R. K. (2009). Echo chambers online?: Politically motivated selective exposure among Internet news users. *Journal of Computer-Mediated Communication*, 14(2), 265-285.
- Ghenai, A., & Mejova, Y. (2018). Fake cures: user-centric modeling of health misinformation in social media. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), 1-20.

- Halberstam, Y., & Knight, B. (2016). Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter. *Journal of Public Economics*, 143, 73-88.
- Kossinets, G., & Watts, D. J. (2009). Origins of homophily in an evolving social network. *American Journal of Sociology*, 115(2), 405-450.
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1), 415-444.
- Moorhead, S. A., Hazlett, D. E., Harrison, L., Carroll, J. K., Irwin, A., & Hoving, C. (2013). A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. *Journal of Medical Internet Research*, 15(4), e85.
- Pennycook, G., & Rand, D. G. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences*, 116(7), 2521-2526.
- Qiu, L., Lin, H., Ramsay, J., & Yang, F. (2012). You are what you tweet: Personality expression and perception on Twitter. *Journal of Research in Personality*, 46(6), 710-718.
- Reimers, N., & Gurevych, I. (2019, November). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 3973-3983).
- Shao, C., Hui, P. M., Wang, L., Jiang, X., Flammini, A., Menczer, F., & Ciampaglia, G. L. (2018). Anatomy of an online misinformation network. *PLoS One*, 13(4), e0196087.
- 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.
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151.

Funding

The research in this publication was made possible by support from the Boston University Social Innovation on Drug Resistance (SIDR) Postdoctoral Program.

Competing interests

The authors declare no competing interests.

Ethics

Institutional review is unnecessary because the authors analyzed public tweets.

Copyright

This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are properly credited.