**ARTICLE**

# Misinformation, manipulation, and abuse on social media in the era of COVID-19

**Emilio Ferrara[1] · Stefano Cresci[2] · Luca Luceri[3]**

## Abstract

The COVID-19 pandemic represented an unprecedented setting for the spread of online misinformation, manipulation, and abuse, with the potential to cause dramatic real-world consequences. The aim of this special issue was to collect contributions investigating issues such as the emergence of infodemics, misinformation, conspiracy theories, automation, and online harassment on the onset of the coronavirus outbreak. Articles in this collection adopt a diverse range of methods and techniques, and focus on the study of the narratives that fueled conspiracy theories, on the diffusion patterns of COVID-19 misinformation, on the global news sentiment, on hate speech and social bot interference, and on multimodal Chinese propaganda. The diversity of the methodological and scientific approaches undertaken in the aforementioned articles demonstrates the interdisciplinarity of these issues. In turn, these crucial endeavors might anticipate a growing trend of studies where diverse theories, models, and techniques will be combined to tackle the different aspects of online misinformation, manipulation, and abuse.

**Keywords** Misinformation · Abuse · Social bots · Infodemics · Social media · COVID-19

✉ Emilio Ferrara
emiliofe@usc.edu

Stefano Cresci
stefano.cresci@iit.cnr.it

Luca Luceri
luca.luceri@supsi.ch

[1] University of Southern California, Los Angeles, CA 90007, USA

[2] Institute of Informatics and Telematics, National Research Council (IIT-CNR), 56124 Pisa, Italy

[3] University of Applied Sciences and Arts of Southern Switzerland (SUPSI), Manno, Switzerland

# Introduction

Malicious and abusive behaviors on social media have elicited massive concerns for the negative repercussions that online activity can have on personal and collective life. The spread of false information [8, 14, 19] and propaganda [10], the rise of AI-manipulated multimedia [3], the presence of AI-powered automated accounts [9, 12], and the emergence of various forms of harmful content are just a few of the several perils that social media users can—even unconsciously—encounter in the online ecosystem. In times of crisis, these issues can only get more pressing, with increased threats for everyday social media users [20]. The ongoing COVID-19 pandemic makes no exception and, due to dramatically increased information needs, represents the ideal setting for the emergence of *infodemics*—situations characterized by the undisciplined spread of information, including a multitude of low-credibility, fake, misleading, and unverified information [24]. In addition, malicious actors thrive on these wild situations and aim to take advantage of the resulting chaos. In such high-stakes scenarios, the downstream effects of misinformation exposure or information landscape manipulation can manifest in attitudes and behaviors with potentially dramatic public health consequences [4, 21].

By affecting the very fabric of our socio-technical systems, these problems are intrinsically interdisciplinary and require joint efforts to investigate and address both the technical (e.g., how to thwart automated accounts and the spread of low-quality information, how to develop algorithms for detecting deception, automation, and manipulation), as well as the socio-cultural aspects (e.g., why do people believe in and share false news, how do interference campaigns evolve over time) [7, 15]. Fortunately, in the case of COVID-19, several open datasets were promptly made available to foster research on the aforementioned matters [1, 2, 6, 16]. Such assets bootstrapped the first wave of studies on the interplay between a global pandemic and online deception, manipulation, and automation.

# Contributions

In light of the previous considerations, the purpose of this special issue was to collect contributions proposing models, methods, empirical findings, and intervention strategies to investigate and tackle the abuse of social media along several dimensions that include (but are not limited to) infodemics, misinformation, automation, online harassment, false information, and conspiracy theories about the COVID-19 outbreak. In particular, to protect the integrity of online discussions on social media, we aimed to stimulate contributions along two interlaced lines. On one hand, we solicited contributions to enhance the understanding on how health misinformation spreads, on the role of social media actors that play a pivotal part in the diffusion of inaccurate information, and on the impact of their interactions with organic users. On the other hand, we sought to stimulate

research on the downstream effects of misinformation and manipulation on user perception of, and reaction to, the wave of questionable information they are exposed to, and on possible strategies to curb the spread of false narratives. From ten submissions, we selected seven high-quality articles that provide important contributions for curbing the spread of misinformation, manipulation, and abuse on social media. In the following, we briefly summarize each of the accepted articles.

The COVID-19 pandemic has been plagued by the pervasive spread of a large number of rumors and conspiracy theories, which even led to dramatic real-world consequences. *"Conspiracy in the Time of Corona: Automatic Detection of Emerging COVID-19 Conspiracy Theories in Social Media and the News"* by Shahsavari, Holur, Wang, Tangherlini, and Roychowdhury grounds on a machine learning approach to automatically discover and investigate the narrative frameworks supporting such rumors and conspiracy theories [17]. Authors uncover how the various narrative frameworks rely on the alignment of otherwise disparate domains of knowledge, and how they attach to the broader reporting on the pandemic. These alignments and attachments are useful for identifying areas in the news that are particularly vulnerable to reinterpretation by conspiracy theorists. Moreover, identifying the narrative frameworks that provide the generative basis for these stories may also contribute to devise methods for disrupting their spread.

The widespread diffusion of rumors and conspiracy theories during the outbreak has also been analyzed in *"Partisan Public Health: How Does Political Ideology Influence Support for COVID-19 Related Misinformation?"* by Nicholas Havey. The author investigates how political leaning influences the participation in the discourse of six COVID-19 misinformation narratives: 5G activating the virus, Bill Gates using the virus to implement a global surveillance project, the "Deep State" causing the virus, bleach, and other disinfectants as ingestible protection against the virus, hydroxychloroquine being a valid treatment for the virus, and the Chinese Communist party intentionally creating the virus [13]. Results show that conservative users dominated most of these discussions and pushed diverse conspiracy theories. The study further highlights how political and informational polarization might affect the adherence to health recommendations and can, thus, have dire consequences for public health.

*"Understanding High and Low Quality URL Sharing on COVID-19 Twitter Streams"* by Singh, Bode, Budak, Kawintiranon, Padden, and Vraga investigate URL sharing patterns during the pandemic, for different categories of websites [18]. Specifically, authors categorize URLs as either related to traditional news outlets, authoritative health sources, or low-quality and misinformation news sources. Then, they build networks of shared URLs (see Fig. 1). They find that both authoritative health sources and low-quality/misinformation ones are shared much less than traditional news sources. However, COVID-19 misinformation is shared at a higher rate than news from authoritative health sources. Moreover, the COVID-19 misinformation network appears to be dense (i.e., tightly connected) and disassortative. These results can pave the way for future intervention strategies aimed at fragmenting networks responsible for the spread of misinformation.
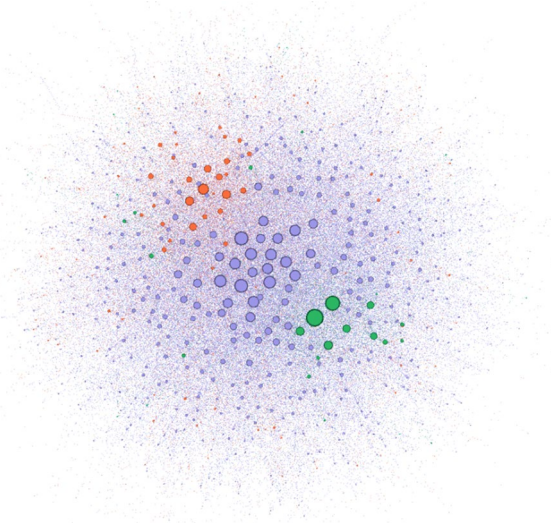
**Fig. 1** Network based on the web-page URLs shared on Twitter from January 16, 2020 to April 15, 2020 [18]. Each node represents a web-page URL, while connections indicate links among web-pages. The purple nodes represent traditional news sources, the orange nodes indicate the low-quality and misinformation news sources, and the green nodes represent authoritative health sources. The edges take the color of the source, while the node size is based on the degree

The relationship between news sentiment and real-world events is a long-studied matter that has serious repercussions for agenda setting and (mis-)information spreading. In *"Around the world in 60 days: An exploratory study of impact of COVID-19 on online global news sentiment"*, Chakraborty and Bose explore this relationship for a large set of worldwide news articles published during the COVID-19 pandemic [5]. They apply unsupervised and transfer learning-based sentiment analysis techniques and they explore correlations between news sentiment scores and the global and local numbers of infected people and deaths. Specific case studies are also conducted for countries, such as China, the US, Italy, and India. Results of the study contribute to identify the key drivers for negative news sentiment during an infodemic, as well as the communication strategies that were used to curb negative sentiment.

Farrell, Gorrell, and Bontcheva investigate one of the most damaging sides of online malicious content: online abuse and hate speech. In *"Vindication, Virtue and Vitriol: A study of online engagement and abuse toward British MPs during the COVID-19 Pandemic"*, they adopt a mixed methods approach to analyze citizen engagement towards British MPs online communications during the pandemic [11]. Among their findings is that certain pressing topics, such as financial concerns, attract the highest levels of engagement, although not necessarily negative. Instead, other topics such as criticism of authorities and subjects like racism and inequality tend to attract higher levels of abuse, depending on factors such as ideology, authority, and affect.

Yet, another aspect of online manipulation—that is, automation and social bot interference—is tackled by Uyheng and Carley in their article *"Bots and online hate during the COVID-19 pandemic: Case studies in the United States and the Philippines"* [22]. Using a combination of machine learning and network science, the authors investigate the interplay between the use of social media automation and the spread of hateful messages. They find that the use of social bots yields more results when targeting dense and isolated communities. While the majority of extant literature frames hate speech as a linguistic phenomenon and, similarly, social bots as an algorithmic one, Uyheng and Carley adopt a more holistic approach by proposing a unified framework that accounts for disinformation, automation, and hate speech as interlinked processes, generating insights by examining their interplay. The study also reflects on the value of taking a global approach to computational social science, particularly in the context of a worldwide pandemic and infodemic, with its universal yet also distinct and unequal impacts on societies.

It has now become clear that text is not the only way to convey online misinformation and propaganda [10]. Instead, images such as those used for memes are being increasingly weaponized for this purpose. Based on this evidence, Wang, Lee, Wu, and Shen investigate US-targeted Chinese COVID propaganda, which happens to rely heavily on text images [23]. In their article *"Influencing Overseas Chinese by Tweets: Text-Images as the Key Tactic of Chinese Propaganda"*, they tracked thousands of Twitter accounts involved in the #USAVirus propaganda campaign. A large percentage ($\simeq$ 38%) of those accounts was later suspended by Twitter, as part of their efforts for contrasting information operations.[1] Authors studied the behavior and content production of suspended accounts. They also experimented with different statistical and machine learning models for understanding which account characteristics mostly determined their suspension by Twitter, finding that the repeated use of text images played a crucial part.

## Conclusion

Overall, the great interest around the COVID-19 infodemic and, more broadly, about research themes such as online manipulation, automation, and abuse, combined with the growing risks of future infodemics, make this special issue a timely endeavor that will contribute to the future development of this crucial area. Given the recent advances and breadth of the topic, as well as the level of interest in related events that followed this special issue—such as dedicated panels, webinars, conferences, workshops, and other special issues in journals—we are confident that the articles selected in this collection will be both highly informative and thought provoking for readers. The diversity of the methodological and scientific approaches undertaken in the aforementioned articles demonstrates the interdisciplinarity of these issues, which demand renewed and joint efforts from different computer science fields, as well as from other related disciplines such as the social, political, and psychological

---

[1] https://blog.twitter.com/en_us/topics/company/2020/information-operations-june-2020.html.

sciences. To this regard, the articles in this collection testify and anticipate a growing trend of interdisciplinary studies where diverse theories, models, and techniques will be combined to tackle the different aspects at the core of online misinformation, manipulation, and abuse.

## References

1. Alqurashi, S., Alhindi, A., & Alanazi, E. (2020). Large Arabic Twitter dataset on COVID-19. arXiv preprint arXiv:2004.04315.
2. Banda, J.M., Tekumalla, R., Wang, G., Yu, J., Liu, T., Ding, Y., & Chowell, G. (2020). A large-scale COVID-19 Twitter chatter dataset for open scientific research—An international collaboration. arXiv preprint arXiv:2004.03688.
3. Boneh, D., Grotto, A. J., McDaniel, P., & Papernot, N. (2019). How relevant is the Turing test in the age of sophisbots? *IEEE Security & Privacy,17*(6), 64–71.
4. 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.
5. Chakraborty, A., & Bose, S. (2020). Around the world in sixty days: An exploratory study of impact of COVID-19 on online global news sentiment. *Journal of Computational Social Science*.
6. Chen, E., Lerman, K., & Ferrara, E. (2020). Tracking social media discourse about the COVID-19 pandemic: Development of a public coronavirus Twitter data set. *JMIR Public Health and Surveillance,6*(2), e19273.
7. Ciampaglia, G. L. (2018). Fighting fake news: A role for computational social science in the fight against digital misinformation. *Journal of Computational Social Science,1*(1), 147–153.
8. Cinelli, M., Cresci, S., Galeazzi, A., Quattrociocchi, W., & Tesconi, M. (2020). The limited reach of fake news on Twitter during 2019 European elections. *PLoS One,15*(6), e0234689.
9. Cresci, S. (2020). A decade of social bot detection. *Communications of the ACM,63*(10), 61–72.
10. Da San M., G., Cresci, S., Barrón-Cedeño, A., Yu, S., Di Pietro, R., & Nakov, P. (2020). A survey on computational propaganda detection. In: The 29th International Joint Conference on Artificial Intelligence (IJCAI'20), pp. 4826–4832.
11. Farrell, T., Gorrell, G., & Bontcheva, K. (2020). Vindication, virtue and vitriol: A study of online engagement and abuse toward British MPs during the COVID-19 Pandemic. *Journal of Computational Social Science*.
12. Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM,59*(7), 96–104.
13. Havey, N. (2020). Partisan public health: How does political ideology influence support for COVID-19 related misinformation?. *Journal of Computational Social Science*.
14. 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.
15. Luceri, L., Deb, A., Giordano, S., & Ferrara, E. (2019). Evolution of bot and human behavior during elections. *First Monday,24*, 9.
16. Qazi, U., Imran, M., & Ofli, F. (2020). GeoCoV19: a dataset of hundreds of millions of multilingual COVID-19 tweets with location information. *ACM SIGSPATIAL Special,12*(1), 6–15.
17. Shahsavari, S., Holur, P., Wang, T., Tangherlini, T. R., & Roychowdhury, V. (2020). Conspiracy in the time of corona: Automatic detection of emerging COVID-19 conspiracy theories in social media and the news. *Journal of Computational Social Science*.
18. Singh, L., Bode, L., Budak, C., Kawintiranon, K., Padden, C., & Vraga, E. (2020). Understanding high and low quality URL sharing on COVID-19 Twitter streams. *Journal of Computational Social Science*.
19. Starbird, K. (2019). Disinformation's spread: Bots, trolls and all of us. *Nature,571*(7766), 449–450.
20. Starbird, K., Dailey, D., Mohamed, O., Lee, G., & Spiro, E.S. (2018). Engage early, correct more: How journalists participate in false rumors online during crisis events. In: Proceedings of the 2018 ACM CHI Conference on Human Factors in Computing Systems (CHI'18), pp. 1–12. ACM.

21. Swire-Thompson, B., & Lazer, D. (2020). Public health and online misinformation: challenges and recommendations. *Annual Review of Public Health,41*, 433–451.
22. Uyheng, J., & Carley, K. M. (2020). Bots and online hate during the COVID-19 pandemic: Case studies in the United States and the Philippines. *Journal of Computational Social Science*.
23. Wang, A. H. E., Lee, M. C., Wu, M. H., & Shen, P. (2020). Influencing overseas Chinese by tweets: Text-Images as the key tactic of Chinese propaganda. *Journal of Computational Social Science*.
24. Zarocostas, J. (2020). How to fight an infodemic. *The Lancet, 395*(10225), 676.

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