

Studying the COVID-19 infodemic at scale

Big Data & Society
January–June: 1–6
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20539517211021115
journals.sagepub.com/home/bds



Anatoliy Gruzd¹ , Manlio De Domenico², Pier Luigi Sacco^{3,4} and Sylvie Briand⁵

Abstract

This special theme issue of *Big Data & Society* presents leading-edge, interdisciplinary research that focuses on examining how health-related (mis-)information is circulating on social media. In particular, we are focusing on how computational and Big Data approaches can help to provide a better understanding of the ongoing COVID-19 infodemic (overexposure to both accurate and misleading information on a health topic) and to develop effective strategies to combat it.

Keywords

Infodemic, COVID-19, social media, misinformation, pandemic

This article is a part of special theme on Studying the COVID-19 Infodemic at Scale. To see a full list of all articles in this special theme, please click here: <https://journals.sagepub.com/page/bds/collections/studyinginfodemicatscale>

Introduction

A year after the World Health Organization (WHO) characterized the COVID-19 outbreak as a global pandemic on 11 March 2020, more than 110 million people worldwide have contracted the novel coronavirus (SARS-CoV-2) and over 2.5 million have died due to the disease or associated complications as of 11 March 2021 (WHO, 2021). With several COVID-19 vaccines developed in record time, countries have started immunization campaigns (*American Journal of Managed Care*, 2021). However, as manufacturing and distribution of the vaccines are ramping up, false and misleading information about vaccines efficacy, safety and side-effects have also increased on social media. This is reflected in the number of vaccine-related claims being debunked by international fact-checking organizations. For example, between 11 March 2020 and 10 March 2021, there have been over 1000 COVID-19 related fact-checked claims mentioning “vaccine” or “vaccines” as tracked by Google Fact-Check Tools API (see Figure 1, Ryerson University Social Media Lab, 2021). The proliferation of false and misleading claims about COVID-19 vaccines is a major challenge

to public health authorities around the world, as such claims sow uncertainty, contribute to vaccine hesitancy and generally undermine the efforts to vaccinate as many people as possible (Abul-Fottouh et al., 2020; Wouters et al., 2021).

But, false and misleading COVID-19 claims are not unique to vaccine-related content. In fact, since the onset of this pandemic, social media has been a key vector in the spread of misinformation about the virus including how it is transmitted and how to treat it. The prevalence of COVID-19 related misinformation on social media contributes to the phenomenon

¹Ted Rogers School of Management, Ryerson University, Toronto, Canada

²CoMuNe Lab, Fondazione Bruno Kessler, Povo (TN), Italy

³Department of Humanities, IULM University, Milan, Italy

⁴metaLAB (at) Harvard, Cambridge, MA, USA

⁵Global Infectious Hazards Preparedness Emergency Preparedness, World Health Organization, Geneva, Switzerland

Corresponding author:

Anatoliy Gruzd, Ryerson University, 350 Victoria Street, Toronto, Ontario M5B 2K3, Canada.

Email: gruzd@ryerson.ca



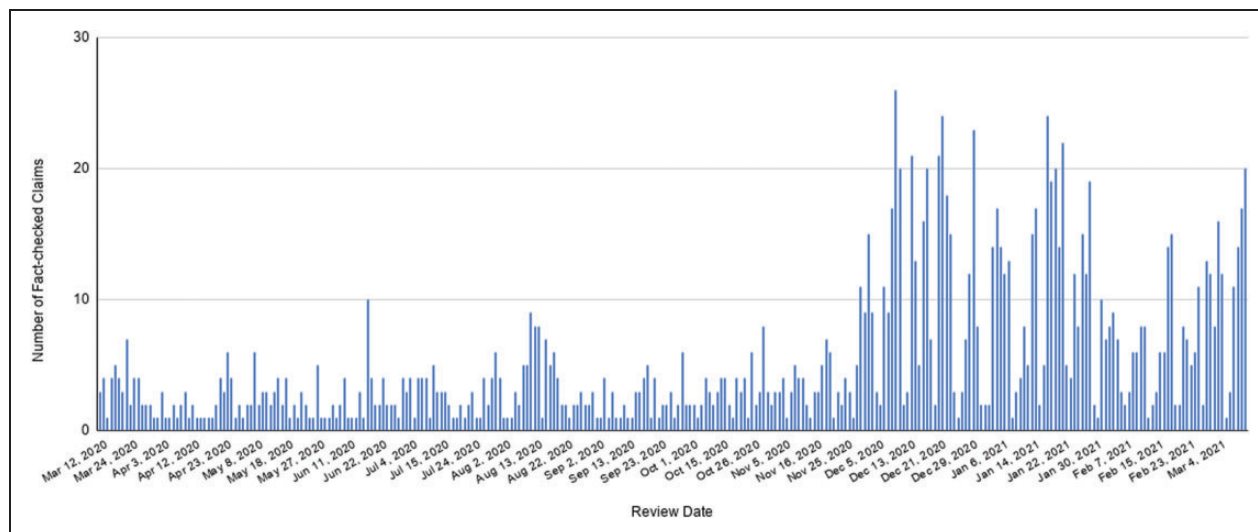


Figure 1. The number of fact-checked claims mentioning “vaccine(s)” as tracked by the Covid19MisInfo portal (<https://covid19misinfo.org/>) with data from Google Fact-Check Tools API.

called “infodemic,” when people are exposed to large quantities of both accurate and misleading information related to a health topic (Buchanan, 2020; Eysenbach, 2020). An infodemic makes it challenging for people to know what or whom to trust, especially when faced with conflicting claims or information. This is exactly the scenario that WHO’s Director-General Tedros Adhanom Ghebreyesus warned the public about in his speech in February of 2020, when he stated, “We’re not just fighting an epidemic; we’re fighting an infodemic,” referring to the proliferation of false and misleading information about SARS-CoV-2 on the internet (Ghebreyesus, 2020).

Combatting the infodemic

Fueled by the lack of knowledge about this new virus, the COVID-19 pandemic became enmeshed in an information ecosystem highly susceptible to the spread of false and misleading claims about the virus, its origin, transmission pathways, and treatment options. Indeed, since the beginning of 2020, Google Fact-Check Tools have recorded over 7000 false, misleading and unproven COVID-related claims debunked by fact-checking organizations such as AFP, Snopes, BOOM, and others (Gruzd and Mai, n.d.). Such rapidly accelerating production and propagation of inaccurate or misleading information calls for new, sophisticated approaches that elicit appropriate responses from the public to limit the uptake of dangerous, misinformation-induced attitudes, habits and behaviors.

While fact-checking has been shown to be an effective measure to educate the public and counter the spread of falsehoods, at least based on the assessment

of short-term effects of fact-checks on one’s beliefs (Dias and Sippitt, 2020; Zhang et al., 2021), the labor intensive manual review process of fact-checking is time consuming and is not scalable due to the sheer volume of both accurate and false posts related to the pandemic circulating on social media (Ceron et al., 2021). Furthermore, the use of automated or semi-automated social media accounts to spread misinformation, also known as social bots (Bessi and Ferrara, 2016; González-Bailón and De Domenico, 2021; Shao et al., 2018; Stella et al., 2018, 2019) has substantially contributed to the scale of this problem. For example, using the COVID-19 Infodemic Observatory, a platform for the large-scale analysis of publicly available discourse on Twitter (Figure 2), Gallotti et al. (2020) analyzed more than 1.6 billion COVID-19 related tweets and estimated that around 40% of them are shared by social bots.

Moreover, while some false claims are easier to spot and fact-check, claims that are politically motivated are much more challenging to refute as they are often shared by partisan actors with a large audience base (Gruzd and Mai, 2020a). As a result, many social media users might still be exposed to false and misleading COVID-19 related claims. For example, according to a census-balanced online survey of 1500 online Canadian adults, 80% of Facebook users and approximately 70% of Reddit, Twitter, TikTok, and YouTube users in Canada reported encountering COVID-19 misinformation on these platforms (Gruzd and Mai, 2020b).

Another aspect that makes it challenging to combat COVID-19 related misinformation is that like pandemics, misinformation also comes in waves, which is often

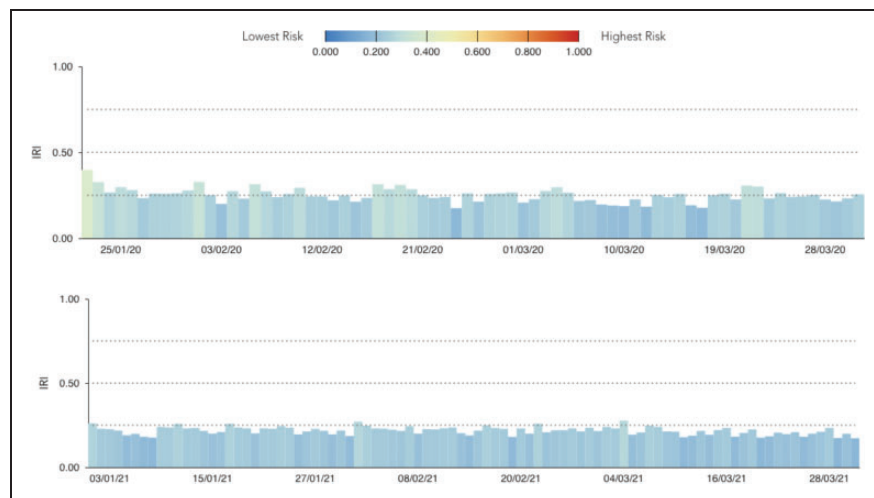


Figure 2. Infodemic Risk Index estimated based on 1.6 billion tweets, from 239 countries, in the first quarter of 2020 (top) and 2021 (bottom), as tracked by the COVID-19 Infodemic Observatory (<https://covid19obs.fbk.eu/>).

triggered by major news and developments in the fight against SARS-CoV-2. For instance, as mentioned above, the most recent wave of falsehoods circulating on social media has largely focused on COVID-19 vaccines, questioning their effectiveness and undermining the public health vaccination campaigns. But unlike the infectious disease, no travel restrictions or lockdown orders can slow down misinformation shared on social media. Because social media platforms are inherently designed to support information virality, false and misleading claims about the pandemic can easily cross international borders, time zones, and sites. A tweet with false information about the virus can be easily repackaged and reshared as a post on Facebook, YouTube or TikTok, even if the original message was fact-checked or removed from Twitter. Furthermore, conspiracy theorists and adversarial social parties may exploit the frustration and disorientation that is brought about by restrictive public interventions to gain more attention and visibility for their claims.

Overview of this special theme on the infodemic

To address the challenges of detecting and combating the spread of COVID-19 misinformation on social media and to contribute to the rapidly growing area of infodemiology (Tangcharoensathien et al., 2020), we are pleased to present this special theme on “Studying the COVID-19 Infodemic at Scale”. This special theme in *Big Data & Society* provides a space for original research articles and commentaries at the intersection of infodemiology, Big Data, and COVID-related dis/misinformation studies that explore

questions such as: What are key terminologies and different computational approaches currently used to study and combat the spread of misinformation on social media? How can we use social media data to estimate the effects of the infodemic on individuals and society in general? And more specifically, how can we assess and mitigate the infodemic risks and consequences using Big Data?

The special theme builds on a successful series of public events and consultations organized by the WHO Information Network for Epidemics (EPI-WIN), including the first technical consultation on responding to the infodemic related to the COVID-19 pandemic held on 7–8 April of 2020, and the first WHO Infodemiology Conference and working group meetings held in June and July of 2020. We are also building on the Big Data & Society symposium called “Viral Data” edited by Leszczynski and Zook (2020) which examined Big Data practices and specifically the notion of data virality as related to the pandemic at the midpoint of 2020. All together the special theme features six research articles and four commentaries by 57 authors from 23 institutions in six countries.

The first article that is part of this theme is a comprehensive explainer-type article by Gradoń et al. (2021) that provides an in-depth review of key terminologies and different computational approaches currently used in studying and combating the spread of misinformation on social media. The techniques discussed in the article include machine learning, network analysis, agent-based modeling, and text mining.

The next piece features an empirically grounded work by Yang et al. (2021) that examines the credibility of information sources about the pandemic shared on two major social media platforms: Facebook and

Twitter. The authors find evidence of coordination in the spread of pandemic information from low-credibility sources by so-called super-spreaders on both platforms. Their results provide more insights into the types of accounts that engage in spreading COVID-19 misinformation on social media platforms.

Continuing in this vein of research, the third article by Haupt, Li and Mackey (2021) examined the quality of information shared on social media. The authors examined the Twitter discourse around the use of the antimalarial drug hydroxychloroquine for treating COVID-19. The researchers found two alarming trends in their data. First, tweets supporting COVID-19 treatment relied on carefully crafted narratives that misused and mischaracterized scientific literature as a ploy to increase tweets' perceived credibility. Second, the authors found that many accounts that shared false claims on this topic appeared to be supporters of former US President Donald Trump and that medical doctors or scientists who engaged in debunking false claims about hydroxychloroquine were less influential on Twitter than the proponents of the drug.

The fourth article by Green et al. (2021) examined the public discourse and information sharing practices on Twitter, but instead of considering opinions about a specific treatment, the authors analyzed a large sample of tweets ($n > 2M$) to test the impact of the first lockdown order in the UK on the subsequent spread of COVID-19 related misinformation (based on shared links to websites known to spread misinformation). The researchers found an increase in the volume of tweets following the lockdown announcement in general, but interestingly misinformation-related tweets were not any more prominent after the announcement than before.

Pascual-Ferrá et al. (2021) examined the expressions of dissatisfaction regarding a particular governmental COVID-19 policy. Specifically, the research team explored the role of toxicity and verbal aggression in public discussions on Twitter about mandatory mask wearing. The results showed that anti-mask related messages were more likely to include toxic exchanges in comparison to messages supporting this practice. The only exception was tweets condemning the anti-mask supporters. The study shows that the presence of toxic language in the conversations around mask wearing in public may create challenges for health agencies when trying to communicate the latest science-based evidence on the effectiveness of facial masks in slowing down the spread of the virus. The study also demonstrates that public health authorities could use the toxicity analysis to monitor and potentially predict the public's uptake of different health advice, directions, and mandates.

While the previous articles took a case-study approach to examine the prevalence and trends of COVID-19 related misinformation on popular social media platforms such as Twitter and Facebook and potentially reducing its prevalence by blocking, fact-checking, and limiting its spread, the article by Basol et al. (2021) took a different approach to understanding and combating health-related misinformation. Based on two user studies, the authors demonstrated how a 5-minute online game can equip social media users with critical information literacy skills to spot misinformation and stop themselves from resharing it.

In addition to the six full articles, this special theme features four commentaries that offer further insight into possible solutions to combating COVID-19 related misinformation on social media. Dunn et al. (2021) call for the development and curation of an open library of misinformation case studies which can be used by scholars and practitioners to support work in this area, especially in improving the consistency and reliability of study designs and reporting on this topic. Godinho et al. (2021) outline a framework to support policy development based on data generated by citizen e-participation in the knowledge co-creation process during a pandemic. Finally, Bloomfield et al. (2021) and Bonnevie et al. (2021) share their insights from studying large datasets of public opinions shared on social media in response to various COVID-19 related public announcements, and implications for public health communicators.

In sum, the articles and commentaries presented in this special theme (a) expand our understanding of the current and future trends in the area of infodemic management, (b) present techniques on how to study the public discourse in response to a global health emergency in real-time and at scale, (c) provide practical guidelines on how health communicators can more effectively engage with the public on social media, and (d) offer insights to social media platforms and policy makers on effective intervention strategies and information literacy skills to combat COVID-19 misinformation.

Acknowledgments

We would like to extend our appreciation to all the reviewers of the special theme who carefully peer-reviewed submissions and provided helpful comments to the authors. We would also like to thank Philip Mai, Alyssa Saiphoo, Jessica Patterson, Rebeca Chacon Cabrera, Paige Bagby, Tim Nguyen, Tina D Purnat, Christine Czerniak, Melinda Frost, Petros Gikonyo, Jamie Guth, Sarah Hess, Vicky Houssiere, Ramona Ludolph, Jianfang Liu, Shi Han Liu, Alexandra Leigh Mcphedran, Avichal Mahajan, Thomas Moran, Lynette Phuong, Romana Rauf, Katherine Sheridan, Sally L Smith, Heather Saul, Aicha Taybi, Judith

Van Holten, Stefano Burzo, Nikki Shindo, Rosamund Lewis, and Tamar Zalk for their feedback and editorial assistance during the preparation of this special theme. Finally, we are grateful to Matthew Zook and Age Poom from the Big Data & Society editorial team for assisting in moving the special theme through to publication. Sylvie Briand is a staff member of the World Health Organization (WHO). Sylvie Briand alone is responsible for the views expressed in this publication and they do not represent the views of the WHO.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work is supported in part by a Canadian Institutes of Health Research (CIHR) grant, “Inoculating Against an Infodemic: Microlearning Interventions to Address CoV Misinformation” (PIs: George Veletsianos, Jaigris Hodson, Anatoliy Gruzd).

ORCID iDs

Anatoliy Gruzd  <https://orcid.org/0000-0003-2366-5163>

Pier Luigi Sacco  <https://orcid.org/0000-0002-5559-2889>

References

- Abul-Fottouh D, Song MY and Gruzd A (2020) Examining algorithmic biases in YouTube’s recommendations of vaccine videos. *International Journal of Medical Informatics* 140: 104175. <https://doi.org/10.1016/j.ijmedinf.2020.104175>
- American Journal of Managed Care* (2021) A timeline of COVID-19 vaccine developments in 2021. Retrieved May 20, 2021, from <https://www.ajmc.com/view/a-time-line-of-covid-19-vaccine-developments-in-2021>
- Basol M, Roozenbeek J, Berriche M, Uenal F, McClanahan WP and Linden S van der. (2021) Towards psychological herd immunity: Cross-cultural evidence for two prebunking interventions against COVID-19 misinformation. *Big Data & Society* 8(1): 20539517211013868. <https://doi.org/10.1177/20539517211013868>
- Bessi A and Ferrara E (2016) Social bots distort the 2016 US presidential election online discussion. *First Monday* 21(11-7).
- Bloomfield P, Walsh M, Magnusson J, et al. (2021) Communicating public health during COVID-19, implications for vaccine rollout. *Big Data & Society*.
- Bonnevie E, Sittig J and Smyser J (2021) The case for tracking misinformation the way we track disease. *Big Data & Society* 8(1): 20539517211013868. <https://doi.org/10.1177/20539517211013868>
- Buchanan M (2020) Managing the infodemic. *Nature Physics* 16(9): 894–894. <https://doi.org/10.1038/s41567-020-01039-5>
- Ceron W, de-Lima-Santos M-F and Quiles MG (2021) Fake news agenda in the era of COVID-19: Identifying trends through fact-checking content. *Online Social Networks and Media* 21: 100116. <https://doi.org/10.1016/j.osnem.2020.100116>
- Dias N and Sippitt A (2020) Researching fact checking: Present limitations and future opportunities. *The Political Quarterly* 91(3): 605–613. <https://doi.org/10.1111/1467-923X.12892>
- Dunn A, Steffens M, Dyda A, et al. (2021) Knowing when to act: A call for an open misinformation library to guide actionable surveillance. *Big Data & Society*.
- Eysenbach G (2020) How to fight an infodemic: The four pillars of infodemic management. *Journal of Medical Internet Research* 22(6): e21820. <https://doi.org/10.2196/21820>
- Gallotti R, Valle F, Castaldo N, et al. (2020) Assessing the risks of ‘infodemics’ in response to COVID-19 epidemics. *Nature Human Behaviour* 4(12): 1285–1293. <https://doi.org/10.1038/s41562-020-00994-6>
- Ghebreyesus TA (2020) *Presentation given at the 56th Munich Security Conference*. Munich, Germany. Retrieved May 20, 2021, from <https://www.who.int/director-general/speeches/detail/munich-security-conference>
- Godinho MA, Borda A, Kariotis T, et al. (2021) Knowledge co-creation in participatory policy & practice: Building community through data-driven direct democracy. *Big Data & Society*.
- González-Bailón S and De Domenico M (2021) Bots are less central than verified accounts during contentious political events. *Proceedings of the National Academy of Sciences* 118(11): 1–8. <https://doi.org/10.1073/pnas.2013443118>
- Gradon K, Hołyst J, Moy W, et al. (2021) Countering misinformation: A multidisciplinary approach. *Big Data & Society* 8(1): 20539517211013850. <https://doi.org/10.1177/20539517211013848>
- Green M, Musi E, Rowe F, et al. (2021) Identifying how COVID-19 related misinformation reacts to the announcement of the UK national lockdown: An interrupted time-series study. *Big Data & Society* 8(1): 20539517211013868. <https://doi.org/10.1177/20539517211013869>
- Gruzd A and Mai P (2020a) Going viral: How a single tweet spawned a COVID-19 conspiracy theory on Twitter. *Big Data & Society* 7(2): 2053951720938405. <https://doi.org/10.1177/2053951720938405>
- Gruzd A and Mai P (2020b) *Inoculating against an infodemic: A Canada-wide COVID-19 news, social media, and misinformation survey*. Report, Ryerson University, Canada. Scholars Portal Dataverse. Retrieved May 20, 2021, from <https://doi.org/10.5683/SP2/JLULYA>
- Gruzd A and Mai P (n.d.). *COVID-19 misinformation portal – A rapid response project*. Ryerson University. Available at: <https://covid19misinfo.org/> (accessed 15 March 2021).
- Haupt MR, Li J and Mackey TK (2021) Identifying and characterizing scientific authority-related misinformation discourse about hydroxychloroquine on twitter using unsupervised machine learning. *Big Data & Society* 8(1): 20539517211013844. <https://doi.org/10.1177/20539517211013843>

- Leszczynski A and Zook M (2020) Viral data. *Big Data & Society* 7(2): 2. 053951720971009. <https://doi.org/10.1177/2053951720971009>
- Pascual-Ferrá P, Alperstein N, Barnett D, et al. (2021) Toxicity and verbal aggression on social media: Polarized discourse on wearing face masks during the COVID-19 pandemic. *Big Data & Society*.
- Ryerson University Social Media Lab (2021) *COVIDGlobal Misinformation Dashboard*. Google Data Studio. Available at: <https://covid19misinfo.org/misinfowatch/> (accessed 11 March 2021).
- Shao C, Ciampaglia GL, Varol O, et al. (2018) The spread of low content by social bots. *Nature Communications* 9(1): 1–9. <https://doi.org/10.1038/s41467-018-06930-7>
- Stella M, Cristoforetti M and De Domenico M (2019) Influence of augmented humans in online interactions during voting events. *PloS One* 14(5): e0214210. <https://doi.org/10.1371/journal.pone.0214210>
- Stella M, Ferrara E and De Domenico M (2018) Bots increase exposure to negative and inflammatory content in online social systems. *Proceedings of the National Academy of Sciences* 115(49): 12435–12440. <https://doi.org/10.1073/pnas.1803470115>
- Tangcharoensathien V, Calleja N, Nguyen T, et al. (2020) Framework for managing the COVID-19 infodemic: Methods and results of an online, crowdsourced WHO technical consultation. *Journal of Medical Internet Research* 22(6): e19659. <https://doi.org/10.2196/19659>
- World Health Organization (11 March 2021) WHO coronavirus (COVID-19) dashboard. Available at: <https://covid19.who.int> (accessed 11 March 2021).
- Wouters OJ, Shadlen KC, Salcher-Konrad M, et al. (2021) Challenges in ensuring global access to COVID-19 vaccines: Production, affordability, allocation, and deployment. *The Lancet* 397(10278): 1023–1034. [https://doi.org/10.1016/S0140-6736\(21\)00306-8](https://doi.org/10.1016/S0140-6736(21)00306-8)
- Yang K-C, Pierri F, Hui P-M, et al. (2021) The COVID-19 infodemic: Twitter versus Facebook. *Big Data & Society* 8(1): 20539517211013860. <https://doi.org/10.1177/20539517211013861>.
- Zhang J, Featherstone JD, Calabrese C, et al. (2021) Effects of fact-checking social media vaccine misinformation on attitudes toward vaccines. *Preventive Medicine* 145: 106408. <https://doi.org/10.1016/j.ypmed.2020.106408>