Delving into Data Science Methods in Response to the COVID-19 Infodemic

Chong, Miyoung	University of South Florida, USA | mc17@usf.edu
Shah, Chirag	University of Washington, USA | chirags@uw.edu
Shu, Kai	Illinois Institute of Technology, USA | kshu@iit.edu
He, Jiangen	University of Tennessee, USA | jiangen@utk.edu
Hagen, Loni	University of South Florida | lonihagen@usf.edu

ABSTRACT

The circulation of myriad of information from diverse digital platforms during the COVID-19 pandemic caused the unprecedented infodemic. Along with the increased case numbers, the shared information accelerated exponentially, especially via social media, and a large proportion of the daily distributed information was blended with myth, rumors, pseudoscience, or modified facts. Uncovering viral mis- and disinformation narratives and information voids is essential to a swift and effective response on delivering public health information and policy by the governments during a public health emergency. Although many studies have examined how information was circulated and shared during the COVID-19 pandemic era, large gaps in literature exist as to how effectively to track, describe, and answer it. In this panel, the panelists propose and discuss data science methods to analyze the COVID-19 infodemic. We hope our panel contribute to exploring more effective and applicable data science methods to investigate infodemic in crises.

KEYWORDS

Infodemic; COVID-19; data science; method; public health

INTRODUCTION

The COVID-19 pandemic was dubbed as a "digital pandemic" because of the circulation of myriad of information from diverse digital platforms (Banerjee & Meena, 2021; Chong & Park, 2021). With more than three billion regular users, social media has become the dominant source of information and main instruments for communication during the prolonged pandemic crisis (Chong & Park, 2021; Tankovska, 2022, January 28). Along with the increased COVID-19 case numbers, the shared information accelerated exponentially, especially via social media, and a large proportion of the daily and digitally distributed information was blended with myth, rumors, pseudoscience, or modified facts that might snowball the effects when they are in wrong hands (Banerjee & Meena, 2021; Chou, Gaysynsky, & Vanderpool, 2021).

The Director–General of the World Health Organization (WHO) said, "We're not just fighting an epidemic; we're fighting an infodemic." An infodemic is an information overload including inaccurate or misleading information in digital and physical environments during a disease outbreak (World Health Organization). Uncovering viral mis- and disinformation narratives and information voids (deficits of quality information) in real-time data is essential to a swift and effective response on delivering public health information and policy by the governments during a public health emergency. The unparalleled scale of the COVID-19 pandemic caused the unprecedented infodemic regarding the origin of the new coronavirus (SARS-CoV-2), the response to the virus, including vaccination and extensive public health and social regulations (Banerjee & Meena, 2021).

A recent study revealed that the chronic isolation and anxiety during the crisis can cause panic and mental discomfort as well as excessive digital media consumption, and this social disruption resulting from unnecessary media usage can likely contribute to the fabrication of information boosting the rampant "misinfodemic" (Banerjee, 2020). The circulation of misinformation has presented a precarious impact on the target audiences for them to grasp the pandemic in nations, such as Italy, Spain, and the United States (Bavel et al., 2020). For example, in a national sample survey, 47% of Americans read the news that COVID-19 was "completely made up" (Pew Research Center, 2021, March 25). A more recent study on misinformation and public health discovered misinformation could affect public health in the same way as it influences "terrorism, immigration, sexism and racism" in the U.S. (Mitchell, Gottfried, Stocking, Walker, & Fedeli, 2021, November 12). Another study found misinformation can prompt confusion and lead to risk-taking behaviors, which can harm health, diminish the pandemic response, trust in public health agencies (Purnat et al., 2021).

Many studies examined how information was circulated and shared during the COVID-19 pandemic era (Banerjee & Meena, 2021; Chong & Park, 2021). However, large gaps in literature exist as to how effectively to track, describe, and answer it (Mahajan et al., 2021; Purnat et al., 2021). Infodemic research has primarily focused on social media that allow easier data access, such as Twitter and YouTube, while misinfodemic is prevailing across the

85th Annual Meeting of the Association for Information Science & Technology | Oct. 29 – Nov. 1, 2022 | Pittsburgh, PA.

Internet as well as offline information ecosystems. Researchers have been making extensive efforts to find ways to better observe and alleviate phenomena and consequences resulting from malinformation in relation to the COVID-19 pandemic crisis (Purnat & Wilhelm, 2020). Aligning the endeavors, this panel focuses on exploring computational data science methods to detect infodemic signals, including identifying and predicting emerging spots of concern, information voids, and mis- and disinformation narratives in the internet ecosystem. To this end, our panelists will propose applicable data science methods from their areas of expertise, which are artificial intelligence, machine learning, data visualization, and social media analytics.

PANEL STRUCTURE AND PARTICIPANT ENGAGEMENT

The duration of the panel session is 90 minutes. In the beginning of the session, the panel moderator will introduce the structure and the panel. In the first portion of the session, each panelist will give a 13-to-15-minute presentation about his or her research and suggest methods to examine malinformation and infodemic cues in online information ecosystems (for about 60 minutes). Thereafter, the moderator will ask the audience to join a 30-minute interactive discussion facilitated by a set of questions based on the panelists' presentations. After the discussion with the audience, the panelists will address their methodological implications in relation to information science and interdisciplinary disciplines by giving one-minute reflection. Directly afterwards, the moderator will summarize the outcomes from the panel discussion and end the session with a final remark.

PANELISTS AND THEIR CONTRIBUTIONS

Shah, Chirag

Misinformation and the Downward Spiral of Information Access Systems

It is increasingly difficult to access any information today that is not mediated by some algorithm behind the scenes—whether it's a search engine or a recommender system. Since people often trust these systems and the information they provide, their actions continue feeding into such systems as validating signals. This creates a positive feedback loop that is hard to break, creating problems of, misinformation and disinformation as well as misrepresentation of people and misalignment of values.

He, Jiangen

User and Machine biases in Artificial Intelligence Systems

Our information and decision-making behaviors when interacting with information systems are affected by both machine and user biases. Dr. He will present different types of biases from both human and machine sides and how HCI design and visualization techniques influence our perception of information and alleviate the negative impacts of the bias.

Shu, Kai

Combating Disinformation on Social Media: A Computational Perspective

Recent advancements of computational approaches on detecting disinformation have shown some early promising results. Novel challenges are still abundant due to its complexity, diversity, dynamics, multi-modality, and costs of fact-checking or annotation. Dr. Shu will present some lessons learned when exploring effective and explainable detecting strategies of fake news, and discuss some imperative challenges in combating disinformation and the need for interdisciplinary research.

Hagen, Loni

Close Human Supervision of Data Science Projects

Dr. Hagen will discuss how to conduct mixed methods studies using machine learning and qualitative human investigation. She uses computational methods, closely supervised by humans, to identify useful insights from social media. Using multiple case studies that adopted bot detection algorithm, social network analysis, and inductive coding, she will discuss the methodological challenges, and share some knowledge on how to get over these challenges.

RELEVANCE TO ASIST AND SIGNIFICANCE TO THE FIELD

ASIS&T Members and the 2022 ASIS&T Annual Meeting attendees who are interested in exploring and learning data science methods to examine infodemic and malinformation, such as misinformation, disinformation, conspiracy theories, and fake news, in the internet ecosystem, including social media, especially during the global-scale crises will find our topics applicable to their research. We anticipate that this panel will provide directions for those works and foster related interdisciplinary research.

CONCLUSION

In this panel session, we will suggest and discuss up-to-date data science approaches to address infodemic tracking and forecasting. However, there is dearth in research or no empirical evidence on how the infodemic studies can be employed to design practical applications for a crisis response by the government and public health organizations.

More partnerships among scholars, public health practitioners, policy makers are required to fill this gap. We hope this panel contribute to filling this gap at our best.

ACKNOWLEDGMENTS

This panel is sponsored by SIG Social Media (SIG SM). SIG SM has hosted a series of panels at the ASIS&T Annual Meeting to support social media research as an independent field of information science.

AUTHOR BIOS

Chong, Miyoung

Miyoung Chong is an Assistant Professor in the Department of Journalism and Digital Communication at the University of South Florida. Her research centers on the intersections of information, minority, and power in digital, social, and news media settings with a critical informatics approach. Her larger body of research is characterized by computational social science and applied data science to investigate digital media and online community engagement. She published her studies in *Government Information Quarterly*, *Journal of Medical Internet Research* (*JMIR*), and *Journal of Business Anthropology*. She wrote a book chapter titled with "Social Media Analytics" and received the SIG SM research award (runner-up) at the ASIS&T Annual Meeting in 2021. She currently serves as the communication chair of the SIG SM at ASIS&T.

Shah, Chirag

Chirag Shah is an Associate Professor in the iSchool at University of Washington. He is the Founding Co-Director of Responsibility in AI Systems & Experiences (RAISE), Founding Chair of ASIS&T SIG AI, and Founding Editor-in-Chief of Information Matters. His research involves designing intelligent information systems that are fair and transparent, which is funded through various grants and gifts from NSF, IMLS, Google, and Amazon. Dr. Shah is the author of "A Hands-On Introduction to Data Science", a textbook published by Cambridge University Press. His new textbook on machine learning is currently under production.

Shu, Kai

Kai Shu is a Gladwin Development Chair Assistant Professor in the Department of Computer Science at Illinois Institute of Technology. He obtained his Ph.D. in Computer Science at Arizona State University and was the recipient of the 2020 ASU Dean's Dissertation Award. His research addresses challenges varying from big data, to social media, and to AI on issues of fake news detection, explainable machine learning, and trustworthy social computing. He is the leading author of a monograph, Detecting Fake News on Social Media (2019), and the leading editor of a book, Disinformation, Misinformation, and Fake News in Social Media (2020). He has published innovative works in highly ranked journals and top conference proceedings such as *ACM KDD*, *SIGIR*, *WSDM*, *WWW*, *CIKM*, *IEEE ICDM*, *IJCAI*, and *AAAI*. He was an invited panelist at the first Global WHO Infodemiology Conference. He presented two tutorials at top data mining conferences, co-organized conference workshops, guest-edited journal special issues, all related to his research on disinformation.

He, Jiangen

Jiangen He is an Assistant Professor in the School of Information Sciences at the University of Tennessee, Knoxville. His research interests include visual analytics, data-driven predictions, and metrics in science of science and social factors in human-AI interaction. He focuses on the development of quantitative and qualitative approaches to the study of how science advances, especially through information visualization and machine learning techniques, and the examination of how socialness in artificial intelligence affects people's decision-making behavior. His research in science of science is currently supported by National Science Foundation (#2122691). He extensively published in venues in information science and related areas, such *JASIST*, *IPM*, *IUI*, and *ISSI*.

Hagen, Loni

Loni Hagen is an assistant professor at University of South Florida. She uses data mining techniques and visualizations to create actionable insights from big data. Her current research interests are in the domains of social media, cybersecurity, public health crisis, and undergraduate data science education. She is an active member of the SIG SM. She has published in *Government Information Quarterly*, *Information Processing and Management*, *New Media and Society*, *Social Science Computer Review, JASIST*, *ACM DGOV*, and *JMIR Infodemeology*. Her research was supported by the National Research Foundation of Korea and Cyber Florida.

REFERENCES

Banerjee, D. (2020). How COVID-19 is overwhelming our mental health.

Banerjee, D., & Meena, K. S. (2021). COVID-19 as an "infodemic" in public health: critical role of the social media. *Frontiers in Public Health, 9*(231). doi:10.3389/fpubh.2021.610623

Bavel, J. J. V., Baicker, K., Boggio, P. S., Capraro, V., Cichocka, A., Cikara, M., ... & Willer, R. (2020). Using social and behavioural science to support COVID-19 pandemic response. *Nature human behaviour, 4*(5), 460-471.

Chong, M., & Park, H. W. (2021). COVID-19 in the Twitterverse, from epidemic to pandemic: information-sharing behavior and Twitter as an information carrier. *Scientometrics, 126*(8), 6479-6503.

Chou, W. Y. S., Gaysynsky, A., & Vanderpool, R. C. (2021). The COVID-19 misinfodemic: Moving beyond fact-checking. *Health Education & Behavior, 48*(1), 9-13.

Mahajan, A., Czerniak, C., Lamichhane, J., Phuong, L., Purnat, T., Nguyen, T., & Briand, S. (2021). WHO public health research agenda for managing infodemics. *European Journal of Public Health, 31*(Supplement_3), ckab164-030.

Mitchell, A., Gottfried, J., Stocking, G., Walker, M., & Fedeli, S. (2021, November 12). Many Americans say made-up news is a critical problem that needs to be fixed. Pew Research Center's Journalism Project. Retrieved April 13, 2022, from https://www.pewresearch.org/journalism/2019/06/05/many-americans-say-made-up-news-is-a-critical-problem-that-needs-to-be-fixed/

Pew Research Center. (2021, March 25). American News Pathways. Pew Research Center. Retrieved April 13, 2022, from https://www.pewresearch.org/pathways-2020/covidcover2/main_source_of_election_news

/us_adults/

Purnat, T. D., Vacca, P., Czerniak, C., Ball, S., Burzo, S., Zecchin, T., ... & Nguyen, T. (2021). Infodemic signal detection during the COVID-19 pandemic: development of a methodology for identifying potential information voids in online conversations. *JMIR infodemiology, 1*(1), e30971.

Purnat, T., & Wilhelm, E. (2020). Building systems for respond to infodemics and build resilience to misinformation. LinkedIn.

Tankovska, H. (2022, January 28). Number of social media users 2025. Statista. Retrieved April 13, 2022, from https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

World Health Organization. (n.d.). Infodemic. World Health Organization. Retrieved April 13, 2022, from https://www.who.int/health-topics/infodemic#tab=tab_1