Pierri et al. *EPJ Data Science* (2024) 13:65
https://doi.org/10.1140/epjds/s13688-024-00504-y

EPJ Data Science
a SpringerOpen Journal

EPJ.org

EDITORIAL Open Access

Computational approaches for cyber social threats

Francesco Pierri[1*], Matthew R. DeVerna[2], Kai-Cheng Yang[3], Jeremy Blackburn[4] and Ugur Kursuncu[5]

*Correspondence:
francesco.pierri@polimi.it
[1]Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy
Full list of author information is available at the end of the article

Abstract

This topical issue, *Computational Approaches for Cyber Social Threats*, showcases cutting-edge research that employs computational methods to tackle the pressing challenges of cyber social threats—such as fake news, disinformation, cyberbullying, hate speech, and online radicalization. In an increasingly interconnected digital landscape, these threats significantly jeopardize societal stability by eroding public trust, intensifying polarization, and widening social divides. The theme "Information Integrity During Crises" is highlighted to underscore the critical role of reliable information during global crises, when the spread of misinformation and disinformation is particularly pervasive. With a collection of nine outstanding papers, this topical issue advances our comprehension of how computational tools can address cyber social threats and protect information integrity during crises.

Keywords: Cyber social threats; Disinformation; Hate speech; Misinformation; Online social media; Computational approaches

1 Introduction

Online platforms have increasingly been used to disseminate harmful content and behaviors, including misinformation, extremism, harassment, human trafficking, and gender-based violence [1, 2], with significant societal repercussions. These issues can lead to real-world harm, erode trust in institutions, polarize public discourse, and deepen social divides [3]. Recent global crises—such as the COVID-19 pandemic [4, 5] and the Russian invasion of Ukraine [6, 7]—highlight the urgent need for reliable information on these platforms. Given the widespread impact on individuals and communities, a focused research agenda is necessary to better understand, detect, and mitigate these complex online threats.

Our International Workshop series, "Cyber Social Threats" (CySoc)—now in its fifth edition[1]—fosters multidisciplinary research that explores multi-faceted aspects of harmful content while leading the discussion on building novel computational methods to reliably detect, derive meaning of, interpret, understand and counter them. The fourth edi-

[1]https://cy-soc.github.io/2024/.

Springer

tion,[2] held at the 2023 ACM Web Conference [8], emphasized the importance of accessing reliable information during global emergencies, with a focus on the theme "Information Integrity During Crises." Such situations create fertile ground for cyber social threats, including the rapid dissemination of disinformation, fake news, and manipulated content on social media platforms, which exploit heightened emotions and uncertainty to amplify their influence, undermine trust in credible sources, and destabilize social and political systems [9, 10].

After hosting the fourth edition of our workshop, we curated this topical issue to invite contributions aimed at advancing our understanding of how computational methods can be harnessed to combat cyber social threats and maintain information integrity during crises. The resulting collection, titled "Computational Approaches for Cyber Social Threats," received 15 submissions, nine of which have been accepted for publication after peer reviews. In this foreword, we summarize the included papers and discuss recent advancements in computational methods for analyzing harmful online communication and cyber social threats.

2 Contributions

Singh et al. [11] introduce an unsupervised method (UTDRM) for retrieving debunked claims in fact-checking without needing human-annotated pairs. This approach generates synthetic claims from fact-checking articles and uses a neural retrieval model, showing competitive or superior performance to state-of-the-art methods across seven datasets.

Ng and Carley [12] present a methodology for identifying three types of bots on Twitter during a 2023 diplomatic incident between the U.S. and China involving a balloon spotted in U.S. airspace. Their analysis reveals that U.S.-based bots primarily discussed the balloon's location while China-based bots emphasized escalating tensions. The study highlights the role of bots in shaping narratives and perceptions, offering insights for policymakers on how automated agents are used to influence political communication and project national images on social media.

Xia et al. [13] examine how the Russian invasion of Ukraine impacted polarization on Finnish Twitter regarding NATO membership. Prior to the invasion, Finnish public opinion was polarized along partisan lines, with distinct pro-NATO, left-wing anti-NATO, and conspiracy-driven anti-NATO groups. After the invasion, the left-wing anti-NATO group began engaging with the pro-NATO group, united by shared condemnation of Russia and democratic values, while the conspiracy-driven group remained isolated. The findings suggest that external threats can reduce partisan polarization, but conspiracy-fueled divisions may persist despite significant external pressure.

Truong et al. [14] introduce methods to detect low-credibility social media accounts by analyzing information diffusion patterns before their content can be classified as misinformation. The study focuses on two networks: the reshare network (accounts trusting other accounts) and the bipartite account-source network (accounts trusting media sources). Using extended centrality measures and graph embedding techniques, the authors show that these networks provide strong signals for estimating account credibility with high accuracy. The findings indicate that accounts tend to share similar credibility with users they

[2] https://cy-soc.github.io/2023/.

reshare or that follow similar sources, reflecting strong patterns of homophily in misinformation spread.

Bertani et al. [15] study how online emotional responses differ during the COVID-19 vaccination campaign compared to a neutral baseline, focusing on socially sensitive and polarizing topics. They find that online discussions during the pandemic evoke a broader range of emotions, influenced by user characteristics and the type of information shared. The study also highlights the role of political orientation in driving the circulation of news, as emotionally charged posts serve to reinforce group affiliations within online communities.

Pratelli et al. [16] analyze Twitter activity during the 2020 U.S. presidential pre-election debate, focusing on the spread of disinformation, particularly in swing states. They find that 88% of online traffic was linked to swing states, where disinformation, often shared by automated accounts, was more prevalent. The debate discussions were led by two main communities—one predominantly Republican, where most disinformation originated, and another with more diverse political affiliations. The findings highlight the heightened disinformation activity in politically contested regions under the winner-take-all electoral system.

Tardelli et al. [17] investigate coordinated behaviors on Twitter during the 2020 U.S. presidential election, highlighting the various groups that participated in online political debates. Utilizing advanced network science methods, the authors identify three main categories of coordinated users: moderate groups genuinely engaged in the electoral discussion, conspiratorial groups spreading misinformation, and foreign influence networks attempting to disrupt or exploit the debate. The findings indicate extensive automation within far-right and conspiratorial communities, while left-leaning users exhibited less coordination and focused on factual communication. The study also assesses Twitter's effectiveness in countering some coordinated activities, contributing to a deeper understanding of online interactions and strategies to mitigate cyber social threats.

Muñoz et al. [18] analyze political polarization on Twitter in Spain from 2011 to 2019, focusing on official political party accounts during various election phases. They perform an extensive comparative analysis of algorithms designed to measure polarization on microblogging platforms. This analysis results in the creation of a new algorithm specifically tailored to capture polarization during political events, which is subsequently validated using real-world data. This study makes substantial contributions to political science, social network analysis, and computational social science by providing a practical method for analyzing polarization in online political discourse.

Alieva et al. [19] study Russia's propaganda on Twitter during the 2022 invasion of Ukraine, focusing on the "fascism/Nazism" narrative. Using network analysis, natural language processing, and qualitative methods, the authors identify key communities and influential actors, as well as the main topics and impactful messages associated with this disinformation. The study enhances understanding of how propaganda spreads on social media and sheds light on the narratives and communities driving disinformation during the invasion.

3 Conclusion

The contributions included in this topical issue provide valuable insights into the dynamics of online platforms, showcasing methodologies for analyzing polarization, propaganda, and disinformation narratives during critical events such as elections and crises.

Utilizing network analysis and natural language processing techniques, these studies deepen our comprehension of how online interactions influence public opinion and affect democratic processes in a highly polarized context. Furthermore, they emphasize the importance of information integrity during crises, as accurate and reliable information is essential for informed decision-making and maintaining public trust in democratic institutions.

Author contributions
All authors read and approved the final manuscript.

Declarations

Competing interests
F.P. declares he is an associate editor of EPJ Data Science.

Author details
[1]Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy. [2]Observatory on Social Media, Indiana University, Bloomington, IN, USA. [3]Network Science Institute, Northeastern University, Boston, MA, USA. [4]Department of Computer Science, Binghamton University, Binghamton, NY, USA. [5]Institute for Insight, Georgia State University, Atlanta, GA, USA.

Accepted: 22 October 2024 Published online: 28 October 2024

References

1. Pierri F (2020) The diffusion of mainstream and disinformation news on Twitter: the case of Italy and France. In: Companion proceedings of the web conference 2020, pp 617–622
2. Sheth A, Shalin VL, Kursuncu U (2022) Defining and detecting toxicity on social media: context and knowledge are key. Neurocomputing 490:312–318
3. Sahneh ES, Nogara G, DeVerna MR, Liu N, Luceri L, Menczer F, Pierri F, Giordano S (2024) The dawn of decentralized social media: an exploration of bluesky's public opening. In: 2024 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM)
4. Pierri F, Tocchetti A, Corti L, Di Giovanni M, Pavanetto S, Brambilla M, Ceri S (2021) VaccinItaly: monitoring Italian conversations around vaccines on Twitter and Facebook. In: Workshop proceedings of the 15th international AAAI conference on web and social, Media
5. Nogara G, Pierri F, Cresci S, Luceri L, Giordano S (2024) Misinformation and polarization around COVID-19 vaccines in France, Germany, and Italy. In: Proceedings of the 16th ACM web science conference, pp 119–128
6. La Gatta V, Wei C, Luceri L, Pierri F, Ferrara E (2023) Retrieving false claims on Twitter during the Russia-Ukraine conflict. In: Companion proceedings of the ACM web conference 2023, pp 1317–1323
7. Bawa A, Kursuncu U, Achilov D, Shalin VL (2024) the adaptive strategies of anti-kremlin digital dissent in telegram during the Russian invasion of Ukraine. ArXiv preprint. arXiv:2408.07135
8. Kursuncu U, Yang K-C, Pierri F, Deverna MR, Mejova Y, Blackburn J (2023) CySoc 2023: 4th international workshop on cyber social threats. In: Companion proceedings of the ACM web conference 2023, pp 1307–1307
9. Pierri F, Perry BL, DeVerna MR, Yang K-C, Flammini A, Menczer F, Bryden J (2022) Online misinformation is linked to early COVID-19 vaccination hesitancy and refusal. Sci Rep 12(1):5966
10. Yang K-C, Pierri F, Hui P-M, Axelrod D, Torres-Lugo C, Bryden J, Menczer F (2021) The COVID-19 infodemic: Twitter versus Facebook. Big Data Soc 8(1):20539517211013861. https://doi.org/10.1177/20539517211013861
11. Singh I, Scarton C, Bontcheva K (2023) Utdrm: unsupervised method for training debunked-narrative retrieval models. EPJ Data Sci 12(1):59
12. Ng LHX, Carley KM (2023) Deflating the Chinese balloon: types of Twitter bots in US-China balloon incident. EPJ Data Sci 12(1):63
13. Xia Y, Gronow A, Malkamäki A, Ylä-Anttila T, Keller B, Kivelä M (2024) The Russian invasion of Ukraine selectively depolarized the Finnish NATO discussion on Twitter. EPJ Data Sci 13(1):1
14. Truong BT, Allen OM, Menczer F (2024) Account credibility inference based on news-sharing networks. EPJ Data Sci 13(1):10
15. Bertani A, Gallotti R, Menini S, Sacco P, De Domenico M (2024) Large-scale digital signatures of emotional response to the COVID-19 vaccination campaign. EPJ Data Sci 13(1):20
16. Pratelli M, Petrocchi M, Saracco F, De Nicola R (2024) Online disinformation in the 2020 US election: swing vs. safe states. EPJ Data Sci 13(1):25
17. Tardelli S, Nizzoli L, Avvenuti M, Cresci S, Tesconi M (2024) Multifaceted online coordinated behavior in the 2020 US presidential election. EPJ Data Sci 13(1):33
18. Muñoz P, Bellogín A, Barba-Rojas R, Díez F (2024) Quantifying polarization in online political discourse. EPJ Data Sci 13(1):39
19. Alieva I, Kloo I, Carley KM (2024) Analyzing Russia's propaganda tactics on Twitter using mixed methods network analysis and natural language processing: a case study of the 2022 invasion of Ukraine. EPJ Data Sci 13(1):42

Publisher's Note