OHARS: Second Workshop on Online Misinformation- and Harm-Aware Recommender Systems

Antonela Tommasel
antonela.tommasel@isistan.unicen.edu.ar
ISISTAN (CONICET/UNCPBA)
Tandil, Argentina

Daniela Godoy
daniela.godoy@isistan.unicen.edu.ar
ISISTAN (CONICET/UNCPBA)
Tandil, Argentina

Arkaitz Zubiaga
a.zubiaga@qmul.ac.uk
Queen Mary University of London
London, UK

ABSTRACT

Recommender systems play a central role in online information consumption and user decision-making by leveraging user-generated information at scale to assist users in finding relevant information and establishing new social relationships. Just as recommendation techniques have become powerful tools that are inserted in most social platforms, they could also involuntarily spread unwanted content and other types of online harms. The same fundamental concepts on which these techniques rely make them facilitators of such unwanted diffusion. To increase the user-perceived quality of recommender systems and mitigating the negative effects of the multiple forms of online harms, it is essential to provide recommender systems with harm-aware mechanisms. To further research in this direction, this Second edition of the Workshop on Online Misinformation- and Harm-Aware Recommender Systems (OHARS 2021) aimed at fostering research in recommender systems that can mitigate the negative effects of online harms by fostering the recommendation of safe content and trustworthy users, with a special interest in research tackling the negative effects of the propagation of harmful content referring to the COVID-19 crisis.

CCS CONCEPTS

• **Information systems → Recommender systems**.

KEYWORDS

Recommender systems, online harms, misinformation, hate speech

ACM Reference Format:
Antonela Tommasel, Daniela Godoy, and Arkaitz Zubiaga. 2021. OHARS: Second Workshop on Online Misinformation- and Harm-Aware Recommender Systems. In *Fifthteenth ACM Conference on Recommender Systems (RecSys '21), September 27-October 1, 2021, Amsterdam, Netherlands.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3460231.3470941

1 INTRODUCTION

In recent years, there has been an increase in the dissemination of false news, rumors, deception and other forms of misinformation, as well as abusive language, incitements of violence, harassment and other forms of hate speech, throughout online platforms. These unwanted behaviours lead to online harms [6] which have become a serious problem with several negative consequences, ranging from public health issues to the disruption of democratic systems [8]. While these phenomena are widely observed in social media, they affect the experience of users on multiple online platforms. For example, collaborative filtering approaches in e-commerce sites are vulnerable to low-quality reviews, manipulation and attacks. In this regard, Amazon has been criticized for allowing vendors to promote white supremacist and anti-Semitic merchandise, which can foster hate crime [7]. Moreover, PayPal monitors users' transactions to avoid providing services to users promoting hateful actions, regardless of whether their activities are illegal [10].

In the last year, the COVID-19 pandemic generated an increased need for information as a response to a highly emotional and uncertain situation. Consequently, cases of misinformation linked to health recommendations have been reported during the COVID-19 pandemic (for example, different media outlets, and even politicians, recommended consuming hot beverages and chlorine dioxide for preventing the disease), which undermines the individual responses to COVID-19, compromises the efficacy of evidence-based policy interventions, and affects the credibility of scientific expertise with potentially longer-term (and even deadly) consequences [3]. At the same time, actions were demanded to control the "tsunami" of hate speech which is rife during the COVID-19 pandemic[1].

Recommender systems play a central role in online information consumption and user decision-making by leveraging user-generated information at scale. As a result, they are affected by different forms of online harms, which may hinder the accuracy of predictions while, at the same time, become unintended means for their spread and amplification. In fact, these systems have recently gone under heavy criticism for promoting the creation of filter bubbles, which contribute to lowering the diversity of the information users are exposed to and the social contacts they create [2]. Some of these issues relate to the concepts and underlying assumptions on which recommender systems are based. For example, the homophily principle according to which like-minded users have a tendency to express an interest in the same items, might lead to information that users are already likely to know or agree with, contributing to the filter bubble effect. These assumptions can be naïve and exclusionary in the era of fake news and ideological uniformity [4]. In their attempt to deliver relevant and engaging suggestions, recommendation algorithms are prone to introduce biases [5], and further foster phenomena such as filter bubbles, echo chambers and opinion manipulation. Similarly, users vulnerability to misinformation and disinformation can be fostered by data, algorithm and interaction biases [1], which contribute to limiting the exposure of users to multiple and diverse points of view.

RecSys '21, September 27-October 1, 2021, Amsterdam, Netherlands

ACM ISBN 978-1-4503-8458-2/21/09.
https://doi.org/10.1145/3460231.3470941

[1] https://news.un.org/en/story/2020/05/1063542

To increase the user-perceived quality of recommender systems and mitigating the negative effects of unwanted content and other forms of online harms, it is essential to design and implement harm-aware mechanism to be integrated into recommender systems. In this regard, recommendation diversification techniques, model-level disruption and explainability techniques could help users to effectively identify the different forms of online harm and make informed decisions regarding what they share and with whom they interact, among other possibilities.

Following the first edition in 2020 [9], OHARS 2021 was the second edition of the Workshop on Online Misinformation- and Harm-Aware Recommender Systems[2]. In this second edition, the workshop aimed at furthering research in recommender systems that can circumvent the negative effects of online harms by promoting the recommendation of safe content and users, with a special interest in research tackling the negative effects of recommending fake or harmful content linked to the COVID-19 crisis. The end goal was to facilitate the discussion about the major challenges and opportunities that will shape future research.

2 WORKSHOP FORMAT AND TOPICS

OHARS was organized as an interactive half-day workshop in conjunction with RecSys 2021 hybrid event. The workshop programme included short presentations and a keynote, with the aim of discussing on the different aspects of harm-aware recommender systems in relation to experiences from the practice of social computing sciences (e.g., specific problems, conceptual models, use cases). The workshop was organized with the aim of fostering the exchange of experiences and research working from different fields but on related problems. Contributions were invited in all topics related to misinformation- and harm-aware recommender systems, focusing on:

- Reducing misinformation effects (e.g. echo-chambers, filter bubbles)
- Hate speech detection and countermeasures
- Online harms dynamics and prevalence
- Computational models for multi-modal and multi-lingual harm detection and countermeasures
- User/content trustworthiness
- Bias detection and mitigation in data/algorithms
- Fairness, interpretability and transparency in recommendations
- Explainable models of recommendations
- Dataset collection and processing
- Design of specific evaluation metrics
- Applications and case studies of misinformation- and harm-aware recommender systems
- The appropriateness of countermeasures for tackling online harms in recommender systems.
- Mitigation strategies against coronavirus-fueled hate speech and COVID-related misinformation propagation.
- Ethical and social implications of monitoring, tackling and moderating online harms.
- Online harm engagement, propagation and attacks in recommender systems.
- Privacy preserving recommender systems.
- Attack prevention in collaborative filtering recommender systems

We encouraged works focused on mitigating online harms in domains beyond social media, such as effects in collaborative filtering settings, e-commerce platforms, news-media, video platforms (e.g. YouTube or Vimeo) or opinion mining applications, among other possibilities. Works specifically analyzing any of the previous topics in the context of the COVID-19 crisis were also welcome, as well as works based on social networks other that *Twitter* and *Facebook*, such as *Tik-Tok*, *Reddit*, *Snapchat* and *Instagram*.

OHARS accepted contributions in the form of research papers, presenting novel contributions describing methodology and experimental results (although possibly preliminary) in detail; position papers, introducing novel points of view in the workshop topics or summarizing research experiences; and practice and experience reports, describing real-world scenarios that present harm-aware recommender systems. In addition, submissions providing dataset descriptions, public data collections that could be used to explore or develop harm-aware recommender systems, as well as demo proposals of recommender systems to be demonstrated to the workshop attendees, were welcome.

3 WEBSITE AND PROCEEDINGS

The workshop material (list of accepted papers, invited talks, and the workshop schedule) can be found on the OHARS 2021 workshop website at https://ohars-recsys.isistan.unicen.edu.ar. The proceedings will be made publicly available. A special section of the Online Social Networks and Media journal will consider extended versions of selected papers from the workshop.

ACKNOWLEDGMENTS

We thank the RecSys 2021 organizing committee for giving us the opportunity to host this workshop in conjunction with RecSys 2021. We would also like to thank the authors and members of the Program Committee for their valuable contributions. The organizers are in part supported by the CONICET–Royal Society International Exchange (IEC\R2\192019) and by the United Kingdom's Engineering and Physical Sciences Research Council (grant EP/V048597/1).

REFERENCES

[1] Ricardo Baeza-Yates. 2020. Bias in Search and Recommender Systems. In *Proceedings of the 14th ACM Conference on Recommender Systems (RecSys '20)*. ACM, Virtual Event, Brazil, 2. https://doi.org/10.1145/3383313.3418435

[2] Miriam Fernandez and Alejandro Bellogín. 2020. Recommender Systems and Misinformation: The Problem or the Solution?. In *Proceedings of the Workshop on Online Misinformation- and Harm-Aware Recommender Systems (OHARS 2020)*. CEUR, Virtual Event, Brazil, 40–50.

[3] Kris Hartley and Minh Khuong Vu. 2020. Fighting fake news in the COVID-19 era: policy insights from an equilibrium model. *Policy Sciences* 53, 4 (2020), 735–758.

[4] Taha Hassan. 2019. Trust and Trustworthiness in Social Recommender Systems. In *Companion Proceedings of The 2019 World Wide Web Conference* (San Francisco, USA) *(WWW '19)*. ACM, New York, NY, USA, 529–532. https://doi.org/10.1145/3308560.3317596

[5] Dimitar Nikolov, Mounia Lalmas, Alessandro Flammini, and Filippo Menczer. 2019. Quantifying Biases in Online Information Exposure. *Journal of the Association for Information Science and Technology* 70, 3 (2019), 218–229.

[6] House of Commons. 2019. Disinformation and "Fake news": Final Report. *UK Parliament* (2019). https://doi.org/pa/cm201719/cmselect/cmcumeds/1791/1791.pdf Accessed 20-Feb-2021.

[7] Partnership for Working Families. 2020. DELIVERING HATE: How Amazon's Platforms Are Used to Spread White Supremacy,

[2] https://ohars-recsys.isistan.unicen.edu.ar/

Anti-Semitism, and Islamophobia and How Amazon Can Stop It. https://www.forworkingfamilies.org/resources/publications/d-e-l-i-v-e-r-i-n-g-h-e-how-amazon%E2%80%99s-platforms-are-used-spread-white. (accessed March 26, 2020).

[8] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. 2018. The spread of low-credibility content by social bots. *Nature communications* 9, 1 (2018), 1–9.

[9] Antonela Tommasel, Daniela Godoy, and Arkaitz Zubiaga. 2020. Workshop on Online Misinformation- and Harm-Aware Recommender Systems. In *Fourteenth ACM Conference on Recommender Systems* (Virtual Event, Brazil) *(RecSys '20)*. Association for Computing Machinery, New York, NY, USA, 638–639. https://doi.org/10.1145/3383313.3411537

[10] Natasha Tusikov. 2019. Defunding hate: PayPal's regulation of hate groups. *Surveillance & Society* 17, 1/2 (2019), 46–53.