

# Workshop on Online Misinformation- and Harm-Aware Recommender Systems

ANTONELA TOMMASEL, antonela.tommassel@isistan.unicen.edu.ar ISISTAN Research Institute (CONICET/UNCPBA), Argentina

DANIELA GODOY, daniela.godoy@isistan.unicen.edu.ar ISISTAN Research Institute (CONICET/UNCPBA), Argentina

ARKAITZ ZUBIAGA, a.zubiaga@qmul.ac.uk Queen Mary University of London, UK

Recommender systems play an important role in the dissemination and propagation of information. This is particularly true for large scale platforms such as social media, where recommender systems assist users in facilitating access to massive user-generated content by finding relevant information and establishing new social relationships. Just as recommendation techniques are designed to become powerful tools, they could in turn spread online harm. Some of these issues stem from the core concepts and assumptions of recommender systems. Harnessing recommender systems with misinformation- and harm-awareness mechanisms becomes essential not only to mitigate the negative effects of the propagation of harmful content, but also to increase the quality and diversity of recommender systems. To further research in this direction, the Workshop on Online Misinformation- and Harm-Aware Recommender Systems (OHARS 2020) aimed at fostering research in recommender systems that can circumvent the negative effects of online harms by promoting the recommendation of safe content and users.

CCS Concepts: • **Information systems** → **Recommender systems**.

Additional Key Words and Phrases: Recommender systems, online harms, misinformation, hate speech

## ACM Reference Format:

Antonela Tommasel, Daniela Godoy, and Arkaitz Zubiaga. 2020. Workshop on Online Misinformation- and Harm-Aware Recommender Systems. In *Fourteenth ACM Conference on Recommender Systems (RecSys '20), September 21–26, 2020, Virtual Event, Brazil*. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/3383313.3411537>

## 1 INTRODUCTION

Social media platforms have become an integral part of everyday life and activities of most people, providing new forms of communication and interaction. These sites allow their users to share information and opinions (in the form of photos, short texts and comments), while at the same time promote the formation of links and social relationships (friendships, follower/followee relations). One of the most valuable features of social platforms is the potential for the dissemination of information at scale. Recommender systems play an important role in this process as they leverage the massive user-generated content to assist users in finding relevant information as well as establishing new social relationships.

The adoption of social media, however, also exposes users to some risks [3], which can have a damaging effect on individuals and society at large. The unmoderated nature of social media sites often results in the appearance and distribution of false, misleading content (for example, hoaxes, conspiracy theories, false news and even satires), or even harmful content such as abusive, discriminatory and offensive comments, and incitement to acts of violence. In fact, the proliferation of misinformation and hate speech online has become a serious problem with several negative consequences, ranging from public health issues to the disruption of democratic systems [2].

---

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2020 Copyright held by the owner/author(s).

Manuscript submitted to ACM

As mediators of online information consumption, recommender systems are both affected by the proliferation of low-quality content in social media, which hinders their capacity of achieving accurate predictions, and, at the same time, become unintended means for the amplification and massive distribution of online harm. Some of these issues stem from the core concepts and assumptions recommender systems are based on. For example, the homophily principle according to which similar users are likely to be interested in the same items, might lead to information that users are already likely to know or agree with, yielding to the so-called "echo chambers". Assumptions like these can be naïve and exclusionary in the era of fake news and ideological uniformity [1]

While attempting to deliver relevant and engaging suggestions about content/users, recommendation algorithms are bound to introduce biases [4]. For example, popularity and homogeneity bias are based on the reliance in popular sources and social networks of like-minded individuals, correspondingly. These common biases limit the exposure of users to diverse points of view and make them vulnerable to manipulation by disinformation. Likewise, recommender systems can be affected by biases in the data (stemming from imbalanced data), the algorithms, and the user interaction or observation – with a focus on the biases related to relevance feedback loops (e.g., ranking) [5].

Harnessing recommender systems with misinformation- and harm-awareness mechanisms becomes essential not only to mitigate the negative effects of the diffusion of unwanted content, but also to increase the user-perceived quality of recommender systems. Novel strategies like the diversification of recommendations, bias mitigation, model-level disruption, explainability and interpretation, among others, can help users in informed decision-making in the context of online misinformation, hate speech and other forms of online harm.

OHARS 2020 was the first edition of the Workshop on Online Misinformation- and Harm-Aware Recommender Systems<sup>1</sup>. The aim of this workshop was to bring together a community of researchers interested in tackling online harms and mitigating their impact on recommender systems and, thus, to facilitate discussion about the major challenges and opportunities that will shape future research.

## 2 WORKSHOP FORMAT AND TOPICS

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
- User/content trustworthiness
- Bias detection and mitigation in data/algorithms
- Fairness 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

OHARS was organized as an interactive half-day workshop in conjunction with RecSys 2020 online 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.

The workshop 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 accepted.

<sup>1</sup><https://ohars-recsys2020.isistan.unicen.edu.ar/>

### 3 WORKSHOP SUMMARY

Over 10 papers were submitted to the workshop and were each refereed by three members of the Program Committee, based on their novelty, technical quality, potential impact, insight, depth, clarity, and reproducibility. Most of the submitted papers focused on technical aspects related to harm-aware recommender systems, and the detection of harmful content and its credibility. For example, related to harm-aware recommender systems, the papers proposed tackling the filter bubble phenomenon in news recommendation systems, reducing harmful or unhealthy recommendations whilst maintaining user satisfaction based on user feedback, and a recommender for privacy preserving practices in social media. In addition, one of the papers proposed an analysis of whether and how recommender systems foster the propagation of misinformation and the creation of filter bubbles, and what strategies could be followed to include diversity in recommendations aiming at widening users' access to recommendations. Finally, three papers tackled the ethics and human rights aspects of the existence of bias in recommender systems, and the adoption of regulation measures to track, record and remove harmful content in social media.

### 4 WEBSITE AND PROCEEDINGS

The workshop material (list of accepted papers, invited talks, and the workshop schedule) can be found on the OHARS 2020 workshop website at <https://ohars-recsys2020.isistan.unicen.edu.ar>. The proceedings will be made publicly available. A special issue based on the topics of the Workshop is also being guest edited in the Personal and Ubiquitous Computing journal.

### ACKNOWLEDGMENTS

We thank the RecSys 2020 organizing committee for giving us the opportunity to host this workshop in conjunction with RecSys 2020. 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).

### REFERENCES

- [1] Taha Hassan. 2019. Trust and Trustworthiness in Social Recommender Systems. In *Companion Proceedings of The 2019 World Wide Web Conference (WWW '19)*. ACM, San Francisco, USA, 529–532.
- [2] House of Commons. 2019. Disinformation and “Fake news”: Final Report. *UK Parliament* (2019). <https://doi.org/pa/cm201719/cmsselect/cmcomeds/1791/1791.pdf> Accessed 19-Feb-2020.
- [3] David M. J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts, and Jonathan L. Zittrain. 2018. The science of fake news. *Science* 359, 6380 (2018), 1094–1096.
- [4] 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.
- [5] Ricardo Baeza-Yates. 2019. Biases in Search and Recommender Systems. <https://www.searchenginejournal.com/biases-search-recommender-systems/339319/close>. Online; accessed 13 July 2020.