Recommender Systems and Misinformation: The Problem or the Solution?

MIRIAM FERNANDEZ, Open University
ALEJANDRO BELLOGÍN*, Universidad Autónoma de Madrid

Recommender Systems have been pointed as one of the major culprits of misinformation spreading in the digital sphere. These systems have recently gone under heavy criticism for promoting the creation of filter bubbles, lowering the diversity of information users are exposed to and the social contacts they create. This influences the dynamics of social news sharing, and particularly the ways misinformation initiates and propagates. However, while Recommender Systems have been accused of fuelling the spread of misinformation, it is still unclear which particular types of recommender algorithms are more prone to recommend misinforming news, and if, and how, existing recommendation algorithms and evaluation metrics, can be modified or adapted to mitigate the misinformation spreading effect. In this position paper, we describe some of the key challenges behind assessing and measuring the effect of existing recommendation algorithms on the recommendation of misinforming articles and how such algorithms could be adapted, modified, and evaluated to counter this effect based on existing social science and psychology research.

Additional Key Words and Phrases: misinformation, news, recommender systems

1 INTRODUCTION

Misinformation has become a common part of our digital media environments, and it is compromising the ability of our societies to form informed opinions [14]. It generates misperceptions, which affect our decision making processes in many domains, including economy, health, environment, or elections. In 2016, post-truth was chosen by the Oxford Dictionary as the word of the year, after achieving a 2000% increase "in the context of the EU referendum in the United Kingdom and the presidential election in the United States". Today, in the context of a global pandemic, misinformation has led to tragic results, including links to assaults, arson and deaths.[1] Although misinformation is a common problem in all media, it is exacerbated in digital social media due to the speed and ease in which posts are spread. The social web enables people to spread information rapidly without confirmation of truth, and to paraphrase this information to fit their intentions and present beliefs [22].

Multiple factors influence the spread of misinformation online including: (i) the ways in which **information** is constructed and presented [12, 34], (ii) the **users**' personality, values and emotions [20, 32], (iii) the architectural characteristics of the **digital platforms** where such information is spread (i.e., the structure of the social networks, constraints on the type of messages or sharing permissions, etc.) [1] and (iv) the **algorithms** that power the recommendation of information within those platforms.[2] While multiple works have concentrated on studying

*Both authors contributed equally to this research.
[1]https://www.bbc.co.uk/news/stories-52731624
[2]https://www.wired.com/story/creating-ethical-recommendation-engines/, https://www.buzzfeednews.com/article/craigsilverman/how-facebook-groups-are-being-exploited-to-spread

OHARS 2020, Sept 22–26, 2020, Rio de Janeiro, BR

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

the effect of different types of information, users and digital platforms on the spread of misinformation, we argue in this paper that there is a need to further explore the effect of existing algorithms on the recommendation of false and misleading information.

As mediators of online information consumption, recommendation algorithms have been strongly criticised for becoming unintended means for the amplification and distribution of misinformation [25].[3] This problem is rooted in the core design and principles in which these algorithms are based. The assumption that users are interested in items that are similar to the ones for which they expressed a preference in the past, or in items that are liked by users that are similar to them, helps build up and boost the so-called "echo-chambers". Moreover, in their attempt to deliver relevant suggestions, recommendation algorithms are prone to amplify biases, such as popularity and homogeneity biases [3, 18]. Echo-chambers and biases may limit the exposure of users to diverse points of view, potentially making them vulnerable to misinformation.

Aiming to break these echo-chambers and to reduce the spread of misinformation, different online platforms are applying different strategies. Twitter, for example, started recommending popular tweets into the feeds of people who did not subscribe to the accounts that posted them. This approach, of providing popular opposing views, was however heavily criticised for amplifying inflammatory political rhetoric and misinformation.[4] Additionally, research indicates that presenting people with corrective information is likely to fail in changing their salient beliefs and opinions, or may, even, reinforce them [16]. People often struggle to change their beliefs even after finding out that the information they already accepted is incorrect or misleading. Nevertheless, some strategies have been found to be effective in correcting misperceptions, such as exposing users to related but disconfirming stories [4], or revealing the demographic similarity of the opposing group [16].

In this paper we argue that, while recommendation algorithms have been heavily criticised for promoting the spread of misinformation, a more in-depth investigation is needed to better understand which of these algorithms are more prone or susceptible of spreading misinformation, under which circumstances, and how the internal functioning of such systems could be modified, or adapted, to counter their misinformation recommendation behaviour. The next section presents our proposed research vision, and the different research building blocks we envision are needed to target this problem.

2 RESEARCH DIMENSIONS AND CHALLENGES

Understanding which recommendation algorithms are more prone to spread misinformation, and under which circumstances is not a trivial problem. Misinformation spreading is a problem with a high number of dimensions that interrelate to one another, some of them also affecting what recommendation algorithms learn and therefore, how they will later on behave. Similarly, adapting such algorithms to counter their misinformation recommendation behaviour requires an in-depth understanding not only of the internal mechanisms of such algorithms, but also of the data they manipulate, the users they serve, and the platforms they operate in. In this section, we present our vision on how to address these problems based on four key building blocks or research dimensions, and we discuss the challenges associated to each of them. The proposed research dimensions are represented in Figure 1, together with how they interrelate with one another.

The first one, *Misinformation: Problem Dimensions*, aims to understand what are the different dimensions of the misinformation problem, and within them, the aspects that may affect the behaviour of recommendation algorithms (e.g., the users, the type of information, etc.). The second one, *Analysis of Recommendation Algorithms*, refers to the need of conducting in-depth investigations of the different existing recommendation algorithms (content-based, collaborative filtering such as matrix factorisation, demographic- and knowledge-based techniques,

[3]https://www.niemanlab.org/2020/01/youtubes-algorithm-is-pushing-climate-misinformation-videos-and-their-creators-are-profiting-from-it/

[4]https://edition.cnn.com/2019/03/22/tech/twitter-algorithm-political-rhetoric/index.html

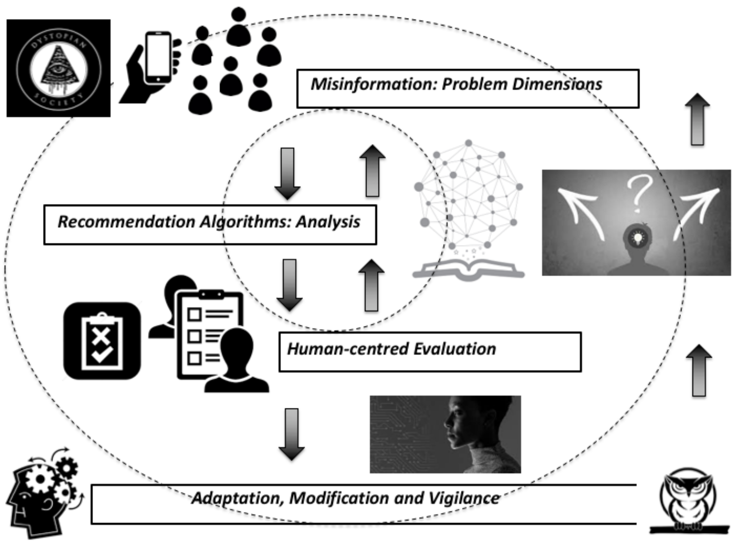

Fig. 1. Research Dimensions

and so on) and how their internal mechanisms may improve or worsen the spread of misinformation. The third one, *Human-centred Evaluation*, refers to the need of modifying existing evaluation methods and metrics to target not only the enhancement of content and/or user similarity, or user satisfaction, but maybe promoting some degree of dissimilarity and cognitive dissonance that could help users to break their filter bubbles. The fourth dimension, *Adaptation, Modification and Vigilance*, refers to the investigation of the different ways in which existing algorithms could be modified and adapted to counter their misinformation recommendation behaviour. Vigilance refers to the need of constant monitoring to ensure that: (i) the dynamics of misinformation are captured over time, and that algorithms are adapted accordingly and, (ii) that the proposed adaptations do not back-fire or introduce any additional ethical issues (e.g., algorithmic adaptations that may reduce the recommendation of misinformation, but that tend to promote misinformation of a more harmful nature should not be considered successful).

As we can see in Figure 1, these four research dimensions closely interrelate with one another. The inner circle of the figure touches on the first three research dimensions. These dimensions encapsulate the first part of the investigation, understanding the effect of recommendation algorithms on the spread of misinformation. The outer circle also includes the fourth dimension, i.e, investigating how these algorithms could be modified to counter their harmful behaviour. Arrows in and out of these dimensions indicate how they influence one another. Understanding the different variables that affect the problem of misinformation (the types of content, the types of users, etc.) is key to assess which of those variables may also affect the development, training and learning of recommendation algorithms. Similarly, understanding the internal mechanisms of such algorithms, in conjunction with the problem of misinformation, is necessary to design appropriate evaluation protocols that, while effectively satisfying the users' information needs, also palliate the recommendation of misleading information. Research centred in all these four dimensions is needed to better comprehend if, and how, we could improve existing recommendation mechanisms.

2.1 Misinformation: Problem Dimensions

As mentioned earlier, misinformation is a problem with multiple dimensions (human, sociological, technological). We will mention here some of the most prominent ones closely related to the development of recommendation algorithms. Note that this does not aim to be an exhaustive list, and more research is needed to identify the various dimensions that may intersect between recommendation and misinformation.

Content is an important dimension of the misinformation problem, and also a key one to consider in the creation, assessment and adaptation of recommendation algorithms. In the case of our proposed vision, we consider that the items to be recommended are online news. News can be present in various *forms* (as news paper articles, blog posts, social media posts, etc.) and discuss a wide range of *topics* (health, elections, disasters, etc.). These items are not only textual, but sometimes include information in different *formats*, like images or videos. Note that combinations of these formats are frequently used to propagate misinformation (e.g., a news title linked with an image from a different place, or from a different time). The *framing* of misinforming articles also varies between false news, rumours, conspiracy theories, misleading content [34]. Other important elements to consider about content are their *origin* (news outlets, social contacts, public figures, etc.) as well as the time when they are posted. Note that recency is particularly relevant on the recommendation of news items. All these aspects of the content need to be considered in the design and adaptation of news recommendation algorithms.

Users are a key dimension of the misinformation problem, and a core one of the functioning of recommendation algorithms. Multiple works have therefore studied the effect of different *motivations* [9], *personalities* [36], *values* [26], and *emotions* [32] and their effect on misinformation, as well as the *susceptibility* of users [33]. For example, extroverts and individuals with high cooperativeness and high reward dependence are founded more prone to share misinformation [9]. Psychology also shows that individuals with higher anxiety levels are more likely to spread misinformation [17]. These aspects have demonstrated to influence users in the spreading of misinformation, hence it would be important to consider and capture them, as much as possible, during the construction of user profiles for recommendation.

Platform and Network features. Platforms that distribute online information are designed differently and therefore facilitate the spread of misinformation in different ways. *Content limitations* (e.g., Twitter and its 280 character limitation for posts), *ability to share information* and select the subsets of users with whom such information is shared (*sharing permissions*), the ability to *vote* (e.g., Reddit) or to *express emotions* for content (e.g. Facebook), are important aspects of platform design that may shape the content, the way information spreads, and the *social network structure*. The typology and topology of the network structure is also a key factor of misinformation dynamics [14].

Other dimensions of the misinformation problem that may affect the development of recommendation algorithms include: *the global events* or the developments happening in the world in a particular moment in time, *the ethics* of information (tensions between privacy and security, censorship, cultural differences, etc.), *the presence of malicious actors* including bots or crowdturfing (where crowdworkers are hired to support and propagate arguments or claims, simulating grassroots social movement), or *the presence of checked facts* within the information space.

2.2 Analysis of Recommendation Algorithms

A key aspect of our research is to understand which recommendation techniques are more prone to suggest misinformative items to users. For this, we need to select a representative pool of algorithms to test against appropriate datasets, among the well-known collaborative filtering (CF), content-based (CB), and hybrid techniques [28].

The recommender systems literature is mostly focused on CF, since these methods can be applied to any domain, only requiring user-item interactions, not needing additional item features or metadata. Studying these methods may help us to better understand whether the user-item interactions could help, by themselves, to spread or avoid misinformative items, since these models neglect any information about the items and their misinformative features. Our main hypothesis is that, since CF algorithms tend to reproduce the tastes of the majority [7], they will probably follow the trend (either spreading or avoiding misinformative items) for those topics where an opinion is already established by most of the community. Nonetheless, it should also be noted that, since the user's previous activity is also considered, this effect may not be so clear, paving the way for dynamic approaches that combine personalised with global models to avoid propagating misinformation.

While CB techniques are less common in the area, techniques based on tags or 'short texts' (such as reviews, or other user generated content) [10, 31] have attracted attention due to their ability to consider the content of the items and adapt in more detail to the domain at hand. In this particular context, where news items are the ones to be recommended, it is important to analyse their textual content. This requests to study recommendation algorithms dealing with natural language and its inherent subtleties (synonyms, negation, sarcasm, etc.). An in-depth algorithmic survey is therefore required to better understand the impact of these techniques in the recommendation of misinformation. This includes classical and hybrid collaborative algorithms [15, 19], and more recent methods aimed at understanding the natural language by, for instance, using Neural Networks [2, 11].

When analysing recommendation algorithms it is also particularly important to study how to define user and item profiles [10]. Multiple representations could be considered: using the full content of the news items, a summary, or even some tags or categories assigned to each item. Explicit modelling of whether an item contains misinformation or not, although helpful, might put these techniques in an unfair position with respect to the CF methods described before. This information should therefore not be included in the studied models, at least in the first stages, so that a fairer assessment of the recommendation algorithms, and their tendency to spread misinformative items, can be conducted.

To the best of our knowledge, the current state-of-the-art on recommender systems does not answer important questions around whether, how, and which, recommender systems tend to spread misinformation. Some connections can be made with the popularity-diversity bias analysed in the context of CF [35], where algorithms optimised for accuracy tend to be biased towards popular items, while returning not novel and not diverse recommendations. Similarly, CB algorithms are well-known for their portfolio effect and content overspecialisation [6]. It is therefore expected that if a user tends to consume misinformative items these algorithms might reinforce such content. However, since this may occur in a user basis, its global effect in the community needs to be properly analysed.

2.3 Human-centred Evaluation

Traditionally, "relevancy" has been considered as the primary dimension to determine the quality of recommendations. We hypothesise that 'relevancy', and **metrics** that target user satisfaction, may not be the most effective ones when aiming to reduce the impact and spread of misinformation. Algorithms promoting a certain degree of cognitive dissonance and metrics that focus on computing a balanced degree of user satisfaction and discomfort may be more suitable to assess and combat misperceptions.

Moreover, in order to detect and contrast the spread of misinformation, we need **datasets** where some *items* have already been labelled or identified as such. Examples of these datasets include the *NELA-GT-2018 dataset*[5], which contains 713 articles collected between 02/2018-11/2018 directly from 194 news and media outlets including mainstream, hyper-partisan, and conspiracy sources, or the more recent COVID-19 dataset generated by fact-checkers in more than 70 countries.[6] However, since we want to apply personalisation algorithms, we also need *user profiles* and *ratings*, which are not available in these data sources. Considering that the research on news recommendation is increasing in the last years, an alternative that could be considered is to combine

[5]https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/ULHLCB

[6]https://www.poynter.org/ifcn-covid-19-misinformation/

public datasets containing user profiles (such as NewsReel [21]) with the previously mentioned labelled datasets. However, it is very likely that the coverage and overlap between those datasets is not large, it might be even minimal. As we can see, the creation of datasets for the evaluation of recommender systems in the context of misinformation spreading requires very careful consideration for the selection and construction of items, profiles and ratings.

Besides the problem of acquiring ground truth data, in order to derive user-centric evaluation metrics that could capture, at the same time, the degree of user satisfaction and cognitive discomfort, we need to come up with a satisfactory answer to the following question: *how can we measure when the task has been successfully addressed?* This translates into assessing that the misinformation has decreased, but, at the same time, that relevant and hopefully novel, diverse, and out-of-the-bubble content is still presented to the user. We foresee a combination of metrics and evaluation dimensions should be analysed to somehow overcome this problem, in particular, from recent works focused on fairness and transparency [13], information dynamics [29], beyond-accuracy metrics [8], and other biases and characteristics that should be considered [5, 24].

2.4 Adaptation, Modification and Vigilance

Combating misinformation is a complex task, and there is consensus in psychology literature that simply presenting people with corrective information is likely to fail in changing their salient beliefs and opinions, or may, even, reinforce them [23]. People often struggle to change their beliefs even after finding out that the information they already accepted is incorrect [30]. Nevertheless, some strategies have been found to be effective in correcting misperceptions, such as providing an explanation rather than a simple refute, exposing users to related but disconfirming stories, or revealing similarities with the opposing group [14].

We believe that the adaptation of existing algorithms to counter the misinformation recommendation problem should build on existing social science and psychology theory, such as the one presented above. We therefore need to better profile users in order to capture their motivations and behaviours when spreading misinformation. We also need to adapt recommendation algorithms in a way that recommendations are not solely based on reinforcing similarity, but on introducing small degrees of opposing views, or content from opposing groups, so that similarities between those groups are also highlighted, and explanations can be provided.

More specifically, we could propose recommendation techniques that, instead of being focused on the notion of similarity between users and content, they could be based on the notion of similarity (for users) and dissimilarity (for content), i.e., on the idea of providing divergent views from similar users. This is based on the hypothesis that sharing disconfirming stories from users that hold a degree of similarity could help correcting misperceptions [4, 16]. Furthermore, assuming we would observe a potential trade-off between personalisation (or at least, some bias coming from some families of techniques) and misinformation propagation, we could propose hybrid recommender systems that address these issues, probably at the expense of lower accuracy or strength of user preferences.

3 DISCUSSION

Breaking the cycle that occurs when misinforming news break and are spread through social and digital platforms is paramount. One of the key components of this problem are the recommendation algorithms that power these digital platforms, often accused of feeding and reinforcing such cycle – because of their important role in exposing users to filtered subsets of information –. We argue in this paper that, instead of being solely part of the problem, recommendation algorithms could become part of the solution. This requires a better understanding of how their existing internal mechanisms reinforce the problem of misinformation, and how such mechanisms could be adapted to counter it. Our hypothesis is that this adaptation should be built on existing social science and psychology theory. Studies from these fields have investigated this problem for years, and the behaviours and motivations that are more commonly associated with the spread of misinformation, as well as the potential mechanisms to palliate this effect. With this goal in mind, our research vision revolves around the four building blocks presented before: understanding the dimensions of the misinformation problem, analysing the impact of different recommendation strategies on spreading misinformation, adapting and modifying evaluation methods and metrics to properly assess these effects on users, and investigating how the algorithms could be modified to counter their misinformation behaviour.

While our proposal is related with the principles of Fairness, Accountability and Transparency (FAT), it provides a step forward existing works, which are mainly focused on reducing biases and discrimination. Our aim is to address the problem of misinformation, but not by means of fact-checked information (which is expensive to obtain), or by detecting and containing malicious accounts and messages (which keep propagating), but by translating successful misinformation management strategies from social science research [4, 16] into computational recommendation models, while allowing to break the filter bubbles that tend to appear when using recommender systems [27].

ACKNOWLEDGMENTS

This work has been co-funded by H2020 Co-Inform (ID:770302) and HERoS (ID:101003606) projects and the Ministerio de Ciencia e Innovación (project reference: PID2019-108965GB-I00).

REFERENCES

[1] Hunt Allcott, Matthew Gentzkow, and Chuan Yu. 2019. Trends in the diffusion of misinformation on social media. *Research & Politics* 6, 2 (2019), 2053168019848554.

[2] Nastaran Babanejad, Ameeta Agrawal, Heidar Davoudi, Aijun An, and Manos Papagelis. 2019. Leveraging Emotion Features in News Recommendations. In *Proceedings of the 7th International Workshop on News Recommendation and Analytics in conjunction with 13th ACM Conference on Recommender Systems, INRA@RecSys 2019, Copenhagen, Denmark, September 20, 2019 (CEUR Workshop Proceedings, Vol. 2554)*, Özlem Özgöbek, Benjamin Kille, Jon Atle Gulla, and Andreas Lommatzsch (Eds.). CEUR-WS.org, 70–78. http://ceur-ws.org/Vol-2554/paper_10.pdf

[3] Alejandro Bellogín, Pablo Castells, and Iván Cantador. 2017. Statistical biases in Information Retrieval metrics for recommender systems. *Information Retrieval Journal* 20, 6 (2017), 606–634.

[4] Leticia Bode and Emily K Vraga. 2015. In related news, that was wrong: The correction of misinformation through related stories functionality in social media. *Journal of Communication* 65, 4 (2015), 619–638.

[5] Ludovico Boratto and Mirko Marras. 2020. Hands on Data and Algorithmic Bias in Recommender Systems. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization, UMAP 2020, Genoa, Italy, July 12-18, 2020*, Tsvi Kuflik, Ilaria Torre, Robin Burke, and Cristina Gena (Eds.). ACM, 388–389. https://doi.org/10.1145/3340631.3398669

[6] Robin D. Burke. 2002. Hybrid Recommender Systems: Survey and Experiments. *User Model. User Adapt. Interact.* 12, 4 (2002), 331–370. https://doi.org/10.1023/A:1021240730564

[7] Rocío Cañamares and Pablo Castells. 2018. Should I Follow the Crowd?: A Probabilistic Analysis of the Effectiveness of Popularity in Recommender Systems. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR 2018, Ann Arbor, MI, USA, July 08-12, 2018*, Kevyn Collins-Thompson, Qiaozhu Mei, Brian D. Davison, Yiqun Liu, and Emine Yilmaz (Eds.). ACM, 415–424. https://doi.org/10.1145/3209978.3210014

[8] Pablo Castells, Neil J. Hurley, and Saul Vargas. 2015. Novelty and Diversity in Recommender Systems. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer, 881–918. https://doi.org/10.1007/978-1-4899-7637-6_26

[9] Xinran Chen and Sei-Ching Joanna Sin. 2013. 'Misinformation? What of it?'Motivations and individual differences in misinformation sharing on social media. *Proceedings of the American Society for Information Science and Technology* 50, 1 (2013), 1–4.

[10] Marco de Gemmis, Pasquale Lops, Cataldo Musto, Fedelucio Narducci, and Giovanni Semeraro. 2015. Semantics-Aware Content-Based Recommender Systems. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer, 119–159. https://doi.org/10.1007/978-1-4899-7637-6_4

[11] Gabriel de Souza Pereira Moreira, Dietmar Jannach, and Adilson Marques da Cunha. 2019. On the Importance of News Content Representation in Hybrid Neural Session-based Recommender Systems. In *Proceedings of the 7th International Workshop on News Recommendation and Analytics in conjunction with 13th ACM Conference on Recommender Systems, INRA@RecSys 2019, Copenhagen, Denmark, September 20, 2019 (CEUR Workshop Proceedings, Vol. 2554)*, Özlem Özgöbek, Benjamin Kille, Jon Atle Gulla, and Andreas Lommatzsch (Eds.). CEUR-WS.org, 18–23. http://ceur-ws.org/Vol-2554/paper_03.pdf

[12] Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H Eugene Stanley, and Walter Quattrociocchi. 2016. The spreading of misinformation online. *Proceedings of the National Academy of Sciences* 113, 3 (2016), 554–559.

[13] Michael D. Ekstrand, Robin Burke, and Fernando Diaz. 2019. Fairness and discrimination in recommendation and retrieval. In *Proceedings of the 13th ACM Conference on Recommender Systems, RecSys 2019, Copenhagen, Denmark, September 16-20, 2019*, Toine Bogers, Alan Said, Peter Brusilovsky, and Domonkos Tikk (Eds.). ACM, 576–577. https://doi.org/10.1145/3298689.3346964

[14] Miriam Fernandez and Harith Alani. 2018. Online misinformation: Challenges and future directions. In *Companion Proceedings of the The Web Conference 2018*. 595–602.

[15] Florent Garcin, Boi Faltings, Olivier Donatsch, Ayar Alazzawi, Christophe Bruttin, and Amr Huber. 2014. Offline and online evaluation of news recommender systems at swissinfo.ch. In *Eighth ACM Conference on Recommender Systems, RecSys '14, Foster City, Silicon Valley, CA, USA - October 06 - 10, 2014*, Alfred Kobsa, Michelle X. Zhou, Martin Ester, and Yehuda Koren (Eds.). ACM, 169–176. https://doi.org/10.1145/2645710.2645745

[16] R Kelly Garrett, Erik C Nisbet, and Emily K Lynch. 2013. Undermining the corrective effects of media-based political fact checking? The role of contextual cues and naïve theory. *Journal of Communication* 63, 4 (2013), 617–637.

[17] Marianne E Jaeger, Susan Anthony, and Ralph L Rosnow. 1980. Who hears what from whom and with what effect: A study of rumor. *Personality and Social Psychology Bulletin* 6, 3 (1980), 473–478.

[18] Dietmar Jannach, Lukas Lerche, Iman Kamehkhosh, and Michael Jugovac. 2015. What recommenders recommend: an analysis of recommendation biases and possible countermeasures. *User Modeling and User-Adapted Interaction* 25, 5 (2015), 427–491.

[19] Mozhgan Karimi, Dietmar Jannach, and Michael Jugovac. 2018. News recommender systems - Survey and roads ahead. *Inf. Process. Manag.* 54, 6 (2018), 1203–1227. https://doi.org/10.1016/j.ipm.2018.04.008

[20] Natascha A Karlova and Karen E Fisher. 2013. A social diffusion model of misinformation and disinformation for understanding human information behaviour. (2013).

[21] Andreas Lommatzsch, Benjamin Kille, Frank Hopfgartner, and Leif Ramming. 2018. NewsREEL Multimedia at MediaEval 2018: News Recommendation with Image and Text Content. In *Working Notes Proceedings of the MediaEval 2018 Workshop, Sophia Antipolis, France, 29-31 October 2018 (CEUR Workshop Proceedings, Vol. 2283)*, Martha A. Larson, Piyush Arora, Claire-Hélène Demarty, Michael Riegler, Benjamin Bischke, Emmanuel Dellandréa, Mathias Lux, Alastair Porter, and Gareth J. F. Jones (Eds.). CEUR-WS.org. http://ceur-ws.org/Vol-2283/MediaEval_18_paper_5.pdf

[22] Dung T Nguyen, Nam P Nguyen, and My T Thai. 2012. Sources of misinformation in Online Social Networks: Who to suspect?. In *MILCOM 2012-2012 IEEE Military Communications Conference*. IEEE, 1–6.

[23] Brendan Nyhan and Jason Reifler. 2013. Which corrections work. *Research results and practice recommendation* (2013).

[24] Alexandra Olteanu, Carlos Castillo, Fernando Diaz, and Emre Kiciman. 2019. Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. *Frontiers Big Data* 2 (2019), 13. https://doi.org/10.3389/fdata.2019.00013

[25] Eli Pariser. 2011. *The filter bubble: What the Internet is hiding from you*. Penguin UK.

[26] Lara SG Piccolo, Alisson Puska, Roberto Pereira, and Tracie Farrell. 2020. Pathway to a Human-Values Based Approach to Tackle Misinformation Online. In *International Conference on Human-Computer Interaction*. Springer, 510–522.

[27] Paul Resnick, R. Kelly Garrett, Travis Kriplean, Sean A. Munson, and Natalie Jomini Stroud. 2013. Bursting your (filter) bubble: strategies for promoting diverse exposure. In *Computer Supported Cooperative Work, CSCW 2013, San Antonio, TX, USA, February 23-27, 2013, Companion Volume*, Amy Bruckman, Scott Counts, Cliff Lampe, and Loren G. Terveen (Eds.). ACM, 95–100. https://doi.org/10.1145/2441955.2441981

[28] Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). 2015. *Recommender Systems Handbook*. Springer. https://doi.org/10.1007/978-1-4899-7637-6

[29] Javier Sanz-Cruzado and Pablo Castells. 2018. Enhancing structural diversity in social networks by recommending weak ties. In *Proceedings of the 12th ACM Conference on Recommender Systems, RecSys 2018, Vancouver, BC, Canada, October 2-7, 2018*, Sole Pera, Michael D. Ekstrand, Xavier Amatriain, and John O'Donovan (Eds.). ACM, 233–241. https://doi.org/10.1145/3240323.3240371

[30] Emily Thorson. 2016. Belief echoes: The persistent effects of corrected misinformation. *Political Communication* 33, 3 (2016), 460–480.

[31] Antonela Tommasel and Daniela Godoy. 2018. Short-text feature construction and selection in social media data: a survey. *Artif. Intell. Rev.* 49, 3 (2018), 301–338. https://doi.org/10.1007/s10462-016-9528-0

[32] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359, 6380 (2018), 1146–1151.

[33] Claudia Wagner, Silvia Mitter, Christian Körner, and Markus Strohmaier. 2012. When Social Bots Attack: Modeling Susceptibility of Users in Online Social Networks.. In *# MSM*. 41–48.

[34] Claire Wardle and Hossein Derakhshan. 2017. Information disorder: Toward an interdisciplinary framework for research and policy making. *Council of Europe report* 27 (2017).

[35] Tao Zhou, Zoltán Kuscsik, Jian-Guo Liu, Matúš Medo, Joseph Rushton Wakeling, and Yi-Cheng Zhang. 2010. Solving the apparent diversity-accuracy dilemma of recommender systems. *Proceedings of the National Academy of Sciences* 107, 10 (2010), 4511–4515.

[36] Bi Zhu, Chuansheng Chen, Elizabeth F Loftus, Chongde Lin, Qinghua He, Chunhui Chen, He Li, Gui Xue, Zhonglin Lu, and Qi Dong. 2010. Individual differences in false memory from misinformation: Cognitive factors. *Memory* 18, 5 (2010), 543–555.