11-2023

Understanding the Contribution of Recommendation Algorithms on Misinformation Recommendation and Misinformation Dissemination on Social Networks

Royal Pathak
Boise State University

Francesca Spezzano
Boise State University

Maria Soledad Pera
Technische Universiteit Delft

Understanding the Contribution of Recommendation Algorithms on Misinformation Recommendation and Misinformation Dissemination on Social Networks

ROYAL PATHAK and FRANCESCA SPEZZANO, Boise State University, USA
MARIA SOLEDAD PERA, Technische Universiteit Delft, The Netherlands

Social networks are a platform for individuals and organizations to connect with each other and inform, advertise, spread ideas, and ultimately influence opinions. These platforms have been known to propel misinformation. We argue that this could be compounded by the recommender algorithms that these platforms use to suggest items potentially of interest to their users, given the known biases and filter bubbles issues affecting recommender systems. While much has been studied about misinformation on social networks, the potential exacerbation that could result from recommender algorithms in this environment is in its infancy. In this manuscript, we present the result of an in-depth analysis conducted on two datasets (Politifact FakeNewsNet dataset and HealthStory FakeHealth dataset) in order to deepen our understanding of the interconnection between recommender algorithms and misinformation spread on Twitter. In particular, we explore the degree to which well-known recommendation algorithms are prone to be impacted by misinformation. Via simulation, we also study misinformation diffusion on social networks, as triggered by suggestions produced by these recommendation algorithms. Outcomes from this work evidence that misinformation does not equally affect all recommendation algorithms. Popularity-based and network-based recommender algorithms contribute the most to misinformation diffusion. Users who are known to be superspreaders are known to directly impact algorithmic performance and misinformation spread in specific scenarios. Findings emerging from our exploration result in a number of implications for researchers and practitioners to consider when designing and deploying recommender algorithms in social networks.

CCS Concepts: • **Information systems** → **Recommender systems**; • **Networks** → **Social media networks;**

Additional Key Words and Phrases: Misinformation, recommendation algorithms, social networks, diffusion, news, Twitter

ACM Reference format:
Royal Pathak, Francesca Spezzano, and Maria Soledad Pera. 2023. Understanding the Contribution of Recommendation Algorithms on Misinformation Recommendation and Misinformation Dissemination on Social Networks. *ACM Trans. Web* 17, 4, Article 35 (October 2023), 26 pages.
https://doi.org/10.1145/3616088

This research has been supported by the National Science Foundation under Award no. 1943370.

Authors' addresses: R. Pathak and F. Spezzano, Boise State University, Boise, Idaho; Boise State University, Boise, Idaho; emails: royalpathak@u.boisestate.edu, francescaspezzano@boisestate.edu; M. S. Pera, Technische Universiteit Delft, Delft, The Netherlands; email: M.S.Pera@tudelft.nl.

1559-1131/2023/10-ART35 $15.00
https://doi.org/10.1145/3616088

1 INTRODUCTION

Misinformation[1] spread in social networks has drastically increased, from the 2016 U.S. Presidential Election to the current Covid-19 pandemic [4, 9], with consequences that have impacted health, politics, economy, and response to natural disasters [22, 35, 52, 96]. Misinformation spread has compromised people's ability to access correct information and have informed opinions. It has also led to people reprimanding individuals and corporations for broadcasting, amplifying, and disseminating untrustworthy, inaccurate, and misleading information. Nevertheless, the reach of misinformation remains prevalent, and the impact of misinformation spread, particularly on social networks, is non-trivial.

Misinformation spread in social networks is influenced by various factors: (i) content consumers (or users) and their behaviors (e.g., engagement with misinformation through tweets and retweets), (ii) content creators who present and construct the information, (iii) the users' adaptive behavior, personality, values, emotions, and susceptibility, (iv) bots and malicious accounts, and (v) recommendation algorithms that the social networks use to present information to their users [21, 76]. Recent research endeavors have addressed bot detection and features that characterize users, content, and context as means to identify misinformation in social networks and countermeasure the first four factors that cause the misinformation to spread in social networks [10, 97]. However, there is a lack of work centered on investigating and understanding the impact that **recommendation algorithms** (**RAs**) can have on misinformation spread in social networks. With social networking sites being prone to influence users' perspectives as a consequence of echo chambers and filter bubbles[2] [27, 86], exacerbated by RAs being "powerful tools that are inserted in most social platforms, they could also involuntarily spread unwanted content and other types of online harm" [84], deepening research in this area becomes a must.

In their vast majority, RAs leverage users' historical data–from the content and topics of the stories they have read to other factors related to engagement like clicking on an item or sharing stories [67, 94]. In their quest for personalization, and given the dynamic nature of user modeling (as users can alter their preferences over time), RAs must maintain up-to-date recommendation models that cater to users' latest interests [89]. Excessive personalization, however, can lead to filter bubbles with an exaggeration of misinformation diffusion as one of the unintended consequences [6, 13, 15, 19, 81]. This outcome can be difficult to correct as RAs are not designed to keep users informed; rather, their primary intent is to keep users engaged and satisfied [98]. Such selective exposure could foster cognitive bias [61], popularity and demographic biases [18], confirmation bias [59] and possibly create a fertile ground for misinformation recommendation and propagation [21, 78]. Despite researchers recognizing that RAs can have some responsibilities for misinformation diffusion in social networks [84], there is not any available framework–considering both user and RA behavior–to study and quantify the impact of RAs in spreading misinformation.

In this manuscript, we investigate how RAs could contribute to misinformation recommendation and propagation in social networks and propose a *new* framework to measure the impacts of RAs on misinformation diffusion. To control scope, we focus on Twitter, fake news, and

[1] Misinformation is incorrect or misleading information that can be shared accidentally, causing different levels of harm. Misinformation comes in many forms, such as satire, propaganda, hoax, rumor, conspiracy, and fake news. This latter in particular, explicitly implies false or misleading information spread deliberately to deceive and is one form of misinformation [55]. Although they can have slightly different connotations in specific contexts, misinformation and fake news are often used interchangeably in popular discourse. In this manuscript, we use the terms synonymously to refer to the dissemination of false or misleading news.

[2] As discussed in [24, 86], echo chambers refer to users' consumption of resources aligned with their views, whereas filter bubbles are deemed "personalization traps" as users are presented only with similar content.

commonly-used RAs and information diffusion models. With our work, we address the following research questions:

RQ1: How do different types of RAs contribute towards misinformation recommendation? Is this influenced by how users engage with misinformation?

RQ2: Which RAs contribute the most to misinformation propagation in social networks? Does user behavior pertaining to engagement with misinformation impact the propagation of misinformation?

We build upon the seminal work of Fernández et al. [21] and *expand* their analysis of collaborative-based RAs to also consider additional RAs, content- and social network-based, and analyze trade-offs between RAs' performance and misinformation recommendations. We use two datasets, which differ from the one considered in [21]: the Politifact FakeNewsNet dataset [74] and HealthStory FakeHealth dataset [11]. Both datasets contain explicit information on user-news item shares (for both real and fake news) and user follower-following relationships. More importantly, we study what happens after recommendations are sent out to the platform users and *simulate* how misinformation diffuses, i.e., how people that are recommended some fake news items are able to propagate it to their friends, friends of friends, and so on. This is the novelty of our work.

To the best of our knowledge, we present the first work examining the role of RAs in the misinformation recommendation and propagation process by also coupling it with users' behavior in the presence of misinformation. Our research bridges social network analysis and recommender systems communities, whereas the findings derived from our study contribute to the definition of guidelines for researchers and practitioners to consider when designing RAs that need to work in the presence of misinformation and users who spread a huge amount of misinformation.

In the rest of this manuscript, we first discuss related literature informing our work (Section 2). We then describe the experimental setup, including datasets, metrics, and RAs, which we use in the empirical explorations conducted to answer our two RQs (Section 3). This is followed by an in-depth analysis of produced results (Section 4). Lastly, we offer concluding remarks, limitations, and open directions for future work (Section 5).

2 BACKGROUND AND RELATED WORK

This section presents the background and related literature that informs our work. We first discuss the connection between recommender systems and misinformation (Section 2.1). After that, we bring attention to fundamental work in the context of social networks pertaining to misinformation diffusion (Section 2.2).

2.1 Recommender Algorithms and Misinformation

Literature about recommender systems is prolific [66]. It addresses not only novel algorithms that enable the recommendation of products and services to a broad range of users in various domains but also considers the consequences of interactions between the users and the recommender systems themselves. In the case of the latter, salient discussions revolve around topics like shilling attacks, filter bubbles, and biases (e.g., popularity bias amplification) [16, 23, 42, 49, 50, 56, 72, 80]. Shrestha et al. [72] examine the robustness of RAs to shilling attacks; that is the impact of malicious users who insert fake reviews or ratings to manipulate the recommendations process. Edizel et al. [16] bring attention to the fact that data used for training RAs can be biased and, in turn, it is likely that the RA will be presented with biased suggestions, creating a "self-perpetuating loop which progressively strengthens the filter bubbles we live in". In the context of recommender and filtering algorithms, filter bubbles are known to lead users to be "over-exposed to ideas that conform with their preexisting perceptions and beliefs, prompting intellectual isolation" [42, 61].

At the same time, filter bubbles are not always caused by RAs. As reported by Möller et al. [56], user interactions with recommender items can, in turn, cause RAs to narrow down the item set considered for recommendation purposes. These are just some examples that spotlight the power of and influence of RAs in a real-world scenario.

RAs do not exist in isolation; instead, they are trained or model data that might be non-representative or already model biases existing "in the wild." External factors undoubtedly impact the performance of RAs (malicious users, but also conflicting requirements/goals for the recommendations [3, 28, 41, 72]). Moreover, with RAs relying on ongoing user-system interactions (likes, shares, follows) to modify presented suggestions, their reach is far beyond individual users (i.e., they have a "transformative impact on society" [53]). Researchers and industry practitioners have increasingly allocated efforts to study the impact of the issues mentioned above on the RAs themselves and their user bases as well as mitigate these issues. Nevertheless, the study of a particular topic remains in its infancy: misinformation. Specifically, there is a gap in the literature regarding understanding the implications of misinformation–amplification of the spread of false and misleading information–inherent to RAs, which is the focus of this work.

Given that recommender systems' users are exposed to resources selected by RAs [61], their consumption of items deemed to be misinformation could "strengthen the further presence of such content in recommendations" [82]. Furthermore, as previously stated, the performance of RAs is often impacted by popularity biases and filter bubbles, which can make "users more vulnerable to misinformation" [21]. For these reasons, it becomes imperative to deepen our understanding of the impact RAs have on propagating misinformation. Elahi et al. [19] emphasize that, while unintentionally, RAs can contribute to both misinformation and disinformation; in turn, this menaces the communities that each RA serves. To date, misinformation has attracted the attention of the recommender system community. This is evidenced by the emergence of workshops like **Online Misinformation-and Harm-Aware Recommender Systems** (**OHARS**) [83, 84], which shine a light and build a community around this important topic; along with the availability of datasets [51, 60, 74] that enable further studying the impact that misinformation has on recommender systems.

Among research specifically focused on misinformation and recommender systems, we find, for instance, the works by Lo et al. [43, 44], which describe intervention models meant to augment recommender system functionality to mitigate the impact of misinformation. Hussein et al. [31] introduce audit experiments that can be conducted to investigate different dimensions (e.g., age, gender) that can contribute towards amplifying misinformation on YouTube. Other authors instead propose tackling misinformation in social media platforms by directly focusing on the culprits: misinformation spreaders [87], or misinformation items themselves [34, 90].

Bellogín and Deldjoo [5] bring to the attention of researchers and industry practitioners the fact that simulation enables explorations of diverse conditions (e.g., profile size, misinformation seeds)–beyond those captured at data collection time–to advance understanding of the influence of misinformation propagation have on RA. This is echoed by Tommasel and Menczer [85] who depend upon simulations to explore different scenarios that enable understanding of recommender systems in social networks and their effect on misinformation spreading behavior. Perhaps the work most closely related to ours is the empirical exploration conducted by Fernández et al. [21]. The authors survey existing work related to the different dimensions of the misinformation ecosystem (e.g., users, content, platforms, and algorithms) and describe the (dis)advantages of existing datasets that can be used to enable research in this area. They also introduce a novel dataset consisting of 2,921 Twitter users, approximately a million tweets, more than a million interactions, and misinformation labels, i.e., false claims related to COVID-19 included in the tweets. Using this novel dataset, they perform an in-depth analysis of how well-known RAs contribute to the amplification of misinformation. Reported findings reveal that the popularity-based algorithm is the most prone to spread misinformation. Other salient insights emerging from simulations of recommendation generation under different conditions (i.e., varied ratios of misinformation items) include the fact that a small number of misinformation items get popular very quickly, methods based on neighbors tend to spread less misinformation, and, in general, the number of factors or neighbors do not seem to impact misinformation spread significantly.

With our work, we built upon the work of Fernández et al. [21] by considering two other datasets, expanding the RAs examined, and exploring the propagation of misinformation as a result of generated recommendations via simulation. We do not, however, directly compare our findings with those reported in [21], as they probe RAs by simulating the recommendation process with varied ratios of misinformation items. Instead, we consider misinformation as originally captured in the datasets.

2.2 Social Networks and Misinformation

In studying misinformation diffusion in social networks, several existing studies have focused on modeling rumor propagation [100], while few have modeled the diffusion of fake news. The proposed models for misinformation diffusion address the problem as the spread of infectious disease among a group of people with social connection (epidemiological models) [36, 73, 79] or as a Hawkes process [58]. The main drawback of these models is that they assume the social network to be implicit, i.e., the connections among individuals are unknown. As a consequence, these models are more suitable for studying global patterns, such as trends and ratios of people sharing a given fake news story, but they cannot work with local node-to-node diffusion patterns. Conversely, classical information diffusion models such as the Independent Cascade and the Linear Threshold models can also be used to model fake news spread [39]. As these models work by explicitly considering how users are connected in the social network (explicit network), they can be combined with RAs that provide individual recommendations. Specifically, we can establish which users get recommended with a given fake news item and, starting from them, we can study how the fake news item gets propagated to their friends, friends of friends, and so on, and know, at the end of the diffusion process, which nodes have been infected. Suppose we instead use a model using an implicit network. In that case, we can only simulate, given a percentage of users that get recommended a given fake news item, the percentage of nodes in the network that will also be influenced by the fake news item.

Beyond studying information diffusion, other works in the social network and social science domains have addressed the problem of studying fake news spreading and characteristics of people keen to spread misinformation and used findings to tackle the problem of detecting fake news spreaders [69]. Vosoughi et al. [91] found out that the profile of fake news spreaders deviates from one of the other users as the former have, on average, significantly fewer followers, followed significantly fewer people, and were significantly less active on Twitter. They also showed that, with respect to the social media platform Twitter, although bots also contribute to spreading fake news, the dissemination of fake news on Twitter is mainly caused by human activity. Similar findings have also been shown by Shu et al. [75], who also reported that, on average, users who share fake news tend to be registered for a shorter time than the ones who share real news. Furthermore, real news spreaders are more likely to be more popular than fake news spreaders, and older people and females are more likely to spread fake news.

Guess et al. [29] analyzed user demographics as predictors of fake news sharing behavior on Facebook and found political orientation, age, and social media usage to be the most relevant. The researchers found that users who leaned to the political right were more likely to share those fake news items, perhaps because the majority of fake news items included for analysis were from 2016 and pro-Trump. Additionally, they observed that seniors tended to share more fake news, probably because this age group has lower digital media literacy skills necessary to assess the veracity of online news. Finally, the researchers found that the more news people post on social media, the less likely they are to share fake news, which can be the case because those users would be more familiar with the platform and what they share.

Recently, the author profiling shared task at the PAN 2020 conference focused on determining whether or not the author of a Twitter feed was keen to spread fake news [63]. Here, participants addressed the problem by considering different linguistic features, including (a) n-grams, (b) writing style, (c) personality and emotions expressed in users' timeline tweets, and (d) word embeddings. Using the Politifact FakeNewsNet dataset considered in this manuscript, Shrestha and Spezzano [71] showed that user personality traits, emotions, and writing style are strong predictors of fake news spreaders, and that, in combination with demographics, behavioral, and network features, can be used to classify fake news spreaders. This approach also outperforms the best models proposed at PAN 2020 fake news spreaders profiling shared task. Similarly, recent work by Giachanou et al. [25, 26] has shown how user psycholinguistic characteristics are useful in differentiating between fake news spreaders and fact-checkers and the importance of considering emotional signals for news claim credibility assessment.

We can see common insights from the literature: newly registered accounts are more likely to be involved in misinformation spread; user-based features such as demographics (age, gender, and political ideology) are essential predictors of fake news detection. Similarly, user emotional signals (positive and negative emotions) such as happiness, joy, anger, fear, sadness, and disgust are strong features to distinguish fake news spreaders.

3 EXPERIMENTAL FRAMEWORK

In this section, we discuss the datasets, the RAs, the information diffusion models, and the metrics we consider to carry out our study. In addition, we describe the experimental protocol we defined to measure the impact of RAs on misinformation diffusion in social networks and answer the research questions presented in Section 1.

3.1 Datasets

In our exploration, we use two datasets: Politifact FakeNewsNet dataset and HealthStory FakeHealth dataset. The FakeNewsNet dataset [74] consists of two datasets, PolitiFact and GossipCop, from two different domains, i.e., politics and entertainment gossip, respectively. Each of these datasets contains details about news content, publisher, social engagement information (news sharing extracted from Twitter), and the users' social network (who follows whom on Twitter). GossipCop focuses on gossip, which is related to a different form of misinformation. As the scope of our work is on fake news, we only use the PolitiFact dataset, which consists of news items with known ground truth labels collected from the fact-checking website PolitiFact[3] where journalists and domain experts fact-checked the news items as fake or real. Overall, the Politifact FakeNewsNet dataset contains 295,469 users (after removing self-claimed bot accounts) sharing 701 news items via tweets and retweets. However, to have enough user information to train RAs, we considered only users who shared at least eight news for our analysis. It results in 1,028 unique Twitter users, 3,021 following relationships (network edges) among users, 542 unique news (322 fake and 220 real), and 20,265 user-news interactions (11,442 with fake news items and 8,823 with real news items), i.e., news items shared by the users in their tweets.

The FakeHealth [11] consists of two datasets, HealthStory and HealthRelease. HealthStory contains news stories reported by news media like Reuters Health. HealthRelease corresponds to news

[3] https://www.politifact.com/

releases from various institutes, including universities, research centers, and companies. Each of these datasets contains details about news contents (with newstext, source publishers, and image links), news reviews, social engagements, user networks, and ground truth (a rating score ranging from 0 to 5 where news pieces whose scores lower than 3 are considered misinformation or fake news). HealthRelease does not contain enough user-item interactions after excluding users who shared less than eight news (only 659 users). Consequently, we only used HealthStory. This results in the FakeHealth HealthStory dataset, which includes 5,406 unique Twitter users, 1,690 unique news (472 fake and 1,218 real), 4,102 following relationships among users, and 120,124 user-news interactions (29,726 with fake news items and 90,398 with real news items).

To the best of our knowledge, these two are the only publicly available datasets containing information about user-item interaction (user-news sharing) and user following. The latter is needed to simulate misinformation dissemination on the social network. Note that we considered including in our analysis other datasets, such as ReCOVery[99]. However, we ultimately excluded them from our experiments due to their lack of enough user-item interactions (only 568 users shared at least eight news items).

3.2 Recommendation Algorithms

Inspired by Fernández et al. [21], in our exploration we consider RAs in different categories. In addition to baselines and common collaborative filtering algorithms (implemented using LensKit [17]), we also probe content-based and network-based RAs (implemented as detailed below). We operated each RA in implicit feedback mode. Overall, our study considers a sample of classical RAs that are common nowadays in the literature focused on misinformation [1, 20, 85]. Also, it is important to note that classical algorithms, particularly those based on collaborative filtering [21], are well-studied in the literature and they are common one-commerce platforms, given that these methods can be "applied to any domain, only requiring user-item interactions, not needing additional item features or metadata" [20]. Content-based algorithms are also of interest since they have the ability to "consider the content of the items and adapt in more detail to the domain at hand." Lastly, network-based strategies were also chosen given the context of our work (social network platforms like Twitter). Recall that our focus is not to explore state-of-the-art RAs' performance but instead scrutinize tradeoffs of performance with respect to misinformation recommendation, as well as misinformation spread (the latter being one of the main contributions of this work).

Collaborative-filtering RAs:

UU User-based collaborative filtering algorithm [65] exploits similar-minded users to produce recommendations (neighborhood size = 10; affinity based on cosine similarity).

II Item-based collaborative filtering [14, 70] utilizes an item-item matrix to determine the similarity between the target item and other items (neighborhood size = 10; affinity based on cosine similarity).

ALS This matrix factorization-based algorithm has been designed to improve RAs' performance in large-scale collaborative filtering problems [30](latent factors = 40; damping factors = 5; 150 iterations).

Content-based RA:

CB This content-based RA models user profiles based on the content of items known to be of interest to the corresponding user [45]. In particular, we use the TF-IDF vector representations of user profiles and news items (with news content tokenized, lowercased, stopwords removed, and stemmed) and cosine similarity as a similarity measure to identify news suggestions.

Non-personalized RA baselines:

- **Rnd** The Random Item RA disregards the interactions between users and items and instead randomly selects items to recommend [8].
- **Pop** This non-personalized RA suggests the most frequently-consumed items [33]. In our case, we use Lenskit's TopN algorithm, which recommends the most-shared news items (popularity="quantile").

Notably, these two non-personalized RAs continue to be deemed suitable baselines on recent studies analyzing RAs [2].

Network-based RA:

- **SMF** This trust-based matrix factorization RA leverages a network of trust relationships among users [32]. SMF learns the latent feature vectors of users and items so that each feature vector is dependent on the feature vectors of direct neighbors in the social network. SMF handles the transitivity of trust and trust propagation, which is not captured by the other trust-based RA, such as the STE Model [47]. In our case, we follow the implementation described in the article where the algorithm was originally introduced [32], with K = 10, $\lambda_U = \lambda_V = 0.1$, and $\lambda_T = 5$.

3.3 Information Diffusion Models

An information diffusion model describes the process by which a piece of information (a fake news item in our case) is spread and reaches users through interactions. In a social network, users typically re-share content shared by other users, usually their friends (or users they follow in the case of Twitter). Initially, a set of users initiate the diffusion process by sharing a piece of information in the network for the first time. These users are called the *seed users*. Next, the followers of the seed users have the possibility to re-share the same piece of information, followed by the followers of seed users' followers, and so on until no one else is further sharing the piece of information and the diffusion process stops. We consider three widely-used models of information diffusion (implemented as in [68]). We chose the **Independent Cascade model** (**ICM**), the **Linear Threshold model** (**LTM**), and the **Node Profile Threshold model** (**NPTM**) to model the diffusion of fake news in social networks because they explicitly use social network information and can be easily combined with individual recommendations as explained in Section 2.2.

The **ICM** is a stochastic information diffusion model [39]. In this case, nodes can have two states: *active*, meaning that the node is already influenced by the information in diffusion, and *inactive* when the node is unaware of the information or not influenced by the information in diffusion. At each step, a newly active node u has the chance to influence an inactive neighbor v according to an influence probability p_{uv}. Each probability p_{uv} is independent of the others. Usually, these probabilities are set as $p_{uv} = 1/|N_{in}(v)|$), where $N_{in}(v) = \{w|(w, v) \in E\}$ is the set of v's incoming neighbors [40].

In the **LTM** each edge (u, v) is associated with a weight b_{uv} and each node u has a threshold value τ_u [39]. Threshold values in a [0,1] interval are often assigned uniformly at random. At each step i, a node v will become active if $\sum_{u \in N_{in}(v), u \in A_{i-1}} b_{uv} \geq \tau_v$, where E is the set of edges in the network, $N_{in}(v)$ is the set of v's incoming neighbors, and A_{i-1} is the set of nodes that are active in the previous step. Edge weights b_{uv} are typically set to be the inverse of the in-degree of node v [40].

The **NPTM** supports the mixed behaviors of the classical LTM and the Node Profile model [54]. Each node v has a profile γ_v, which describes the likelihood of spreading content similar to the ones they had already spread in the past. At each step i, for each node v, there is an evaluation if $\sum_{u \in N_{in}(v), u \in A_{i-1}} b_{uv} \geq \tau_v$, as in the LTM. If the above evaluation is satisfied, the model evaluates the node profile, i.e., a random value q in [0,1] is extracted, and if $q \geq \gamma_v$, the node adopts the content; otherwise, the node refuses to adopt. In our case, the node profile (γ_v) for each node v is computed as the average cosine similarity between the news article shared by the node v in the test set and a news article shared by v in the training set. Also, at every iteration, δ percentage (adopter rate) of nodes spontaneously become infected due to endogenous effects. We used δ = 0.001 in our implementation.

Neighbors' influence, user preference, or endogenous effects are all realistic assumptions to consider in modeling information diffusion. In our experiments, we used the implementation provided by Rossetti et al. [68] for both LTM and NPTM where the condition $\sum_{u \in N_{in}(v), u \in A_{i-1}} b_{uv} \geq \tau_v$ is implemented as checking whether the percentage of v's neighbors that are active is greater than τ_v.

3.4 Metrics

In our empirical exploration, we turn to metrics that capture RA performance as well as the impact of misinformation.

Given that we examine Top-N RAs, much like in the recent analysis of Top-N RAs undertaken by Anelli et al. [2], we use **Mean Reciprocal Rank (MRR)**, a common assessment metric, to quantify RA performance. MRR reflects the average ranking of the first relevant recommendation in the Top-10 list produced by each RA under study.

To quantify the amount of misinformation recommended by a RA, we use **Misinformation Count (MC)** and **Misinformation Ratio Difference (MRD)**. MC measures the count of misinformation items recommended to each user [21]. In other words, MC is the proportion of recommended items known to be misinformation over the length of the recommendation list. In our case, we examine the top-10 recommendations produced by each RA under study. MC values range between 0 and 1; the closer to 1, the more misinformation items are included among the generated recommendations. MRD (in Equation (1)) measures the average difference between the misinformation ratios in the train and recommendation list across all the users [21].

$$MRD@N = \frac{1}{T}\sum_{u=1}^{T} M_t^u - M_r^u, \tag{1}$$

where M_t^u is the ratio of misinformation items for the user u with respect to what is observed in training, M_r^u is the ratio of misinformation items present within the top-N recommendations for the user u (N = 10 in our case), and T is the total number of users. MRD values range between −1 and 1. A negative MRD value indicates the ratio of misinformation is larger in the recommendation list; a positive MRD value indicates the ratio of misinformation is larger in the training set.

To quantify the effect of misinformation on the whole social network, Twitter in our case, we use the **Expected Spread (E_{Spread})**. Given a social network, a diffusion model, and a piece of information n, the *spread* is defined as the percentage of infected nodes (users who shared n) at the end of the diffusion process [39]. Because the diffusion models we consider have a random component (the influence probabilities for ICM and the node threshold values for LTM and NPTM), these models are usually run many times, and an expected spread is computed, i.e., the average spread among all the runs. In our experiments, we ran each model 100 times. The seed nodes are counted when the (expected) spread is computed.

To characterize vertices that are influential in diffusing fake news in the social network, we rely on the concept of **diffusion centrality** [38]. The diffusion centrality measures how well a node (or a set of nodes) can diffuse a property p, in our case a given fake news item, given the structure of the social network and a diffusion model for the property p. The diffusion centrality of a node w (or a set of nodes W) is computed as the expected spread achieved when the node w (or a set of

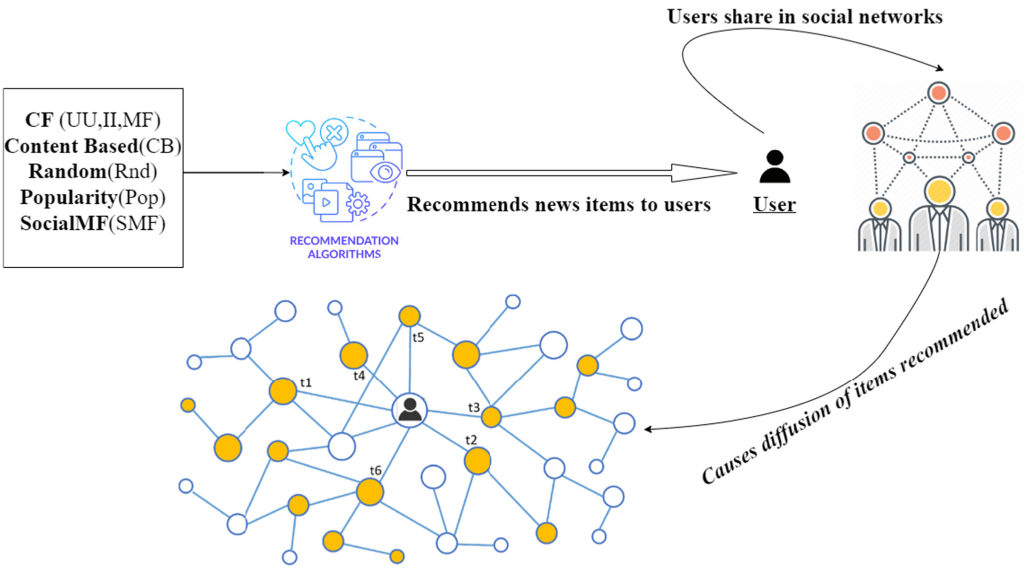
CF (UU,II,MF)
Content Based(CB)
Random(Rnd)
Popularity(Pop)
SocialMF(SMF)
RECOMMENDATION ALGORITHMS
Recommends news items to users
User
Users share in social networks
Causes diffusion of items recommended
t1
t2
t3
t4
t5
t6

Fig. 1. Scenario for misinformation during recommendation and propagation.

nodes W) is included in the initial set of seeds minus the expected spread achieved when the node w (or a set of nodes W) is *not* included in the initial set of seeds.

3.5 Experimental Protocol

We conduct two experiments to study how RAs contribute to misinformation recommendation and dissemination in social networks. These experiments are designed to answer the research questions presented in the Introduction.

Experiment 1. While the performance of RAs is not the primary focus of our study, in the first experiment, we study the potential connection between RAs and their propensity to enable misinformation as a result of the generated recommendations, as shown in Figure 1. To do so, we used a temporal leave-one-out strategy to split the dataset into training and test set[4] and compute MRR, MC, and MRD based on the top-10 recommendations generated for each user in the PolitiFact FakeNewsNet dataset and the HealthStory FakeHealth dataset, respectively.

Experiment 2. We study how each considered RAs contributes to the diffusion of fake news on Twitter based on the recommendation lists produced for each RA in Experiment 1. The dissemination usually happens after a RA recommends a news item to its users, and each user shares (likes, tweets, retweets, comments) the news items in the social network, causing the propagation of news items suggested by the RA (see Figure 1). This is measured by the expected spread generated after some users are initially recommended a given news item by the considered RA. More specifically, given a RA R and a fake news item n, we let the R trigger the diffusion of n by computing the seed users as those whose top-10 recommendations include n. Then, we let the diffusion model simulate the diffusion of the fake news item among the remaining users (i.e., the ones who are not originally suggested n, but can share it because they are influenced by the people they follow). We consider the fake news items present in the test set and treat them as a single piece of information

[4] We select the user-item interaction with the latest timestamp for each of the 1,028 unique users in the PolitiFact FakeNewsNet dataset and each of the 5,406 individual users in the HealthStory FakeHealth dataset as the test set for each dataset, i.e., ground truth; the remaining user-item interactions comprise the training set.

that diffuses in the Twitter social network. We then *simulate* its diffusion in the network by using the independent cascade, the linear threshold, and the node profile threshold models. We want to emphasize that we aim at replicating a real deployed system and understanding the propagation of misinformation as a result of the latest recommendation presented to a user in a low-latency system such as Twitter. Hence, we used the fake news items present in the test set, which are the latest timestamped news items. It is worth noting that this simulation can be carried out for each RA independently of whether they use or not social network information to compute the recommendation list. In fact, each RA is used to determine the seed users to pass to the diffusion model.

User types. We are interested in understanding whether users' aptitude towards misinformation impact RA performance and users' likelihood of spreading misinformation. For this reason, we extend Experiments 1 and 2 by juxtaposing results generated by different user groups: superspreader and non-superspreader users. Given a threshold value θ, we define a user as a *superspreader* if the percentage of shared fake news items in the training set is at least θ, and performed our experiments for different values of θ in the set $\{50\%, 60\%, 70\%, 80\%, 90\%, 100\%\}$. We report in Section 4 the experiment results corresponding to the representative threshold of $\theta = 60\%$ and discuss whether trends generalize across all the considered thresholds. We include detailed results for all the threshold values, along with the number of superspreaders corresponding to the respective value for the threshold θ (Table 1) [5] in the Online Appendix [62].

When a superspreader is a user whose percentage of shared fake news items in the training set is at least 60%, we have 636 distinct users who are superspreaders and 392 distinct users who are non-superspreaders in the Politifact FakeNewsNet dataset. Similarly, we have 92 distinct users who are superspreaders and 5,314 distinct users who are non-superspreaders in HealthStory FakeHealth dataset. This results in two different scenarios, one where there is a balance between superspreader and non-superspreader (Politifact FakeNewsNet dataset), and another where the superspreaders are highly unbalanced as compared to non-superspreaders (HealthStory FakeHealth dataset).

As discussed in Section 2, previous research has characterized users who are keen to spread fake news as likely to have fewer followers, to be more leaned to the political right and to be somewhat older. We do not have information regarding the political leaning or the age of the users in our considered datasets, but we were able to examine their number of followers. As reported in Tables 2 and 3 in the Online Appendix [62], the number of followers is generally higher for the users we defined as superspreaders than the other group of users, with differences being statistically significant in the Politifact FakeNewsNet dataset for all the threshold values except for $\theta = 90\%$.[6] It is worth noting that this finding does not contradict previous work where fake news spreaders are often considered as users who spread at least one fake news item, so it is possible to have many inactive or inexperienced users (i.e., they have few connections or they are more senior) who spread some news items by mistake. In our work, we are trying to capture more active or experienced fake news spreaders who intentionally spread fake news.

4 RESULTS AND DISCUSSION

In this section, we present the experiments' results to address the research questions driving our study. *We use results from Experiment 1 to answer RQ1 and outcomes from Experiment 2 to answer*

[5] Note that there are no superspreaders for thresholds $\theta = 90\%$ and $\theta = 100\%$ in the HealthStory FakeHealth dataset.

[6] We do not have a representative number of superspreaders as compared to the number of non-superspreaders in the HealthStory FakeHealth dataset to check whether the differences in the number of followers are statistically significant also in this dataset.

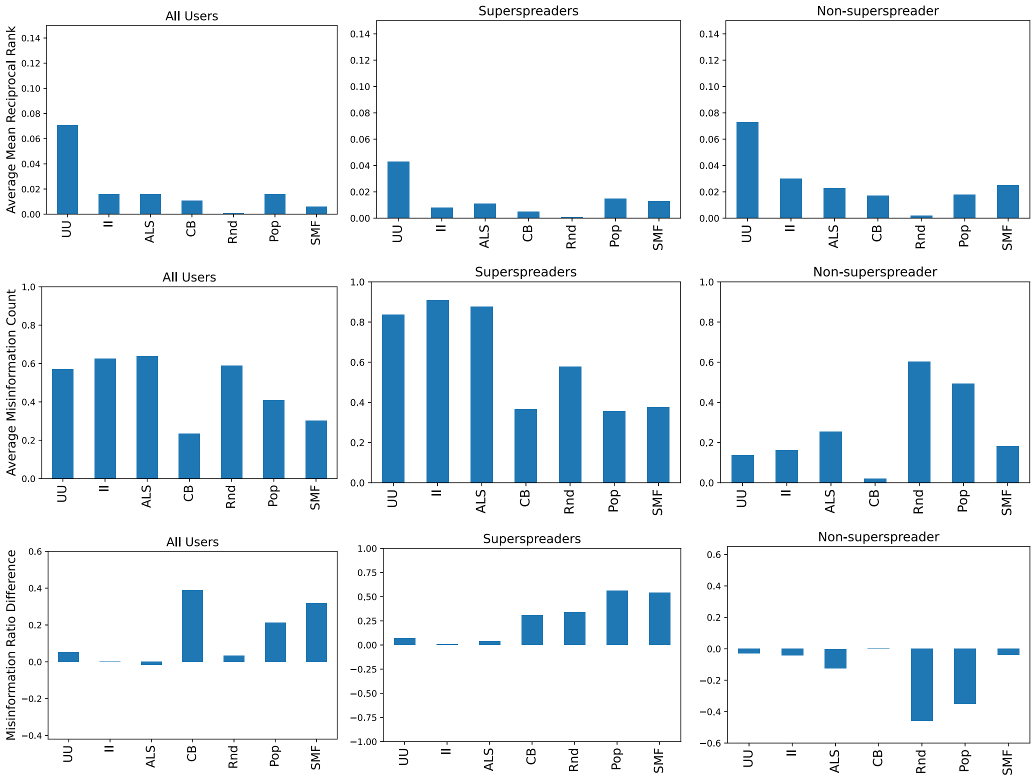

Fig. 2. Average MRR, Average MC, and MRD for each of RA, for all users, superspreaders (defined by $\theta = 60\%$), and non-superspreaders in the POLITIFACT FAKENEWSNET DATASET.

RQ2. We also discuss potential implications emerging from our findings regarding the real-world design and deployment of RAs for social networks in the presence of misinformation.

Whenever we say significant, this is based on a paired t-test when comparing RAs within the same user group (with Bonferroni correction, $n = 7$) and a non-paired t-test when comparing across user groups, and a p-value less than 0.05.

4.1 Experiment 1: RAs, Performance, and Misinformation

4.1.1 All Users Represented in the Respective Datasets. When we look at the general user group, we see from Figures 2 and 3 that regardless of the dataset examined, collaborative-filtering RAs (and Pop in the case of POLITIFACT FAKENEWSNET DATASET) are the ones that fare best based on MRR, i.e., include higher within the top-10 suggestions items that are relevant to their users. MRR is significantly lower for CB and SMF than that computed for collaborative-filtering RAs. The fact that CB underperforms–compared to collaborative-filtering counterparts–is not unexpected, as other works have also reported similar outcomes in the news domains [46]. When looking at the prominence of fake news within the top-10 recommendations, we see that in both datasets the average MC is significantly higher for UU, II, and ALS than Pop, CB, and SMF. In practice, this means that collaborative-filtering RAs tend to include more misinformation among their recommendations generally. Fernández et al. [21] reported on a similar pattern but on the COVID-19 dataset. Noticeable, MC score for Rnd is comparable to those of collaborative filtering RAs, which we attribute to the random nature of the produced recommendations. Furthermore, we posit that

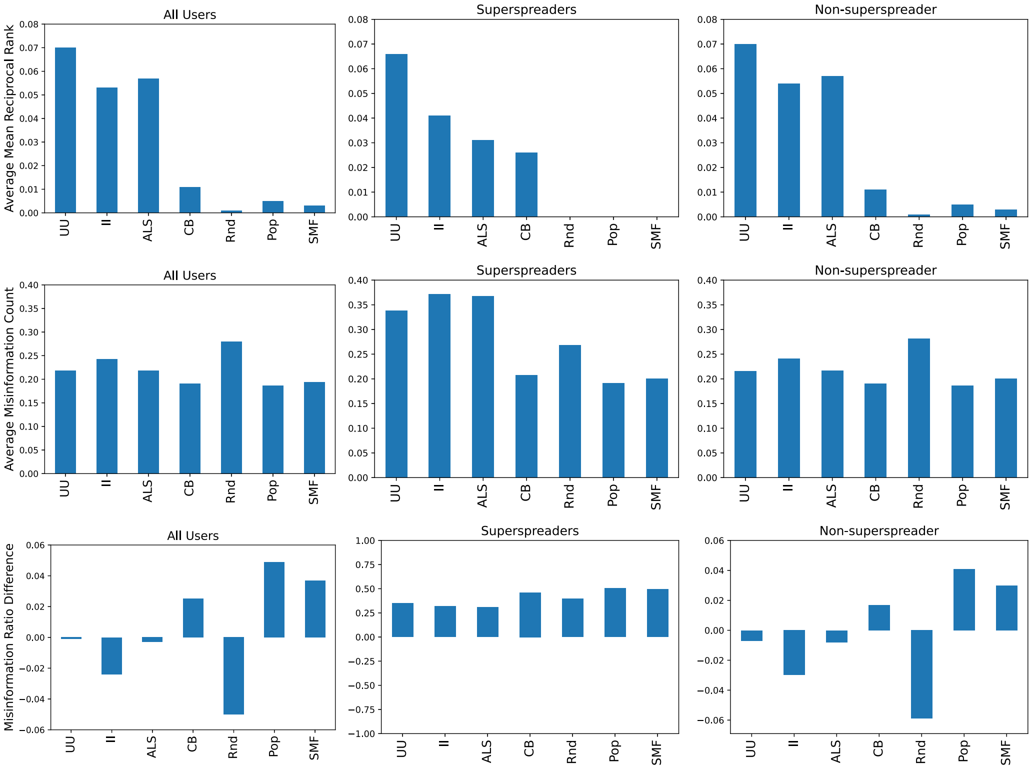

Fig. 3. Average MRR, Average MC, and MRD for each of RA, for all users, superspreaders (defined by $\theta = 60\%$), and non-superspreaders in the HealthStory FakeHealth dataset.

the significantly lower MC score produced by CB is due to how recommendations are generated, i.e., based on content matching. Based on MRD, we see that scores for UU, II, ALS, and Rnd remain close to 0 (akin to results reported in Fernández et al. [21]), indicating that the ratio of misinformation remains fairly stable across training and recommendation lists. CB yields the highest MRD score, closely followed by SMF (although visibly more prominent in Politifact FakeNewsNet dataset than HealthStory FakeHealth dataset); this indicates that the ratio of misinformation is larger in training lists than in recommendation lists. These results seem to demonstrate that CB and SMF are less prone to recommending misinformation, in the sense that items known to be fake news appear less among recommended items, even if present in users' historical interactions prompting the recommendations.

It emerges from Figures 2 and 3 that trends observed for Pop in Politifact FakeNewsNet dataset and HealthStory FakeHealth dataset are not alike. A closer look at the item distributions in these two datasets could explain these disparities. In the case of MRR, the number of unique items that serve as candidate recommendations is larger for HealthStory FakeHealth dataset (1,690 news items) than the Politifact FakeNewsNet dataset (542 news items). The reduced number in candidate items makes it so that the recommendation task for the former is inherently more difficult given the potentially broad range of options that might be relevant and yet not explicitly noted as so by users. In the case of MC and MRD, we attribute disparities to the noticeable difference in ratios of fake items in both datasets and the proportion of popular misinformation items this algorithm presents its users in the top-10 recommendation list (3.42% for Politifact FakeNewsNet dataset, decreasing to 1.5% for its counterpart dataset).

Simultaneously examining MRR, MC, and MRD lead us to conclude that CB and SMF are the best RA alternatives if the emphasis is on curtailing misinformation. Their MRR is indeed significantly lower than collaborative-filtering counterparts. MC scores are significantly lower for CB and SMF than for other RAs and their proportion of suggested items that are deemed fake news are less impacted by users' historical interactions, indicating that they are less likely to suggest misinformation (i.e., items known to be fake news in our case).

4.1.2 The Impact of Superspreaders. We are interested in exploring whether users' aptitude towards misinformation (i.e., sharing tweets spreading fake news) impacts RAs' performance and their likelihood of contributing to misinformation. For this, we compare and contrast MRR, MC, and MRD computed for superspreaders and non-superspreaders in both the considered datasets, which correspond to two different scenarios: one represented by the Politifact FakeNewsNet dataset where superspreaders and non-superspreaders are closer in number (for instance, we have 636 superspreaders and 392 non-superspreaders for threshold $\theta = 60\%$), and the one represented by the HealthStory FakeHealth dataset where the ratio of superspreaders vs. non-superspreaders in approximately 1:100.

Regarding the first case, i.e., the Politifact FakeNewsNet dataset, we see in Figure 2 (for superspreaders defined by $\theta = 60\%$) and Figures 9, and 11–14 in the Online Appendix [62] (for superspreaders defined by other threshold values) that trends observed for superspreaders deviate from those seen for non-superspreaders.[7] Among the non-superspreaders, UU, II, ALS, and SMF yield the highest MRR across all the considered thresholds. Furthermore, MRD for UU, II, and ALS remains close to 0 for all the considered thresholds. This is also true for SMF up to threshold $\theta = 80\%$, while it becomes larger and positive for thresholds $\theta = 90\%$ and $\theta = 100\%$ (which means that the amount of recommended misinformation is lower in the test than in the training set). The MC scores are low, i.e., up to around 0.3 for all but thresholds $\theta = 90\%$ and $\theta = 100\%$ where the values increase up to around 0.5 for UU, II, and ALS. Differing from the findings reported in the prior section for all users, it becomes apparent that when superspreaders are excluded, UU, II, ALS, and SMF are able to present relevant items higher in the recommendation list more often while minimizing the presentation of fake news (compared to a content-based RA) and keeping the ratio between misinformation considered for user modeling vs. predictions (training vs. testing sets) fairly unchanged or lower than the one in the training set. Given the tradeoff between user satisfaction (higher MRR) and less misinformation (lower MC and MRD close to 0), we argue that collaborative-filtering RAs and SMF generally suit non-superspreaders better. Instead, among superspreaders (and across all the considered thresholds), user satisfaction is high for UU, II, ALS, and SMF but so is the average number of fake news present on top-10 recommendations for collaborative-based RAs, which we attribute to the high number of fake news items in their user profiles. SMF and CB are, in turn, more resilient to users known to be superspreaders from the perspective of misinformation (both have lower MC scores and positive MRD), at the cost of user satisfaction (lower MRR for CB). Regarding the general trend for the behavior of Pop across all the considered thresholds, we observe that its MRR and MC are higher for non-superspreaders, while MRD is higher for superspreaders. This could be explained by the likelihood of similar news items being part of superspreaders' user models (i.e., shared items), decreasing the chances of users being exposed to novel fake news as a result of recommendations.

In the case of the HealthStory FakeHealth dataset, trends observed for superspreaders and non-superspreaders are similar. Figure 3 refers to the case when superspreaders are defined by

[7] In the Politifact FakeNewsNet dataset, differences in MRD and MC scores between superspreaders and non-superspreaders are significant. So are differences in MRR for UU, II, ALS, and CB.

$\theta = 60\%$ and illustrate that UU, II, and ALS yield the highest MRR and MC for both superspreaders and non-superspreaders. Hence, CB is the preferred RA in this scenario as its MC is among the lowest values and its MRR is higher than Pop and SMF, which have comparable MC. MRD values are positive for superspreaders and close to zero for non-superspreaders. All the above trends remain true regardless of the threshold used to define superspreaders (cf. Tables 15–18 in the Online Appendix [62]).

Overall, the differences in performance we have observed in the two considered datasets for superspreaders and non-superspreaders could be attributed to the different distribution of these two user types between the two datasets.

4.2 Experiment 2: Study of Misinformation Diffusion in the Social Network After Recommendations

4.2.1 All Users Represented in the Respective Datasets. Figures 4 and 5 show the average expected spread[8] for each of the examined RAs and diffusion models when we look at the general user group. We observe that both Pop and SMF are highly prone to disseminate misinformation in the user network. Looking at Politifact FakeNewsNet dataset, Pop achieves the highest average expected spread with all the three considered diffusion models (52.36% with LTM, 73.31% with ICM, and 69.78% with NPTM). This is not unexpected, given the prevalence of fake news in this dataset, along with the fact that user-item interactions (sharing of tweets) involving fake news comprise more than half of our dataset. As popularity-based RAs naturally leverage sharing rates in their recommendation strategy, the amplification of misinformation is anticipated. Notably, the volume of fake news on this dataset aligns with that often seen on real-world news sites and social network sites like Twitter. For example, volume News[9] reported that false news stories are more popular, and they are 70% more likely to be retweeted than true stories. When considering results computed using HealthStory FakeHealth dataset, Pop achieves the second highest average expected spread with all three considered diffusion models (28.04% with LTM, 46.03% with ICM, and 54.51% with NPTM). Recall that the volume of fake news in this dataset is lower than in the Politifact FakeNewsNet dataset.

Among non-baseline algorithms, SMF contributes to the highest average expected spread (36.02% with LTM, 64.34% with ICM, and 60.53% with NPTM) in Politifact FakeNewsNet dataset, and 99% in HealthStory FakeHealth dataset,[10] followed by collaborative-based RAs (UU, II, and ALS) and CB. This trend is observed in all of the diffusion models considered, even if more prominent with LTM than ICM and NPTM. This is evident due to the nature of trust-based RAs like SMF. In SMF, a user's latent feature vector is dependent on the direct neighbors' latent feature vectors; hence there is a high likelihood that a node and its neighbors are suggested the same news items, resulting in more seeds at the beginning and higher overall spread in the social network [32].

Looking at Politifact FakeNewsNet dataset, there is an average expected spread of 14.94% with ALS, 13.86% with UU, 11.97% with II, and 7.25% with CB when the information diffusion is simulated with the LTM. Moreover, when the information diffusion is simulated with the ICM, there is an average expected spread of 56.34% with ALS, 54.57% with UU, 52.44% with II and 48.02% with CB. If instead NPTM is used, then there is an average expected spread of 51.27% with ALS, 51.62%

[8] Since we treated each fake news item as an individual property that spreads through the social network, we report the average expected spread across all the fake news items in the test set.

[9] https://www.volumenews.com/health/health-news/fake-news-lies-spread-faster-social-media-truth-does-n854896/

[10] Of note, news articles recommended by SMF are the most popular ones in the HealthStory FakeHealth dataset. It is then unsurprising that they are suggested to almost all the users in their top-10 list of recommended items.

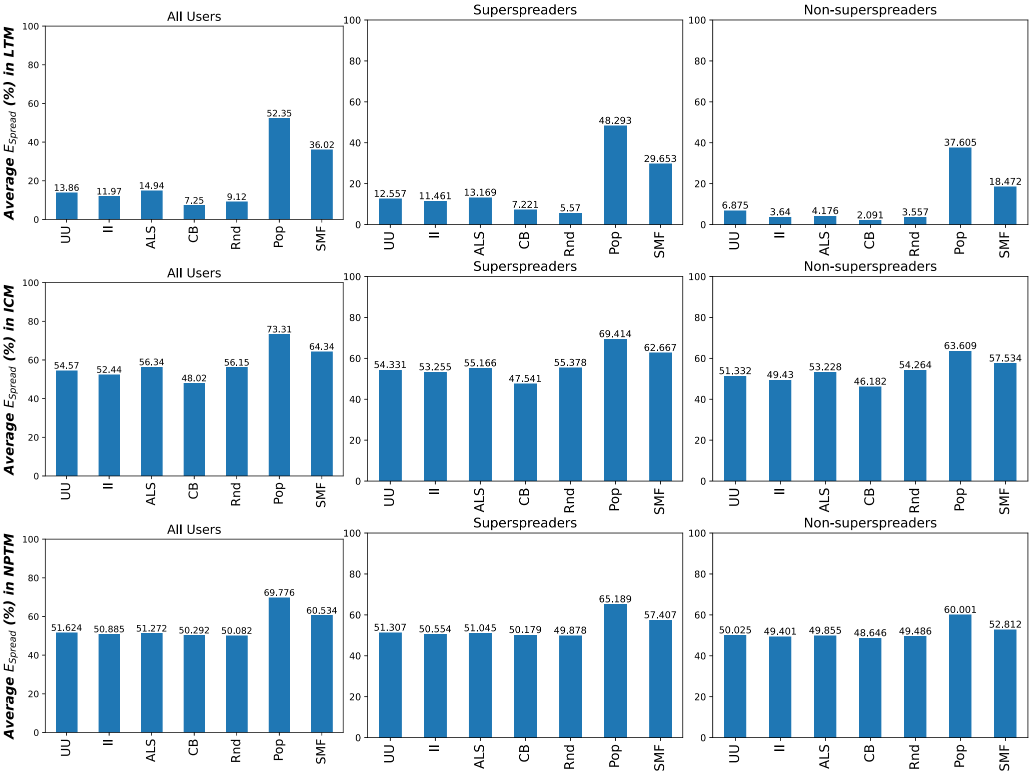

Fig. 4. Average expected spread for each of RA, for all users, superspreaders (defined by $\theta = 60\%$), and non-superspreaders in the Politifact FakeNewsNet dataset according to the Linear Threshold model (LTM—top row), Independent Cascade model (ICM—middle row), and the Node Profile Threshold Model (NPTM—bottom row).

with UU, 50.89% with II, and 50.29% with CB. In the case of HealthStory FakeHealth dataset, there is an average expected spread of 1.87% with ALS, 1.82% with UU, 1.84% with II, and 1.66% with CB when the information diffusion is simulated with the LTM, an average expected spread of 40.83% with ALS, 41.59% with UU, 38.35% with II and 36.73% with CB when the information diffusion is simulated with the ICM, and an average expected spread of 38.14% with ALS, 38.20% with UU, 38.18% with II, and 38.12% with CB when the information diffusion is simulated with the NPTM. Recall that the percentage of expected spread also depends on the number of seeds. As we can see in Figure 6, Pop presents at least one fake news item to 417 users in their recommendation lists in Politifact FakeNewsNet dataset, and 1,437 users in their recommendation lists in HealthStory FakeHealth dataset, SMF to 188 and 5,256 users, respectively, followed by collaborative-based RAs, CB, and Rnd. The higher the number of seeds, the higher the expected spread. Much like for CB, misinformation spread is also low for Rnd, regardless of the dataset considered. Nevertheless, Rnd is merely studied as a baseline, i.e., as another way to contextualize findings, not as a RA prominently used on social networks. This is because Rnd disregards interactions between users and items in producing recommendations and instead arbitrarily selects items from within the candidate item set.

When the diffusion of fake news in the network is simulated with ICM a higher spread is achieved as compared to LTM and NPTM, regardless of the dataset. This is due to the fact that

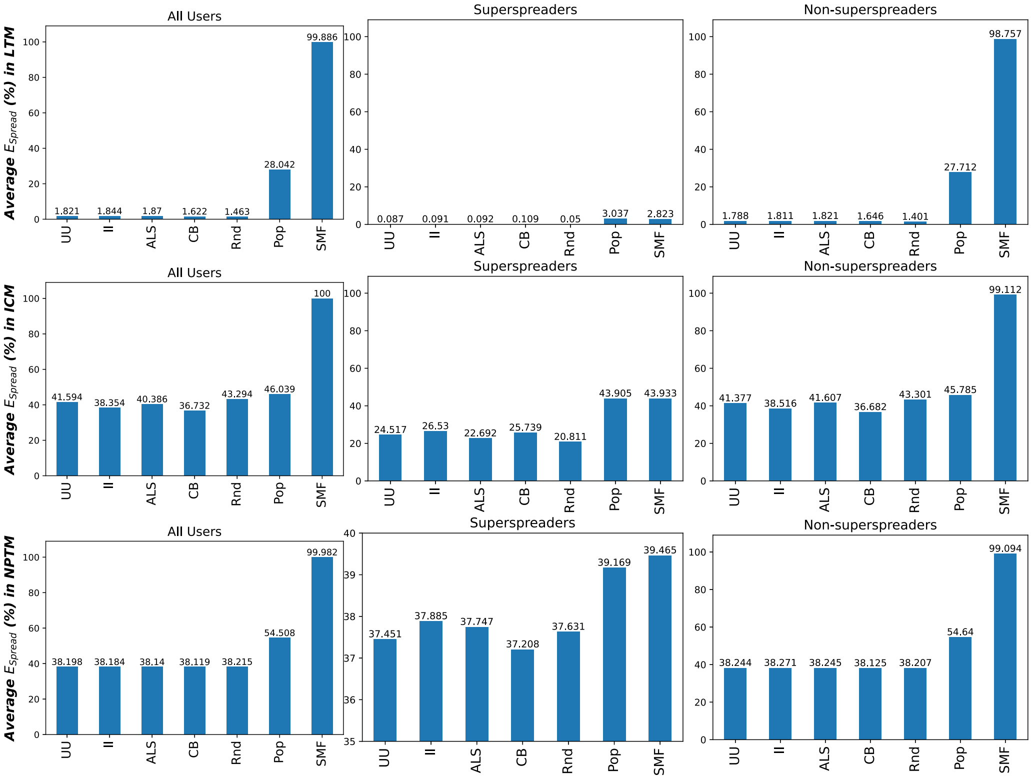

Fig. 5. Average expected spread for each of RA, for all users, superspreaders (defined by $\theta = 60\%$), and non-superspreaders in the HealthStory FakeHealth dataset according to the Linear Threshold model (LTM—top row), Independent Cascade model (ICM—middle row), and the Node Profile Threshold Model (NPTM—bottom row).

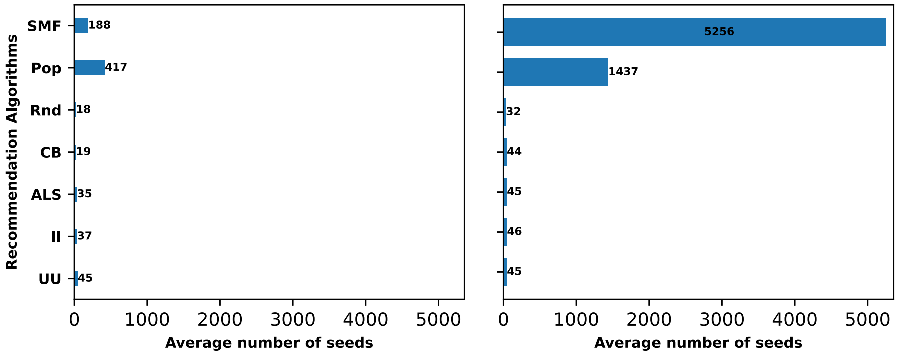

Fig. 6. Average number of seeds as determined by different RAs in Politifact FakeNewsNet dataset (left) and HealthStory FakeHealth dataset (right)

in the ICM, each node can be independently influenced by each of its incoming neighbors (hence has more chances of being infected), while in the linear and node profile threshold models, each node becomes infected if the percentage of infected incoming neighbors is above the node threshold.

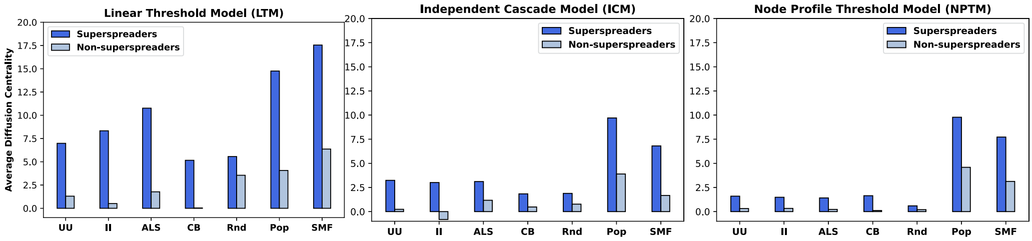

Fig. 7. Average Diffusion Centrality for different user types in the Politifact FakeNewsNet dataset. Superspreaders are defined by $\theta = 60\%$.

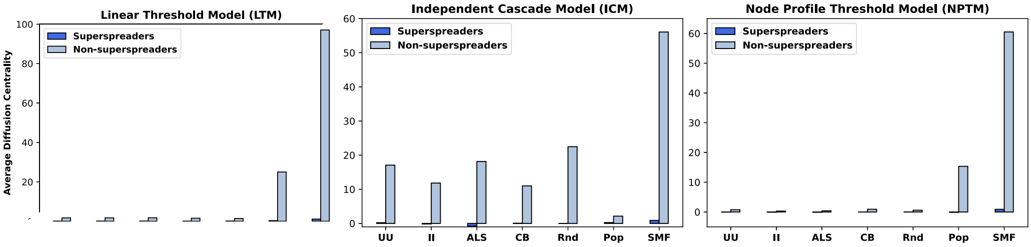

Fig. 8. Average Diffusion Centrality for different user types in the HealthStory FakeHealth dataset. Superspreaders are defined by $\theta = 60\%$.

4.2.2 The Impact of Superspreaders. We extended our analysis by studying the impact of superspreaders in the diffusion process and computed and compared the diffusion centrality of superspreader seeds vs. the diffusion centrality of non-superspreader seeds. Here, the diffusion centrality of the set of superspreader seeds is computed as the expected spread achieved when all the seeds are considered (both super- and non-superspreaders) minus the expected spread achieved when only non-superspreader seeds are considered (i.e., superspreader seeds are removed from the set of seeds). Similarly, the diffusion centrality of the set of non-superspreader seeds is computed as the expected spread achieved when all the seeds are considered (both super- and non-superspreaders) minus the expected spread achieved when only superspreader seeds are considered (i.e., non-superspreader seeds are removed from the set of seeds).

In Figures 7 and 8, we compare the average diffusion centrality (across all the fake news items in the test set) of superspreader (defined by the threshold $\theta = 60\%$) seeds vs. non-superspreader seeds. The average diffusion centrality of both superspreaders and non-super-spreaders seeds is positive in all the cases,[11] meaning that both user types have an impact on spreading misinformation on the social network. Moreover, the seeds generated by SMF and Pop in the Politifact FakeNewsNet dataset and SMF in the HealthStory FakeHealth dataset yield the highest values of diffusion centrality for both user types; as shown in Figures 29–31 in the Online Appendix [62], this holds true regardless of the threshold used to define superspreaders.

It is evident from Figure 7 that, in the case of Politifact FakeNewsNet dataset, the average diffusion centrality of superspreader seeds is significantly higher than the average diffusion centrality of non-superspreader seeds (and this trend remains up to threshold $\theta = 80\%$– cf. Figures 29 and 30 in the Online Appendix [62]). This indicates that in a setting where the two user types

[11] With the exception of II with non-superspreader seeds and ICM in Politifact FakeNewsNet dataset. This holds true for all the other considered thresholds in the Politifact FakeNewsNet dataset except $\theta = 80\%$ (cf. Figures 29 and 30 in the Online Appendix [62]).

are almost balanced, superspreaders have more responsibility in diffusing misinformation among social network users. This also translates into a higher average expected spread achieved by superspreaders as compared to non-superspreaders for all thresholds up to threshold $\theta = 80\%$ (cf. Figures 19–24 in the Online Appendix [62]). For instance, when we consider non-baseline RAs, Figure 4 shows that the average expected spread increment for threshold $\theta = 60\%$ ranges from 5.1% with CB to 11.2% with SMF when the LTM is considered, from 1.4% with CB to 5.1% with SMF when the ICM is considered, and from 1.15% with II to 4.59% with SMF when the NPTM is considered. Furthermore, for both user types, the trend of average expected spread is the same as the trend reported for all users. It is important to note that not all superspreaders are malicious. People may spread misinformation on social media platforms unintentionally because of several factors, including having difficulties at discerning news veracity [77] and not being familiar with the platform features [29]. On the other end, users may want to intentionally spread fake news because it is "funny" and generates engagement among their friends [95].

Looking at Figure 8, which depicts the case where superspreaders are defined by threshold $\theta = 60\%$ in the HealthStory FakeHealth dataset, we see that the average diffusion centrality of non-superspreaders seeds is significantly higher than the average diffusion centrality of superspreaders seeds, indicating that non-superspreaders have more responsibility in diffusing misinformation among the social network users. This is understandable as in the scenario represented by the HealthStory FakeHealth dataset, the ratio of superspreaders to non-superspreaders is close to 1:100.[12] This also translates into a higher average expected spread achieved by non-superspreaders as compared to superspreaders for each RAs as reported in Figure 5 for $\theta = 60\%$. As shown in Figures 25–28 in the Online Appendix [62], these trends generally remain true regardless of the threshold used to define superspreaders. Moreover, similar results emerge on the Politifact FakeNewsNet dataset when we define superspreaders using $\theta \in \{90\%, 100\%\}$ (cf. Figures 23 and 24 in the Online Appendix [62]) as the number of non-superspreaders starts to become bigger than the number of superspreaders.

Overall, the results reported in this section highlight that the user group containing most users has a higher influence on fake news spread. Also, experiments show that the diffusion patterns are similar regardless of the dataset and diffusion model considered or threshold chosen to define the superspreaders. Specifically, Pop and SMF lead to the maximum misinformation spread, while CB leads to the least misinformation spread. Hence, we conclude that misinformation spread is mainly influenced by the RA chosen rather than the diffusion model or definition of superspreader.

4.3 Practical Implications of Our Study

Findings resulting from the experiments and analysis presented in Sections 4.1 and 4.2 reveal implications related to misinformation and fake news recommendations on social networks such as Twitter. We argue that these implications serve as lessons learned to inform researchers and practitioners' understanding of the connection between misinformation and the use of recommender systems in social networks.

When *prioritizing user satisfaction*, collaborative filtering alternatives are to be favored. This, however, is at the cost of misinformation amplification at the user and network levels, i.e., these strategies are more prone to recommend more fake news and propagate it in the social network by generating a high expected spread–particularly when fake news diffusion is described by ICM.

[12] Due to the high imbalance between superspreaders and non-superspreaders in the HealthStory FakeHealth dataset, we also computed the normalized diffusion centrality (the diffusion centrality for each user type is divided by the number of users of that type) for this dataset. We observed trends that align with those pertaining to diffusion centrality, as reported in Figures 8 and 31 in the Online Appendix [62], i.e., non-superspreaders have higher diffusion centrality than superspreaders. Hence, we decided to keep the original figures in the article.

If the priority is instead to *lessen the impact of misinformation*, then the most favorable RAs appear to be SMF and CB. The latter, in particular, since misinformation propagation is relatively controlled at the user- and network- level (in terms of MC, MRD, and E(Spread), respectively).

We posit that the most useful scenario is when user satisfaction and misinformation are *simultaneously accounted for*. In the case of the Politifact FakeNewsNet dataset where most news items are fake, non-personalized RAs–explicitly Pop–arose as the preferred alternative. Pop resulted in high MRR, and it best-contained misinformation in the immediate (user) range (high and positive MRD and low MC w.r.t. collaborative-filtering counterparts). Unfortunately, this comes at a cost in the long-term: Pop negatively impacts the social network regarding misinformation spread. Based on the outcomes emerging from the HealthStory FakeHealth dataset, where the fake news items are fewer than the real ones, CB is the preferred RA as it offers more relevant recommendations to its users while simultaneously lessening the impact of misinformation recommendations.

Perhaps the most unexpected takeaway shines a light on *user behavior*. When superspreader and non-superspreader volumes are comparable, as in the case of Politifact FakeNewsNet dataset, non-superspreaders are best served by collaborative-filtering strategies. In this case, a social network platform would not only increase user satisfaction while minimizing misinformation recommendations among their non-superspreader users, but it would also ensure that long-range misinformation diffusion across the social network is contained. On the other end, superspreaders are best served by CB as their trends are more similar to the ones discussed above for all users. Hence, if it is possible to discern among users based on their misinformation behavior, then using different RAs is the best choice. In this case, we argue that social network platforms should combine recommendation strategies with detection algorithms that are able to identify superspreader users from others [63, 71]. Instead, when the number of superspreaders is very low, as in the case of HealthStory FakeHealth dataset, both user types are best served by CB.

The main goal of a RA is user satisfaction which can lead to recommending misinformation to the users. Hence, there are always tradeoffs between user satisfaction and misinformation recommendation. The current literature only presents intervention models on top of RAs to increase user exposure rate towards diverse verified news and change user beliefs to mitigate misinformation[43, 44]. Wang et al. [93] proposed a model, Rec4Mit, that uses a news veracity classifier to recommend only true news to the users. However, in their experiments, the authors only used a test set composed of users that interacted with real news, so it is unclear how much users that interacted with fake news may be satisfied by this solution. More generally, researchers and industry practitioners should direct their efforts to develop RAs that are explicitly designed to prevent and mitigate fake news recommendations (and later its dissemination in the network).

5 CONCLUSIONS, LIMITATIONS, AND FUTURE WORK

In this manuscript, we investigated the strength of the connection between misinformation and RAs. For this, we first examined performance vs. misinformation recommendation tradeoffs. Among all the considered RAs, we saw that CB, followed by SMF, is the least prone to misinformation during recommendation, regardless of user behavior, but at the expense of user satisfaction. Conversely, collaborative filtering approaches (UU, II, ALS) are better alternatives for prioritizing user satisfaction, but at the cost of misinformation recommendations. Next, we studied the misinformation spread triggered by RAs by simulating misinformation propagation via well-known information diffusion models. We discovered that SMF and Pop are the most prone to contribute to misinformation diffusion in social networks, while CB results the least prone.

Along the way, we scrutinized users and how their habits (i.e., sharing fake news in their tweets) could potentially influence RA outcomes and further contribute to misinformation spread. With this in mind, we compared and contrasted trends emerging from users deemed superspreaders and non-superspreaders. Overall, we saw that both superspreaders and non-superspreaders are responsible for spreading misinformation. At the same time, when there is a balance in superspreaders and non-superspreaders, like in Politifact FakeNewsNet dataset, superspreaders have more responsibility in diffusing misinformation among social network users. Hence, we also showed the benefits of using different RAs for different user types. However, when superspreaders are not prominent in the social network (as in the case of HealthStory FakeHealth dataset), then inevitably, non-superspreaders take more accountability in disseminating misinformation. Findings from our study suggest that RAs and user engagement with misinformation are responsible for misinformation propagation in social networks. Our conclusions can directly impact the news domain, so prominent in the literature nowadays [64, 88, 92]. For instance, news sites like CNN and NBC can empower the strength of popularity based (and their modifications) recommendation strategies for news diffusion to many users. Especially among the non-superspreaders, Pop brings the right balance in tradeoffs between performance and misinformation recommendation. On the other hand, the presence of misinformation in real life is noticeable. Thus, our work brings to the attention of the RecSys community the need to redesign news RAs to be attentive to the perils of misinformation and proper profiling of superspreaders' and non-superspreaders' presence.

Our study is not free from limitations. First of all, it has been conducted on only two datasets of limited size, but, to the best of our knowledge, the Politifact FakeNewsNet dataset and HealthStory FakeHealth dataset are the only publicly available datasets containing all the information and sufficient number of user-news interactions needed to carry out our study. Second, the independent cascade, the linear threshold, and the node profile threshold models are widely used to model information diffusion and are easy to apply. Still, they may not be the best models to use to model misinformation spread. Other studies have highlighted how only considering network information is not enough, but user and news characteristics should also be considered when modeling misinformation spread in social networks [37, 48]. Given the limited information available in the considered datasets, we incorporated the user profile in the news sharing process via the Node Profile Threshold Model. However, if available, many other news and user characteristics should be considered, e.g., the emotions contained in the news [91]. Furthermore, even though the RAs considered in our study expand upon those initially studied in the seminal work by Fernández et al. [21], the list is not comprehensive. Much like we did for diffusion models, we selected well-known yet baselines RAs. In the future, we plan to extend our analysis to other hybrid [7, 57] and deep-learning-based RAs [12]; we will also consider state-of-the-art diffusion models [37].

The study of the connections between social networks, misinformation, and RAs is just the beginning. We have investigated a foundational aspect of misinformation in social networks by looking at the contribution to misinformation propagation due to RAs used by social network platforms. Important questions remain unanswered and thus prompt new research paths to explore. These include redesigning RAs to explicitly account for misinformation and developing new evaluation metrics that treat misinformation as a dimension to assess recommender systems and could help characterize, detect, and mitigate misinformation diffusion in social media platforms. Recommender systems are not the only information access platforms via which users are exposed to misinformation. Consequently, extending research in this area to further probe and contain misinformation in search engine responses is also critical.

REFERENCES

[1] Enrique Amigó, Yashar Deldjoo, Stefano Mizzaro, and Alejandro Bellogín. 2023. A unifying and general account of fairness measurement in recommender systems. *Information Processing and Management* 60, 1 (2023), 103115.

[2] Vito Walter Anelli, Alejandro Bellogín, Tommaso Di Noia, Dietmar Jannach, and Claudio Pomo. 2022. Top-N recommendation algorithms: A quest for the state-of-the-art. In *Proceedings of the 30th ACM Conference on User*

Modeling, Adaptation, and Personalization (UMAP'22). Association for Computing Machinery, 121–131. DOI:https://doi.org/10.1145/3503252.3531292

[3] Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, Eugenio Di Sciascio, and Felice Antonio Merra. 2020. Sasha: Semantic-aware shilling attacks on recommender systems exploiting knowledge graphs. In *Proceedings of the European Semantic Web Conference.* Springer, 307–323.

[4] Pablo Barberá. 2018. Explaining the spread of misinformation on social media: Evidence from the 2016 US presidential election. In *Proceedings of the Symposium: Fake News and the Politics of Misinformation. APSA.*

[5] Alejandro Bellogín and Yashar Deldjoo. 2021. Simulations for novel problems in recommendation: Analyzing misinformation and data characteristics. In *Proceedings of the SimuRec '21: The SimuRec Workshop Held in Conjunction with the 15th ACM Conference on Recommender Systems (RecSys).*

[6] Alessandro Bessi. 2016. Personality traits and echo chambers on Facebook. *Computers in Human Behavior* 65, C (2016), 319–324. DOI:https://doi.org/10.1016/j.chb.2016.08.016

[7] Michel Capelle, Marnix Moerland, Frederik Hogenboom, Flavius Frasincar, and Damir Vandic. 2015. Bing-SF-IDF+: A hybrid semantics-driven news recommender. In *Proceedings of the 30th Annual ACM Symposium on Applied Computing (SAC'15).* Association for Computing Machinery, 732–739. DOI:https://doi.org/10.1145/2695664.2695700

[8] Pablo Castells, Neil Hurley, and Saul Vargas. 2021. Novelty and diversity in recommender systems. In *Proceedings of the Recommender Systems Handbook.* Springer, 603–646.

[9] Mingxi Cheng, Chenzhong Yin, Shahin Nazarian, and Paul Bogdan. 2021. Deciphering the laws of social network-transcendent COVID-19 misinformation dynamics and implications for combating misinformation phenomena. *Scientific Reports* 11, 1 (2021), 1–14.

[10] Stefano Cresci. 2020. A decade of social bot detection. *Communications of the ACM* 63, 10 (2020), 72–83. DOI:https://doi.org/10.1145/3409116

[11] Enyan Dai, Yiwei Sun, and Suhang Wang. 2020. Ginger Cannot Cure Cancer: Battling Fake Health News with a Comprehensive Data Repository. In *Proceedings of the International AAAI Conference on Web and Social Media.* DOI:https://doi.org/10.48550/ARXIV.2002.00837

[12] Gabriel de Souza Pereira Moreira. 2018. CHAMELEON: A deep learning meta-architecture for news recommender systems. In *Proceedings of the 12th ACM Conference on Recommender Systems (RecSys'18).* Association for Computing Machinery, 578–583. DOI:https://doi.org/10.1145/3240323.3240331

[13] 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.

[14] Mukund Deshpande and George Karypis. 2004. Item-based top-N recommendation algorithms. *ACM Transactions on Information Systems* 22, 1 (2004), 143–177. DOI:https://doi.org/10.1145/963770.963776

[15] Dominic DiFranzo and Kristine Gloria-Garcia. 2017. Filter bubbles and fake news. *XRDS: Crossroads, the ACM Magazine for Students* 23, 3 (2017), 32–35.

[16] Bora Edizel, Francesco Bonchi, Sara Hajian, André Panisson, and Tamir Tassa. 2020. FaiRecSys: Mitigating algorithmic bias in recommender systems. *International Journal of Data Science and Analytics* 9, 2 (2020), 197–213.

[17] Michael D. Ekstrand. 2020. LensKit for python: Next-generation software for recommender systems experiments. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management.* DOI:https://doi.org/10.1145/3340531.3412778

[18] Michael D. Ekstrand, Mucun Tian, Ion Madrazo Azpiazu, Jennifer D. Ekstrand, Oghenemaro Anuyah, David McNeill, and Maria Soledad Pera. 2018. All the cool kids, how do they fit in?: Popularity and demographic biases in recommender evaluation and effectiveness. In *Proceedings of the 1st Conference on Fairness, Accountability, and Transparency.* Sorelle A. Friedler and Christo Wilson (Eds.), Proceedings of Machine Learning Research, Vol. 81, PMLR, 172–186. Retrieved from https://proceedings.mlr.press/v81/ekstrand18b.html

[19] Mehdi Elahi, Dietmar Jannach, Lars Skjærven, Erik Knudsen, Helle Sjøvaag, Kristian Tolonen, Øyvind Holmstad, Igor Pipkin, Eivind Throndsen, Agnes Stenbom, et al. 2021. Towards responsible media recommendation. *AI and Ethics* (2021), 1–12.

[20] Miriam Fernández and Alejandro Bellogín. 2020. Recommender systems and misinformation: The problem or the solution?. In *Proceedings of the OHARS Workshop, Co-located with the 14th ACM Conference on Recommender Systems.*

[21] Miriam Fernández, Alejandro Bellogín, and Iván Cantador. 2021. Analysing the effect of recommendation algorithms on the amplification of misinformation. arXiv:2103.14748. Retrieved from https://arxiv.org/abs/2103.14748

[22] Marco Furini, Silvia Mirri, Manuela Montangero, and Catia Prandi. 2020. Untangling between fake-news and truth in social media to understand the Covid-19 coronavirus. In *Proceedings of the 2020 IEEE Symposium on Computers and Communications (ISCC).* IEEE, 1–6.

[23] Ruoyuan Gao and Chirag Shah. 2020. Counteracting bias and increasing fairness in search and recommender systems. In *Proceedings of the 14th ACM Conference on Recommender Systems.* 745–747.

[24] Yingqiang Ge, Shuya Zhao, Honglu Zhou, Changhua Pei, Fei Sun, Wenwu Ou, and Yongfeng Zhang. 2020. Understanding echo chambers in e-commerce recommender systems. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2261–2270.

[25] Anastasia Giachanou, Bilal Ghanem, Esteban A. RÃŋssola, Paolo Rosso, Fabio Crestani, and Daniel Oberski. 2022. The impact of psycholinguistic patterns in discriminating between fake news spreaders and fact checkers. *Data and Knowledge Engineering* 138 (2022), 101960.

[26] Anastasia Giachanou, Paolo Rosso, and Fabio Crestani. 2021. The impact of emotional signals on credibility assessment. *Journal of the Association for Information Science and Technology* 72, 9 (2021), 1117–1132. DOI: https://doi.org/10.1002/asi.24480

[27] Quentin Grossetti, Cédric Du Mouza, and Nicolas Travers. 2019. Community-based recommendations on Twitter: Avoiding the filter bubble. In *Proceedings of the Web Information Systems Engineering–WISE 2019: 20th International Conference*. Springer, 212–227.

[28] Yulong Gu, Zhuoye Ding, Shuaiqiang Wang, Lixin Zou, Yiding Liu, and Dawei Yin. 2020. Deep multifaceted transformers for multi-objective ranking in large-scale e-commerce recommender systems. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management*. 2493–2500.

[29] Andrew Guess, Jonathan Nagler, and Joshua Tucker. 2019. Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances* 5, 1 (2019), eaau4586.

[30] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative filtering for implicit feedback datasets. In *Proceedings of the 2008 8th IEEE International Conference on Data Mining*. 263–272. DOI: https://doi.org/10.1109/ICDM.2008.22

[31] Eslam Hussein, Prerna Juneja, and Tanushree Mitra. 2020. Measuring misinformation in video search platforms: An audit study on YouTube. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–27.

[32] Mohsen Jamali and Martin Ester. 2010. A matrix factorization technique with trust propagation for recommendation in social networks. In *Proceedings of the 4th ACM Conference on Recommender Systems (RecSys'10)*. Association for Computing Machinery, 135–142. DOI: https://doi.org/10.1145/1864708.1864736

[33] Yitong Ji, Aixin Sun, Jie Zhang, and Chenliang Li. 2020. A re-visit of the popularity baseline in recommender systems. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1749–1752.

[34] Shan Jiang and Christo Wilson. 2021. Structurizing misinformation stories via rationalizing fact-checks. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 617–631.

[35] Fang Jin, Edward Dougherty, Parang Saraf, Yang Cao, and Naren Ramakrishnan. 2013. Epidemiological modeling of news and rumors on Twitter. In *Proceedings of the 7th Workshop on Social Network Mining and Analysis (SNAKDD'13)*. Association for Computing Machinery, 9 pages. DOI: https://doi.org/10.1145/2501025.2501027

[36] Fang Jin, Edward Dougherty, Parang Saraf, Yang Cao, and Naren Ramakrishnan. 2013. Epidemiological modeling of news and rumors on Twitter. In *Proceedings of the 7th Workshop on Social Network Mining and Analysis*. 1–9.

[37] Abishai Joy, Anu Shrestha, and Francesca Spezzano. 2021. Are you influenced?: Modeling the diffusion of fake news in social media. In *Proceedings of the ASONAM'21: International Conference on Advances in Social Networks Analysis and Mining*. Michele Coscia, Alfredo Cuzzocrea, Kai Shu, Ralf Klamma, Sharyn O'Halloran, and Jon G. Rokne (Eds.), ACM, 184–188. DOI: https://doi.org/10.1145/3487351.3488345

[38] Chanhyun Kang, Sarit Kraus, Cristian Molinaro, Francesca Spezzano, and V. S. Subrahmanian. 2016. Diffusion centrality: A paradigm to maximize spread in social networks. *Artificial Intelligence* 239 (2016), 70–96. DOI: https://doi.org/10.1016/j.artint.2016.06.008

[39] David Kempe, Jon Kleinberg, and Éva Tardos. 2003. Maximizing the spread of influence through a social network. In *Proceedings of the SIGKDD*. 137–146.

[40] Yuchen Li, Ju Fan, Yanhao Wang, and Kian-Lee Tan. 2018. Influence maximization on social graphs: A survey. *IEEE Transactions on Knowledge and Data Engineering* 30, 10 (2018), 1852–1872.

[41] Chen Lin, Si Chen, Hui Li, Yanghua Xiao, Lianyun Li, and Qian Yang. 2020. Attacking recommender systems with augmented user profiles. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management*. 855–864.

[42] Ping Liu, Karthik Shivaram, Aron Culotta, Matthew A Shapiro, and Mustafa Bilgic. 2021. The interaction between political typology and filter bubbles in news recommendation algorithms. In *Proceedings of the Web Conference 2021*. 3791–3801.

[43] Kuan-Chieh Lo, Shih-Chieh Dai, Aiping Xiong, Jing Jiang, and Lun-Wei Ku. 2021. All the wiser: Fake news intervention using user reading preferences. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*. 1069–1072.

[44] Kuan-Chieh Lo, Shih-Chieh Dai, Aiping Xiong, Jing Jiang, and Lun-Wei Ku. 2022. VICTOR: An implicit approach to mitigate misinformation via continuous verification reading. In *Proceedings of the ACM Web Conference 2022 (WWW'22)*. Association for Computing Machinery, 3511–3519. DOI: https://doi.org/10.1145/3485447.3512246

[45] Pasquale Lops, Marco de Gemmis, and Giovanni Semeraro. 2011. Content-based recommender systems: State-of-the-art and trends. *Recommender Systems Handbook* (2011), 73–105.

[46] Zhongqi Lu, Zhicheng Dou, Jianxun Lian, Xing Xie, and Qiang Yang. 2015. Content-based collaborative filtering for news topic recommendation. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence.*

[47] Hao Ma, Irwin King, and Michael R. Lyu. 2009. Learning to recommend with social trust ensemble. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'09).* Association for Computing Machinery, 203–210. DOI:https://doi.org/10.1145/1571941.1571978

[48] Long Ma, Chei Sian Lee, and Dion H. Goh. 2013. Understanding news sharing in social media from the diffusion of innovations perspective. In *Proceedings of the 2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical, and Social Computing.* IEEE, 1013–1020.

[49] Masoud Mansoury, Himan Abdollahpouri, Mykola Pechenizkiy, Bamshad Mobasher, and Robin Burke. 2020. Feedback loop and bias amplification in recommender systems. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management.* 2145–2148.

[50] Masoud Mansoury, Himan Abdollahpouri, Mykola Pechenizkiy, Bamshad Mobasher, and Robin Burke. 2021. A graph-based approach for mitigating multi-sided exposure bias in recommender systems. *ACM Transactions on Information Systems* 40, 2 (2021), 1–31.

[51] Shahan Ali Memon and Kathleen M. Carley. 2020. Characterizing covid-19 misinformation communities using a novel Twitter dataset. In *Proceedings of the 5th International Workshop on Mining Actionable Insights from Social Networks (MAISoN 2020), Co-located with CIKM.*

[52] Marcelo Mendoza, Barbara Poblete, and Carlos Castillo. 2010. Twitter under Crisis: Can we trust what we RT?. In *Proceedings of the 1st Workshop on Social Media Analytics (SOMA'10).* Association for Computing Machinery, 71–79. DOI:https://doi.org/10.1145/1964858.1964869

[53] Silvia Milano, Mariarosaria Taddeo, and Luciano Floridi. 2020. Recommender systems and their ethical challenges. *AI and Society* 35, 4 (2020), 957–967.

[54] Letizia Milli, Giulio Rossetti, Dino Pedreschi, and Fosca Giannotti. 2018. Active and passive diffusion processes in complex networks. *Applied Network Science* 3, 1 (2018). DOI:https://doi.org/10.1007/s41109-018-0100-5

[55] Maria D. Molina, S. Shyam Sundar, Thai Le, and Dongwon Lee. 2021. âĂIJFake NewsâĂİ Is Not simply false information: A concept explication and taxonomy of online content. *American Behavioral Scientist* 65, 2 (2021), 180–212. DOI:https://doi.org/10.1177/0002764219878224

[56] Judith Möller, Damian Trilling, Natali Helberger, and Bram van Es. 2018. Do not blame it on the algorithm: An empirical assessment of multiple recommender systems and their impact on content diversity. *Information, Communication and Society* 21, 7 (2018), 959–977.

[57] Gabriel De Souza P. Moreira, Dietmar Jannach, and Adilson Marques Da Cunha. 2019. Contextual hybrid session-based news recommendation with recurrent neural networks. *IEEE Access* 7 (2019), 169185–169203. DOI:https://doi.org/10.1109/ACCESS.2019.2954957

[58] Taichi Murayama, Shoko Wakamiya, Eiji Aramaki, and Ryota Kobayashi. 2021. Modeling the spread of fake news on Twitter. *PLOS ONE* 16, 4 (2021), 1–16. DOI:https://doi.org/10.1371/journal.pone.0250419

[59] Raymond S. Nickerson. 1998. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology* 2, 2 (1998), 175–220. DOI:https://doi.org/10.1037/1089-2680.2.2.175

[60] Jeppe Nørregaard, Benjamin D. Horne, and Sibel Adalı. 2019. NELA-GT-2018: A large multi-labelled news dataset for the study of misinformation in news articles. In *Proceedings of the International AAAI Conference on Web and Social Media.* 630–638.

[61] Eli Pariser. 2011. *The Filter Bubble: What the Internet is Hiding from You.* Penguin UK.

[62] Royal Pathak, Francesca Spezzano, and Maria Soledad Pera. 2023. [ONLINE APPENDIX] Understanding the Contribution of Recommendation Algorithms on Misinformation Recommendation and Misinformation Dissemination on Social Networks. Retrieved from https://github.com/royalpk/Recsys_Misinformation

[63] Francisco Rangel, Anastasia Giachanou, Bilal Hisham Hasan Ghanem, and Paolo Rosso. 2020. Overview of the 8th author profiling task at pan 2020: Profiling fake news spreaders on Twitter. In *Proceedings of the CEUR Workshop Proceedings.* Sun SITE Central Europe, 1–18.

[64] Shaina Raza and Chen Ding. 2021. News recommender system: A review of recent progress, challenges, and opportunities. *Artificial Intelligence Review* (2021), 1–52.

[65] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. 1994. GroupLens: An open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work (CSCW'94).* Association for Computing Machinery, 175–186. DOI:https://doi.org/10.1145/192844.192905

[66] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2011. Introduction to recommender systems handbook. In *Proceedings of the Recommender Systems Handbook.* Springer, 1–35.

[67] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2015. Recommender systems: Introduction and challenges. In *Proceedings of the Recommender Systems Handbook.* Springer, 1–34.

[68] Giulio Rossetti, Letizia Milli, Salvatore Rinzivillo, Alina Sîrbu, Dino Pedreschi, and Fosca Giannotti. 2018. NDlib: A python library to model and analyze diffusion processes over complex networks. *International Journal of Data Science and Analytics* 5, 1 (2018), 61–79.

[69] Giancarlo Ruffo, Alfonso Semeraro, Anastasia Giachanou, and Paolo Rosso. 2023. Studying fake news spreading, polarisation dynamics, and manipulation by bots: A tale of networks and language. *Computer Science Review* 47 (2023), 100531.

[70] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. 2001. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th International Conference on World Wide Web (WWW'01)*. Association for Computing Machinery, 285–295. DOI: https://doi.org/10.1145/371920.372071

[71] Anu Shrestha and Francesca Spezzano. 2022. Characterizing and predicting fake news spreaders in social networks. *International Journal of Data Science and Analytics* 13, 4 (2022), 385–398.

[72] Anu Shrestha, Francesca Spezzano, and Maria Soledad Pera. 2021. An empirical analysis of collaborative recommender systems robustness to shilling attacks. In *Proceedings of the 2nd Workshop on Online Misinformation-and Harm-Aware Recommender Systems (OHARS 2021)*.

[73] Gulshan Shrivastava, Prabhat Kumar, Rudra Pratap Ojha, Pramod Kumar Srivastava, Senthilkumar Mohan, and Gautam Srivastava. 2020. Defensive Modeling of Fake News Through online social networks. *IEEE Transactions on Computational Social Systems* 7, 5 (2020), 1159–1167. DOI: https://doi.org/10.1109/TCSS.2020.3014135

[74] Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. 2020. Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big Data* 8, 3 (2020), 171–188.

[75] Kai Shu, Xinyi Zhou, Suhang Wang, Reza Zafarani, and Huan Liu. 2019. The role of user profiles for fake news detection. In *Proceedings of the ASONAM'19: International Conference on Advances in Social Networks Analysis and Mining*. ACM, 436–439.

[76] Jakub Simko, Matus Tomlein, Branislav Pecher, Robert Moro, Ivan Srba, Elena Stefancova, Andrea Hrckova, Michal Kompan, Juraj Podrouzek, and Maria Bielikova. 2021. Towards continuous automatic audits of social media adaptive behavior and its role in misinformation spreading. In *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*. Association for Computing Machinery, 411–414. DOI: https://doi.org/10.1145/3450614.3463353

[77] Francesca Spezzano, Anu Shrestha, Jerry Alan Fails, and Brian W. Stone. 2021. That's fake news! reliability of news when provided title, image, source bias and full article. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–19. DOI: https://doi.org/10.1145/3449183

[78] Larissa Spinelli and Mark Crovella. 2020. How YouTube leads privacy-seeking users away from reliable information. In *Proceedings of the Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization (UMAP'20 Adjunct)*. Association for Computing Machinery, 244–251. DOI: https://doi.org/10.1145/3386392.3399566

[79] Marcella Tambuscio, Giancarlo Ruffo, Alessandro Flammini, and Filippo Menczer. 2015. Fact-checking effect on viral hoaxes: A model of misinformation spread in social networks. In *Proceedings of the 24th International Conference on World Wide Web*. 977–982.

[80] Nava Tintarev, Shahin Rostami, and Barry Smyth. 2018. Knowing the unknown: Visualising consumption blind-spots in recommender systems. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*. 1396–1399.

[81] Matus Tomlein, Branislav Pecher, Jakub Simko, Ivan Srba, Robert Moro, Elena Stefancova, Michal Kompan, Andrea Hrckova, Juraj Podrouzek, and Maria Bielikova. 2021. An audit of misinformation filter bubbles on YouTube: Bubble bursting and recent behavior changes. In *Proceedings of the 15th ACM Conference on Recommender Systems*. Association for Computing Machinery, 1–11. DOI: https://doi.org/10.1145/3460231.3474241

[82] Matus Tomlein, Branislav Pecher, Jakub Simko, Ivan Srba, Robert Moro, Elena Stefancova, Michal Kompan, Andrea Hrckova, Juraj Podrouzek, and Maria Bielikova. 2021. An audit of misinformation filter bubbles on YouTube: Bubble bursting and recent behavior changes. In *Proceedings of the 15th ACM Conference on Recommender Systems*. 1–11.

[83] Antonela Tommasel, Daniela Godoy, and Arkaitz Zubiaga. 2020. Workshop on online misinformation-and harm-aware recommender systems. In *Proceedings of the 14th ACM Conference on Recommender Systems*. 638–639.

[84] Antonela Tommasel, Daniela Godoy, and Arkaitz Zubiaga. 2021. OHARS: Second workshop on online misinformation-and harm-aware recommender systems. In *Proceedings of the 15th ACM Conference on Recommender Systems*. 789–791.

[85] Antonela Tommasel and Filippo Menczer. 2022. Do recommender systems make social media more susceptible to misinformation spreaders?. In *Proceedings of the 16th ACM Conference on Recommender Systems*. 550–555.

[86] Antonela Tommasel, Juan Manuel Rodriguez, and Daniela Godoy. 2021. I want to break free! recommending friends from outside the echo chamber. In *Proceedings of the 15th ACM Conference on Recommender Systems*. 23–33.

[87] Antonela Tommasel, Juan Manuel Rodriguez, and Filippo Menczer. 2022. Following the trail of fake news spreaders in social media: A deep learning model. In *Proceedings of the Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation, and Personalization*. 29–34.

[88] Christoph Trattner, Dietmar Jannach, Enrico Motta, Irene Costera Meijer, Nicholas Diakopoulos, Mehdi Elahi, Andreas L. Opdahl, Bjørnar Tessem, Njål Borch, Morten Fjeld, et al. 2021. Responsible media technology and AI: Challenges and research directions. *AI and Ethics* (2021), 1–10.

[89] João Vinagre, Alípio Mário Jorge, Marie Al-Ghossein, and Albert Bifet. 2020. ORSUM-workshop on online recommender systems and user modeling. In *Proceedings of the 14th ACM Conference on Recommender Systems*. 619–620.

[90] Nguyen Vo and Kyumin Lee. 2018. The rise of guardians: Fact-checking url recommendation to combat fake news. In *Proceedings of the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval*. 275–284.

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

[92] Sanne Vrijenhoek, Mesut Kaya, Nadia Metoui, Judith Möller, Daan Odijk, and Natali Helberger. 2021. Recommenders with a mission: Assessing diversity in news recommendations. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*. 173–183.

[93] Shoujin Wang, Xiaofei Xu, Xiuzhen Zhang, Yan Wang, and Wenzhuo Song. 2022. Veracity-aware and event-driven personalized news recommendation for fake news mitigation. In *Proceedings of the ACM Web Conference 2022 (WWW'22)*. Association for Computing Machinery, 3673–3684. DOI: https://doi.org/10.1145/3485447.3512263

[94] Chuhan Wu, Fangzhao Wu, Yongfeng Huang, and Xing Xie. 2022. Personalized news recommendation: Methods and challenges. *ACM Transactions on Information Systems* (2022). DOI: https://doi.org/10.1145/3530257

[95] Waheeb Yaqub, Otari Kakhidze, Morgan L. Brockman, Nasir Memon, and Sameer Patil. 2020. Effects of credibility indicators on social media news sharing intent. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 1–14. DOI: https://doi.org/10.1145/3313831.3376213

[96] Huiling Zhang, Md Abdul Alim, Xiang Li, My T. Thai, and Hien T. Nguyen. 2016. Misinformation in online social networks: Detect them all with a limited budget. *ACM Transactions on Information Systems* 34, 3, (2016), 24 pages. DOI: https://doi.org/10.1145/2885494

[97] Xichen Zhang and Ali A. Ghorbani. 2020. An overview of online fake news: Characterization, detection, and discussion. *Information Processing and Management* 57, 2 (2020), 102025. DOI: https://doi.org/10.1016/j.ipm.2019.03.004

[98] Qian Zhao, F. Maxwell Harper, Gediminas Adomavicius, and Joseph A. Konstan. 2018. Explicit or implicit feedback? Engagement or satisfaction? A field experiment on machine-learning-based recommender systems. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*. 1331–1340.

[99] Xinyi Zhou, Apurva Mulay, Emilio Ferrara, and Reza Zafarani. 2020. ReCOVery: A multimodal repository for COVID-19 news credibility research. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management*. 3205–3212.

[100] Arkaitz Zubiaga, Ahmet Aker, Kalina Bontcheva, Maria Liakata, and Rob Procter. 2018. Detection and resolution of rumours in social media: A survey. *ACM Computing Surveys* 51, 2 (2018), 1–36.

Received 13 September 2022; revised 21 June 2023; accepted 15 July 2023