Highlights

The Role of Recommendation Algorithms in the Formation of Disinformation Networks

Pau Muñoz,Raúl Barba-Rojas,Fernando Díez,Alejandro Bellogín

- Dense and cohesive network structures facilitate rapid disinformation spread
- Higher accuracy algorithms correlate with more cohesive disinformation networks
- DeepRank and Word2Vec contribute to formation and consolidation of disinformation networks
- New metrics reveal disinformation amplification in legitimate networks by certain algorithms
- Promoting diversity in recommenders can reduce the formation of disinformation networks

The Role of Recommendation Algorithms in the Formation of Disinformation Networks ⋆

Pau Muñoz[a], Raúl Barba-Rojas[b], Fernando Díez[a] and Alejandro Bellogín[a,∗]

[a]*Universidad Autónoma de Madrid, Escuela Politécnica Superior, Madrid, Spain*
[b]*Universidad de Castilla-La Mancha, Escuela Superior de Informática, Ciudad Real, Spain*

ARTICLE INFO

Keywords:
Social networks
Disinformation
Content diffusion
Accountability

ABSTRACT

Disinformation on social networks, especially those that share media content, remains a critical issue with far-reaching societal implications. Although extensive research has addressed the prevalence and mitigation of false information, the specific impact of recommendation algorithms on the creation and consolidation of disinformation networks has not been thoroughly examined. In this work, we bridge this gap by simulating how various recommendation techniques—ranging from basic yet foundational approaches such as popularity-based and content-based methods—shape network dynamics and facilitate disinformation spread. These classical algorithms are essential building blocks of modern hybrid and task-specific recommender systems; understanding their effects is thus crucial for assessing systemic risks. Using a dataset comprising tweets from 275 disinformation agents and 275 legitimate journalism agents, we conduct a realistic simulation grounded in probabilistic click models of user behavior and real-world social media data. Our findings reveal that certain recommendation approaches can significantly reinforce the cohesion and visibility of disinformation networks, thereby amplifying their reach. These results underscore the necessity for algorithmic accountability and the design of ethically responsible recommender systems to maintain information integrity on social platforms.

1. Introduction

In today's digital world, social media has risen to become the main channel for spreading information. This marks a key change from the old days when traditional media, with its centralized control and gatekeeping, was the most popular option. Now, we are seeing a more user-driven model that is decentralized and allows anyone to both consume and create content at an unprecedented scale (Mazzara, Biselli, Greco, Dragoni, Marraffa, Qamar and de Nicola, 2013).

However, this shift has not emerged without challenges. A major issue is the departure from traditional editorial oversight towards algorithmic curation, in particular via recommendation algorithms. These systems, designed to maximize user engagement by personalizing content, can unintentionally create echo chambers and facilitate the spread of disinformation (Wu, Morstatter, Carley and Liu, 2019; Tomlein, Pecher, Simko, Srba, Móro, Stefancova, Kompan, Hrckova, Podrouzek and Bieliková, 2021). Such phenomena may reshape public opinion and have substantial impacts on critical societal issues such as politics, economy, and social causes (Stefanone, Lackaff and Rosen, 2010; Cantador, Cortés-Cediel and Fernández, 2020). Hence, the way content is recommended on social media plays a key role in shaping the digital information ecosystem.

Content recommendation mechanisms are governed by diverse algorithmic approaches (Anandhan, Shuib, Ismail and Mujtaba, 2018), which can be broadly categorized into collaborative filtering (CF), content-based (CB), deep learning (DL), matrix factorization (MF), and reinforcement learning (RL) algorithms. Modern recommender systems, although highly complex, often integrate or enhance these core methodologies. CF recommends content based on user similarity patterns (Nikolakopoulos, Ning, Desrosiers and Karypis, 2022), whereas CB aligns recommendations with content attributes such as metadata and textual similarity (Musto, de Gemmis, Lops, Narducci and Semeraro,

⋆This work has been supported by grant PID2022-139131NB-I00 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe".

∗Corresponding author

pau.munnozp@estudiante.uam.es (P. Muñoz); raul.barba@alu.uclm.es (R. Barba-Rojas); fernando.diez@uam.es (F. Díez); alejandro.bellogin@uam.es (A. Bellogín)

ORCID(s): 0009-0007-9255-4295 (P. Muñoz); 0000-0002-4549-0483 (R. Barba-Rojas); 0000-0001-7098-5659 (F. Díez); 0000-0001-6368-2510 (A. Bellogín)

2022). MF identifies latent factors from user-item interactions to refine predictions, while DL models leverage neural architectures to capture nuanced user behaviors and content features (Zhang, Yao, Sun and Tay, 2019). Content-based deep learning methods process textual and multimedia data, while graph-based approaches better capture structural and relational dynamics within networks. Finally, RL adopts a dynamic trial-and-error strategy, continuously adapting recommendations based on user feedback (Afsar, Crump and Far, 2023).

These algorithmic families represent the essential foundations of modern social media recommendation systems. While contemporary models may refine or hybridize these approaches, the underlying principles remain consistent. Despite the considerable body of research on disinformation and recommendation algorithms individually, the specific impact of distinct recommendation strategies on the formation and consolidation of disinformation networks has not been systematically explored. Most existing works address either disinformation mitigation at a general level or improvements in recommendation techniques, but do not analyze how core algorithmic choices shape network structures susceptible to disinformation. This gap motivates our study.

Conceptual framing. Modern recommendation systems, although highly complex today, are fundamentally built upon a few core conceptual strategies: collaborative filtering, content-based filtering, optimization through deep learning, and reinforcement learning. These foundational methodologies are conceptually simple, clearly distinguishable, and easy to understand. In our research, we deliberately focus on these basic strategies rather than on composite state-of-the-art systems. This allows us to isolate and analyze how each foundational approach individually contributes to the formation and reinforcement of disinformation networks. By doing so, we aim to identify which core mechanisms require more cautious treatment when designing future recommendation systems that seek to balance accuracy and information integrity.

Our work. We aim to bridge this gap by critically evaluating how different recommendation algorithms influence the spread of disinformation, focusing on their contribution to the creation and reinforcement of disinformation networks. Leveraging a real-world dataset tracking disinformation and legitimate actors over three years, we simulate recommendation scenarios with 12 representative algorithms and analyze the resulting network formations and dynamics.

Research questions. We formulate the following research questions:

RQ1: Do different recommendation algorithms generate different networks from a content diffusion perspective?

RQ2: What is the impact of recommendation algorithms in disinformation propagation?

RQ3: Is there any relation between disinformation propagation and accuracy of recommendation algorithms?

Main contributions. Our main contributions are:

- A comparative analysis showing how networks generated by different recommendation algorithms differ in terms of content diffusion and structural properties.
- Empirical evidence identifying which algorithms facilitate the emergence and consolidation of disinformation networks.
- An exploration of the tradeoff between recommendation accuracy and disinformation amplification, highlighting cases where both objectives can be partially balanced.

Implications. Through this comparative analysis, we uncover insights into how recommendation strategies may inadvertently favor the formation and expansion of disinformation networks. Ultimately, our findings aim to inform the design of innovative models that maintain recommendation accuracy while constraining disinformation spread by fostering more diverse, healthy, and trustworthy information ecosystems (Wang, Zhang, Wang and Ricci, 2024). This contributes both to the academic discourse on social media impacts and to practical guidelines for ethical recommender system design.

2. Formulation and background

2.1. Fundamental concepts

Here, we introduce a series of definitions that are of interest and necessary both, to establish the context in which we will develop the research, and to narrow down the technical meaning of the networks to which we will refer from now on.

Definition 1 (Online Social Network)**.** An online social network, or simply a social network, in the context of this work, is an online structure consisting of a set of nodes, representing users, and edges, representing relationships or interactions between those users, where the flow of information and content occurs within a digital environment. An example of such a network is *X*.

Definition 2 (Information Network)**.** An information network is a structure composed of nodes and edges dedicated to the distribution and circulation of information. The nodes represent actors or information sources, while the edges represent the channels or media through which the information flows.

Definition 3 (Disinformation Network)**.** A disinformation network is a sub-network within a social network, characterized by formal and/or informal coordination between users aimed at distributing disinformation on specific topics (Muñoz, Díez and Bellogín, 2024).

Definition 4 (Legitimate Network)**.** In the context of our research, a legitimate network is a sub-network within an online social network that forms naturally or organically, without being characterized by the distribution of disinformation, and consists of genuine interactions between users.

Definition 5 (Journalism Network)**.** A journalism network is a sub-network within an online social network composed of journalists and media professionals, characterized by interactions and the distribution of journalistic content. By default, it would be considered an example of a legitimate network.

Definition 6 (Recommendation Network)**.** A recommendation network in the context of this work is an online social network where a recommendation system operates, suggesting content or connections to users based on prior interactions within the network.

Definition 7 (Resulting Network)**.** The resulting network is the network that emerges after simulating the actions of a recommendation system on the original social network, highlighting the newly suggested connections.

2.2. Background

In the current digital era, social media platforms are overtaking traditional media outlets, e.g., newspapers and TV, as the foremost information source for people (Hanna, Rohm and Crittenden, 2011; Rajendran and Thesinghraja, 2014; Yoon and Kim, 2001). This shift is attributed to the accessibility and immediacy social media offers, enabling users to receive and disseminate news at unprecedented speed.

Social media platforms are fundamentally distinct from traditional media in terms of their structural and operational paradigms. Unlike the centralized, gate-kept nature of traditional media, social media is characterized by its decentralization and omnipresence (Kelly, 2010). Users are not limited by geographical or temporal constraints when accessing these platforms, thanks to the ubiquitous presence of smartphones and internet connectivity. This universal access empowers users not only to consume content, but also to create and share their own information (Rajendran and Thesinghraja, 2014; Jankowski and Van Selm, 2000). This democratization of content creation and distribution does, however, introduce a unique challenge: the bypassing of traditional editorial oversight. In the realm of social media, content is disseminated into the network without undergoing a rigorous human editorial process[1] (Jankowski and Van Selm, 2000; Yoon and Kim, 2001).

In the absence of a conventional editorial filter, the responsibility of content curation and distribution largely falls upon algorithmic recommendation systems (Jankowski and Van Selm, 2000; Yoon and Kim, 2001; Pariser, 2011). These systems, by leveraging complex algorithms, determine the visibility and reach of content shared on the platform. By analyzing user behavior and interactions, recommendation algorithms (RAs) curate a personalized feed for each user, thereby becoming the principal mechanism for content distribution on social media (Anandhan et al., 2018). This shift from human editorial control to algorithm-driven curation marks a significant transformation in how information is disseminated and consumed, raising questions about the implications for information quality and the formation of echo chambers or disinformation networks (Anandhan et al., 2018; Tomlein et al., 2021).

The unparalleled ease of access provided by social media platforms has led to their growing influence on public opinion formation (Hanna et al., 2011). Individuals increasingly rely on these platforms for engaging in discussions and debates on critical issues, ranging from national and international politics, including shaping voting preferences,

[1] Although this applies in general, there are some content categories that, in most social media, tend to be explicitly prohibited, e.g., material related to terrorism or certain types of criminal activity against minors, where platforms enforce removal policies.

to economic debates and social causes (Rajendran and Thesinghraja, 2014). This trend underscores social media's significant role in modern public sphere.

The democratization of publishing, afforded by the minimal algorithmic filtering on social media, significantly contributes to the rise of online affective polarization, radicalization, and the proliferation of disinformation (Wu et al., 2019; Guess and Lyons, 2020; Aïmeur, Amri and Brassard, 2023). This environment, where short pieces of information are rapidly shared in real-time, is ripe for manipulation by networks of bots and trolls, either state-sponsored or spontaneous, exacerbating these issues (Guess and Lyons, 2020).

Networks characterized by an "echo chamber" structure are particularly potent in this regard (Villa, Pasi and Viviani, 2021; Pariser, 2011). These echo chambers are networks that are highly dense, focusing intensely on a narrowly selected range of topics that are deeply infused with political significance (Valle, Broersma and Ponsioen, 2021; Posegga and Jungherr, 2019). The dynamics within these echo chambers ensure that disinformation is not only circulated but also amplified, as users within these networks repeatedly share and reinforce the same narratives without exposure to dissenting or opposing views (Villa et al., 2021).

These phenomena are indicative of the powerful role that network structure and user engagement play in the dissemination and amplification of content on social media (Muñoz et al., 2024). The dense, self-reinforcing nature of echo chambers significantly magnifies the spread of disinformation, making it a formidable challenge to counteract within the digital public square.

Various studies have underscored the role of recommendation systems in social networks as catalysts for the formation of echo chambers (Pariser, 2011; Van den Bogaert, Geerts and Harambam, 2022; Raza and Ding, 2022; Pathak, Spezzano and Pera, 2023; Fernández, Bellogín and Cantador, 2024). It is acknowledged that these systems, by tailoring content based on user preferences, inadvertently contribute to the creation of digital environments where users are predominantly exposed to viewpoints that align with their own. This lack of exposure to diverse perspectives perpetuates and reinforces existing beliefs, further isolating users within their ideological bubbles (Pariser, 2011). Moreover, in the field of shilling or poisoning attacks on recommendation systems, previous studies have demonstrated that even a small amount of false information can significantly influence recommendation outcomes, leading to the formation of disinformation networks. Vulnerabilities in both the data and model training phases can be exploited, allowing attackers to inject biased or fake interactions into the system, which then promotes disinformation by distorting the recommendations it generates (Nguyen et al., 2024). These manipulations can compromise the integrity of the system, amplifying disinformation and causing clusters of users to form around such content. Importantly, this manipulation can occur either actively, through users intentionally seeking to influence the algorithmic process, or passively, when users share false information due to personal conviction or unawareness.

Therefore, disinformation is not limited to blatant fake news but extends to the creation and dissemination of altered realities grounded in half-truths or misleading narratives, often to establish or reinforce certain political or ideological positions. In this work, we focus on understanding the broader emergence and persistence of disinformation networks, regardless of whether they form through deliberate manipulation or unconscious propagation of false information. The subtle yet profound impact of small amounts of false data on recommendation systems underscores the role of these networks in shaping user engagement and content consumption (Nguyen et al., 2024).

Despite the importance and recent interests in this topic, how specifically these recommendation algorithms affect the (dis)information networks through feedback loops, and whether different families produce different behaviors remains unexplored. Throughout the rest of this paper, we aim to address these questions and provide answers towards a better understanding of the role of these algorithms in this context.

3. Methodology

In this work, we are interested in the study of recommendation algorithms and their impact on the generation of disinformation networks. For that reason, we considered X/Twitter as the online social network to be analyzed. Following the literature (Barash, Fink, Cameron, Schmidt, Dong, Macy, Kelly and Deshpande, 2020; Linvill and Warren, 2020; Muñoz et al., 2024), we use a dataset extracted from this social network that contains accounts explicitly marked as disinformation combined with other more neutral accounts regarding their behavior towards disinformation. Section 4.1 provides more details about the data collection step.

Studying the spread of disinformation on social media platforms like X (formerly Twitter) has commonly relied on simulations, either through agent-based models (ABMs) replicating user interactions or by leveraging real-world datasets. For instance, ABMs have been used to simulate the effects of disinformation strategies, such as bots amplifying

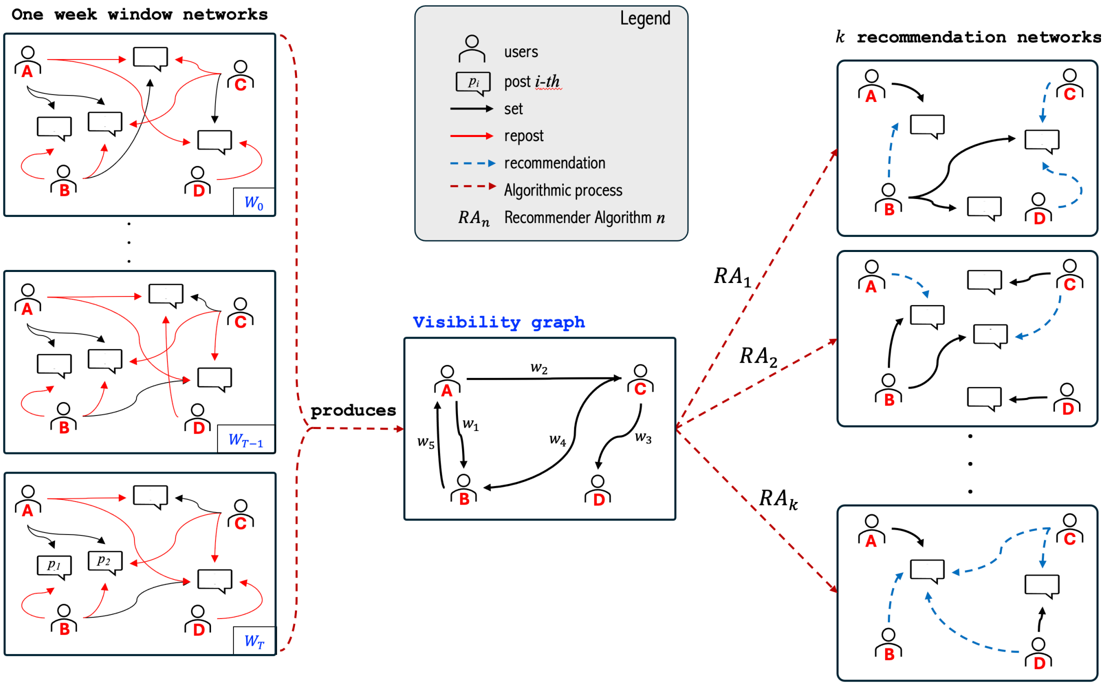

Figure 1: Visual depiction of generation of visibility graph and recommendation networks for current window W_T, while also considering previous T windows $W_0, \cdots, W_{T-1}$. The creation of the visibility graph employs T time windows. In each window, we have a set of users $(A, B, C, D, ...)$ who post and repost content among themselves. The *visibility graph* represents the activity followed by the users in terms of the relationship between them, with the graph's arcs denoting the corresponding weights obtained from the intensity of the relationship. From the visibility graph, recommendation networks are simulated as detailed in Figure 2.

key influencers or bridging distinct communities (Beskow and Carley, 2019). Similarly, simulations based on contact recommendation systems have been employed to understand how algorithms affect network structures and information flow, contributing to phenomena like filter bubbles or just to study how different algorithmic approaches affect network properties (Sanz-Cruzado and Castells, 2018). Our approach builds on these established methodologies, particularly on simulation based on actual data by focusing on a dataset extracted from X that includes both disinformation actors and neutral accounts, as has been done in prior research.

In our methodology, we divide the dataset into weekly time windows, consistent with the temporal nature of post visibility on X, where content often has a lifespan of about a day but may persist longer if it gains traction (Beskow and Carley, 2019). For each week, we simulate the behavior of various recommendation algorithms by using past windows to train the model and the current window to evaluate its performance.

Now, we describe the methodology used to carry out our study throughout simulations, an acknowledged tool in the area to quantify novel assumptions (Ferro, Fuhr, Grefenstette, Konstan, Castells, Daly, Declerck, Ekstrand, Geyer, Gonzalo, Kuflik, Lindén, Magnini, Nie, Perego, Shapira, Soboroff, Tintarev, Verspoor, Willemsen and Zobel, 2018): first, we describe the methods we used to perform simulations with the selected recommendation algorithms (Section 3.1); then, we present our approach for generating the networks that will be analyzed (Section 3.2). In both approaches (depicted in Figure 1), in order to consider the temporal evolution of the information and the algorithm effects, we divided the dataset in sessions or windows of 7 days. This is consistent with the literature, as posts in X have a lifespan of about a day, although certain viral content and news may last for a more extended period (Orellana-Rodriguez and Keane, 2018). A weekly cycle is particularly effective in capturing user behavior patterns, as certain news are introduced in the network, whereas old trending contents are still spread across. This temporal frame allows for the incorporation of the most recent interactions and preferences, thereby enhancing the algorithm's responsiveness to changes in user interests.

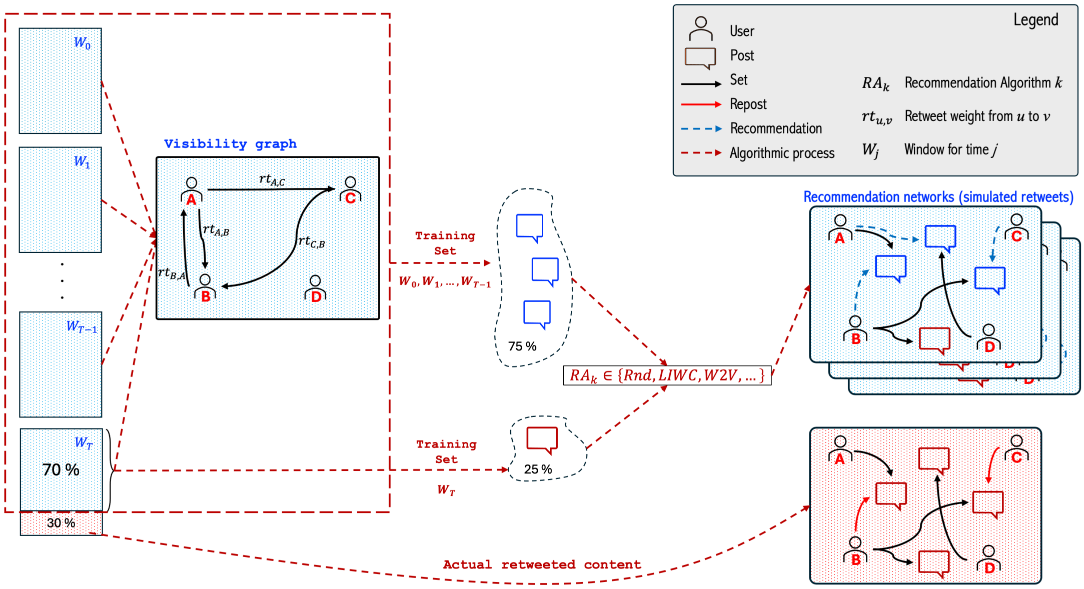

Figure 2: Visual representation of the followed experimental methodology, highlighting that, in a given time window W_T, the visibility graph is created according to all the content of the previous $T-1$ windows and the training subset (70%) of the current window W_T, considering the number of retweeted posts between users as the connection weight. The same information is used to train the recommendation algorithms and generate the recommendation networks, but for discovery purposes, the candidate items to be recommended are drawn from previous and current windows in a 75-25 ratio. Finally, the evaluation is assessed with respect to the test set (30%) of the current window.

3.1. Simulation of recommendation algorithms

In line with our conceptual framework, we evaluate representative algorithms from each fundamental family: collaborative filtering, content-based methods, deep learning models, and reinforcement learning approaches. Rather than focusing on highly engineered or composite commercial systems, we select core representatives to provide a clean and interpretable analysis of how basic algorithmic mechanisms influence network dynamics and disinformation propagation.

For each implemented RA, we simulated its behavior as follows. First, a percentage of posts from the current window W_n are considered, particularly 70% as training data as it is commonly employed in the literature (Muraina, 2022); however, to better deal with sparsity issues while exploiting past interests of users, past T windows might also be considered (in this work, we considered $T = 2$). Retweet information corresponding to user retweets up to W_{n-1} (to avoid skewed recommendations by considering the current window) is used to build what we call *visibility graph* (see left part of Figure 2). This graph determines, for each user, the content that is visible to such user, based on her previous experience. Hence, it uses data from previous windows to exploit the retweet relationships between each pair of users. Indeed, a user (user A) is connected to another user (user B) in the visibility graph ($A \rightarrow B$) if and only if user A retweeted content from user B in the previous T windows. Thus, the visibility graph is a directed and weighted graph, where the edge weight is related to the number of retweeted posts, indicating the strength of that retweet relationship from one user to another. Naturally, the visibility graph provides a useful and intuitive representation of the content visible to each user. It is used to filter the content (e.g., tweets or posts) the algorithm can recommend to a user.

Formally, consider we have a network G with $|G| = N$ users and $W_\tau, W_1, ..., W_{\tau+T}$ time windows. We define the set $P_i^t = \{p_k(u_i, W_t); k \in \mathbb{N}\}$, $1 \leq i \leq N$, $\tau \leq t \leq \tau + T$, being $p_k(u_i, W_t)$ the k-th post of i-th user at t-th window ($\tau \leq t \leq \tau + T$). Similarly, we define the set $R_{j,i}^t = \{r(u_j, p_k(u_i, W_t)); 1 \leq j \leq N, k \in \mathbb{N}\}$ being $r(u_j, p_k(u_i, W_t))$ the repost of j-th user to the post of i-th user at t-th window ($\tau \leq t \leq \tau + T$). For example, at W_T in Figure 1 (assuming

$\tau = 0$), we have: $P_A^T = \{p_1(A, W_T), p_2(A, W_T)\}$ (in this case $u_i = A$). The corresponding set of reposts for user B (in this case $u_j = B$) is $R_{B,A}^T = \{r(B, p_1(A, W_T)), r(B, p_2(A, W_T))\}$.

Under these definitions, to define the *visibility graph*, we focus on the current time window W_n; that is, we set $\tau + T = n$, hence the starting window will have index $\tau = n - T$. In this situation, we will set a connection between user u_i (namely A) and user u_j (namely B), represented as $(A \rightarrow B) \iff \exists\ r(u_j, p_k(u_i, W_n)) \wedge \forall t \in [\tau, n-1] \mid \exists p_k(u_i, W_t) \implies \exists r(u_j, p(u_i, W_t))$. This definition is equivalent to set $|P_i^t| = |R_{j,i}^t|$; $t = \tau, ..., n-1$. In other words, we mean that by looking at the relationship in the n-th window, for each post by user A that has been reposted by user B, these relationships post-repost are also reproduced in the previous windows (up to T windows, i.e., up to window with index $\tau = n - T$). The resulting *visibility graph* represents all those relations between users that post and repost consistently during a time split in previous T windows with respect to the current n-th window.

Then, we leverage the use of the visibility graph to determine which posts have to be ranked by each algorithm in a window for every user. In this scenario, the recommender has to provide N recommendations to each user. Considering a general formulation of a recommendation algorithm, it would produce a score for a given (user, item) pair, so it is critical which items are requested from the algorithm (Said and Bellogín, 2014). Thus, to enhance the discovery of timeless content while mitigating recency bias often observed in recommender systems, we carefully allocate the distribution of recommendations to include 25% from the current window and 75% from the past content, with the rationale grounded in optimizing user engagement and content relevance, thereby offering a more diversified and enriching user experience.

This allocation is informed by the dual objectives of ensuring users are kept abreast of the latest trends and discussions, which the 25% current window allocation achieves by surfacing recent content that might be of immediate interest (Bountouridis, Harambam, Makhortykh, Marrero, Tintarev and Hauff, 2019; Sanz-Cruzado and Castells, 2022). Concurrently, the 75% allocation for past content leverages the rich repository of historical data, acknowledging that valuable and engaging content is not solely confined to the present moment (Bountouridis et al., 2019; Sanz-Cruzado and Castells, 2022). While this decision may deserve further analysis, since it affects all RAs equally, we consider it is out of the scope of this work and aim to analyze its effect in our experiments in the future.

Regarding the number of recommendations to produce, we decided to set $N = 10$ as a deliberate choice aimed at optimizing the user experience by providing a manageable number of options that is enough to offer variety without overwhelming users, ensuring engagement while maintaining decision simplicity (Sanz-Cruzado, Pepa and Castells, 2018; Sanz-Cruzado and Castells, 2022).

Once recommendations are generated for each user and window, a *retweet simulation* process starts. This process aims at simulating the behavior of a user who may (or may not) retweet the provided recommendations by the system. Since users, as in any interactive system, tend to give more attention and interact (i.e., retweet) more with recommendations at the top of the presented list, we model this behavior through a rank-based click model (Chuklin, Markov and de Rijke, 2015), where the retweeting probability decreases with the ranking position. We decided to employ this strategy, where the user is much more likely to interact with the first item they visually encounter in the recommendation list, as this perspective is well-supported by existing literature. Additionally, by using the same approach across all tested algorithms, we ensure an effective comparison between them, as the interaction model remains consistent for each recommendation system (Joachims, Granka, Pan, Hembrooke and Gay, 2017; Cheng, Adamic, Dow, Kleinberg and Leskovec, 2014; Epstein and Robertson, 2015).

Finally, these *simulated retweets* are then compared with the actual retweeted content to evaluate the accuracy (the remaining 30% as test dataset, see Figure 2). This information is then used to understand the relationship between aspects that can foster the generation and consolidation of (dis)information networks and accuracy metrics.

3.2. Recommendation network generation

As an outcome of the simulation, a *recommendation network* is created for every algorithm in each time window. In this way, recommendation algorithms are used to simulate recommendations for each user, which builds a network specifically related to the algorithm considered in the simulation.

To create this recommendation network, during the simulation, as described before, retweets are recommended by a given algorithm for every user. The network is a directed and weighted graph, whose nodes represent users and whose edges $(A \rightarrow B)$ represent the existence of a simulated retweet in a given direction (see right part of Figures 1 and 2): A retweets a post from B from the recommendations performed by the specific algorithm. Indeed, the recommendation network represents a simulation of the information flow from the nodes in the original network, and it allows us to leverage Social Network Analysis (SNA) techniques to analyze and understand such network and its properties.

Table 1
Dataset publications per user type and year.

	Collected tweets			
	2019	2020	2021	2022
Disinformation	1,584,460	2,521,810	1,826,461	2,075,601
Journalism	1,256,867	1,200,606	984,451	796,121

Table 2
Properties of the complete networks (using all data from the entire 2019-2022 period), with 275 users in both journalists and disinformation actors categories.

Property	Journalists	Disinformation Actors
Users	275	275
Posts	3,906,047	7,194,766
Negative Posts	118,951	513,566
Nodes	953	815
Edges	30,235	20,180
Average Degree	60.15	50.08
Avg Retweets	361.11	1,867.52

Thus, as part of our methodology, we create recommendation networks for each time window in our dataset, considering every recommendation algorithm. Moreover, when necessary, we apply this process either to the entire user community or only to a subset of the users, to better understand the information dynamics and consolidation of networks between sub-communities (in particular, to isolate the disinformation accounts).

4. Experiments

To address the research questions formulated in Section 1, we designed a systematic experimental framework simulating the dynamics of content recommendation and network formation over time. This section details the experimental pipeline, describing how the data was obtained and processed (Section 4.1), the recommendation algorithms implemented (Section 4.2), and the network-based metrics employed to evaluate the effects of the different algorithms (Section 4.3).

Our experimental process follows three main stages: (i) preprocessing and selection of data from social networks; (ii) simulation of content recommendations using multiple algorithms across sliding time windows; and (iii) analysis of the resulting networks to assess structural properties and disinformation amplification tendencies. Each step has been carefully designed to maintain comparability across algorithms and to isolate their specific influence on network dynamics.

4.1. Data

In this research, we employ the disinformation and legitimate agents dataset crafted by the authors in (Muñoz et al., 2024). Such dataset contains tweet information from 275 disinformation accounts and 275 legitimate accounts spanning from 2019 to August 2022. In order to identify the disinformation agents, authors employed fact-checking websites. On the other hand, journalist accounts were obtained based on articles published on most browsed online media outlets. We decided to employ this dataset as it allows us to compare real disinformation networks with legitimate networks, which offers a suitable playground for determining the role that recommendation algorithms play in the formation and consolidation of disinformation networks. Statistics about the publications in the dataset can be found in Table 1, and specific properties of each network in Table 2.

While the dataset provides detailed tweet information (including its content) and user interactions, it lacks explicit user-tweet interaction data such as likes, and user-follow relationships for each temporal window. This limitation is primarily due to the strict download limits imposed by the academic API of X (Twitter at the time of the data set creation, between October 2022 and January 2023), which restricted the number of requests for accessing such interactions[2]. These constraints made it extremely challenging to recover this type of information comprehensively in large-scale datasets. To address this, we constructed a visibility graph based on retweet behavior, which serves as a proxy for user-item interactions. Although this approach is widely used in similar studies, it may not fully replicate the complexity of user behavior on the platform. Nevertheless, the large volume of data employed in this study compensates for these limitations, providing robust insights into the dynamics of recommendation algorithms in social networks.

As specified in the previous section, we divide the dataset on a fixed-time-window basis: we collect in a window W_n all the interactions that occur within 7 days, including the active users in that period and the tweets being shared (i.e., retweeted). It should be noted that this dataset does not contain information regarding a very basic and important relationship in social networks: the follower/followee relationship at every window step (only at the end of the entire period). Such a relationship is relevant because it gives an idea of the content that is *available* (visible) to a user during a specific window of time. To overcome the absence of this relationship in our dataset, we propose to compute the *visibility graph* described in the previous section for every time window, to determine the content available to each user in a specific moment.

4.2. Recommendation algorithms

We adopt a multifaceted approach to recommender systems, implementing algorithms from well-known families, each chosen for its unique strengths and ability to complement the others, while covering as much as possible the full spectrum of available techniques.

In particular, the Random (**Rnd**) approach serves as a baseline to compare against more sophisticated algorithms, as it does not implement an intelligent policy to perform content recommendation to the users in the social network. Collaborative Filtering (**CF**), specifically user-based with KNN (K-Nearest Neighbors) (Nikolakopoulos et al., 2022), taps into the wisdom of the crowd, making recommendations based on user similarity metrics. Additionally, we implemented a Word2Vec-based (**W2V**) algorithm that only uses the Word2Vec encoding of the tweets (and user preferences); and a variation (**MD-W2V**) that also leverages the use of additional user information, such as their popularity (number of followers and followees) and other pertinent characteristics, enriching the contextual backdrop against which recommendations are made. A multi-armed bandit recommender (Agrawal and Goyal, 2012; Sanz-Cruzado, Castells and López, 2019), employing Thompson Sampling (**TS**), suggests users to each user and then weights their tweets based on neighborhood visibility, enhancing the relevance of tweet recommendations based on the network's structure and the positive/negative feedback given to the agent that is in charge of creating recommendations for each user. Furthermore, we implemented another hybrid method by mixing the bandit approach with the content-based approach to compare with the individual advantages of these algorithms and analyze possible differences (**TS-W2V**).

DeepRank (Chen and Zhou, 2020) (**DR**) represents our candidate for the deep learning family, offering a hybrid approach that combines the strengths of both collaborative and content-based methods, by recommending content based on both topology (i.e., user interconnections) and content (i.e., the meaning of the posts). Neural Matrix Factorization (**NeuMF**) (He, Liao, Zhang, Nie, Hu and Chua, 2017) expands the capabilities of collaborative filtering models by replacing the traditional dot product with a deep neural network, combining linear and nonlinear user-item interactions. This approach enables the modeling of more complex interaction functions and has proven effective in tasks involving implicit feedback.

Linguistic Inquiry and Word Count (**LIWC**) (Pennebaker, Boyd, Jordan and Blackburn, 2015) CB recommendation delves into psychological and linguistic cues within user preferences relying heavily on text topics encoded in LIWC categories instead of the overall wording, since LIWC is a research tool used for text analysis to determine the percentage of words reflecting different emotions, thinking styles, social concerns, and even parts of speech (Pennebaker et al., 2015), and it has already been used in the literature for other applications (Alharthi and Inkpen, 2019; Berbatova, 2019; Yang, Nikolenko, Huang and Farseev, 2022). Finally, to exploit the social component of the network, we include two algorithms: one based on how “similar” pairs of users are and generate link recommendations based on it (Adamic and Adar, 2003) (**Friendship**), and a structure-based algorithm that provides link recommendations based on the probability of a random walk between two users (Lichtenwalter, Lussier and Chawla, 2010) (**PropFlow**). ConsisRec (**CR**) (Yang, He, Song, Liu and Tang, 2021) uses graph neural networks (GNNs) to address the problem of social inconsistency in social recommendation. This approach selects and weights consistent neighbors through a combination of sampling and attention strategies, maximizing the relevance of the aggregated information to improve the accuracy of rating predictions.

[2] This period, however, was before the acquisition of Twitter by Elon Musk, where third-party access to its API ended.

In total, we have implemented and reported results for 12 recommendation algorithms, covering the different approaches that comprehend the core *essential* components of the recommender systems that constitute the state-of-the-art in the field (Anandhan et al., 2018). While other, more complex techniques could be tested, our goal was to consider as much as possible techniques that work in isolation, to be able to understand which components of the recommendation algorithms and/or of the data impact more or less on the formation of disinformation networks. Indeed, this diversified approach allows us to explore the breadth of recommendation system strategies and to (rigorously) address the proposed research questions.

While other, more complex techniques could be tested, our goal was to focus on techniques that operate in isolation, allowing us to analyze the impact of specific components of recommendation algorithms on the formation of disinformation networks. The algorithms we selected represent the core families of recommendation strategies used in social networks. These families encompass the main approaches to recommending content based on the information available within a social network system. Although modern recommendation systems are more advanced, they primarily build upon or combine the foundational strategies we have implemented here. Conceptually, these advanced systems only refine or integrate the core approaches, typically relying on more extensive datasets and improved training methods. However, the underlying principles remain the same.

Furthermore, simulating more sophisticated algorithms, such as those used by platforms like X (Twitter), is extremely challenging due to the unavailability of their internal machine learning models or data, particularly the deep learning components critical to their functionality. Given the vast variety of recommendation algorithms in the literature, covering only deep learning-based systems would be impractical. By focusing on the essential algorithms, we aim to understand which recommendation strategies are more influential in contributing to, or mitigating, the phenomena observed in disinformation networks. This approach allows us to compare core strategies under the same conditions, helping to identify the components within these more advanced systems that are responsible for shaping the outcomes studied here.

4.3. Network analysis

To assess the characteristics of the different networks being generated by the recommendation algorithms, we specifically leverage the use of the following techniques, which aim to measure efficiency and density (to account for how quickly information is spread) together with other properties from SNA (Newman, 2010; Shao, Hui, Wang, Jiang, Flammini, Menczer and Ciampaglia, 2018).

On the one hand, we use density (Δ), a measure of edge presence compared to the maximum possible; efficiency (E), which assesses how effectively information travels across the network based on the shortest paths between node pairs; and average degree ($\overline{Deg}$), indicating the mean number of connections per node. Additionally, modularity (Q) is used to evaluate the strength of community structures by comparing actual edge distribution within communities against a random distribution model. The average clustering coefficient ($\overline{CC}$) measures the tendency of nodes to cluster, reflecting the graph's local cohesiveness. To capture node influence and relevance, we analyze the average eigenvector centrality ($\overline{EVC}$) and the PageRank ($\overline{PR}$) of the generated networks, to consider not only the number of connections but also the significance of connected nodes. These metrics collectively offer a comprehensive view of the network's topology, efficiency, and community dynamics.

Beyond standard measurements from SNA, we introduce more specific metrics to measure well-known aspects of disinformation networks, such as the capability of treating very similar topics (high resistance to topic changes), their ability to maintain a subset of users as the *core* of the disinformation network (user persistence from window to window), and to increase the relevance of these users as windows go by. In particular, we define $Topics$ as a measure of the average number of topics mentioned by the users of a (disinformation) recommendation network in a window, as disinformation networks are expected to discuss a reduced number of topics when compared to legitimate networks. More specifically, we calculate $\%TP$ as the percentage of topics in window W_{n-1} that are still mentioned in window W_n. Similarly, we calculated the user persistence ($\%UP$), as the ratio of users in the recommendation network generated by an algorithm for a window W_{n-1} that also belong to the recommendation network generated for window W_n. In order to provide a measure of the content amplification, we calculate $\overline{CA}$, as the average number of retweets from users that were recommended by the algorithm during each window. Finally, we provide another measure of content amplification, URL, which is the average number of shared URLs considering all the users that conform the (disinformation) recommendation network created by a RA for a given window.

To further analyze the impact of RAs in the generation and consolidation of disinformation networks, we employed statistical tests to provide, under a given confidence threshold, that our claims are valid. Statistical tests have been

Table 3
Average content diffusion metrics per full recommendation network of each RA. Highest and lowest values per row (metric) are highlighted in blue and red, respectively.

Metric	Rnd	LIWC	W2V	MD-W2V	CF	TS	TS-W2V	DR	Friendship	PropFlow	NeuMF	ConsisRec
$\overline{d_r}$	6	4	5	2	5	2	2	5	2	2	4	3
$\overline{DLL}$	3225	1290	3211	3159	2515	2488	2468	2376	3219	3232	2378	2590
URL_{EC}	15582	12779	14146	10541	15369	10243	10302	12563	11759	13347	13820	11295
$Post_{EC}$	27215	21725	24922	20123	26882	16086	16314	20896	19486	24149	24813	20045
SDI	0.683	0.819	0.749	1.813	0.710	1.156	1.118	0.704	1.701	1.293	0.850	1.920

widely applied in the context of SNA under similar circumstances and objectives (Vieira, Jerônimo, Campelo and Marinho, 2020; Sitaula, Mohan, Grygiel, Zhou and Zafarani, 2020; Morais and Cruz, 2020). In this work, we use the Mann-Whitney U test on three of the most relevant network metrics for disinformation networks: efficiency, density, and modularity (Newman, 2010). With these tests, we seek to understand whether the hypotheses related to the corresponding network metric are statistically significant. More specifically, our null hypothesis H_0 is that there is no significant difference between the network metric (efficiency/density/modularity) distribution generated by one algorithm that enhances the creation and consolidation of disinformation networks and the one generated by another algorithm (we would run these tests in a pairwise fashion). The alternative hypothesis, H_1, is, therefore exactly the opposite. In our experiments, we perform this test using 95% confidence ($\alpha = 0.05$).

Additionally, to study content dynamics from the disinformation sub-part of a recommendation network to its legitimate sub-part, we propose the following metrics aimed at reliably measuring these content dynamics. First, we calculated the average distance from the journalism network to the disinformation network, d_r. Similarly, we also computed the number of directional edges that interconnect journalists to disinformation actors. This metric, that we called *Disinformation-to-Legitimate Linkage* ($\overline{DLL}$), provides an intuition of which algorithms bolster the propagation of disinformation through the legitimate network. Furthermore, we also computed URL_{EC} and $Post_{EC}$, which refer to the URL Exposure Count and the Tweet Exposure Count from journalists to (the content of) disinformation actors. Indeed, these two measures complement the previous ones in the intuition of which algorithms favor the expansion of disinformation through legitimate networks. Last, we computed the Structural Diversity Index (*SDI*) (Musso and Helbing, 2024). Such index is used to understand the diversity in a social network from a topology-based perspective. In this case, it allows us to understand whether an algorithm favors the creation of more diverse ecosystems (shared by both disinformation actors and journalists).

5. Results

Here we present the results obtained according to the methodology introduced in Section 3 using the settings described in Section 4.

To provide a clear structure, the results are organized into three main parts, corresponding to our research questions: the formation of content networks (Section 5.1), the impact on disinformation dynamics (Section 5.2), and the relation between accuracy and disinformation propagation (Section 5.3).

5.1. Formation of content networks through RAs

To address **RQ1**, we analyze how different recommendation algorithms generate networks of content diffusion, focusing on the structural properties and information flow between disinformation and legitimate actors. In this experiment, we use the recommendation algorithms and their content recommendations to trace and quantify the flow of information from users belonging to the disinformation network to users belonging to the legitimate network (journalists).

Our analysis is focused on identifying patterns and pathways through which disinformation infiltrates and propagates through these legitimate networks. This exploration is crucial for understanding how disinformation gains credibility and spreads across different segments of the public discourse, thereby influencing societal perceptions and behavior. Considering the content diffusion metrics presented in Section 4.3, we obtain the results shown in Table 3.

Considering these results, we observe how CF stands out as one of the best algorithms for enhancing content dynamics from the disinformation network to the journalism network, as it creates a significant number of links (2515) from journalists to disinformation actors (journalists receive content recommendations from disinformation actors),

Table 4
Network-based metrics computed on recommendation networks restricted to disinformation nodes.

Metric	Rnd	LIWC	W2V	MD-W2V	CF	TS	TS-W2V	DR	Friendship	PropFlow	NeuMF	ConsisRec
Nodes	416	408	416	407	357	361	360	338	402	404	390	400
Edges	875	648	838	383	805	541	551	758	427	469	720	580
$\overline{Deg}$	8.430	3.440	8.412	8.495	7.854	7.714	7.670	7.837	8.700	8.738	7.965	7.850
D	13	8	11	3	12	3	4	10	3	3	9	6
E	0.059	0.018	0.047	0.025	0.061	0.026	0.026	0.055	0.030	0.020	0.045	0.035
Δ	0.005	0.004	0.005	0.002	0.006	0.004	0.004	0.007	0.004	0.003	0.006	0.005
Q	0.829	0.846	0.858	0.971	0.801	0.903	0.898	0.793	0.942	0.947	0.860	0.900
$\overline{CC}$	0.017	0.018	0.020	0.004	0.021	0.021	0.022	0.024	0.007	0.007	0.015	0.014
$\overline{EVC}$	0.002	0.002	0.002	0.002	0.003	0.003	0.003	0.003	0.003	0.002	0.003	0.002
$\overline{PR}$	$25e^{-4}$	$26e^{-4}$	$25e^{-4}$	$26e^{-4}$	$31e^{-4}$	$30e^{-4}$	$30e^{-4}$	$33e^{-4}$	$26e^{-4}$	$26e^{-4}$	$28e^{-4}$	$27e^{-4}$
$\overline{IncPR_w}$	$2.83e^{-5}$	$1.40e^{-5}$	$1.11e^{-5}$	$4.24e^{-6}$	$4.73e^{-5}$	$3.33e^{-5}$	$2.65e^{-5}$	$4.22e^{-5}$	$1.69e^{-5}$	$1.06e^{-5}$	$3.12e^{-5}$	$2.45e^{-5}$

Table 5
Content-based metrics computed on recommendation networks restricted to disinformation nodes.

Metric	Rnd	LIWC	W2V	MD-W2V	CF	TS	TS-W2V	DR	Friendship	PropFlow	NeuMF	ConsisRec
Topics	1361	1268	1315	1207	1348	1119	1121	1204	1190	1302	1280	1355
%TP	0.575	0.572	0.578	0.570	0.574	0.565	0.567	0.572	0.569	0.574	0.572	0.567
URL	49.901	48.857	46.421	43.495	53.626	40.939	41.070	52.178	50.036	51.026	51.167	41.086
%UP	0.810	0.770	0.819	0.758	0.795	0.776	0.777	0.770	0.763	0.794	0.785	0.770
$\overline{CA}$	41.452	36.548	38.302	42.349	44.814	30.440	30.657	38.345	41.220	49.839	39.031	33.189

while also amplifying significantly the content of the disinformation network that reaches the legitimate network. Compared to the Random baseline, we observe how CF creates networks with a similar influence of the disinformation network on the legitimate network. It is worth noting that, due to Rnd's behavior, the resulting networks maximize the impact of disinformation actors on legitimate users, as random links are created between users that would never be connected in a real scenario. This is coherent with other conclusions derived in our study, as CF creates networks with properties that resemble the network properties of a disinformation network (by enhancing efficiency and density while moderating fragmentation).

On the other hand, algorithms like DR, W2V, and NeuMF also create networks that maximize the impact of the disinformation network on the legitimate network. NeuMF, by combining collaborative filtering and deep learning, demonstrates a strong capability to identify user-item interactions, resulting in dense and efficient networks that amplify disinformation dynamics. Similarly, all these algorithms that seem to enhance the dissemination of disinformation content into the legitimate network are characterized by the creation of networks with a small SDI, which also characterizes disinformation networks, as it implies that information can readily travel through the network without requiring traveling long distances to reach a specific part of it.

ConsisRec, in contrast, tends to foster the creation of structurally diverse networks by prioritizing heterogeneous user connections. This structural diversity reduces the likelihood of dense, insular clusters forming, thereby limiting the impact of disinformation dissemination. At the same time, we also observe that specific algorithms are not useful for content dissemination from the disinformation network to the legitimate network, thus making it hard for creating more diverse networks. For instance, LIWC tends to create networks with a very small number of connections from journalists to disinformation actors, thus reducing the capability of sharing content from the disinformation network with the legitimate network.

Similarly, TS, TS-W2V, and MD-W2V appear as approaches with reduced content dynamics from the disinformation to the legitimate network. This observation can be derived from the reduced content amplification that is performed in this network (low values of URL_{EC} and $Post_{EC}$ metrics) and also from the high SDI metric values that suggest the creation of networks that foster structural diversity, as opposed to the algorithms that seemed to facilitate content diffusion into the legitimate network (CF, W2V, DR, NeuMF, and Rnd).

In summary, we observe that different algorithms lead to the creation of different recommendation networks, which have a significant impact on how disinformation can potentially spread from the disinformation network to the legitimate network.

These results confirm that different recommendation strategies generate distinct diffusion networks, validating that algorithmic choices significantly influence content flow patterns (answering **RQ1**).

5.2. Impact on disinformation of RAs

To address **RQ2**, we restrict the analysis to recommendations among disinformation actors, aiming to evaluate how algorithms contribute to the formation and reinforcement of disinformation networks. In this experiment, we limit the recommendations to users identified as disinformation actors, analyzing the resulting networks and dynamics. Thus, Tables 4 and 5 present the average values of the network and content metrics (respectively) associated to the recommendation networks generated by each algorithm.

Considering these results, we observe there are significant differences between the algorithms. For instance, the *Nodes* metric varies between 338 (DR) and 416 (Rnd), which makes a difference of almost 100 users. This difference can also be noticed in other metrics, such as the diameter, which takes values varying between 3 (MD-W2V, TS) and 13 (Rnd), or even the edges metric, which represents the recommendations between different pairs of users. Hence, we observe that, in general, there are very different values for most of the metrics in both tables.

As a consequence, we conclude that different approaches produce networks with different values for the analyzed properties and metrics. Some approaches tend to generate more recommendations between different users, such as Rnd, which explains the high number of nodes, edges, and degree. However, other algorithms tend to focus on alternative aspects; for example, W2V leads to the creation of networks with fewer nodes, edges, and degree, as a result of less exploration and more exploitation in its results when compared to other algorithms, in particular, Rnd.

Moreover, we observe that certain algorithms facilitate the generation of disinformation networks. Disinformation networks form when specific topics and narratives attract users who actively seek to promote them. These networks become densely connected as users engage extensively around these narratives, generating highly concentrated clusters. There are clear algorithmic factors that facilitate these processes. Broadly speaking, our analysis indicates that the consolidation of disinformation networks is influenced by two key factors: the recommendation of content tied to the specific narrative, and the recommendation of highly active contacts centered on these topics.

Different algorithms exhibit varying tendencies to facilitate these processes. Algorithms based on textual recommendations, such as W2V in our case, allow users to quickly access resonant narratives. Once established, collaborative filtering algorithms have the potential to further consolidate these networks by reinforcing connections among users involved in these narratives. Similarly, NeuMF creates dense networks with high efficiency and relatively low modularity, resembling the behavior of DR and W2V. The deep learning component in NeuMF enhances its precision, allowing it to capture complex user-item interactions, which results in effective but potentially risky recommendations when narratives align with disinformation. Its lower SDI compared to ConsisRec indicates a structure that can facilitate rapid information flow, making it susceptible to amplifying disinformation when improperly managed. In contrast to these algorithms, ConsisRec demonstrates a strong focus on structural diversity, as reflected in its high SDI value (1.920, the highest among all algorithms). By fostering heterogeneous connections, ConsisRec significantly reduces the likelihood of forming dense, insular clusters, which are characteristic of disinformation networks. This approach aligns with the mitigation of disinformation spread by encouraging network structures that are less prone to concentrated amplification.

Based on these results, DR stands as the algorithm that contributes the most to the generation of disinformation networks. First, we observe that DR tends to create recommendation networks where fewer users are recommended by the algorithm, because it tends to focus always on the same kind of users. In particular, those users that make the algorithm to lower the loss function, which are also users with really high relevance in the generated recommendation network (highest average network PageRank) and users that tend to increase their relevance as time windows go by. We also observe how this algorithm creates a decent number of edges between the nodes, i.e., the algorithm provides a moderately high number of connections between a relatively small number of users, which increases the density (Δ) of the recommendation networks (known to be a key factor in disinformation networks generation (Muñoz et al., 2024)). Furthermore, DR generates very efficient networks (third algorithm with more average efficiency, E), while creating networks with less modularity and larger clustering coefficient. Indeed, the last observation allows us to understand that DR tends to create networks where information spreads rapidly, as there are less apparent communities with higher cohesion, when compared to the (disinformation) recommendation networks generated by other approaches. Moreover, we also observe that DR is one of the algorithms that creates networks with less topics, while also being one of the algorithms that tends to keep more topics from one window to another ($\%TP = 0.572$).

In summary, DR is an algorithm that strongly contributes to the generation of disinformation networks, as it creates very efficient and dense recommendation networks, where information spreads rapidly, due to a smaller network fragmentation and more cohesion between their users, whose influence tends to evolve both positively and more quickly over time when compared to other approaches. Therefore, DeepRank is an algorithm that significantly contributes to

Table 6
Average accuracy metrics per recommendation network generated by each RA (best values in bold).

Metric	Rnd	LIWC	W2V	MD-W2V	CF	TS	TS-W2V	DR	Friendship	PropFlow	NeuMF	ConsisRec
Accuracy	0.037	0.178	**0.211**	0.138	0.043	0.081	0.105	0.083	0.038	0.052	0.198	0.203
Recall	0.001	0.002	**0.003**	0.002	0.001	0.001	0.001	0.001	0.001	0.001	**0.003**	**0.003**
F1	0.001	0.004	**0.005**	0.004	0.001	0.002	0.002	0.002	0.001	0.001	**0.005**	**0.005**
F2	0.001	0.003	**0.004**	0.002	0.001	0.001	0.001	0.001	0.001	0.001	**0.004**	**0.004**

the formation of disinformation networks. This is largely due to its deep learning foundation, which combines a strong emphasis on textual content with the user connections within the network. This combination makes DR particularly susceptible to poisoning or shilling attacks, where the introduction of false information can distort the recommendation process and subsequently influence the overall information dynamics. Similar to the Word2Vec algorithm, which also scores highly in this regard, DR's sensitivity to textual inputs allows for the easy insertion of slogans or false news into the system. Disinformative users are able to connect and reinforce each other's content, leading to dense and highly cohesive networks that facilitate the rapid spread of disinformation. This phenomenon is consistent with findings in the literature, where algorithms that prioritize both content and network structure tend to amplify disinformation under certain conditions (Nguyen et al., 2024).

A similar observation, although to a smaller extent, is found for CF, since it tends to recommend a small number of users, with many connections, very high network efficiency, density, and clustering coefficient with a low modularity, suggesting the generation of (disinformation) recommendation networks where information spreads swiftly and efficiently. However, according to the pairwise Mann-Whitney U test explained in Section 4.3, with 95% confidence the null hypothesis (there is no significant difference between two distributions) was rejected, for the three most relevant metrics that characterize disinformation networks (efficiency, density, and modularity); therefore, the actual produced distributions are different, even though their overall behavior might be similar.

Thus, our findings demonstrate that certain recommendation algorithms -— especially those leveraging textual similarity and deep learning -— are more prone to facilitate the creation of disinformation networks, providing an empirical answer to **RQ2**.

5.3. Relation between accuracy and disinformation propagation of RAs

To address **RQ3**, we investigate here the relationship between recommendation accuracy and the structural properties associated with disinformation propagation. With this goal in mind, we analyze the impact of certain network properties and metrics known to affect disinformation (as already described in previous sections) on the accuracy of the algorithms. Hence, we aim to study whether certain kinds of networks, generated by specific approaches, favor accuracy, or there is no correlation between them.

It is important to note that the relatively low accuracy values reported in our simulations are consistent with expectations, given the inherent limitations of the dataset and the experimental conditions. First, our models do not have access to the complete dataset representing the full Twitter/X social network, nor do we have insights into its internal architecture, thereby restricting the global visibility of user interactions. This constraint, often encountered in the literature, is particularly relevant in the context of small-world networks (Amaral, Scala, Barthelemy and Stanley, 2000), where the absence of the entire network structure can significantly impact the performance of recommendation systems. Additionally, our evaluation focuses exclusively on retweet behavior, which constitutes only a fraction of the broader range of user activities within the platform, though it is the most relevant fraction when it comes to the spread of information and the generation and consolidation of user networks, as the action of retweet is a clear indication of endorsement. Therefore, the observed accuracy is in line with prior research, where similar studies using partial datasets of real-world social networks have reported comparable, or even lower, accuracy levels under such conditions.

Also, low accuracy is just a result of the users not retweeting every single publication they get recommended but, instead, just restrict their decision to a few tweets that really represent their perspective and, most of the times, due to anti-homophily (Macskassy and Michelson, 2011). Our study is fully justified in comparing different algorithmic approaches under identical conditions, despite the accuracy values, as they are relative. Thus our study allows for a fair evaluation of their relative performance. Our primary aim is to assess the effect of essential algorithmic strategies in environments where data and network visibility are limited, providing valuable insights into their efficacy under such constraints.

For this, Table 6 shows the average accuracy results obtained for each recommendation algorithm across all temporal windows. We observe that W2V is highlighted as the algorithm with the highest values for the reported accuracy metrics. In other words, W2V is the algorithm that is more capable of delivering to each user the content that such user expects to be recommended (content that follows its preferences). Naturally, this observation is logical, as W2V uses NLP techniques to encode the content of each tweet and the preferences of the user. Similarly, algorithms following a similar approach, like LIWC or MD-W2V, are capable of achieving high accuracy values, although both of them are still below W2V, which seems to solve the content recommendation problem to users in the best possible way, in comparison with the other algorithms. It is noteworthy that some algorithms do not significantly improve the accuracy results with respect to Rnd, in particular CF and Friendship.

When considering NeuMF, we observe that it achieves relatively high accuracy values, close to those of LIWC and MD-W2V. This is expected, given NeuMF's deep learning foundation, which effectively combines matrix factorization and neural networks to model user-item interactions. However, its accuracy is still below W2V, indicating that while NeuMF captures complex patterns effectively, it does not outperform algorithms explicitly designed for textual content.

ConsisRec, on the other hand, achieves a balance between accuracy and structural diversity. Its accuracy is slightly below W2V but comparable to NeuMF, reflecting its focus on ensuring diversity in recommendations without heavily compromising precision. This balance suggests that ConsisRec is a promising approach for reducing disinformation network formation, as it avoids reinforcing insular clusters while maintaining competitive accuracy levels.

When considering all the presented results in combination, we conclude that the algorithms that seemed to facilitate the generation and consolidation of disinformation networks (DR and W2V) stand as algorithms with high values of network efficiency and medium-high accuracy values. More specifically, W2V is the algorithm that achieves extremely high accuracy values, while keeping a high network efficiency. While W2V duplicates the accuracy value of DR (on average), DR also achieves extremely high network efficiencies with accuracy at a reasonable level. Thus, these algorithms are quite *dangerous*, as they improve accuracy while enhancing the generation and consolidation of disinformation networks, through boosting the network properties that resemble those networks. When considering other algorithms, like LIWC and MD-W2V, these approaches achieve high relatively accuracy values, however, the generated networks are fragmented, compromising their efficiency. If we analyze CF, it might not be seen as a *threat* (compared against W2V and DR), since it compromises accuracy to a large extent, even though it generates extremely efficient networks. Other algorithms like Friendship and PropFlow seem to generate inefficient networks, while having accuracy values that are lower than others.

In summary, the results indicate a partial tradeoff between recommendation accuracy and the propensity for disinformation amplification: algorithms achieving higher accuracy often also create network structures more susceptible to disinformation spread (thus answering **RQ3**).

6. Conclusions

Our findings highlight the particularly concerning role for CB recommendation algorithms, especially those adept at processing and amplifying certain features like hashtags and URLs, exemplified specifically in the Word2Vec approach. While achieving high levels of recommendation accuracy, this algorithm family significantly bolsters the formation and perpetuation of disinformation networks, posing a substantial threat to the integrity of the online conversation.

In this section, we summarize the outcomes of our study to answer all the RQs we pose, discuss the limitations of the work we have presented here, and outline possible future work on mitigating the adverse effects of recommendation in social media.

6.1. Discussion of research questions

RQ1: Do different recommendation algorithms generate different networks from a content diffusion perspective?

We have observed that different algorithms lead to the creation of different recommendation networks. While some methods create networks more similar to each other (e.g., CF creates very efficient (highest E metric), highly dense networks that experience little fragmentation and are very cohesive (relatively low $\overline{CC}$); while DR produces similar networks, TS and TS-W2V generate inefficient (higher E), sparse, and centralized networks with small diameter and high average degree), in general, the characteristics and inherent properties of such networks are different enough to produce non-comparable dynamics and capabilities for content diffusion among algorithms.

The answer to this research question (i.e., each algorithm generates different networks to spread content) is in agreement with previous observations found in the literature, although mostly focused on misinformation rather than disinformation as in our study (Tommasel and Menczer, 2022; Pathak et al., 2023; Fernández et al., 2024), even though in those works a narrower set of algorithms was tested in smaller datasets (since they were limited to explicit fact-checked posts within the social network).

In particular, NeuMF produces networks with a moderate number of edges (720) and density (0.006), sitting between CF and DR in terms of cohesiveness and efficiency. While it avoids the extreme fragmentation seen in some algorithms like TS (with a Q value of 0.903), it also does not achieve the high density of DR (whose D value is 10). ConsisRec, on the other hand, generates networks with reduced density and higher structural diversity, reflected in its relatively higher diameter (its D value is 6) and lower clustering coefficient (value of 0.014). This behavior makes ConsisRec more resistant to the formation of dense clusters typical of disinformation networks, while maintaining a balanced level of connectivity.

RQ2: What is the impact of recommendation algorithms in disinformation propagation?

Our research confirms that specific recommendation algorithms, particularly Word2Vec and its derivatives, play a critical role in fostering the formation of dense, efficient, and moderately modular disinformation networks. These algorithms, through their high precision in recommending content (best one according to Table 6), enable the rapid dissemination of narratives that align with the interests of users promoting disinformation. This creates a paradox inherent in the design of social media platforms: while these algorithms enhance user engagement and platform utility, they simultaneously facilitate the spread of disinformation by promoting content that aligns with specific, often harmful, narratives.

Disinformation networks typically emerge when topics and narratives of interest attract users motivated to disseminate such content. As users connect and engage around these narratives, they generate highly concentrated and dense clusters. From our analysis, we observe that the consolidation of these networks is modulated by two primary factors: first, the recommendation of content directly tied to the disinformation narrative; and second, the recommendation of highly active users who are heavily involved in promoting this content. Algorithms that recommend textually similar content—such as Word2Vec—allow users to quickly access resonant narratives, thus accelerating the process of network formation. Once these narratives are established, collaborative filtering algorithms further solidify these networks by reinforcing connections between users who engage with similar content, effectively consolidating the disinformation networks.

NeuMF, like DR, shows a balance between network cohesion and efficiency ($E = 0.045$ whereas $\overline{CC} = 0.015$), enabling the propagation of disinformation while maintaining moderate fragmentation. This makes it similar in behavior to algorithms that prioritize user preferences but do not explicitly focus on diversity or disinformation mitigation. ConsisRec, with its emphasis on structural diversity, creates less cohesive and denser networks, reducing the likelihood of dense disinformation clusters while still enabling moderate levels of content diffusion.

These results, while not directly comparable with previous work because of their focus on misinformation, achieve similar conclusions in terms of this and the next research question. Namely, that the most accurate recommendation algorithms tend to amplify misinformation (Pathak et al., 2023; Fernández et al., 2024).

RQ3: Is there any relation between disinformation propagation and accuracy of recommendation algorithms? We have observed that there exist recommendation algorithms that facilitate the generation and consolidation of disinformation networks while keeping decent content recommendation accuracy levels, as is the case with DR and W2V. We also observed that different algorithms create recommendation networks whose content dynamics are highly influenced by the different shapes of such networks. While all of them try to maximize recommendation accuracy, the networks they generate can be very different in terms of their shape and properties, as they can have a varying range of densities and efficiencies (see Table 4), or be highly or poorly fragmented. As a result, not all the algorithms create optimal networks for the creation and consolidation of (dis)information networks and content diffusion.

NeuMF achieves good accuracy values while maintaining moderately efficient networks, positioning it closer to DR and W2V in terms of its potential to consolidate disinformation. ConsisRec, however, trades some accuracy for structural diversity, which disrupts the formation of dense disinformation clusters. While this comes at the cost of slightly reduced recommendation precision, the resulting network dynamics align more closely with the goals of mitigating disinformation proliferation.

6.2. The role of core recommendation strategies

Our analysis reveals that the susceptibility of a recommendation system to foster disinformation networks does not solely depend on the complexity or novelty of the algorithm used. Instead, it fundamentally stems from the basic recommendation strategies—or components—embedded within the system.

Specifically, our findings highlight that:

- **Textual similarity exploitation**, as embodied by Word2Vec-based methods, enables users to quickly access resonant narratives, accelerating the clustering of like-minded individuals and facilitating the formation of dense disinformation communities.
- **Collaborative filtering reinforcement**, characteristic of models like DeepRank and NeuMF, further consolidates these clusters by strengthening the connections between users who engage with similar content, promoting echo chambers and reducing diversity.
- **Efficiency optimization**, pursued by various approaches, fosters network structures that minimize fragmentation and maximize rapid information spread—properties that are advantageous for disinformation dissemination.

Modern recommendation systems typically combine these basic strategies—textual matching, collaborative affinity, and optimization for engagement—in increasingly sophisticated ways. However, our study shows that each component, independently, can contribute to the emergence and reinforcement of disinformation networks.

By isolating the effects of these core mechanisms, we provide critical insights into which aspects of recommendation logic require careful management. Future systems must reconsider how they integrate and balance these components, prioritizing not only accuracy and user engagement but also the promotion of healthy, diverse, and resilient information ecosystems.

6.3. Limitations

Acknowledging the limitations of this research, it is pertinent to note that while the study incorporated state-of-the-art recommendation systems from the four key families—content-based, topology-based, deep learning, and reinforcement learning—the vast and evolving landscape of recommender systems includes more complex algorithms and numerous variants aimed at enhancing accuracy and performance. Due to the practical constraints of this study, it was not feasible to explore every such variation, particularly those refining existing models rather than offering fundamentally new perspectives on network formation.

This choice, while enabling a focused examination of foundational methodologies, inherently limits the breadth of algorithmic nuances considered. For instance, in this work, reinforcement learning-based methods such as multi-armed bandits do not achieve particularly high recommendation accuracy or network consolidation. This outcome may partly result from the simplicity of the implemented models, neglecting more advanced strategies (Silva, Werneck, Silva, Pereira and Rocha, 2022).

Our research also has limitations related to data access. We could not access the full network dataset but only a portion of it. Therefore, we rely on comparisons between disinformation and legitimate networks generated through simulated recommendations, although full dataset availability could enable more comprehensive analyses. Furthermore, the actual recommendation algorithms employed by the platform under study (i.e., X/Twitter) are unknown, and would have influenced user interactions and the visibility graph used for our experiments.

Additionally, the dataset analyzed corresponds to a specific segment of the Spanish-speaking Twitter network, as introduced in prior work. While disinformation dynamics are often universal in online social networks (Villa et al., 2021), there could be socio-cultural or topical biases that limit the generalizability of our findings (Ramponi, Brambilla, Ceri, Daniel and Giovanni, 2020).

Finally, the absence of rich user-item interaction data (e.g., likes, timestamps) and limited temporal granularity in follower relationships constrained the modeling possibilities. Although the visibility graph served as a viable workaround, future studies could benefit from datasets offering richer interaction and metadata.

It is also essential to acknowledge that disinformation dissemination is driven by multiple factors beyond algorithmic recommendations, including user psychology, topic relevance, and the presence of coordinated campaigns (e.g., troll farms). Viral content often gains traction due to timeliness or community mobilization, independent of the underlying recommendation mechanisms. Thus, although recommendation systems influence content distribution, broader societal and contextual factors equally shape disinformation networks.

6.4. Future Work

Looking ahead, several concrete directions emerge to extend this research. First, expanding the analysis to additional social networks—such as Facebook, TikTok, or Reddit—with different content mobilization dynamics could offer comparative insights into platform-specific vulnerabilities.

Second, extending experiments to datasets from diverse cultural, linguistic, and geographic contexts (Corradini, 2024) would help assess the robustness and universality of our findings regarding disinformation propagation.

Third, incorporating richer user-item interaction data, when available, could allow the implementation of more sophisticated recommendation algorithms, such as sequential recommenders (Li, Xie, Zhang, Wang, Zou, Li, Luo and Li, 2024) or emotion-aware models that adapt to user sentiment (Liu, Xu, Zhang and Jiang, 2023). Similarly, approaches integrating social feedback loops (Gan, Wang, Yi and Gu, 2024) could be explored.

Fourth, and most importantly, an immediate avenue involves developing mitigation strategies. Future work could design recommendation algorithms that explicitly balance accuracy and resilience against disinformation amplification. Potential approaches include introducing structural or linguistic diversity in the recommended content (Geng, He, Liang, Niu, Liu and He, 2023) to hinder echo chamber formation.

Finally, incorporating adaptive rule-based mechanisms—such as adjusting recommendation strategies based on observed network polarization metrics—could make the simulation framework more dynamic and reflective of real-world algorithmic interventions, addressing one of the limitations identified in this work.

6.5. Concluding recommendations

Our study reveals that recommendation algorithms used in online social networks play a pivotal role in shaping the structure of social networks, and that this structure significantly influences the formation of disinformation networks and the propagation of the information they disseminate.

In this study, we identified that algorithms like Word2Vec and DeepRank, which excel in accuracy by aligning recommendations with user preferences, also amplify the formation of dense, efficient, and cohesive disinformation networks. These findings highlight a critical trade-off in current recommendation systems: optimizing engagement often comes at the cost of facilitating the spread of harmful narratives.

To mitigate this effect, future algorithms must put more focus in prioritizing structural and textual diversity balanced over accuracy. Introducing mechanisms that recommend connections across different user clusters or content with varied perspectives can naturally prevent the consolidation of disinformation networks. Furthermore, harm-aware recommender systems considering dynamic adjustments based on real-time metrics, such as network density or the spread of specific narratives, could help reduce the propagation of harmful content without sacrificing recommendation quality.

Ultimately, the insights from this research suggest that balancing accuracy with the containment of disinformation is feasible. By integrating strategies that actively disrupt the formation of cohesive disinformation clusters, platforms can design recommendation systems that maintain user satisfaction while limiting the unintended amplification of harmful narratives.

CRediT authorship contribution statement

Pau Muñoz: Conceptualization, Methodology, Software, Investigation, Visualization, Writing - Original draft preparation. **Raúl Barba-Rojas:** Methodology, Software, Writing - Original draft preparation. **Fernando Díez:** Investigation, Supervision, Writing - Original draft preparation. **Alejandro Bellogín:** Investigation, Supervision, Writing - Original draft preparation.

References

Adamic, L.A., Adar, E., 2003. Friends and neighbors on the web. Soc. Networks 25, 211–230. URL: https://doi.org/10.1016/S0378-8733(03)00009-1, doi:10.1016/S0378-8733(03)00009-1.

Afsar, M.M., Crump, T., Far, B.H., 2023. Reinforcement learning based recommender systems: A survey. ACM Comput. Surv. 55, 145:1–145:38. URL: https://doi.org/10.1145/3543846, doi:10.1145/3543846.

Agrawal, S., Goyal, N., 2012. Analysis of thompson sampling for the multi-armed bandit problem, in: Mannor, S., Srebro, N., Williamson, R.C. (Eds.), COLT 2012 - The 25th Annual Conference on Learning Theory, June 25-27, 2012, Edinburgh, Scotland, JMLR.org. pp. 39.1–39.26. URL: http://proceedings.mlr.press/v23/agrawal12/agrawal12.pdf.

Aïmeur, E., Amri, S., Brassard, G., 2023. Fake news, disinformation and misinformation in social media: a review. Soc. Netw. Anal. Min. 13, 30. URL: https://doi.org/10.1007/s13278-023-01028-5, doi:10.1007/S13278-023-01028-5.

Alharthi, H., Inkpen, D., 2019. Study of linguistic features incorporated in a literary book recommender system, in: Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing, Association for Computing Machinery, New York, NY, USA. p. 1027–1034. URL: https://doi.org/10.1145/3297280.3297382, doi:10.1145/3297280.3297382.

Amaral, L.A.N., Scala, A., Barthelemy, M., Stanley, H.E., 2000. Classes of small-world networks. Proceedings of the National Academy of Sciences 97, 11149–11152.

Anandhan, A., Shuib, N.L.M., Ismail, M.A., Mujtaba, G., 2018. Social media recommender systems: Review and open research issues. IEEE Access 6, 15608–15628. URL: https://doi.org/10.1109/ACCESS.2018.2810062, doi:10.1109/ACCESS.2018.2810062.

Barash, V., Fink, C., Cameron, C.J., Schmidt, A.C., Dong, W., Macy, M., Kelly, J., Deshpande, A., 2020. A twitter social contagion monitor, in: Atzmüller, M., Coscia, M., Missaoui, R. (Eds.), IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2020, The Hague, Netherlands, December 7-10, 2020, IEEE. pp. 523–530. URL: https://doi.org/10.1109/ASONAM49781.2020.9381313, doi:10.1109/ASONAM49781.2020.9381313.

Berbatova, M., 2019. Overview on NLP techniques for content-based recommender systems for books, in: Kovatchev, V., Temnikova, I.P., Sandrih, B., Nikolova, I. (Eds.), Proceedings of the Student Research Workshop Associated with RANLP 2019, Varna, Bulgaria, September 2-4, 2019, INCOMA Ltd.. pp. 55–61. URL: https://doi.org/10.26615/issn.2603-2821.2019_009, doi:10.26615/ISSN.2603-2821.2019_009.

Beskow, D.M., Carley, K.M., 2019. Agent based simulation of bot disinformation maneuvers in twitter, in: 2019 Winter Simulation Conference (WSC), IEEE. pp. 854–865.

Van den Bogaert, L., Geerts, D., Harambam, J., 2022. Putting a human face on the algorithm: Co-designing recommender personae to democratize news recommender systems. Digital Journalism , 1–21.

Bountouridis, D., Harambam, J., Makhortykh, M., Marrero, M., Tintarev, N., Hauff, C., 2019. SIREN: A simulation framework for understanding the effects of recommender systems in online news environments, in: danah boyd, Morgenstern, J.H. (Eds.), Proceedings of the Conference on Fairness, Accountability, and Transparency, FAT* 2019, Atlanta, GA, USA, January 29-31, 2019, ACM. pp. 150–159. URL: https://doi.org/10.1145/3287560.3287583, doi:10.1145/3287560.3287583.

Cantador, I., Cortés-Cediel, M.E., Fernández, M., 2020. Exploiting open data to analyze discussion and controversy in online citizen participation. Inf. Process. Manag. 57, 102301. URL: https://doi.org/10.1016/j.ipm.2020.102301, doi:10.1016/J.IPM.2020.102301.

Chen, M., Zhou, X., 2020. Deeprank: Learning to rank with neural networks for recommendation. Knowledge-Based Systems 209, 106478.

Cheng, J., Adamic, L.A., Dow, P.A., Kleinberg, J., Leskovec, J., 2014. Can cascades be predicted?, in: Proceedings of the 23rd international conference on World wide web, pp. 925–936.

Chuklin, A., Markov, I., de Rijke, M., 2015. Click Models for Web Search. Synthesis Lectures on Information Concepts, Retrieval, and Services, Morgan & Claypool Publishers. URL: https://doi.org/10.2200/S00654ED1V01Y201507ICR043, doi:10.2200/S00654ED1V01Y201507ICR043.

Corradini, E., 2024. Deconstructing cultural appropriation in online communities: A multilayer network analysis approach. Inf. Process. Manag. 61, 103662. URL: https://doi.org/10.1016/j.ipm.2024.103662, doi:10.1016/J.IPM.2024.103662.

Epstein, R., Robertson, R.E., 2015. The search engine manipulation effect (seme) and its possible impact on the outcomes of elections. Proceedings of the National Academy of Sciences 112, E4512–E4521.

Fernández, M., Bellogín, A., Cantador, I., 2024. Analysing the effect of recommendation algorithms on the spread of misinformation, in: Aiello, L.M., Mejova, Y., Seneviratne, O., Sun, J., Kaiser, S., Staab, S. (Eds.), Proceedings of the 16th ACM Web Science Conference, WEBSCI 2024, Stuttgart, Germany, May 21-24, 2024, ACM. pp. 159–169. URL: https://doi.org/10.1145/3614419.3644003, doi:10.1145/3614419.3644003.

Ferro, N., Fuhr, N., Grefenstette, G., Konstan, J.A., Castells, P., Daly, E.M., Declerck, T., Ekstrand, M.D., Geyer, W., Gonzalo, J., Kuflik, T., Lindén, K., Magnini, B., Nie, J., Perego, R., Shapira, B., Soboroff, I., Tintarev, N., Verspoor, K., Willemsen, M.C., Zobel, J., 2018. From evaluating to forecasting performance: How to turn information retrieval, natural language processing and recommender systems into predictive sciences (dagstuhl perspectives workshop 17442). Dagstuhl Manifestos 7, 96–139. URL: https://doi.org/10.4230/DagMan.7.1.96, doi:10.4230/DAGMAN.7.1.96.

Gan, M., Wang, C., Yi, L., Gu, H., 2024. Exploiting dynamic social feedback for session-based recommendation. Inf. Process. Manag. 61, 103632. URL: https://doi.org/10.1016/j.ipm.2023.103632, doi:10.1016/J.IPM.2023.103632.

Geng, S., He, X., Liang, G., Niu, B., Liu, S., He, Y., 2023. Accuracy-diversity optimization in personalized recommender system via trajectory reinforcement based bacterial colony optimization. Inf. Process. Manag. 60, 103205. URL: https://doi.org/10.1016/j.ipm.2022.103205, doi:10.1016/J.IPM.2022.103205.

Guess, A.M., Lyons, B.A., 2020. Misinformation, Disinformation, and Online Propaganda. Cambridge University Press. SSRC Anxieties of Democracy, p. 10–33.

Hanna, R., Rohm, A., Crittenden, V.L., 2011. We're all connected: The power of the social media ecosystem. Business Horizons 54, 265–273. URL: https://www.sciencedirect.com/science/article/pii/S0007681311000243, doi:https://doi.org/10.1016/j.bushor.2011.01.007. sPECIAL ISSUE: SOCIAL MEDIA.

He, X., Liao, L., Zhang, H., Nie, L., Hu, X., Chua, T.S., 2017. Neural collaborative filtering, in: Proceedings of the 26th International Conference on World Wide Web (WWW), ACM. pp. 173–182. doi:10.1145/3038912.3052569.

Jankowski, N.W., Van Selm, M., 2000. Traditional news media online: an examination of added-value. Communications 25, 85–102. URL: https://doi.org/10.1515/comm.2000.25.1.85, doi:10.1515/comm.2000.25.1.85.

Joachims, T., Granka, L., Pan, B., Hembrooke, H., Gay, G., 2017. Accurately interpreting clickthrough data as implicit feedback. ACM SIGIR Forum 51, 4–11.

Kelly, J., 2010. Parsing the Online Ecosystem: Journalism, Media, and the Blogosphere. Springer New York, New York, NY. pp. 93–108. URL: https://doi.org/10.1007/978-1-4419-6099-3_7, doi:10.1007/978-1-4419-6099-3_7.

Li, Z., Xie, Y., Zhang, W.E., Wang, P., Zou, L., Li, F., Luo, X., Li, C., 2024. Disentangle interest trend and diversity for sequential recommendation. Inf. Process. Manag. 61, 103619. URL: https://doi.org/10.1016/j.ipm.2023.103619, doi:10.1016/J.IPM.2023.103619.

Lichtenwalter, R., Lussier, J.T., Chawla, N.V., 2010. New perspectives and methods in link prediction, in: Rao, B., Krishnapuram, B., Tomkins, A., Yang, Q. (Eds.), Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, July 25-28, 2010, ACM. pp. 243–252. URL: https://doi.org/10.1145/1835804.1835837, doi:10.1145/1835804.1835837.

Linvill, D.L., Warren, P.L., 2020. Troll factories: Manufacturing specialized disinformation on twitter. Political Communication 37, 447–467. URL: https://doi.org/10.1080/10584609.2020.1718257, doi:10.1080/10584609.2020.1718257.

Liu, Z., Xu, W., Zhang, W., Jiang, Q., 2023. An emotion-based personalized music recommendation framework for emotion improvement. Inf. Process. Manag. 60, 103256. URL: https://doi.org/10.1016/j.ipm.2022.103256, doi:10.1016/J.IPM.2022.103256.

Macskassy, S.A., Michelson, M., 2011. Why do people retweet? anti-homophily wins the day!, in: Proceedings of the International AAAI Conference on Web and Social Media, AAAI. pp. 209–216.

Mazzara, M., Biselli, L., Greco, P.P., Dragoni, N., Marraffa, A., Qamar, N., de Nicola, S., 2013. Social networks and collective intelligence: a return to the agora. IGI Global. pp. 88–113. URL: https://www.igi-global.com/chapter/social-networks-collective-intelligence/75889, doi:10.4018/978-1-4666-3926-3.ch005.

Morais, N., Cruz, M., 2020. Gender perception about fake news and disinformation: Case study with portuguese higher education students, in: EDULEARN20 Proceedings, IATED. pp. 7746–7753. URL: https://doi.org/10.21125/edulearn.2020.1957, doi:10.21125/edulearn.2020.1957.

Muñoz, P., Díez, F., Bellogín, A., 2024. Modeling disinformation networks on twitter: structure, behavior, and impact. Appl. Netw. Sci. 9, 4. URL: https://doi.org/10.1007/s41109-024-00610-w, doi:10.1007/S41109-024-00610-W.

Muraina, I., 2022. Ideal dataset splitting ratios in machine learning algorithms: general concerns for data scientists and data analysts, in: 7th International Mardin Artuklu Scientific Research Conference.

Musso, A., Helbing, D., 2024. How networks shape diversity for better or worse URL: https://doi.org/10.1098/rsos.230505, doi:10.1098/rsos.230505.

Musto, C., de Gemmis, M., Lops, P., Narducci, F., Semeraro, G., 2022. Semantics and content-based recommendations, in: Ricci, F., Rokach, L., Shapira, B. (Eds.), Recommender Systems Handbook. Springer US, pp. 251–298. URL: https://doi.org/10.1007/978-1-0716-2197-4_7, doi:10.1007/978-1-0716-2197-4_7.

Newman, M.E.J., 2010. Networks: an introduction. Oxford University Press, Oxford; New York.

Nguyen, T.T., et al., 2024. Manipulating recommender systems: A survey of poisoning attacks and countermeasures. ACM Computing Surveys .

Nikolakopoulos, A.N., Ning, X., Desrosiers, C., Karypis, G., 2022. Trust your neighbors: A comprehensive survey of neighborhood-based methods for recommender systems, in: Ricci, F., Rokach, L., Shapira, B. (Eds.), Recommender Systems Handbook. Springer US, pp. 39–89. URL: https://doi.org/10.1007/978-1-0716-2197-4_2, doi:10.1007/978-1-0716-2197-4_2.

Orellana-Rodriguez, C., Keane, M.T., 2018. Attention to news and its dissemination on twitter: A survey. Comput. Sci. Rev. 29, 74–94. URL: https://doi.org/10.1016/j.cosrev.2018.07.001, doi:10.1016/J.COSREV.2018.07.001.

Pariser, E., 2011. The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think. Penguin.

Pathak, R., Spezzano, F., Pera, M.S., 2023. Understanding the contribution of recommendation algorithms on misinformation recommendation and misinformation dissemination on social networks. ACM Trans. Web 17, 35:1–35:26. URL: https://doi.org/10.1145/3616088, doi:10.1145/3616088.

Pennebaker, J., Boyd, R., Jordan, K., Blackburn, K., 2015. The development and psychometric properties of LIWC2015. University of Texas at Austin. doi:10.15781/T29G6Z.

Posegga, O., Jungherr, A., 2019. Characterizing political talk on twitter: A comparison between public agenda, media agendas, and the twitter agenda with regard to topics and dynamics, in: Bui, T. (Ed.), 52nd Hawaii International Conference on System Sciences, HICSS 2019, Grand Wailea, Maui, Hawaii, USA, January 8-11, 2019, ScholarSpace. pp. 1–10. URL: https://hdl.handle.net/10125/59697.

Rajendran, L., Thesinghraja, P., 2014. The impact of new media on traditional media. Middle-East Journal of Scientific Research 22, 609–616.

Ramponi, G., Brambilla, M., Ceri, S., Daniel, F., Giovanni, M.D., 2020. Content-based characterization of online social communities. Inf. Process. Manag. 57, 102133. URL: https://doi.org/10.1016/j.ipm.2019.102133, doi:10.1016/J.IPM.2019.102133.

Raza, S., Ding, C., 2022. News recommender system: a review of recent progress, challenges, and opportunities. Artif. Intell. Rev. 55, 749–800. URL: https://doi.org/10.1007/s10462-021-10043-x, doi:10.1007/S10462-021-10043-X.

Said, A., Bellogín, A., 2014. Comparative recommender system evaluation: benchmarking recommendation frameworks, in: Kobsa, A., Zhou, M.X., Ester, M., Koren, Y. (Eds.), Eighth ACM Conference on Recommender Systems, RecSys '14, Foster City, Silicon Valley, CA, USA - October 06 - 10, 2014, ACM. pp. 129–136. URL: https://doi.org/10.1145/2645710.2645746, doi:10.1145/2645710.2645746.

Sanz-Cruzado, J., Castells, P., 2018. Enhancing structural diversity in social networks by recommending weak ties, in: Proceedings of the 12th ACM conference on recommender systems, ACM. pp. 233–241.

Sanz-Cruzado, J., Castells, P., 2022. RELISON: A framework for link recommendation in social networks, in: Amigó, E., Castells, P., Gonzalo, J., Carterette, B., Culpepper, J.S., Kazai, G. (Eds.), SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11 - 15, 2022, ACM. pp. 2992–3002. URL: https://doi.org/10.1145/3477495.3531730, doi:10.1145/3477495.3531730.

Sanz-Cruzado, J., Castells, P., López, E., 2019. A simple multi-armed nearest-neighbor bandit for interactive recommendation, in: Bogers, T., Said, A., Brusilovsky, P., Tikk, D. (Eds.), Proceedings of the 13th ACM Conference on Recommender Systems, RecSys 2019, Copenhagen, Denmark, September 16-20, 2019, ACM. pp. 358–362. URL: https://doi.org/10.1145/3298689.3347040, doi:10.1145/3298689.3347040.

Sanz-Cruzado, J., Pepa, S.M., Castells, P., 2018. Structural novelty and diversity in link prediction, in: Champin, P., Gandon, F., Lalmas, M., Ipeirotis, P.G. (Eds.), Companion of the The Web Conference 2018 on The Web Conference 2018, WWW 2018, Lyon , France, April 23-27, 2018, ACM. pp. 1347–1351. URL: https://doi.org/10.1145/3184558.3191576, doi:10.1145/3184558.3191576.

Shao, C., Hui, P.M., Wang, L., Jiang, X., Flammini, A., Menczer, F., Ciampaglia, G.L., 2018. Anatomy of an online misinformation network. PLOS ONE 13, 1–23. URL: https://doi.org/10.1371/journal.pone.0196087, doi:10.1371/journal.pone.0196087.

Silva, N., Werneck, H., Silva, T., Pereira, A.C.M., Rocha, L., 2022. Multi-armed bandits in recommendation systems: A survey of the state-of-the-art and future directions. Expert Syst. Appl. 197, 116669. URL: https://doi.org/10.1016/j.eswa.2022.116669, doi:10.1016/J.ESWA.2022.116669.

Sitaula, N., Mohan, C.K., Grygiel, J., Zhou, X., Zafarani, R., 2020. Credibility-Based Fake News Detection. Springer International Publishing, Cham. pp. 163–182. URL: https://doi.org/10.1007/978-3-030-42699-6_9, doi:10.1007/978-3-030-42699-6_9.

Stefanone, M.A., Lackaff, D., Rosen, D., 2010. The relationship between traditional mass media and "social media": Reality television as a model for social network site behavior. Journal of Broadcasting & Electronic Media 54, 508–525.

Tomlein, M., Pecher, B., Simko, J., Srba, I., Móro, R., Stefancova, E., Kompan, M., Hrckova, A., Podrouzek, J., Bieliková, M., 2021. An audit of misinformation filter bubbles on youtube: Bubble bursting and recent behavior changes, in: Pampín, H.J.C., Larson, M.A., Willemsen, M.C., Konstan, J.A., McAuley, J.J., Garcia-Gathright, J., Huurnink, B., Oldridge, E. (Eds.), RecSys '21: Fifteenth ACM Conference on Recommender Systems, Amsterdam, The Netherlands, 27 September 2021 - 1 October 2021, ACM. pp. 1–11. URL: https://doi.org/10.1145/3460231.3474241, doi:10.1145/3460231.3474241.

Tommasel, A., Menczer, F., 2022. Do recommender systems make social media more susceptible to misinformation spreaders?, in: Golbeck, J., Harper, F.M., Murdock, V., Ekstrand, M.D., Shapira, B., Basilico, J., Lundgaard, K.T., Oldridge, E. (Eds.), RecSys '22: Sixteenth ACM Conference on Recommender Systems, Seattle, WA, USA, September 18 - 23, 2022, ACM. pp. 550–555. URL: https://doi.org/10.1145/3523227.3551473, doi:10.1145/3523227.3551473.

Valle, M.E.D., Broersma, M., Ponsioen, A., 2021. Political interaction beyond party lines: communication ties and party polarization in parliamentary twitter networks. Social Science Computer Review 40, 736–755. doi:https://doi.org/10.1177/0894439320987569.

Vieira, L.L., Jerônimo, C.L.M., Campelo, C.E.C., Marinho, L.B., 2020. Analysis of the subjectivity level in fake news fragments, in: de Salles Soares Neto, C. (Ed.), WebMedia '20: Brazillian Symposium on Multimedia and the Web, São Luís, Brazil, November 30 - December 4, 2020, ACM. pp. 233–240. URL: https://doi.org/10.1145/3428658.3430978, doi:10.1145/3428658.3430978.

Villa, G., Pasi, G., Viviani, M., 2021. Echo chamber detection and analysis: A topology- and content-based approach in the covid-19 scenario. Social Network Analysis and Mining 11. doi:10.1007/s13278-021-00779-3.

Wang, S., Zhang, X., Wang, Y., Ricci, F., 2024. Trustworthy recommender systems. ACM Trans. Intell. Syst. Technol. 15. URL: https://doi.org/10.1145/3627826, doi:10.1145/3627826.

Wu, L., Morstatter, F., Carley, K.M., Liu, H., 2019. Misinformation in social media: Definition, manipulation, and detection. SIGKDD Explor. 21, 80–90. URL: https://doi.org/10.1145/3373464.3373475, doi:10.1145/3373464.3373475.

Yang, Q., He, X., Song, G., Liu, M., Tang, J.R., 2021. Consisrec: Enhancing gnn for social recommendation via consistency of interest and social connection, in: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), ACM. pp. 2369–2377. doi:10.1145/3447548.3467307.

Yang, Q., Nikolenko, S.I., Huang, A., Farseev, A., 2022. Personality-driven social multimedia content recommendation, in: Magalhães, J., Bimbo, A.D., Satoh, S., Sebe, N., Alameda-Pineda, X., Jin, Q., Oria, V., Toni, L. (Eds.), MM '22: The 30th ACM International Conference on Multimedia, Lisboa, Portugal, October 10 - 14, 2022, ACM. pp. 7290–7299. URL: https://doi.org/10.1145/3503161.3548769, doi:10.1145/3503161.3548769.

Yoon, S.J., Kim, J.H., 2001. Is the internet more effective than traditional media? factors affecting the choice of media. Journal of Advertising Research 41, 53–60. URL: https://www.journalofadvertisingresearch.com/content/41/6/53, doi:10.2501/JAR-41-6-53-60, arXiv:https://www.journalofadvertisingresearch.com/content/41/6/53.full.pdf.

Zhang, S., Yao, L., Sun, A., Tay, Y., 2019. Deep learning based recommender system: A survey and new perspectives. ACM Comput. Surv. 52, 5:1–5:38. URL: https://doi.org/10.1145/3285029, doi:10.1145/3285029.