Applied Network Science

RESEARCH

Open Access

Modeling disinformation networks on Twitter: structure, behavior, and impact

Pau Muñoz[1], Fernando Díez[1] and Alejandro Bellogín[1*]

*Correspondence:
alejandro.bellogin@uam.es

[1] Universidad Autónoma de Madrid, Madrid, Spain

Abstract

The influence and pervasiveness of misinformation on social media platforms such as Twitter have been well-documented in recent years. These platforms' real-time, rapid-fire nature and the personalized, echo-chamber-like environments they foster, often inadvertently, assist in misinformation amplification. To better understand this situation and how to encourage safer and broader narratives, this paper presents a comparative study of the activity of 275 Twitter accounts tagged as disinformation sources and 275 accounts tagged as legitimate journalists over a 3.5-year period in the Spanish context. By employing various modeling techniques, we investigate the structural differences and behavioral patterns between the two groups. Our findings demonstrate that disinformation accounts exhibit a coordinated behavior, among other distinct characteristics, leading to more efficient (dis)information propagation. The implications of these findings for understanding the dynamics of disinformation networks and combating their impact are discussed.

Keywords: Micro-blogging, Disinformation, Social network analysis, Information dynamics

Introduction

The media, including newspapers, radio, and television, has played for a very long time an instrumental role in shaping societal narratives and influencing public opinion on various subjects, particularly political and social matters. This influence is not merely a reflection of the media's role in information dissemination, but also an indicator of its potential as a tool for power (Chaffee and Metzger 2001). Consequently, many political actors have harnessed this tool to their advantage, utilizing media platforms to propagate their viewpoints and ideologies through highly curated narratives (Guarino et al. 2020). This phenomenon is neither new nor transient, as it continues to unfold in the constantly evolving media landscape of the present day (McCombs and Shaw 1972; Iyengar and Kinder 1987).

The evolution of traditional media into digital platforms has expanded the reach of these narratives and complexified their dynamics. In this digital age, the line between the producer and consumer of news has blurred, resulting in a significantly more participatory and less controlled environment (Lazer et al. 2018). This transformation has paved the way for a paradigm shift in influence dynamics, consequently opening doors

to disseminating, not just diverse viewpoints, but also unverified information and disinformation. The accessibility and interactivity of social media platforms, like microblogging sites (including Twitter, Threads, or Mastodon among others), have made them prime platforms for such activities (Zubiaga et al. 2018).

Indeed, social media platforms have democratized access to information, allowing users to both consume and generate content (Tandoc et al. 2018). This paradigm shift has resulted in an unprecedented expansion in the volume of information available to the public, contributing to digital media's ascendance over traditional media. Twitter, among other social media networks, has emerged as a significant player in this new era of information exchange, serving as a real-time source of news, opinions, and discourses (Bastos and Mercea 2019). This evolution not only attests to the dynamic nature of media consumption but also underscores its profound implications for understanding the contemporary digital information ecosystem (Bruns et al. 2018). In addition, in this digital information era, the internet, particularly social media, allows individuals to tailor their information intake according to their preferences (Bazmi et al. 2023). Users have the autonomy to selectively connect with sources they deem credible, trustful, or align with their perspectives, whether these sources are legitimate news outlets, individual experts, influencers, or even sources known for propagating unverified or misleading content. This personalized nature of information consumption represents a double-edged sword in the modern media environment (Flaxman et al. 2016).

The structure of this and other similar platforms fosters a rapid-fire exchange of information, transcending geographical boundaries and establishing an interconnected global community (Vosoughi et al. 2018). While the lack of an editorial filter can enhance the diversity of viewpoints and facilitate the spread of grassroots narratives, it also carries implications for the credibility and veracity of information. The absence of gatekeepers raises questions about the quality of the content shared (Xu et al. 2023), giving rise to phenomena such as misinformation and disinformation, which have become significant concerns in our contemporary digital information ecosystem (Lewandowsky et al. 2012; Saxena et al. 2023), in part, because of the difficulty to detect the so-called fake news (Jing et al. 2023).

As a consequence, online social media became widely consumed in our societies, which in turn has become the dissemination of organized misinformation increasingly pervasive (Zhou et al. 2021). Misinformation is characterized by the deliberate propagation of incorrect or manipulated information, which is often intended to mislead audiences and influence their perspectives or behaviors. This concept should be distinguished from disinformation, although the terms are often used interchangeably. While disinformation also involves the deliberate spread of false information, it is typically orchestrated by individuals or organized groups with a calculated intent to deceive, often with political, financial, or societal objectives in mind (Vosoughi et al. 2018; Magelinski et al. 2022). These groups can coordinate their (dis)informative action both spontaneously or formally (for example, in the case of nation-state-backed disinformation campaigns). When these accounts consistently act in a coordinated way over time, they constitute disinformation networks, which can be cross-platform, as recently characterized in Ng et al. (2022).

Therefore, a disinformation network, particularly in social media like Twitter, is essentially a system of interconnected accounts. These accounts are not just casually connected; they are actively collaborating, either implicitly or explicitly, to disseminate false information or deliberately deceptive narratives. This may occur for various reasons, such as for political gain, to sow social discord, to discredit individuals or organizations, or even to manipulate financial markets among other scenarios (Magelinski et al. 2022; Shao et al. 2018). These malicious actors employ sophisticated strategies to shape narratives and manipulate public opinion. They might present distorted facts, entirely fabricated stories, or decontextualized truths to promote a particular agenda or ideology (Shao et al. 2018). In the digital age, these tactics are not confined to shadowy corners of the internet but are often played out on mainstream social media platforms like Twitter. In these platforms' high-speed, high-volume environment, such content can quickly gain traction, potentially influencing large audiences before corrective measures can be put in place (Shao et al. 2018).

Consequently, studying misinformation and disinformation on social media platforms is not merely a niche academic pursuit but a pressing concern with real-world implications. It is a field that requires rigorous analysis to understand the structure, behavior, and impact of these disinformation networks (Shao et al. 2018; Vosoughi et al. 2018; Flaxman et al. 2016; Lewandowsky et al. 2012). Moreover, the design principles and user behavior patterns that make Twitter a fertile ground for misinformation and disinformation are not unique to this platform (Starbird et al. 2019). Any micro-blogging or social media platform operating under similar mechanisms and attracting a substantial user base could face the same challenges. Such platforms, too, are susceptible to manipulating their features and algorithms by actors aiming to spread misinformation, thereby perpetuating the cycle. This reality suggests that the phenomenon of misinformation is not only a concern for the present, but is likely to persist and potentially expand into new digital arenas in the future (Guess et al. 2019; Vosoughi et al. 2018; Starbird et al. 2019).

Therefore, understanding the modus operandi of disinformation networks on these platforms, the nature of their interaction with legitimate information sources, and the impact they generate is paramount. Our driving hypothesis in this work is that these networks are characterized by their structure and the dynamics of their interactions. The structure can include elements such as the number and arrangement of nodes (individual user accounts) and edges (connections between accounts, primarily by retweeting as the primary mechanism of content sharing), the presence of clusters or tightly-knit groups, and the overall network density (Vosoughi et al. 2018; Guarino et al. 2020). The network dynamics can include factors such as the speed at which information travels through the network, the frequency and patterns of interaction between accounts, and the evolution of these factors over time (Shao et al. 2018). Through this research, we aim to contribute to that understanding by examining the structural differences and behavioral patterns between disinformation and legitimate sources on Twitter. By doing so, we hope to shed light on the mechanisms of disinformation and provide insights to guide future interventions and policies to combat its spread (Starbird et al. 2019; Grinberg et al. 2019; Bastos et al. 2020).

Aims and scope

Our core research objective lies in understanding the network structure and dynamics characteristic of disinformation networks: sets of user accounts that are interconnected through mutual content sharing (retweeting), and actively engaged in creating, sharing, and promoting disinformation (Guarino et al. 2020). In pursuit of this goal, we strive to study the temporal evolution of network properties within disinformation networks during a 3.5-year period (from 2019 to mid-2022), contrasting them with those of networks composed of legitimate information disseminators, both in the context of the Spanish political landscape. Our primary interest lies in determining the efficiency with which informationâ€"or rather disinformationâ€"propagates within these disinformation networks.

With this goal in mind, we will contrast disinformation actors against journalists as legitimate sources of information. Unlike anonymous users or those with obscured identities who might engage in the propagation of disinformation, journalists are publicly identifiable entities, which imparts a certain degree of accountability and transparency to their actions on these platforms (Molyneux et al. 2020). They are tethered to the media outlets they represent, which typically uphold strict editorial standards and scrutiny before releasing content (Nielsen et al. 2020). This adherence to journalistic ethics and the principles of truth, accuracy, objectivity, fairness, and public accountability further distinguishes these professionals from disinformation actors (Molyneux et al. 2020). Furthermore, journalists possess a recognized professional track record, often with a considerable following, influencing public discourse. This visibility and credibility they bring to the platform contrast the often covert, manipulative operations of disinformation actors (Molyneux et al. 2020; Wardle and Derakhshan 2017).

Hence, by comparing and contrasting these two types of accounts—disinformation disseminators and legitimate journalists—this research seeks to uncover their distinctive structural differences, behavioral patterns, and consequent impacts on the Twitter network. Such findings would provide critical insights into the battle against the ongoing disinformation crisis. Moreover, we seek to unravel the factors contributing to forming network structures that facilitate the diffusion of information within disinformation networks. We are especially interested in identifying the conditions under which these disinformation networks manifest increased levels of cohesion and efficiency in their flow of disinformation.

Therefore, our aims can be summarized in the following research questions:

Research questions

The previously described aims can be summarized in the following research questions:

- **RQ1:** How do the disinformation networks behave in comparison to legitimate journalism networks according to the network structure? In particular, we shall consider network structure from connectivity and centrality perspectives (*RQ1a*) and from the community structure and information flow point of views (*RQ1b*).
- **RQ2:** What is the statistical significance of the variations in the temporal patterns of activity between disinformation networks and legitimate journalism networks?

- **RQ3:** How do the information content patterns influence the structure of the disinformation network?

The underlying hypothesis in this work, and upon which the previous research questions rely, is that there are significant differences between networks created by disinformation actors and legitimate ones. We will contrast and confirm this in the rest of the paper, by providing specific answers to these questions.

Background

Despite the increasing recognition of the existence and operation of specific social media accounts, particularly on platforms like Twitter, dedicated solely to the propagation of disinformation, the full extent of their impact and functionality still needs to be expanded. Research has confirmed that these accounts, often part of more extensive orchestrated 'influence campaigns', can operate with networks of automated accounts or 'bots', primarily amplifying a particular narrative (Marwick and Lewis 2017).

However, what remains nebulous is the magnitude to which these disinformation actors can outcompete or outmaneuver legitimate actors on these platforms. While it is evident that disinformation campaigns can significantly shape the discourse (Starbird et al. 2019; Pavlíková et al. 2021), the precise metrics or mechanisms of their influence vis-à-vis authentic voices have yet to be comprehensively examined. For instance, we lack a complete understanding of their reach, spread, or resonance among the audience compared to legitimate information sources.

Furthermore, the level of coordination within these disinformation networks remains an area requiring more empirical scrutiny (Guarino et al. 2020). While a degree of coordination is evident in the concerted distribution of specific narratives, the intricacies of these coordination efforts, such as the command structure, decision-making processes, and synchronization methods, must be thoroughly understood. One approach to address this was recently introduced in Magelinski et al. (2022), where the authors propose a synchronized action framework for detecting automated coordination by constructing and analyzing multi-view networks.

Finally, the consequent effects of these campaigns on shaping public opinion, political attitudes, or behavior are still largely conjectural. While anecdotal evidence and case studies provide insights (Bastos et al. 2020; Grinberg et al. 2019; Törnberg et al. 2020), the field still needs robust empirical evidence to quantify the real-world impact of these coordinated disinformation actors.

Micro-blogging networks as tools for political information

Micro-blogging networks, with Twitter as a foremost example, have become pivotal instruments in producing and consuming political information in the contemporary digital environment. These platforms, characterized by real-time updates, concise post lengths, and wide-reaching network structures, are uniquely suited to shaping political discourse and mobilizing public opinion (Conover et al. 2011; Jungherr et al. 2012).

Twitter, in particular, exhibits several distinct features that make it a potent platform for political information exchange. The platform's real-time nature enables instantaneous reporting and commenting on events, facilitating an active, dynamic political

dialogue (Conover et al. 2011; Jungherr et al. 2012). Its broad reach, enabled by network structures that transcend geographical and political boundaries, allows messages to disseminate widely and rapidly. Furthermore, the platform's capacity to accommodate diverse voices, from official political figures and journalists to activists and everyday citizens, fosters a multifaceted and dynamic political discourse.

This potent mix of accessibility, immediacy, and reach gives Twitter significant influence over political information landscapes (Jungherr et al. 2012). However, these same attributes can also be exploited by actors intending to spread disinformation, leading to manipulations of the political discourse and potential distortions in public understanding and opinion. Recognizing the intricacies of these dynamics within micro-blogging networks is a fundamental step towards effectively addressing the challenges of disinformation in our digital societies (Himelboim et al. 2013).

Propaganda and other related concepts

Propaganda refers to the strategic and orchestrated use of information, often biased or misleading, to shape public opinion or behavior toward a particular ideological, political, or commercial objective. It is typically associated with deliberately manipulating facts, ideas, arguments, or even emotional appeals to influence an audience (Ellul 2021; Henderson 1943; Huckin 2016).

In micro-blogging networks like Twitter, propaganda can take on unique characteristics. Given the brevity of content and the real-time nature of these platforms, propaganda is often tailored to be easily digestible and rapidly disseminated (Ratkiewicz et al. 2011). This can include using sensational or provocative language, visual elements, or hashtags to draw attention and encourage sharing. Moreover, due to the networked structure of these platforms, propaganda can quickly spread beyond its initial audience, reaching and influencing a diverse range of users. Propaganda in these networks is not confined to state actors or organizations; even individuals can become propagators, willingly or otherwise (Starbird et al. 2019).

Disinformation and misinformation

While often used interchangeably, misinformation and disinformation have distinct implications. Misinformation refers to any incorrect or misleading information, regardless of intent. A user might unknowingly spread misinformation, often due to a genuine mistake or misunderstanding (Pérez-Escolar et al. 2023; Wardle and Derakhshan 2017). Disinformation, on the other hand, is a subset of misinformation characterized by intent. It refers to the deliberate creation and sharing of false or manipulated information to deceive audiences, often to achieve specific strategic, political, or commercial goals (Tandoc et al. 2018; Lewandowsky et al. 2017; Wardle and Derakhshan 2017).

In the context of micro-blogging networks, these phenomena become particularly complex. Given the speed at which information spreads on platforms like Twitter, misinformation and disinformation can rapidly reach large audiences. The anonymous or pseudonymous nature of many accounts on these platforms can make it difficult to ascertain the intent behind misleading posts, complicating efforts to distinguish between misinformation and disinformation (Allcott and Gentzkow 2017). Additionally, algorithms that prioritize engagement can inadvertently promote

misinformation and disinformation, as false or sensational content often elicits strong reactions (Lewandowsky et al. 2017; Allcott and Gentzkow 2017).

Understanding the nuances between propaganda, misinformation, and disinformation is crucial in developing effective strategies to combat these phenomena on micro-blogging networks. Each requires a different approach: counteracting propaganda might foster media literacy and critical thinking; addressing misinformation could entail fact-checking and corrective information, while combating disinformation may necessitate platform-level interventions and policy changes (Wardle and Derakhshan 2017).

Disinformation and propaganda, while distinct, are closely intertwined. Both are used as tools to influence public opinion, often toward a specific political, ideological, or commercial end. However, they differ primarily in their relationship with truth and intention (Ha et al. 2021; Wardle and Derakhshan 2017).

Propaganda may utilize factual and false information, but it is mainly characterized by its use of biased or misleading narratives to promote a particular point of view. Disinformation, conversely, involves the deliberate creation and dissemination of false information intending to deceive (Ha et al. 2021). In many cases, disinformation can be a form of propaganda. By creating and spreading false narratives, actors can manipulate public perception and behavior to align with their goals. For instance, a political actor might disseminate disinformation about an opponent's policies or personal life to undermine them and sway public sentiment in their favor (Wardle and Derakhshan 2017).

Disinformation and political polarization

Disinformation also plays a significant role in political polarization, in particular in the so-called *affective polarization*, which refers to the process where a society's attitudes towards political, ideological, or social issues diverge towards extreme opposing positions (Conover et al. 2011; Tucker et al. 2018). Disinformation can exacerbate these divisions by disseminating false narratives, particularly those that play on existing biases, fears, or prejudices. For instance, disinformation that portrays a particular political group as an existential threat to another group can intensify existing animosities, leading to further polarization (Conover et al. 2011; Azzimonti and Fernandes 2018).

Micro-blogging networks like Twitter can amplify these effects due to their structure and algorithms. As users are more likely to interact with content that aligns with their views, platforms may serve them more such content, leading to echo chambers that reinforce and intensify their beliefs (Azzimonti and Fernandes 2018). Disinformation can thrive in these echo chambers, driving polarization by further entrenching users in their existing viewpoints and making them more susceptible to extreme or divisive narratives (Conover et al. 2011; Azzimonti and Fernandes 2018; Tucker et al. 2018).

Therefore, while disinformation is not the sole cause of polarization, it can be a powerful catalyst, leveraging and exacerbating existing divisions for strategic ends (Conover et al. 2011). Understanding this relationship is crucial for developing interventions to counter disinformation and mitigate its impact on societal polarization (Conover et al. 2011; Tucker et al. 2018; Azzimonti and Fernandes 2018).

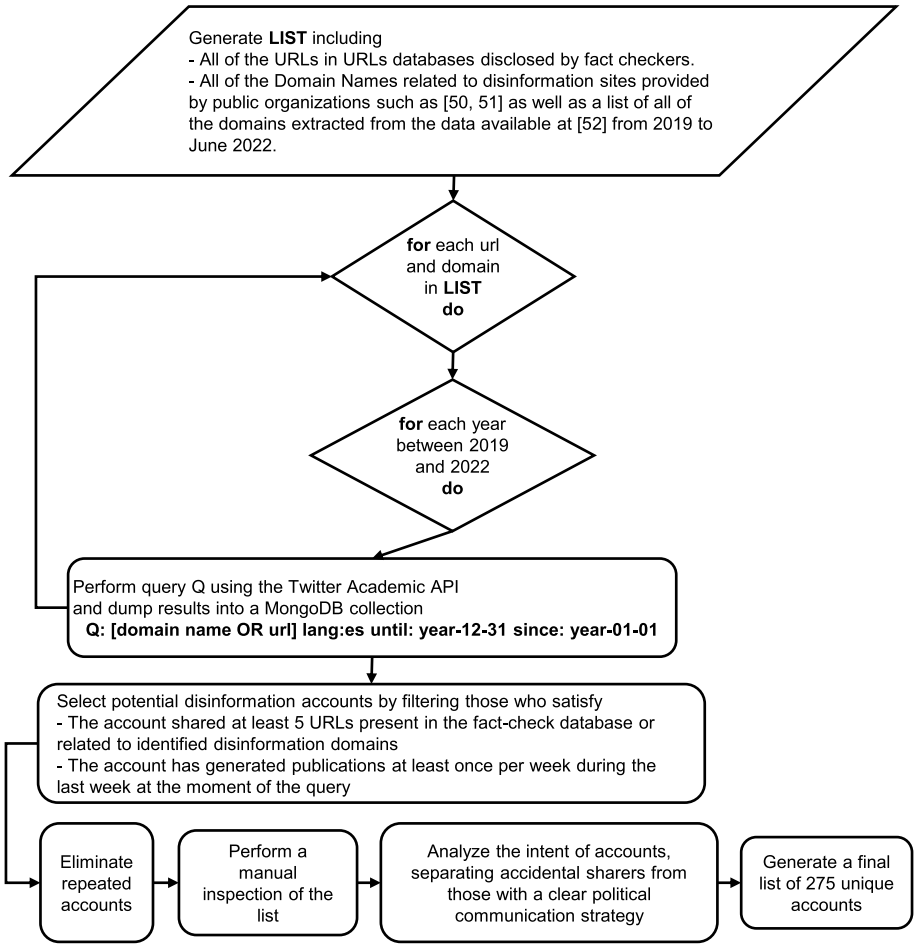


Fig. 1 Diagram flow of the process for generating a data set of Twitter disinformation accounts

Methods and techniques

Our study implements a multidimensional approach to explore the dynamics of information propagation on micro-blogging sites. Our methodology begins with identifying and enumerating Twitter accounts linked to legitimate information distributors and disinformation actors. We gathered and tracked their online activities for 3.5 years, from 2019 to mid-2022, creating a rich dataset for subsequent analysis.

Having established our comprehensive data set, we implemented social network analysis techniques to rigorously examine the intrinsic network properties of legitimate information and disinformation nodes. The core objective of this phase was to discern the main difference between these contrasting networks, thereby yielding insights into the probable pathways of information propagation within each of them.

Data collection

Constructing our data set commenced with identifying "disinformation actors" - accounts that consistently disseminate misleading narratives or counterfeit news. Given the challenging nature of accurately labeling an account as a disinformation agent, we employed the following flexible approach (depicted in Fig. 1), in a thorough and

consistent way to improve its reproducibility. Initially, we consulted verified databases of fictitious news and domain names affiliated with disinformation spreading, courtesy of international organizations such as The European Commission [see the first block in the diagram flow, using Wang (2017); Shu et al. (2018); Challenge (2019); Singer-Vine (2016); Zimdars (2017); D'Ulizia et al. (2021) as databases and (TaskForce 2022; Moroz and Loza 2022, 2022) for domain names]. Subsequently, we sought out those Spanish accounts (through the use of the 'lang:es' parameter in our search queries) that demonstrated the highest interaction levels with such dubious content. Our decision to focus on Spanish accounts was twofold: it capitalized on the authors' proficiency in the Spanish language and the Spanish political landscape. This, in particular, addressed a significant gap in research on Spanish language disinformation.

Our data set generation then incorporated an additional check. We undertook a qualitative review process to mitigate the potential for false positives. The manual filtering process aimed to refine the data set, whereby we favored accounts demonstrating frequent interactions with events or narratives within Spain. This was specified to keep the focus of our research on this regional phenomena, while also allowing us to serve as filtering *experts*, considering our experience and familiarity with the events. Therefore, the creation of our list of accounts associated with disinformation resulted from a multi-step process, as outlined above. This meticulous approach allowed for the compilation of a robust data set in its representation of disinformation activity on Twitter, allowing us to analyze their behavioral patterns in depth. The whole process is depicted in the diagram flow presented in Fig. 1.

The number of unique accounts identified before the manual inspection was 513. Then, a team of three volunteer students with a background in political science, along with the help of the authors, conducted another qualitative assessment and manually extracted 275 unique disinformation accounts (last two blocks in the diagram flow). The process involved selecting the most active and consistent accounts related to disinformation. This was a conscious decision, since these highly active accounts are key in the spread of disinformation, as their primary aim is to achieve maximum possible impact, due to their significantly higher potential for influencing and impacting online conversations. In order to assist the validation of the selected accounts as disinformation actors, they were checked using the "misinfo.me" online service, being all of them flagged as mainly disinformation sharers.

Upon establishing the 275 disinformation-related accounts, we created a commensurate sample size for the comparison group, aiming to identify them as legitimate purveyors of information or more specifically, journalists. We assembled a pool of principal digital media outlets within Spain to construct this sample. This list was substantiated by cross-referencing numerous rankings—such as OJD (2022)[1] and Statista (2022)—to ensure the inclusion of prominent, mainstream outlets. From these sources, we identified individual journalists frequently engaged in public discourse, particularly in socially relevant areas such as politics and society. Indeed, the specific selection of journalists

[1] OJD, https://www.ojd.es, (from Spanish *Oficina de Justificación de la Difusión*, in English *Audit Bureaux of Circulations*) is the Spanish organization that provides, among others, services of control and issuing of dissemination reports as well as data consultation figures via the Internet. It belongs to the International Federation of Audit Bureaux of Circulations (IFABC), http://www.ifabc.org.

who cover politics and society was not arbitrary. Propaganda and disinformation typically orbit around the themes of politics and societal issues, with these subjects often being the prime targets of such misleading campaigns (Fallis 2015; Ruohonen 2021). Due to their contentious nature and potential for social impact, these topics are prime vehicles for the proliferation of disinformation. Furthermore, such focus areas frequently serve as battlegrounds for public opinion, making them fertile grounds for disinformation actors to exploit.

We further distilled our selection based on activity level from this pool of journalists. The accounts exhibiting the highest degree of interaction were selected for inclusion in our data set. This approach—as it was done for disinformation actors—ensures that our analysis is relevant and focused on entities with the greatest potential to influence the online discourse. This selection method, outlined as a diagram flow in Fig. 2, provided a balanced, representative sample for studying the behavior of disinformation networks on Twitter in contrast to their legitimate counterparts. A final list of 275 accounts was generated to be comparable with those obtained through the previous process.

Modeling techniques

In order to analyze the key differences between both networks, we employed content analysis and network analysis techniques. These include the creation of network graphs that depict the complex interplay of interactions between users and provide a visual and quantitative representation of information flow. We divided these graphs into specific temporal segments to capture temporal variations in these dynamics. Complementing this, we dove into the content patterns, using a state-of-the-art deep-learning algorithm for sentiment analysis and observing specific discourse trends. This two-pronged approach enabled us to understand the pathways of information spread and the role shared content plays in these dynamics.

Network generation

We partitioned the activity data gathered for both sets of accounts into temporal segments, which we defined as one week in duration, spanning 2019 to 2022. This decision was grounded in the observation that, on Twitter, news typically has a lifespan of 24 h, except for certain viral news pieces, especially those concerning social and political issues, that may persist for a more extended period (Mohd Shariff et al. 2014; Guenther et al. 2021). In our assessment, a seven-day window aptly captures this fluctuation. Moreover, segregating both networks into associated temporal points allows us to juxtapose their activity patterns over time statistically, as we shall explain in "Methodology".

Therefore, we generated a graph for each temporal segment and data set (i.e., journalists and dis-informers). Within these graphs, nodes correspond to the identified accounts, and directed edges symbolize the retweet action from one account (A) to another (B). The weight of these edges equates to the number of retweets exchanged between accounts during the corresponding time window. This representation offers a quantifiable and visual method of understanding the flow of information and the dynamics within these networks.

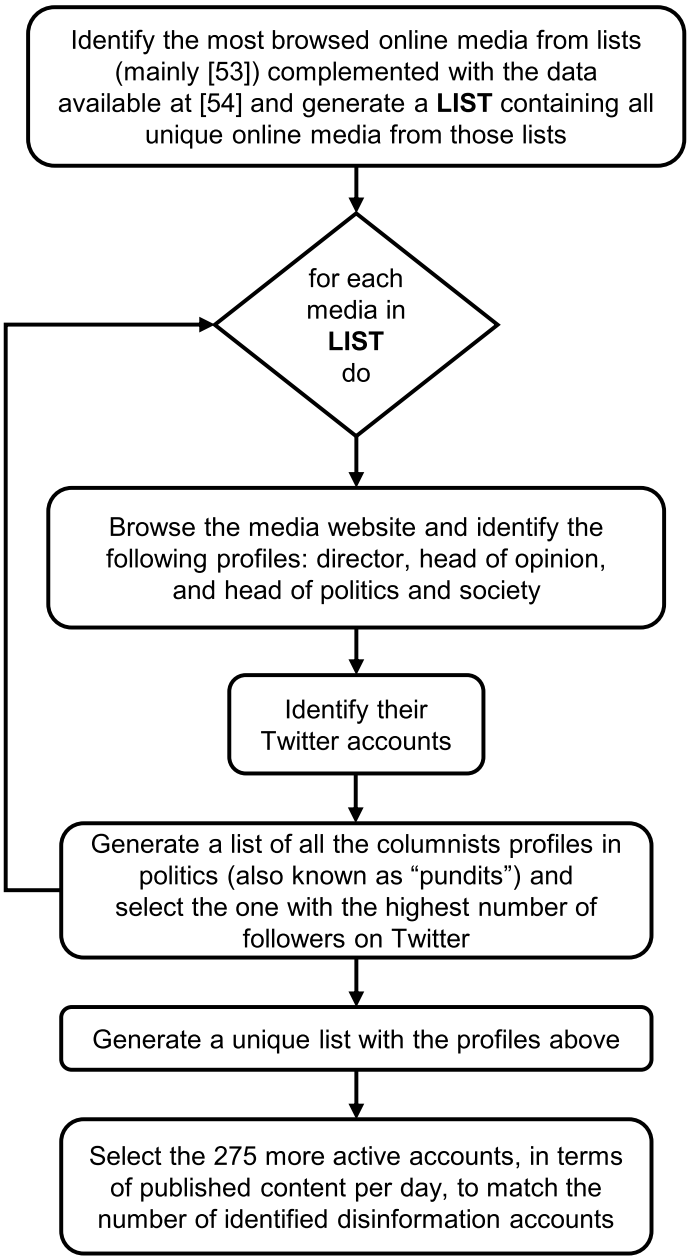


Fig. 2 Diagram flow of the process for identifying influential media-related (journalists) Twitter accounts

Network analysis

After generating graphs corresponding to each temporal window for both data sets, we probed their characteristics utilizing established network metrics. This examination aims to identify the type of network within which information propagates more rapidly and, where feasible, pinpoint contributing factors to this phenomenon. Concurrently, our exploration extended to the patterns of content generation within these networks at each temporal juncture. We sought to comprehend the interplay between the nature and volume of shared content and how it may shape the configuration of the network.

Our analysis was designed to furnish a comparative evaluation of the evolution of both networks over the designated period. Furthermore, it was instrumental in deciphering the dynamics that either accelerate or decelerate the flow of information within the network. Notably, the robustness of the network â€" its resilience to disruptions or perturbations â€" was also a focal point of our investigation.

Through this in-depth analysis, we aimed to unveil the intricate workings of these networks, offering invaluable insights into the behavior of disinformation networks on Twitter and their subsequent impact on the legitimate journalistic landscape. Next, we summarize the network metrics employed in this work, based on well-known concepts from the area (Newman 2010).

Density: The density of a directed and weighted graph is a measure of how many edges are present in the graph compared to the maximum possible number of edges. It quantifies how "connected" the graph is. Given a directed and weighted graph G with N nodes and M edges, the density can be defined as:

$$D(G) = \frac{M}{N(N-1)} \tag{1}$$

Average degree (Weighted): The average degree (also called average degree centrality measure) of a directed and weighted graph is a measure of how many connections, on average, each node has, considering the weights of the edges. The average degree can be calculated separately for in-degree ($\bar{k}_{in}(G)$), that is, the number of edges pointing to the node, and out-degree ($\bar{k}_{out}(G)$), that represents the number of edges starting from the node in a directed graph.

$$\bar{k}_{in}(G) = \frac{1}{N}\sum_{i=1}^{N} k_{in}(i) \tag{2}$$

$$\bar{k}_{out}(G) = \frac{1}{N}\sum_{i=1}^{N} k_{out}(i) \tag{3}$$

Efficiency: The efficiency of a directed and weighted graph is a measure of how efficiently information can be transmitted across the network. Efficiency is calculated as the inverse of the average of the shortest paths between all pairs of nodes in the graph.

$$E(G) = \frac{1}{N(N-1)} \sum_{i \neq j \in G} \frac{1}{d(i,j)} \tag{4}$$

where $d(i, j)$ computes the shortest path distance between nodes i and j.

These metrics can be used to analyze the structural properties of a graph, providing valuable insights into the connectivity, clustering, community structure, and overall efficiency of the network.

Modularity: The modularity of a directed and weighted graph measures the strength of its community structure. It quantifies the difference between the number of edges within communities and the expected number of edges if the edges were distributed randomly, preserving the nodes' in- and out-degree. Thus, given a directed

and weighted graph G with N nodes partitioned into C communities, the modularity can be defined as:

$$Q(G) = \frac{1}{C}\sum_{c=1}^{C}\left[e_c - \frac{k_{in}^{(c)}k_{out}^{(c)}}{C}\right] \tag{5}$$

Although several approaches could be used to partition the network, in this work we use the Louvain method (Blondel et al. 2008).

Average clustering coefficient: The average clustering coefficient of a directed and weighted graph measures the degree to which nodes in the graph tend to cluster together, considering the weights of the edges. It is the average of the local clustering coefficients of all nodes in the graph.

$$\overline{Cc}(G) = \frac{1}{N}\sum_{i=1}^{N}\frac{2E_i}{k_i(k_i-1)} \tag{6}$$

where $\overline{Cc}(G)$ represents the average clustering coefficient of a graph G. It is calculated by summing up the local clustering coefficients for each node i in the graph (computed as the ratio of twice the number of edges E_i between the neighbors of node i to the product of the (in-)degree k_i of node i and its degree minus one), and dividing by the total number of nodes N. This quantity measures the overall tendency of nodes in G to form clusters.

Average eigenvector centrality: The average eigenvector centrality of a directed and weighted graph is a measure of the overall importance or influence of nodes in the network. It takes into account not only the number of connections a node has, but also the importance of the nodes to which it is connected.

$$\overline{EVC}(G) = \frac{1}{N}\sum_{i=1}^{N}EVC(i) \tag{7}$$

where $EVC(i)$ is the eigenvector centrality of a node, a measure of the influence of that particular node in a network. It assigns relative scores to all nodes in the network based on the principle that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes. Given a directed and weighted graph G with N nodes and adjacency matrix A, the eigenvector centrality $EVC(i)$ of node i can be found as the element i of the eigenvector v corresponding to the largest eigenvalue λ of the adjacency matrix, i.e., $Av = \lambda v$. It should be noted that this metric is only defined for undirected networks; hence, to get around this issue and use it in our directed networks, we adapted it by symmetrizing the graph. Specifically, if account A retweeted account B, or vice versa, we treated this as a non-directed link between A and B for the purposes of calculating Eigenvector centrality.

Content analysis

In addition to analyzing the properties of the networks, which remain independent of the specific content shared by each account, we ventured into the examination of the type of content posted within these networks. This deeper dive aimed to discern any

possible correlation between the nature of shared content and the evolving structure of the network over time. Specifically, we sought to identify whether certain types of content or attitudes could contribute to or be associated with increased efficiency or density within the network.

Tweet sentiment: In order to achieve this, we started by analyzing the sentiment associated to each publication. We employed a state-of-the-art deep learning-based algorithm (Pérez et al. 2021) for classifying user posts based on their sentiment. The classification divided posts into those displaying predominantly positive sentiment and those with predominantly negative sentiment. From there, we could compute the average number of predominantly negative tweets during a particular period.

$$\text{sentiment}(tweet) = \begin{cases} 'negative' & \text{if} \quad \text{neg}(tweet) > \text{pos}(tweet) \\ 'positive' & \text{otherwise} \end{cases}$$

References to controversial events per tweet: We also sought to identify references to significant geopolitical events, such as NATO-related activities or the war in Ukraine. These topics have been a focal point within the Spanish communication space during the period under review and have been substantially covered by actors disseminating disinformation, as indicated by existing research (Garat 2023; Smart et al. 2022). Besides, in light of the COVID-19 pandemic's significant presence in major disinformation studies, we also included it in our analysis (Kouzy et al. 2020). These three events, while global in their implications, had direct and significant impacts on Spain: a) Spain hosted the NATO summit in Madrid in 2022, which brought NATO-related discussions and narratives to social media; b) the conflict in Ukraine also held substantial relevance in Spain, both because it was the first major war on European soil since the Balkan conflicts, but also because the Spanish government was divided in their discourses, further increasing the controversy around this event; and c) the impact of COVID-19 on Spain was particularly pronounced, given the country's implementation of a highly restrictive and controversial lockdown policy, making it a central topic of discourse and disinformation within the Spanish Twitter sphere.

In this context, we calculated the average number of references to COVID, NATO, and Ukraine per tweet within the network for each time window studied. In order to do that, we started by calculating the number of references to each of the topics by searching for the words 'COVID', 'NATO', and 'UKRAINE' in each text and then averaging all the occurrences found per tweet.

$$\text{avg\_nato} = \frac{1}{T} \sum_{tweet=1}^{T} \text{references\_nato}(tweet)$$

$$\text{avg\_ukraine} = \frac{1}{T} \sum_{tweet=1}^{T} \text{references\_ukraine}(tweet)$$

$$\text{avg\_covid} = \frac{1}{T} \sum_{tweet=1}^{T} \text{references\_covid}(tweet)$$

where T is the number of tweets in a specific instance of the network.

Table 1 Properties of the complete networks (i.e., using all the data from the entire 2019–2022 period), in both cases, with 275 nodes

	Tweets	RTs	$E(G)$	$Q(G)$	$\overline{Cc}(G)$	$\overline{EVC}(G)$
Journalists	3,906,047	96,551	0.170	0.743	0.334	0.028
Disinformation	7,194,766	513,566	0.268	0.580	0.502	0.036

RTs refers to retweets (edges), and the last four columns correspond to metrics defined in Sect. "Network analysis" (in all cases, metrics are in the [0, 1] range, where a higher value denotes the network is more efficient, modular, or clustered)

URL and hashtags per tweet: In the final phase of our content analysis, we noted the number of URLs and hashtags shared per tweet on average within the network for each studied time window. This allowed us to observe potential patterns or shifts in content sharing behaviors over time.

Experiments and results

In this section, we will address the research questions introduced at the beginning of the paper: to answer **RQ1** (How do the disinformation networks behave in comparison to legitimate journalism networks according to the network structure?) in "Behavior of disinformation networks according to the network structure" we will analyze the output of the structural network metrics presented before (*RQ1a*), while the output of those metrics related to information flow and propagation will be considered for *RQ1b*. Moreover, for all the considered metrics we will compute significance test statistics to address **RQ2** (What is the statistical significance of the variations in the temporal patterns of activity between disinformation networks and legitimate journalism networks?), whereas correlation between the type of content and the structure of the network will be presented in "Behavior of disinformation networks according to the network content" to answer **RQ3** (How do the information content patterns influence the structure of the disinformation network?). Before that, in the following "Methodology" and "Initial analysis of collected data", we introduce the considered methodology to answer these research questions and an initial analysis on the collected data.

Methodology

To address the research questions considered throughout this work, we have processed the data collected as explained in Sect. "Data collection". First, let us recall we create two (sets of) networks by exploiting the retweet action among two subsets of users: those categorized as journalists and those as disinformation actors (see Sect. "Network generation"). A summary of the overall graphs generated when using all this information is presented in Table 1.

Moreover, since the data was collected during 3.5 years (2019–2022[2]), a different network was created for each temporal segment of one week of duration, resulting in 338 different networks. These networks are the ones considered for analysis in this section. In the experiments we present in the following sections, we use this data in several, complementary ways. In some cases, we consider the temporal evolution (time series) of all the network metrics defined in Sects. "Network analysis" and "Content analysis". This

[2] In fact, only half year of 2022 was considered, since that was the most recent data available at request time.

means that those metrics were computed on each network, and the obtained scores were recorded for every instance of the network throughout the 2019–2022 period (once for each temporal segment). These time series will be considered to assess whether any statistically significant difference exists between the two types of networks (journalists and disinformation), as our aim is to delineate the behavioral patterns distinguishing disinformation networks from journalist networks, with a particular focus on the interaction structures within each network.

For this, we initially applied a Mann–Whitney U test, assuming the null hypothesis (H0) is that *there is no difference between the journalists and disinformation actors groups* for each scenario. This test was used because, after checking the normality of the data using the Shapiro-Wilk test, the results indicated that the data were not normally distributed for both groups. Thus, a non-parametric test was chosen for further analysis. For the sake of clarity, these time series are also plotted to allow a visual inspection of the data.

In conjunction with the Mann–Whitney U test, our methodology also incorporated a one-sample t-test on the weekly differences in average metrics between the two groups. This approach enables an examination of whether the observed weekly mean differences in these metrics are significantly distinct from zero, thus offering insights into the dynamic interplay between the networks over time. The integration of this test complements the distributional analysis provided by the Mann–Whitney U test, shedding light on both the distributional differences and the temporal consistency and significance of these differences. However, since it can only be applied to normally distributed data, we applied a logarithmic transformation before running the test.

Furthermore, to augment the robustness of our findings, we employed the Kolmogorov-Smirnov (KS) test for a granular analysis at the individual user level. This was particularly pivotal for our study of eigenvector centrality and degree centrality. For each user in both the disinformation and journalist networks, we computed the weekly averages of these metrics. The KS test was then applied to these data sets to determine if the distributions of eigenvector centrality and degree centrality values for individual users differed significantly between the two networks. This level of detailed analysis allows us to assert with greater confidence whether the observed patterns in network metrics are indeed reflective of underlying differences in the behavioral dynamics of disinformation actors and journalists.

The integration of the Mann–Whitney U test, the one-sample t-test, and the Kolmogorov-Smirnov test in our methodology provides a comprehensive and scientifically rigorous framework. This multifaceted approach enhances the depth of our analysis, as it examines the contrasting behaviors of disinformation and journalist networks across both aggregate and individual levels over a temporal spectrum.

Finally, the other main method used in our experiments consists of a correlation analysis via scatter plots, where a linear fit of the data is attempted, producing a measure of the goodness-of-fit and the probability that the relationship between the two variables is equal to zero (p-value).

Initial analysis of collected data

In our preliminary analysis, journalists and disinformation actors displayed consistent patterns of reciprocal retweeting. This common behavior of sharing each other's content

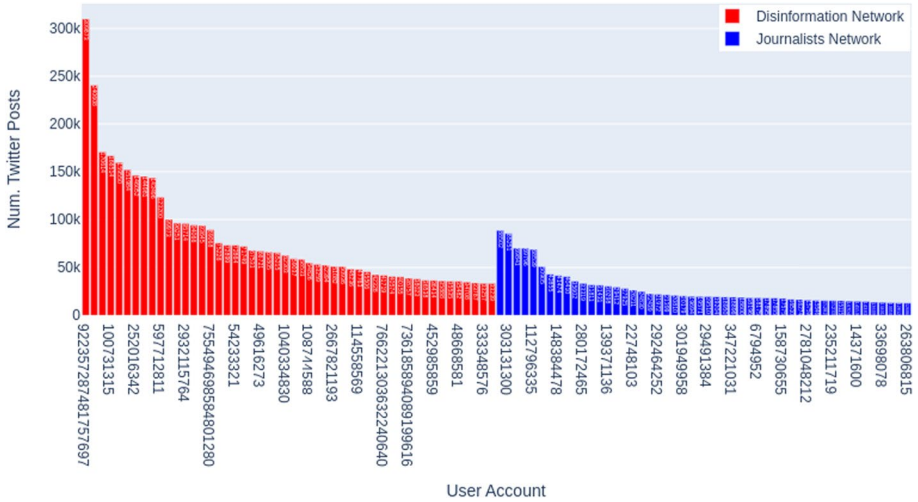


Fig. 3 Number of publications of the 50 top accounts in each network

over time results in the creation of information networks. Within these networks, users receive information from a variety of sources. In turn, any information - news, commentary, slogan, etc. - inserted into a network can circulate within that network via retweets. Given this observed phenomenon, we concluded that building network graphs is the optimal approach to investigate the dynamics of information dispersion within both groups, as done in previous works (Sanz-Cruzado and Castells 2018).

As Table 1 shows, the two analyzed networks evidenced different values of efficiency ($E(G)$), modularity ($Q(G)$), average clustering coefficient ($\overline{Cc}(G)$), and average eigenvector centrality ($\overline{EVC}(G)$), in particular, the disinformation network is more efficient and evidences a higher average clustering coefficient, highlighting its internal cohesion. However, since these values are collected for the entire networks, no fine-grained analysis can be performed—something we shall show later in subsequent sections.

Moreover, upon initial observation of account activity throughout the studied period, we note that the disinformation network and the network of legitimate actors demonstrate patterns where few users are responsible for most posts (see Fig. 3). It is also evident that the disinformation network produced a (total) higher volume of posts over the study period, since the top users of each network produced a remarkably different number of publications: more than 300K for the disinformation network and around 90K for the journalists..

Considering the account creation dates in Fig. 4, we observe that most accounts associated with legitimate actors were established between 2010 and 2012, coinciding with Twitter's rise in popularity in Spain, but also with an electoral period. In contrast, we noted two periods of substantial account creation within the disinformation network, one in 2018 and another in 2020. Interestingly, the latter coincides with the onset of the COVID-19 pandemic.

Behavior of disinformation networks according to the network structure

In this section, we perform different experiments to understand the behavior of disinformation networks (in comparison with the behavior of journalist networks) by

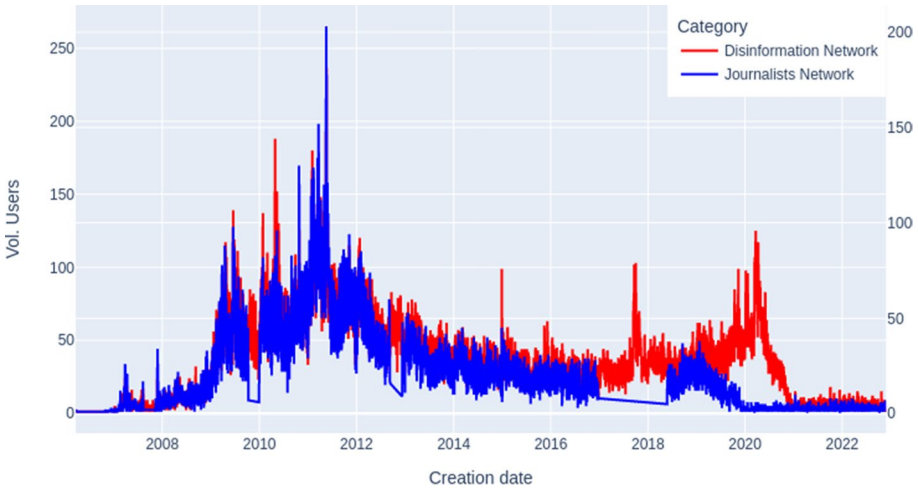
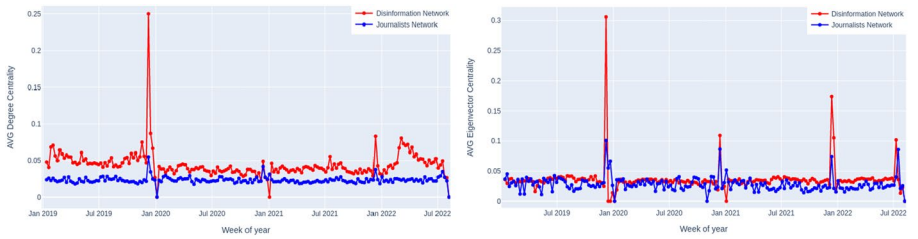


Fig. 4 Number of accounts created per month in the Disinformation and Journalists networks. Including the accounts mentioned, quoted o retweeted by them



(a) Evolution of average degree centrality (b) Evolution of average eigenvector centrality

Fig. 5 Evolution of average degree centrality (left) and average eigenvector centrality (right) on the Disinformation and Journalists Networks

considering the structure of the network derived through the interactions between the nodes in each network. For this, as explained in the methodology, we will contrast and compare those two networks throughout the time dimension, in particular computing the Mann–Whitney U significance test assuming the null hypothesis (H0) is that there is no difference between the journalists and disinformation actors groups.

The results of these tests are presented next, according to the type of metrics being analyzed: connectivity and centrality (Sect. Connectivity and centrality) and community structure and information flow (Sect. Community structure and information flow).

Connectivity and centrality

In this section, we focus on two definitions of centrality: average degree and eigenvector. Figure 5 shows the evolution of these metrics throughout the studied period of 2019-2022 for both networks. The results of running the significance tests on these data are:

1 For the average degree (also called average degree centrality, $\bar{k}_{in}(G)$), the Mann–Whitney U test demonstrated a statistically significant difference between the Journalists and Disinformation actors groups (U = 5373.0, p = 3.48e−40). We reject the null hypothesis (H0) for the average degree centrality, indicating that the median

average degree centrality for both groups is significantly different, being 0.0445 for the disinformation network and 0.0235 for the journalists network.

Additionally, a one-sample t-test on the weekly differences in average degree centrality revealed a t-statistic of −16.44 with a p-value of 5.01e−38, robustly rejecting the null hypothesis (H0). This indicates that the mean difference in average degree centrality between the two groups is significantly different from zero, with the negative t-statistic suggesting a higher average degree centrality in the disinformation network compared to the journalists network. Further enhancing our analysis, the Kolmogorov-Smirnov Test was conducted at the individual user level to compare the weekly averages of degree centrality for each user within both networks over the studied period. This test yielded a KS statistic of 0.78 and a p value of 2.81e−15, confirming that the distributions of degree centrality values are significantly different between individual users of the disinformation and journalists networks.

These comprehensive findings, which include both network-level and individual user-level analyses, strongly support the conclusion that there are not only significant differences in the distribution of average degree centrality values but also a consistent and notable divergence in the average and median values of this metric over time between the two networks.

2 For the eigenvector centrality ($\overline{EVC}(G)$), the Mann-Whitney U test indicated a statistically significant difference between the journalists and disinformation actors groups (U = 13348.0, p = 1.36e−11), suggesting distinct patterns in node influence and connectivity. We reject the null hypothesis (H0) for average eigenvector centrality, indicating that the median average eigenvector centrality for both groups is significantly different, being 0.0362 for the disinformation network and 0.0284 for the journalists network.
Further, a one-sample t-test on the weekly differences in average eigenvector centrality yielded a t-statistic of −4.89 with a p value of 2.21e−06, robustly rejecting the null hypothesis (H0) and indicating a significant mean difference between the groups. The negative t-statistic implies that, on average, the disinformation network exhibits higher eigenvector centrality compared to the journalists network. To deepen our analysis, we again conducted the Kolmogorov-Smirnov Test at the individual user level, comparing the weekly averages of eigenvector centrality for each user within the networks across the studied period. This test resulted in a KS statistic of 0.56 with a p value of 1.45e−07, confirming that the distributions of eigenvector centrality values are significantly different between individual users of the disinformation and journalists networks.
These collective findings, encompassing both network-level and individual user-level analyses, strongly support the conclusion that there are not only significant differences in the distribution of eigenvector centrality values but also a consistent and substantial divergence in the average and median values of this metric over time between the two networks.

According to the tests, when considering the elements of centrality, the nodes within the disinformation network have displayed a sustained pattern of more connections over

time. This attribute feeds into a network structure that fosters and facilitates the rapid spread of information. Furthermore, the higher average eigenvector centrality in the disinformation network reveals that the nodes within these networks are more numerous in their connections and boast superior quality connections. This, in turn, enables a faster distribution of information.

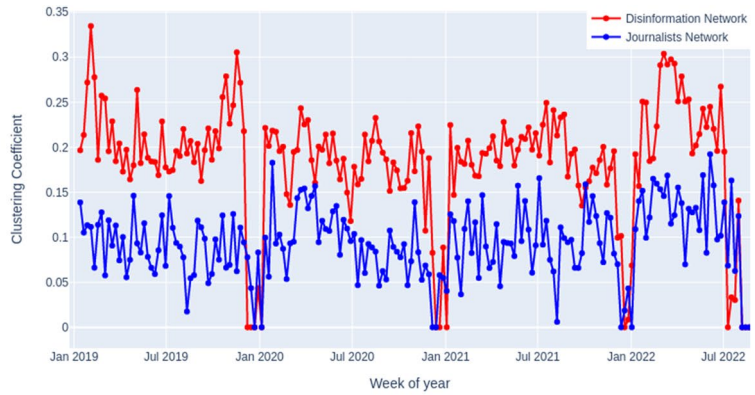
An intriguing aspect revealed by these metrics—in particular, according to EVC—is the potential presence of well-connected 'conversation leaders' within these networks, equivalent to webpages with high PageRank, a classical proxy for authority (Brin and Page 1998). They could be perceived as strategic coordinators or influencers who may help steer the direction of the shared narratives. However, a thorough investigation into this phenomenon would necessitate more detailed research. Identifying and understanding these key actors could be crucial for devising strategies to mitigate the influence of disinformation networks.

Community structure and information flow

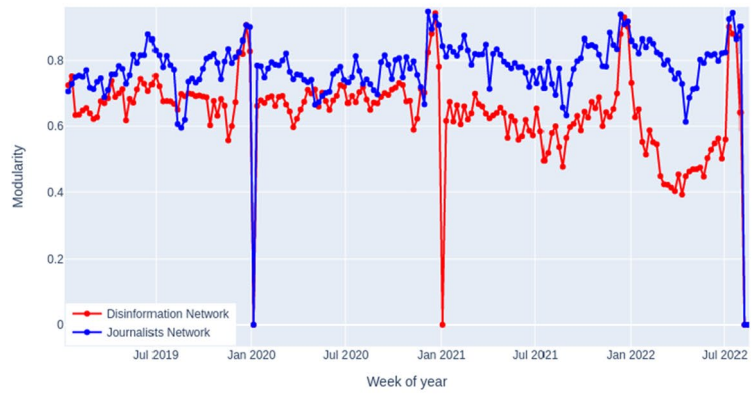
We now focus on efficiency, modularity, and clustering coefficient network metrics, more associated to how the information flows throughout a given network. Figure 6 shows the evolution of these metrics, and the corresponding results when running the tests to contrast the previously defined null hypothesis (there is no difference between the journalists and disinformation actors groups) are:

1 For efficiency ($E(G)$), the Mann–Whitney U test revealed a statistically significant difference between the Journalists and Disinformation actors groups (U = 6959.0, p = 4.80e−33). We reject the null hypothesis (H0) for efficiency, indicating that the mean efficiency for both groups is significantly different, being 0.0416 for the Disinformation network and 0.0154 for the Journalists network.
2 For modularity ($Q(G)$), the Mann–Whitney U test showed a statistically significant difference between the Journalists and Disinformation actors groups (U = 35077.0, p = 5.16e−28). We reject the null hypothesis (H0) for modularity, indicating that the mean modularity for both groups is significantly different, being 0.6451 for the Disinformation network and 0.7807 for the Journalists network.
3 For average clustering coefficient ($\overline{Cc}(G)$), the Mann-Whitney U test revealed a statistically significant difference between the Journalists and Disinformation actors groups (U = 7728.0, p = 6.97e−30). We reject the null hypothesis (H0) for $\overline{Cc}(G)$, indicating that the mean average clustering coefficient for both groups is significantly different, being 0.1871 for the Disinformation network and 0.0956 for the Journalists network.

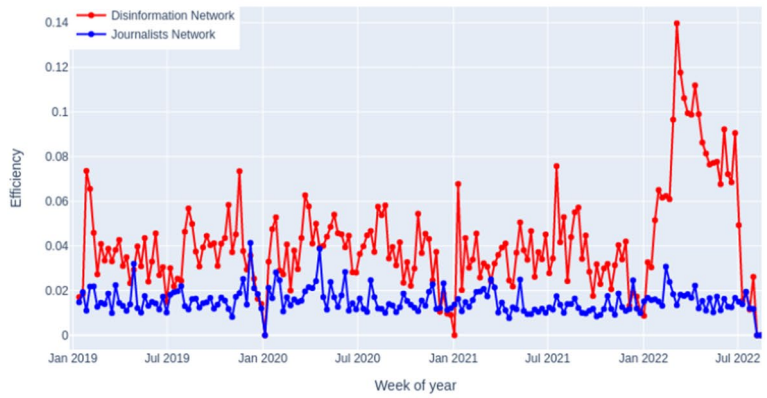
Drawing from our findings, as proven by the statistical tests, it is clear that information tends to flow more swiftly within the disinformation network. This trend of enhanced efficiency in information propagation has been consistent throughout the study periods from 2019 to 2022. Likewise, the observed disparities in the clustering coefficient and modularity indicate a more fragmented structure over time in the network of legitimate informers and a more cohesive structure among disinformation actors. This suggests that these networks, while vital for disseminating truthful and



(a) Evolution of the clustering coefficient



(b) Evolution of the modularity



(c) Evolution of the efficiency

Fig. 6 Evolution of clustering coefficient (top), modularity (center), efficiency (bottom) on the Disinformation and Journalists Networks

reliable information, may not be as interconnected or tightly-knit as their disinformation counterparts. Consequently, this could hinder the speed at which accurate information is disseminated within and across these networks.

There are, however, at least two aspects from Fig. 6 that deserves further explanation. First, the drops in modularity values for both the disinformation and journalist networks,

as shown in Fig. 6b. These are primarily due to distinct periods of reduced activity within these networks. On certain days, the studied accounts, although typically active, exhibited lower levels of engagement. This resulted in smaller networks with fewer retweets, directly impacting the network structure, since modularity, being a measure that hinges on the existence and definition of communities within a network, is sensitive to changes in network size. Second, the substantial spike in efficiency for the disinformation networks in 2022, as indicated in Fig. 6c, correlates with the significant geopolitical events surrounding the Russian campaign and subsequent large-scale land invasion in Ukraine. During such period, these networks exhibited peaks of activity, likely as a response to the unfolding events. This heightened activity led to increased connectivity and coordination among the accounts within the disinformation network, thus resulting in the observed spike in efficiency.

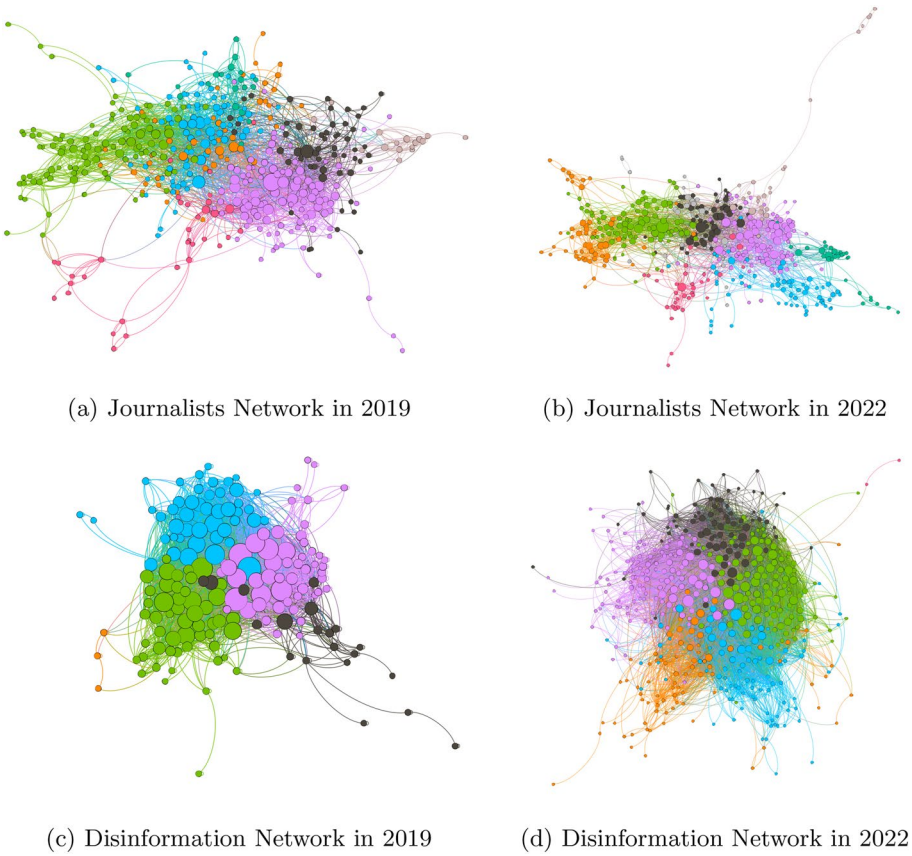
Behavior of disinformation networks according to the network content

Our analysis sought to establish a correlation between the nature of network activity and its structure, emphasizing activities that could hint at coordinated behavior. For this reason, in this experiment we analyze the networks at different moments in time and correlate their content characteristics against their density and efficiency.

First, in Fig. 7 we show the journalist and disinformation networks at different moments in time. Each dot in the graph is a node (Twitter account) of a given network, its size is proportional to the number of retweets it made during that period, and an edge exists if a retweet was made between those nodes. The colors represent communities detected by the Louvain method (Blondel et al. 2008). Based on these graphs, we observe that the disinformation networks tend to be less spread than the journalists networks. There are also more isolated communities in the journalist case, and the size of their nodes (the number of retweets) is smaller, evidencing their lower rate of interaction with the rest of the network.

Second, we found that the disinformation network density tended to be higher during periods with increased post and hashtag volumes, suggesting that the network becomes denser when it resonates or orchestrates a communication campaign. Some examples of this behavior are shown in Fig. 8, where the reported metrics are computed weekly on the disinformation network and plotted against their density. While no clear correlations emerged regarding posts related to COVID-19, we observed that the network displayed increased density when its focus on Ukraine or NATO intensified. This could indicate a coordinated effort by state actors or organized groups around these topics.

Similarly, when contrasting the network density against the negative emotion of the information, we observed periods of increased network density coincided with more negative sentiment. This may imply that these possible campaigns or coordinated actions are deeply emotional, for example, by using more aggressive or strong vocabulary. It is interesting to observe that, among the five variables analyzed with respect to density, this dimension achieved the highest R value, indicating a stronger relation between those variables.



(a) Journalists Network in 2019 (b) Journalists Network in 2022

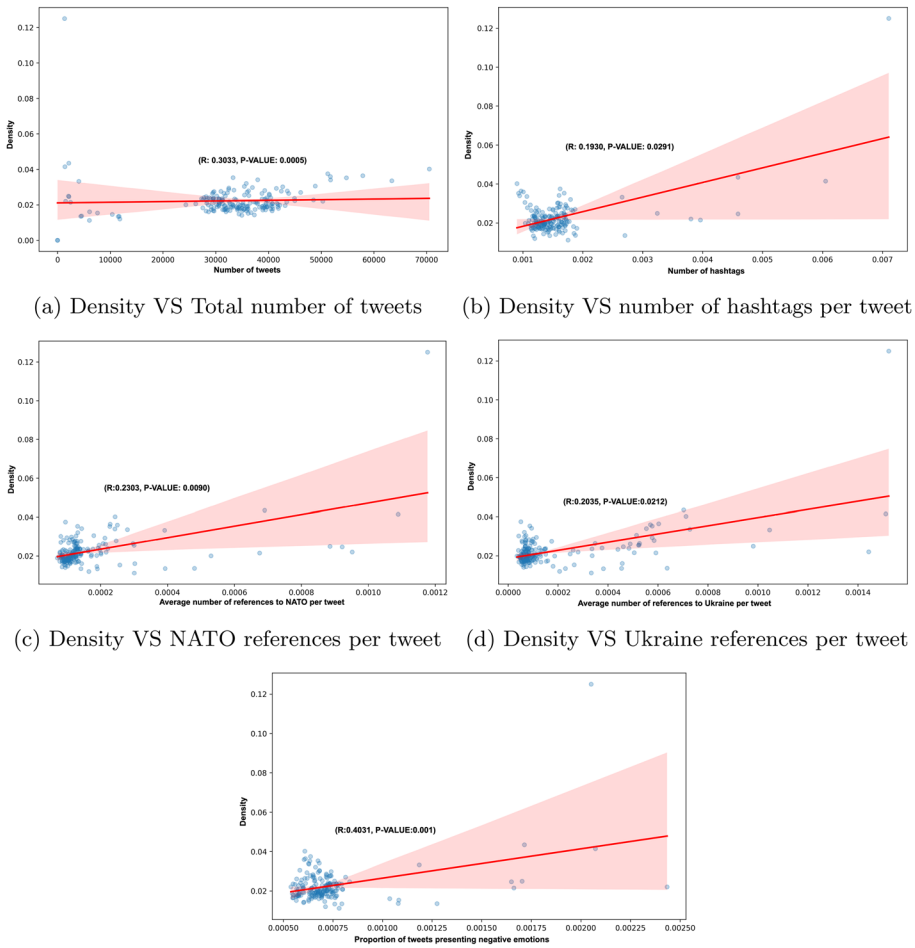
(c) Disinformation Network in 2019 (d) Disinformation Network in 2022

Fig. 7 Retweet graphs of the Journalists and Disinformation networks captured in 2019 and 2022

The correlations we discovered were overall weak (except, to some extent, with respect to the tweet sentiment), making it difficult to conclusively establish a cause-effect relationship between the network's shape and the nature of its content. However, these relationships offer intriguing insights, such as those already discussed. Nonetheless, we noted an improvement in efficiency when the total number of retweets within the network was higher and when a more significant proportion of published tweets contained URLs (see Fig. 9). This suggests that the disinformation network becomes more efficient when it absorbs and disseminates information, demonstrating the remarkable capacity of these networks to facilitate information flow.

Discussion

In the preceding sections, we have examined the defining characteristics of disinformation networks on Twitter and outlined their potential operation strategies within the Spanish communication space. In this section, we will interpret these findings and their implications for understanding and mitigating the impact of



(a) Density VS Total number of tweets (b) Density VS number of hashtags per tweet

(c) Density VS NATO references per tweet (d) Density VS Ukraine references per tweet

(e) Density VS Number of tweets with negative emotion per tweet

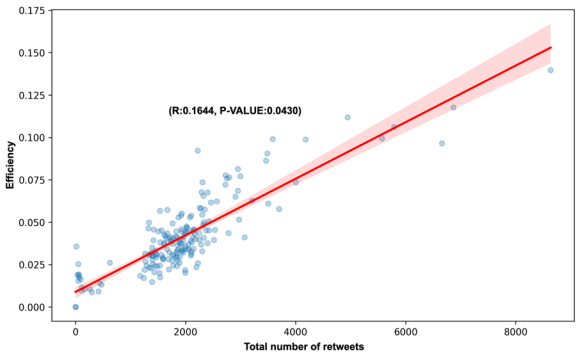
Fig. 8 Scatter plots between density and other variables in the disinformation network, including the p-value and the goodness of fit (R) of a linear fit on such data

disinformation. Furthermore, we will acknowledge the limitations of our study and outline prospective directions for future research in this domain.

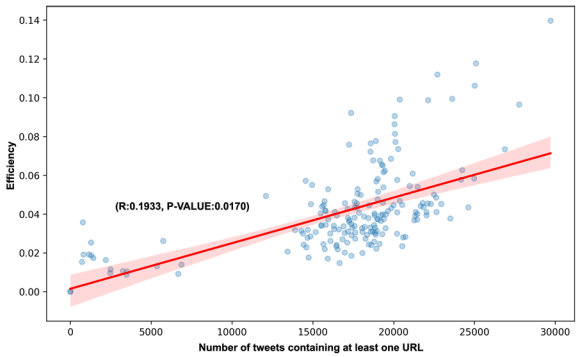
Implications for understanding disinformation networks

Firstly, the actors involved in disseminating disinformation are markedly more active and work intricately within their networks. This heightened level of activity, combined with a strong interconnectedness, allows them to amplify their visibility and draw more attention to their narratives.

Moreover, disinformation actors can coordinate their efforts, particularly during specific campaigns or around contentious topics. Our study showcases such coordination in the case of NATO-related narratives. This phenomenon aligns with substantial research and spotlights nation-states as key disinformation actors within online networks, by using, for example, a network of both human-operated and automated



(a) Efficiency VS total edge weight (total number of retweets between users in the network)



(b) Efficiency VS number of tweets containing URLs

Fig. 9 Scatter plots considering efficiency in the disinformation network

accounts (known as 'bots') to disseminate misleading narratives, amplify divisive content, and create a false impression of grassroots support (or opposition) to specific issues (a tactic known as 'astroturfing') (Stengel 2019). In addition to these overarching strategies such countries employ, there are also standard tools and techniques used in these disinformation campaigns; conspiracy theories, health misinformation, and the propagation of extreme political narratives are among the most frequently observed (Kalathil and Boas 2018).

A noteworthy pattern we observed is the surge in the activity of disinformation networks during times of social crisis. During these periods, they actively disseminate URLs, especially those linked to disinformation media outlets and sources known to spread fake news. By exploiting social vulnerabilities and heightened emotions during crises, they amplify their influence, reaching a broader audience.

Therefore, given their higher density, increased levels of activity, and unique network structure, disinformation networks on Twitter possess a significant potential to captivate users. Once these users fall into the network, they are more likely to be exposed to and receive false or biased information and propaganda faster than legitimate information. Hence, for those users who already belong (probably inadvertently) to such disinformation networks, the spread and impact of misleading narratives

would be exacerbated, in particular, when compared to users belonging to journalists or other neutral actors within the social network.

Strategies for mitigating the impact of disinformation

Our research underscores that disinformation networks distinguish themselves through their notably higher density, specialized structure, and proficient communication flows. Therefore, to effectively mitigate the spread of disinformation, we must devise strategies that target these unique characteristics.

One of the key strategies involves disrupting the intricate web of connections within disinformation networks. These networks function effectively because of the interconnectedness of their actors, who continuously reinforce each other's messages through retweets, participation in hashtags linked to disinformation campaigns, or the widespread sharing of fraudulent news. Interrupting this reinforcement chain would undermine the network's efficiency and reach, limiting its impact. Advanced algorithms can be developed to identify and shut down inauthentic accounts, thereby disrupting these networks at their core.

Simultaneously, it is of high importance to analyze network activity patterns to understand and identify coordinated inauthentic behavior. This analysis will provide a foundation for targeted interventions and astroturfing detection mechanisms. Policy recommendations for social media platforms can be developed to enforce stricter measures against coordinated inauthentic behaviors, enhancing the overall integrity of the information ecosystem.

Reinforcing the links within networks disseminating legitimate information is equally important. This strategy increases the spread and visibility of accurate information and provides a counter-narrative to disinformation. Moreover, strengthening these networks would equip users to resist the influence of disinformation networks and help create a more balanced information ecosystem on social media platforms, even though people may struggle to change their beliefs even after finding out that the presented information is incorrect or misleading (Garrett et al. 2013). Public awareness campaigns and media literacy programs are essential in educating users to recognize and respond to disinformation. Collaboration with independent fact-checkers will aid in quickly debunking false narratives, reducing their spread and impact.

The urgency of these actions is especially pronounced during periods of social unrest, electoral contexts, or events with the potential to disrupt public security significantly. During those volatile times, disinformation networks are often the most active and have the highest potential to cause harm.

The urgency of these actions is especially pronounced during periods of social unrest, electoral contexts, or events with significant potential to disrupt public security. During these volatile times, disinformation networks are often the most active and harmful. In this endeavor, we recognize the pivotal role that recommendation systems play (Ricci et al. 2022). These systems, which are responsible for content distribution on platforms

Table 2 Roles and responsibilities of various actors in preventing disinformation

Actor	Possible actions to prevent disinformation
Social media platforms	Deleting fake accounts Breaking disinformation networks Reinforcing legitimate news sources Implementing advanced algorithms for detecting inauthentic behavior Adjusting recommendation systems to deprioritize disinformation
Governments and institutions	Promoting media literacy campaigns Developing and enforcing policies against coordinated inauthentic behavior Collaborating internationally to tackle cross-border disinformation Funding research on disinformation spread and its impact
Users	Engaging in media literacy education Learning to recognize and respond to disinformation Using critical thinking to assess the credibility of information Reporting suspicious or misleading content to platforms
Independent fact-checkers and NGOs	Identifying and debunking disinformation quickly Collaborating with social media platforms to highlight accurate information Educating the public about identifying fake news
Researchers and academics	Conducting behavioral studies on disinformation spread Developing new tools and methods to detect and analyze disinformation networks Collaborating with platforms and governments to provide insights and recommendations
International bodies and coalitions	Facilitating cross-border cooperation in combating disinformation Establishing global standards and protocols for information integrity Coordinating efforts among member states to address disinformation challenges

like Twitter (Gupta et al. 2013), can be leveraged strategically to minimize the visibility of disinformation. By deprioritizing content from disinformation accounts—especially within the disinformation networks themselves—these systems could weaken the disinformation networks' structure and efficiency.

However, it is essential to recognize that content does not exist in isolation - it circulates within networks. Strategies to combat disinformation should focus not just on individual users, but also on the broader network dynamics. Understanding and altering information flow dynamics at the network level enables more effective impediment of disinformation spread and boosts the spread of accurate, reliable information. Furthermore, international cooperation against cross-border disinformation and investment in research on behavioral patterns of disinformation spread will provide a holistic approach to combating this global issue.

By incorporating these multifaceted strategies, we aim to create a more resilient and informed digital community, equipped to resist the influence of disinformation networks and foster a balanced information ecosystem on social media platforms. Finally, in Table 2 we summarize the roles and responsibilites that different actors may have in preventing disinformation, by implementing the strategies discussed before.

Limitations and future research directions

While our research provides illuminating insights into the behavior of disinformation networks, it is essential to acknowledge certain limitations within our study. Firstly, the scope of our research was primarily concentrated within the Spanish communication space in Spain. As such, the samples studied, albeit systematically and rigorously collected, are specific to this geographic and cultural context. Extending this research to include other linguistic and cultural contexts could provide a more comprehensive understanding of the global dynamics of disinformation.

Additionally, while Twitter remains a widely used platform for information dissemination, it is only one of many social media platforms, each with its unique dynamics. Hence, the behaviors and patterns observed on Twitter may not completely represent disinformation strategies across all platforms. Furthermore, the rapidly evolving social media landscape at the time of writing this article may introduce new dynamics that could either exacerbate or mitigate the disinformation strategies we have studied.

Our study spans more than three years, a substantial period that provides consistent and relevant results. However, it is crucial to consider that some of the accounts studied may have been active before the start of our research period. This prior activity could have influenced the network structure and dynamics we observed, but was not accounted for in our analysis.

Finally, when studying coordination within disinformation networks, it is crucial to understand that such coordination can arise spontaneously due to shared interests or ideologies among individuals consuming and producing content. However, there may also be more calculated and organized strategies being managed on other platforms or in offline environments. Distinguishing between these two types of coordination can be challenging, and our study may need to be extended in the future to capture this complexity fully.

Despite these limitations, our research offers valuable insights into disinformation networks' behavior and strategies, contributing to the broader understanding of how false information spreads and how it can be mitigated. Future research should address these limitations, broadening the scope and deepening the understanding of the dynamics at play.

Future research avenues should investigate disinformation networks across various linguistic and cultural communities. English, Russian, Arabic, and Chinese represent significant sectors of the global internet user base, each with its cultural nuances and potential variations in disinformation dynamics. A comparative analysis across such diverse linguistic and cultural backgrounds would undoubtedly enrich our understanding of the global patterns of disinformation and its impact.

Further research could delve deeper into the types of media shared within these networks by developing a more refined taxonomy. This approach could yield insights into political biases and the most successful disinformation narratives, and identify political groups exhibiting higher levels of coordination and efficiency. Understanding the types of narratives that gain traction within these networks could inform more targeted and effective countermeasures.

The replication of our study on other social media platforms, such as microblogging or general-purpose networks, is another vital avenue to explore. Given the varying dynamics across different platforms, it would be invaluable to ascertain whether the patterns we observed on Twitter are consistent across other platforms or if each platform presents unique challenges and opportunities in combating disinformation.

While our study focused on journalists as primary disseminators of legitimate information, future research could incorporate other influential user categories. These could include politicians, influencers, or cyber activists, whose roles in the information dissemination process could significantly impact the spread of (dis)information and the efficacy of countermeasures.

The most critical focus for future research, however, should be developing and implementing strategies designed to disrupt disinformation networks and enhance the efficiency of legitimate information dissemination. Such strategies could range from redesigning recommendation systems to implementing more sophisticated communication campaigns that target specific areas of these harmful networks.

Developing these strategies necessitates an understanding that tackling disinformation is not merely about fact-checking or debunking individual false narratives. Instead, it requires a strategic shift in the information flows within and between these networks. This includes re-engineering algorithms that govern these flows, changing the incentives for sharing information, and creating an environment that fosters critical information consumption among users.

In essence, the battle against disinformation is a contest over the control and direction of information flows. As such, dismantling disinformation networks involves disrupting the existing harmful flows and proactively shaping beneficial ones. By focusing on these two dimensions, we can disrupt these networks and mitigate their impact. This represents a challenging but essential task for researchers, policymakers, and practitioners committed to preserving the integrity of our information ecosystems.

Conclusions

This research delves into the structure and behavior of Twitter accounts associated with legitimate journalists and disinformation actors. Our data set spans from 2019 to mid-2022, encompassing various accounts and their activities. The focus of our study is to illuminate how these diverse actors form networks within the Twitter platform and engage in distinct dynamics of content production and sharing. Our findings reveal that disinformation actors form considerably denser networks than journalists create, which underscores clear signs of coordination within their information-sharing dynamics. This characteristic is critical as it indicates a calculated, collective approach to disseminating disinformation, contributing to its pervasive nature on the platform.

Moreover, our analysis utilizes network metrics such as efficiency, which measures the speed at which information propagates within a network. We observe that information within disinformation networks flows considerably faster than networks formed by legitimate journalists, which exhibit a higher degree of fragmentation. This faster

propagation of information allows for the rapid and widespread distribution of disinformation, often outpacing the dissemination of corrective or countering information from legitimate sources. This contrast between the behavior of disinformation networks and legitimate information sources offers insights into the challenges of countering disinformation on Twitter. The orchestrated network structure and efficient information dissemination within disinformation networks pose significant obstacles to mitigating the impact of misinformation on the platform. By shedding light on these dynamics, our study contributes valuable insights to the ongoing discourse on tackling the disinformation crisis in the digital age.

In summary, disinformation networks demonstrate a unique capacity for adaptability, elevating their density levels during periods of heightened social controversy, such as the war in Ukraine or debates concerning NATO. This heightened activity often correlates with a more negative sentiment within the network, hinting at possible coordinated actions. Though the correlation is not definitive and further investigation is required, this trend aligns with the discourse typically driven by nation-states around such topics.

This superior efficiency of disinformation networks in communication flow and their adaptive nature underscores the challenges in combating disinformation. Nevertheless, it also points to potential avenues for intervention. For instance, strategies that fracture the efficiency and density of these disinformation networks could be particularly impactful. As such, future research should delve into network activation and coordination mechanisms and expand to include other national and cultural contexts. Another potential intervention in these scenarios may include evidencing reasons or contexts behind specific tweets, as presented recently in Li et al. (2022), probably not to everyone in the network but depending on the characteristic of the information being sent (number of hashtags or URLs) or based on the sender/receiver.

Considering our findings, recommendation systems on platforms like Twitter could be valuable targets for future intervention research. The potential to influence these systems to disrupt the efficiency of disinformation networks while simultaneously enhancing the efficiency of legitimate networks may be a critical component in the fight against disinformation. In the digital age, such strategic interventions are more critical than ever for preserving the integrity of our information ecosystems.

Appendix A: Additional information—disinformation and journalist accounts

In this section, to facilitate the reproducibility of our work, we present in Tables 3 and 4 the ids of the accounts used in this study. Note that, to preserve the safety of the users behind these accounts (journalists, in particular[3]), we avoid sharing the user names, even though they may be obtained through the API.

[3] See https://www.bu.edu/articles/2023/disinformation-researchers-under-attack-by-government-legislators/.

Table 3 Ids of accounts used in our study as disinformation actors

Id	Id	Id	Id	Id
1000025123263524864	1528872169	253020963	3250842464	535153760
1003133438	1533855511	2569400510	330361451	53789862
100731315	1536167761	257859379	333033292	54233321
1009550909540585473	1561703076	259799108	333348576	55279448
102454320	1613828468	259897306	333936316	555521701
1031419921	1617884742	2605337039	3366058611	558462908
1040334830	1623272011	2609524507	3430772032	564138118
1064953765814509569	163141341	262851272	343933160	590421119
106891797	164110029	266012628	345257174	597712811
107177786	165104029	2667821193	3468902956	59894497
1081168493091962880	165127264	266838221	347300732	601657931
108744588	16799023	268035270	348538097	606623283
1089985800069095426	1688470074	2693017699	357590912	612411163
1104344103179960320	1707867426	2716353140	363421116	617894888
1106569081854066690	176058394	2732715937	36674807	700503810
110922804	17636635	273924214	367070806	708482255
1110891412508340224	178852637	2755279044	369846834	714195188667846657
1112478389309505536	183661695	2766498805	37835750	714556579
1118059060098629632	183786438	278248787	381657036	720743812562337798
112134350	185143177	280081621	390387588	72732893
112170559	1884163777	2809357258	391344883	736185894089199616
1121998616	18856867	2824259531	393476699	745339783
112747809	1896481891	282675582	394229561	755494698584801280
113035227	1923495216	2827483187	399275188	761154976076926977
1133334569577586688	193095342	283409352	4035057615	762405116
114558569	193096110	285255977	407754987	762903092983541761
114741363	1931893196	2858434521	411577733	763100287028568064
1150056069022007298	195446876	287786986	411647930	765599356980498432
1154527447766962176	19599446	289894237	413277087	766221303632240640
115660898	199566583	2919036392	414962189	769562616003960832
116831511	200568348	2922924261	415022746	803388691477630976
1179525037	201517097	2932115764	416154050	804748838330335234
1194010389186527233	201957241	2965135588	416876488	810200597685272576
Id	**Id**	**Id**	**Id**	**Id**
119497599	203262579	2982700905	41880514	820497732
1199191479094304768	203555695	298993329	425924139	822016688749154305
1210905474754695168	207208127	299661475	435346412	826044679179362304
123975474	21263335	301045311	45013575	840631711427891200
1252255963	214731619	3022877042	452985859	84427144
1281521971	2242909302	303848470	461900216	845571660090671104
1283507407	229598421	3040732982	465085203	85119380
130376756	2333901440	3040948607	475202064	851492096674541569
130452219	2365896248	305514503	4826563611	852269288
1311971648	2372314050	307558964	4831408433	857303965
131795521	2382387620	3079813761	48351615	862585086050533380
13346352	2394020821	309341660	4838961	867818602791018496
135368243	2401859508	3104949454	48668581	877113807461646336
1355594084	2413234485	3106771385	488097570	881197285769654272
1357033094	2425563233	3131419456	488543082	891599857630236672

Table 3 (continued)

Id	Id	Id	Id	Id
138726004	2435331090	3133111667	49016599	893474107
1392054620	244077566	314429644	49616273	898740373
139903735	247379224	3171668783	502092248	90432924
141027991	247888588	3208050838	505731001	907246319781195776
145336121	2511075531	3214613968	51280043	909465013370413056
14575708	251290516	32169306	52422182	922357287481757697
1474986842	252016342	3239745664	532490808	923106269761851392

Table 4 Ids of accounts used in our study as journalists

Id	Id	Id	Id	Id
351566384	317226440	67383910	105507745	3359284941
210132467	121366287	78863928	275674102	22748103
2942519806	37062760	184831017	6503172	33313641
14932200	268875728	149635747	698104968105095170	2413025666
116908364	268429272	276480123	118032930	381059099
115793824	228483751	215815774	18932906	392670224
224589305	114037455	242606835	89784280	216755042
722817261795287041	245863642	796684728	171962889	928718891181838336
103841173	263780425	899764291385077760	402035518	1032383648
482053121	196623028	18627726	544781860	361497515
54235496	28550047	865821732770373633	174726190	20164993
843937475068346373	618952760	257962682	822000025	134917034
1082598510	270307259	270607088	283930140	875525911
239765900	392177110	231424061	3306429471	41562449
107153756	8076532	161237361	16694719	551405439
85384885	341988494	286949912	151513481	119404032
189102085	188699808	143455500	246764832	227977305
96639908	418123217	384893636	44336530	486855644
26557207	2575293810	316706708	1413746881	46061866
102977300	236421131	19232900	107759816	320863192
292464252	429796168	278205448	426874087	322003135
33698078	353756954	94144199	870518544	58788205
18944456	255652146	367308015	303131300	155242359
65369125	1015573542	769919	218844578	176297919
559055487	226196017	156630555	342171657	618166944
288881933	3131004953	84186668	228687267	252305989
82863268	159979641	731573	840592769320116224	139371136
301306806	47936941	505412617	879011415922704389	194543506
464057783	29491384	1239229933	268234381	16947439
374737533	139767585	263806815	114235426	14600838
583625672	6794952	106220868	46296077	3874812255

Id	Id	Id	Id	Id
95232591	235211719	407913953	3168171	225187854
522523887	780183727318106113	185985009	83808453	220693082
368859006	94026873	154925267	250092838	264816224
14831098	1701969248	912746615986900994	210913028	601338508

Table 4 (continued)

Id	Id	Id	Id	Id
129636906	2196246424	44651546	87815477	219804554
58487243	87768187	371830459	26211178	12822592
205457327	775988391523586048	1660775364	798904559406120960	538978461
280172465	313886137	112796335	135275648	81379216
65609667	149831017	845237932717985793	345371701	19339140
268748579	21121637	127550968	1068208362	401371220
243569143	59087132	23675375	489342588	143422938
11120292	396059470	309705905	144810157	92147974
50598703	161629977	543731127	832986477197983753	266563663
279541793	1222673754	377680686	2781048212	704022213792546816
14371600	88146816	20637082	1186799853378121728	1080173945440141312
562931470	13937642	100502343	60959278	1095068646
16276054	287662628	2345139698	309097402	245847198
224202948	38720717	1387925532	58489786	247031000
15056194	301949958	229259339	191007783	335223244
3245066614	158730655	295854911	16837276	270916490
489848723	87748292	216056423	271519821	1553130476
195909630	106467821	20388141	209239240	17897369
557099769	296763063	607923199	1144954350437113856	531442414
251820322	92377922	428486667	277416367	1175479403868016642

Acknowledgements
The authors thank the reviewers for their thoughtful comments and suggestions.

Author contributions
PM: Conceptualization, Methodology, Software, Validation, Formal analysis, Visualization, Writing—Original draft, Writing—Review and Editing. FD: Conceptualization, Methodology, Formal analysis, Supervision, Writing—Original draft, Writing—Review and Editing. AB: Conceptualization, Methodology, Formal analysis, Supervision, Writing—Original draft, Writing—Review and Editing.

Funding
This work has been supported by grant PID2019-108965GB-I00 funded by Ministerio de Ciencia e Innovación and grant PID2022-139131NB-I00 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe".

Availability of data and materials
The datasets generated and analyzed during the current study are not publicly available to comply with Twitter/X Developer Agreement, but are available from the corresponding author on reasonable request.

Declarations

Competing interests
The authors declare that they have no competing interests.

Received: 23 October 2023 Accepted: 14 January 2024
Published online: 26 January 2024

References

Allcott H, Gentzkow M (2017) Social media and fake news in the 2016 election. J Econ Persp 31(2):211–236
Alliance4Democracy: Hamilton 2.0 Dashboard (2022). Accessed Jul 2023 https://securingdemocracy.gmfus.org/hamilton-dashboard/,
Azzimonti M, Fernandes M (2018) Social media networks, fake news, and polarization. Technical report, National Bureau of Economic Research
Bastos MT, Mercea D (2019) The brexit botnet and user-generated hyperpartisan news. Soc Sci Comput Rev 37(1):38–54
Bastos MT, Mercea D, Baronchelli A (2020) The brexit botnet and user-generated hyperpartisan news. Soc Sci Comput Rev 38(1):38–54
Bazmi P, Asadpour M, Shakery A (2023) Multi-view co-attention network for fake news detection by modeling topic-specific user and news source credibility. Inf Process Manag 60(1):103146. https://doi.org/10.1016/j.ipm.2022.103146

Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech Theory Exp 2008(10):10008. https://doi.org/10.1088/1742-5468/2008/10/P10008
Brin S, Page L (1998) The anatomy of a large-scale hypertextual web search engine. Comput Netw 30(1–7):107–117. https://doi.org/10.1016/S0169-7552(98)00110-X
Bruns A, Harrington S, Hurcombe E (2018) Click, share, send, forget: the dynamics of news diffusion via twitter. J Stud 19(11):1559–1579
Chaffee SH, Metzger MJ (2001) The end of mass communication? Mass Commun Soc 4(4):365–379
Challenge FN (2019) Fake news challenge: a dataset and competition for fake news detection. Accessed Jul 2023 from http://www.fakenewschallenge.org/
Conover MD, Ratkiewicz J, Francisco M, Gonçalves B, Menczer F, Flammini A (2011) Political polarization on twitter. ICWSM 133:89–96
D'Ulizia A, Caschera MC, Ferri F, Grifoni P (2021) Repository of fake news detection datasets. Version 1. 4TU.ResearchData. Dataset. Accessed Jul 2023 from https://doi.org/10.4121/14151755.v1
Ellul J (2021) Propaganda: the formation of men's attitudes. Vintage, New York, NY
Fallis D (2015) What is disinformation? Library Trends 63(3):401–426
Flaxman S, Goel S, Rao JM (2016) Filter bubbles, echo chambers, and online news consumption. Publ Opin Quart 80(S1):298–320
Garat JR (2023) Ucrania: La desinformación como arma de guerra. Cuadernos de pensamiento naval: Suplemento de la revista general de marina 35:33–52
Garrett RK, Nisbet EC, Lynch EK (2013) Undermining the corrective effects of media-based political fact checking? the role of contextual cues and naïve theory. J Commun 63(4):617–637
Grinberg N, Joseph K, Friedland L, Swire-Thompson B, Lazer D (2019) Fake news on twitter during the 2016 us presidential election. Science 363(6425):374–378
Guarino S, Trino N, Celestini A et al (2020) Characterizing networks of propaganda on twitter: a case study. Appl Netw Sci. https://doi.org/10.1007/s41109-020-00286-y
Guenther L, Ruhrmann G, Zaremba MC, Weigelt N (2021) The newsworthiness of the march for science in germany: comparing news factors in journalistic media and on twitter. JCOM 20(02):03
Guess A, Nyhan B, Reifler J (2019) Exposure to untrustworthy websites in the 2016 us election. Nat Hum Behav 3(4):1–9
Gupta P, Goel A, Lin J, Sharma A, Wang D, Zadeh R (2013) WTF: the who to follow service at twitter. In: Schwabe D, Almeida VAF, Glaser H, Baeza-Yates R, Moon SB (eds) 22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13–17, 2013, pp. 505–514. International World Wide Web Conferences Steering Committee / ACM, New York, NY. https://doi.org/10.1145/2488388.2488433
Ha L, Perez LA, Ray R (2021) Mapping recent development in scholarship on fake news and misinformation, 2008 to 2017: Disciplinary contribution, topics, and impact. Am Behav Sci 65(2):290–315
Henderson EH (1943) Toward a definition of propaganda. J Soc Psychol 18(1):71–87
Himelboim I, Smith MA, Shneiderman B, Park S (2013) Birds of a feather tweet together: Integrating network and content analyses to examine cross-ideology exposure on twitter. J Comput Medi Commun 18(2):154–174
Huckin T (2016) Propaganda defined. In: Propaganda and Rhetoric in Democracy: History, Theory, Analysis, pp. 118–136
Iyengar S, Kinder DR (1987) News that Matters: Television and American Opinion. University of Chicago Press, Chicago, IL
Jing J, Wu H, Sun J, Fang X, Zhang H (2023) Multimodal fake news detection via progressive fusion networks. Inf Process Manag 60(1):103120. https://doi.org/10.1016/j.ipm.2022.103120
Jungherr A, Jürgens P, Schoen H (2012) Why the pirate party won the german election of 2009 or the trouble with predictions: A response to (2011) tumasjan, a., sprenger, t. o., sander, p. g., & welpe, i. m. Soc Sci Comput Rev 30(2):229–234
Kalathil S, Boas TC (2018) The rise of digital repression: How technology is reshaping power, politics, and resistance. J Democr 29(3):41–55
Kouzy R et al (2020) Coronavirus goes viral: quantifying the covid-19 misinformation epidemic on twitter. Cureus 12:3
Lazer DM, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, Schudson M (2018) The science of fake news. Science 359(6380):1094–1096
Lewandowsky S, Ecker UK, Cook J (2012) Misinformation and its correction: Continued influence and successful debiasing. Psychol Sci Pub Inter 13(3):106–131
Lewandowsky S, Ecker UK, Cook J (2017) Beyond misinformation: understanding and coping with the post-truth era. J Appl Res Memory Cognit 6(4):353–369
Li Z, Hu H, Wang H, Cai L, Zhang H, Zhang K (2022) Why does the president tweet this? discovering reasons and contexts for politicians' tweets from news articles. Inf Process Manag 59(3):102892. https://doi.org/10.1016/j.ipm.2022.102892
Magelinski T, Ng L, Carley K (2022) A synchronized action framework for detection of coordination on social media. J Online Trust Saf. https://doi.org/10.54501/jots.v1i2.30
Marwick A, Lewis R (2017) Media manipulation and disinformation online. Technical report, Data Society Research Institute
McCombs ME, Shaw DL (1972) The agenda-setting function of mass media. Publ Opin Quarter 36(2):176–187
Mohd Shariff S, Zhang X, Sanderson M (2014) User perception of information credibility of news on twitter. In: de Rijke, M.e.a. (ed) Advances in Information Retrieval. ECIR 2014. Lecture Notes in Computer Science, vol 8416. Springer, Cham
Molyneux L, Vasconcelos AC, Breen L (2020) Journalists as sensemakers: Sensemaking in the context of covid-19. J Stud 21(13):1733–1749
Moroz O, Loza A (2022) YouControl, Database of Russian propagandists. Accessed Jul 2023 https://youcontrol.com.ua/en/articles/database-of-russian-propagandists/
Newman MEJ (2010) Networks: an introduction. Oxford University Press, Oxford
Ng LHX, Cruickshank IJ, Carley KM (2022) Cross-platform information spread during the january 6th capitol riots. Soc Netw Anal Min 12(1):133. https://doi.org/10.1007/S13278-022-00937-1
Nielsen RK, Fletcher R, Newman N, Brennen JS, Howard PN (2020) Navigating the infodemic: How people in six countries access and rate news and information about coronavirus. Reuters Institute for the Study of Journalism

OJD: Principales medios de comunicación en España (2022). Accessed Jul 2023 https://www.ojdinteractiva.es/medios-digitales
Pavlíková M, Šenkýřová B, Drmola J (2021) Propaganda and disinformation go online. In: Challenging Online Propaganda and Disinformation in the 21st Century, pp. 43–74
Pérez JM, Giudici JC, Luque F (2021) pysentimiento: A Python Toolkit for Sentiment Analysis and SocialNLP tasks. arXiv:2106.09462
Pérez-Escolar M, Lilleker D, Tapia-Frade A (2023) A systematic literature review of the phenomenon of disinformation and misinformation. Med Commun 11(2):76–87
Ratkiewicz J, Conover M, Meiss M, Gonçalves B, Patil S, Flammini A, Menczer F (2011) Detecting and tracking political abuse in social media. In: Proceedings of the fifth international AAAI conference on weblogs and social media, pp 297–304
Ricci F, Rokach L, Shapira B (eds) (2022) Recommender systems handbook. Springer, New York. https://doi.org/10.1007/978-1-0716-2197-4
Ruohonen J (2021) A few observations about state-centric online propaganda. CoRR. arXiv:2104.04389
Sanz-Cruzado J, Castells P (2018) Enhancing structural diversity in social networks by recommending weak ties. In: Pera S, Ekstrand MD, Amatriain X, O'Donovan J (eds) Proceedings of the 12th ACM Conference on Recommender Systems, RecSys 2018, Vancouver, BC, Canada, October 2–7, 2018, pp 233–241. ACM, New York, NY. https://doi.org/10.1145/3240323.3240371
Saxena N, Sinha A, Bansal T, Wadhwa A (2023) A statistical approach for reducing misinformation propagation on twitter social media. Inf Process Manag 60(4):103360. https://doi.org/10.1016/j.ipm.2023.103360
Shao C, Ciampaglia GL, Varol O, Yang KC, Flammini A, Menczer F (2018) The spread of low-credibility content by social bots. Nat Commun 9(1):1–10
Shu K, Mahudeswaran D, Wang S, Lee D, Liu H (2018) Fakenewsnet: A data repository with news content, social context and dynamic information for studying fake news on social media. arXiv:1809.01286
Singer-Vine J (2016) BuzzFeedNews: A dataset of fact-checked articles from BuzzFeed News. Accessed Jul 2023 https://github.com/BuzzFeedNews/2016-10-facebook-fact-check
Smart B, et al (2022) # istandwithputin versus# istandwithukraine: the interaction of bots and humans in discussion of the russia/ukraine war. In: International Conference on Social Informatics. Springer, Cham
Starbird K, Spiro E, Arif A, Wilson T (2019) Disinformation as collaborative work: surfacing the participatory nature of strategic information operations. Proc ACM Hum Comput Inter 3:1–27
Statista: Principales medios en España (2022). Accessed Jul 2023 https://es.statista.com/estadisticas/476795/periodicos-diarios-mas-leidos-en-espana/,
Stengel R (2019) Information Wars: How We Lost the Global Battle Against Disinformation and What We Can Do About It. Grove Press, New York, NY
Tandoc EC, Lim ZW, Ling R (2018) Defining fake news. Digit J 6(2):137–153
TaskForce E (2022) EuVSDisinfo, DISINFO DATABASE. Accessed Jul 2023 https://euvsdisinfo.eu/disinformation-cases/
Törnberg P, Carlsson U, Clerwall C (2020) Disinformation and the European parliament election 2019: A case study of the European union Stratcom task force. Med Commun 8(2):411–421
Tucker JA et al (2018) Social media, political polarization, and political disinformation: A review of the scientific literature. Political polarization, and political disinformation: a review of the scientific literature
Vosoughi S, Roy D, Aral S (2018) The spread of true and false news online. Science 359(6380):1146–1151
Wang W (2017) liar, liar pants on fire: A new benchmark dataset for fake news detection. https://doi.org/10.18653/v1/p17-2067
Wardle C, Derakhshan H (2017) Information disorder: Toward an interdisciplinary framework for research and policy making. Council of Europe report
Xu Y, Zhou D, Wang W (2023) Being my own gatekeeper, how I tell the fake and the real - fake news perception between typologies and sources. Inf Process Manag 60(2):103228. https://doi.org/10.1016/j.ipm.2022.103228
Zhou C, Xiu H, Wang Y, Yu X (2021) Characterizing the dissemination of misinformation on social media in health emergencies: An empirical study based on COVID-19. Inf Process Manag 58(4):102554. https://doi.org/10.1016/j.ipm.2021.102554
Zimdars M (2017) BigMcLargeHuge/opensources. Accessed Jul 2023 https://github.com/BigMcLargeHuge/opensources/blob/master/sources/sources.csv
Zubiaga A, Aker A, Bontcheva K, Liakata M, Procter R, Tolmie P (2018) Detection and resolution of rumours in social media: A survey. ACM Comput Surv (CSUR) 51(2):1–36

Publisher's Note