Coordinated amplification, coordinated inauthentic behaviour, orchestrated campaigns

A systematic literature review of coordinated inauthentic content on online social networks

de-Lima-Santos, M.-F.; Ceron, W.

DOI
10.4324/9781003403203-14

Publication date
2024

Document Version
Final published version

Published in
Mapping Lies in the Global Media Sphere

10 Coordinated Amplification, Coordinated Inauthentic Behaviour, Orchestrated Campaigns

A Systematic Literature Review of Coordinated Inauthentic Content on Online Social Networks

Mathias-Felipe de-Lima-Santos

University of Amsterdam,
Federal University of São Paulo

Wilson Ceron

Federal University of São Paulo

Introduction

The internet and online social networks have resulted in dramatic changes in the information landscape. While some optimistic views see this movement as an opportunity to foster participation and diversity, more sceptical perceptions fear that this communication through networks and algorithms has limited exposure to a variety of content by exposing users to pre-existing beliefs, such as echo chambers (Sunstein, 2017) and filter bubble (Pariser, 2012). Furthermore, this scepticism is increased by the rise of false and misleading information on online social networks, disseminated by malicious actors intending to deceive people and discredit democratic processes (Farkas & Schou, 2018; Wardle & Derakhshan, 2017). It is also proved that coordinated efforts have been a fertile ground for mis- and disinformation on different social media platforms (Broniatowski, 2021; Freelon & Wells, 2020; Keller et al., 2020).[1]

Furthermore, there is a boom in the use of artificial intelligence (AI) methods to produce human-like text to deceive people (Huijstee et al., 2021; Ng & Taeihagh, 2021). For example, recent investigations have shown that ChatGPT can produce clean, convincing text that could generate repeated conspiracy theories and misleading narratives (Hsu & Thompson, 2023). Some have warned that AI generative and automated content could be leveraged for mass deception or political overthrow (Hsu & Thompson, 2023). For example, one of the purposes of political astroturfing (i.e., hidden information

DOI: 10.4324/9781003403203-14

campaigns in which actors mimic genuine users' behaviour by incentivising the spread of information online) is to reach and change people's perceptions and conduct as many regular as possible, contingent on a broad reach and simulating an actual campaign (Schoch et al., 2022).

Given human susceptibility to deceiving content, scholars are examining how mis- and disinformation information spread on online social networks, using a variety of methods, such as bot detection (Spiegel Rubin et al., 2021), super spreaders (Yang et al., 2021) and coordinated activities (Giglietto et al., 2020a). Facebook, for example, adopted this "ill-defined" concept of "coordinated inauthentic behaviour" (CIB) to describe this type of content that uses automated tactics instead of establishing a clear boundary between problematic and non-problematic information. The platform has been employing CIB to remove content since 2018. Although there is an inevitable criticism that Facebook attempted to substantiate the relationship between coordinated behaviour and problematic information sharing as a solution to combat mis- and disinformation (Giglietto et al., 2020a), it is a crucial effort to mitigate its effect, as this response seems to work in a combination of different stopgap measures (Bode & Vraga, 2021). Thus, there has been a recent growing of interest within the scholarly community in detecting coordinated campaigns on online social networks, rather than focusing on the small groups responsible for instigating or sustaining these messages (Weber & Neumann, 2021).

Through a systematic literature review, this chapter locates and synthesises related research on coordinated inauthentic content on online social networks, described here as "coordinated campaigns." Our systematic review of existing literature on this topic: (i) Describes the state of this field by identifying the patterns and trends in the conceptual and methodological approaches, topics, and practices; and (ii) sheds light on potentially essential gaps in the field and suggests recommendations for future research.

To reveal the conceptual and empirical evidence of coordinated campaigns, we address the following research questions:

RQ1: How has current scholarship defined CIB or coordinated campaigns?
RQ2: How does the mis/disinformation scholarship address the problem of automated campaigns?
RQ3: What are the main challenges and constraints to detecting automated campaigns?

The collected database consists of 202 materials from Scopus®, 45 from Web of Science, 324 from ACM® and 21 from IEEE®. Publications without the chosen terms in the title, abstract and keywords were excluded. Duplicate studies were also excluded. Another exclusion criterion was scholar publications not written in English. Our final dataset was composed of 92 studies.

Findings show that there is an evolution of the approaches used to detect coordinated activities. While bot detection was the focus in the early years, more recent research focused on using advanced computational methods based on training datasets or identifying coordinated campaigns by timely and similar

content. It is important to highlight there are no perfect solutions, as data are limited, and all methods present certain caveats. Political, health and disinformation topics were predominantly found in these studies. However, coordinated activities could be seen in financial markets and promotion campaigns of artists. Due to the data availability, Twitter is by far the most studied platform, although studies have shown that coordinated activities can be found on other online social network platforms. We conclude by discussing the implications of current approaches and outlining an agenda for future research.

Methodological Approaches

Through organised, transparent and replicable processes that include predefined search strings and standard inclusion and exclusion criteria (Higgins et al., 2011; Mohamed Shaffril et al., 2021), this study provides an overview of this topic, comparing its synonym and how scholars methodologically approach it. Furthermore, it is built on existing evidence, allowing researchers to identify gaps and directions for future research. Qualitative techniques of pattern matching and explanation building have been employed to categorise descriptively these published studies, highlighting their commonalities and disparities using the eyeballing technique (Bhimani et al., 2019). Therefore, a descriptive, rather than a statistical, analysis of results is presented in this chapter.

Tranfield et al. (2003) laid out a three-stage procedure for producing a systematic literature review: Planning, execution and guided reporting. The first step is to set the research objectives that support a broad scan of articles. This study focuses on peer-reviewed journal articles, as they are considered to have the most significant impact on research integrity and retention (Podsakoff et al., 2005), and conference proceedings, which is quite common in computer science scholarship.

The second step was selecting the databases from which the initial list of articles would be retrieved. As this topic lies in computer science, social sciences and humanities, we relied on four databases. The Scopus® and Web of Science® were chosen for the social science and humanities data collection. They offer a broad range of indexed content from thousands of journals because of their relevance to this scientific literature. To cover the computer science academic literature, we included IEEE® and ACM®, the most extensive databases for Science and technology studies (STS).

Third, we relied on a combination of keywords derived from Weber and Neumann (2021)'s work to search for relevant studies. To include the broad range of definitions surrounding CIB, we conducted a snowball collecting other terms in the studies listed in the previous studies. In total, 65 terms were found, such as "URL Sharing Behaviour," "orchestrated campaign," "malicious retweeter," "automate tactic," etc. (see Table 10.1).

Using these keywords described in Table 10.1, we searched them in the title, abstract, keywords and manuscript of the four databases. Our data collection happened in Q2 2022, retrieving a total of 592 articles. As shown in Figure 10.1, we adopted a series of inclusion and exclusion criteria. We

Table 10.1 Terms used in the search strings

Terms	*Alternative*
automate tactic	
automated account	
automated activity	
automated and orchestrated manipulation	
automated behavior	automated behaviour
automated bots	
automated content spreader	
automated manipulation	
automated shell account	
automated social media account	
automated social software	
automated software	
automated-based information campaign	
centralized coordination	
coincidental behavior	coincidental behaviour
coordinated account	
coordinated action	
coordinated activities	coordinated activity
coordinated amplification	
coordinated astroturfing campaign	
coordinated attack	
coordinated behavior	coordinated behaviour
coordinated bot	
coordinated campaign	
coordinated communication	
coordinated disinformation campaign	
coordinated effort	
coordinated free text campaign	
coordinated groups	
coordinated human-run account	
coordinated inauthentic activity	
coordinated inauthentic behavior	coordinated inauthentic behaviour
coordinated influence	
coordinated link	
coordinated malinformation campaign	
coordinated manipulation	
coordinated misinformation campaign	
coordinated network	
coordinated online action	
coordinated political influence	
coordinated retweet activity	
coordinated spam message	
coordinated spread	
coordinated trolling	
coordinated way	
coordinating communication	
coordination detection algorithm	coordination detection algorithm
coordination network	

(*Continued*)

Table 10.1 (Continued)

Terms	*Alternative*
coordination of multiple accounts	
coordination pattern	
coordination strategies	coordination strategy
coordination tactic	
detecting synchronized action	
highly coordinating communities	
inauthentic account	
inauthentic information campaign	
inauthentic online behavior	inauthentic online behaviour
Link Sharing Behavior	Link Sharing Behaviour
malicious organized activities	malicious organized activity
malicious retweeter	
orchestrated bots	
orchestrated campaign	
patterns of coordination	
synchronized action	
URL Sharing Behavior	URL Sharing Behaviour

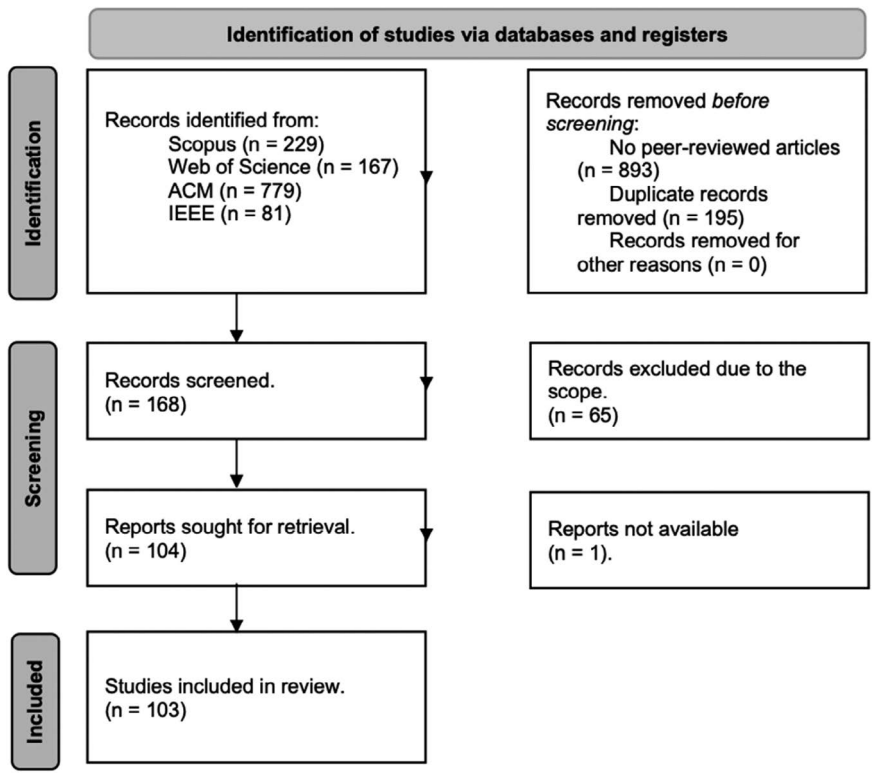

Figure 10.1 Flow diagram depicting the flow of information through the different phases of a systematic review.

excluded duplicate publications. Our exclusion criterion also includes studies not written in English.

Additionally, we performed a screening process where we included only articles that clearly covered the topic of coordinated campaigns in their abstract. It was not necessary to have specific words, but we would be able to identify if the publication includes this discussion.

Having performed these steps, our considered dataset consists of 92 publications, including peer-reviewed articles and conference proceedings. The articles were each coded for some variables, such as the study's objectives, topics covered, scope (online social network platforms), tools and methods used and limitations of such approaches. In a second step, all diversity-related terms appearing in the article were coded – if the topic was substantially discussed (i.e., appearing more than once outside of footnotes or references). After coding, these publications were thematically filtered and reported in descriptive narrative format in the following sections.

Coordinated Campaigns on Social Media Platforms

As a scarce and fluid commodity (Myllylahti, 2020), our attention has become critical to online social networks. Many actors tried to catalyse our attentional spotlight to amplify individual voices above the crowd (Pennycook et al., 2021). These interactions may be artificially inflated by various legitimate and fringe attention-hacking interventions (Giglietto et al., 2018), which are "capable of hijacking conversations, influencing other users, and manipulating content dissemination" (Khaund et al., 2022, p. 530). Using bots and sockpuppet accounts to amplify these individuals' voices above the crowd requires coordination, that is, "the additional information processing performed when multiple, connected actors pursue goals that a single actor pursuing the same goals would not perform" (Malone, 1988, p. 32). Sometimes referred to as the megaphone effect, coordinated action requires conversations to happen in a manner that is distinguishable from human contributions, regularly leaving traces in the form of digital records (Weber & Neumann, 2021) that can be revisited later to detect these coordinated actions.

In the literature, the deceptive practice known as astroturfing is also used to describe orchestrated marketing, public relations or propaganda activities that mask sponsors of a message or organisation to make it appear to have originated from and is supported by grassroots participants (Piña-García & Espinoza, 2022; Schoch et al., 2022). Similarly, Facebook adopted the concept of CIB to describe content that uses automated tactics and potentially causes harm to its users (Gleicher, 2018). According to the company, the use of fake accounts to "artificially boost the popularity of content or engage in behaviours designed to enable other violations under our Community Standards" are not allowed in its platforms (Broniatowski, 2021, p. 2). Scholars have criticised this approach, as the Silicon Valley company, instead

of establishing a clear boundary between problematic and non-problematic information, decided to put everything under the same umbrella indiscriminately, substantiating the relationship between coordinated behaviour and the sharing of problematic information (Giglietto et al., 2020a).

Events on online social networks gain momentum when a large audience is engaged in discussion. Thus, not all orchestrated campaigns have malicious intentions of feeding people with false information. For example, civil society organisations tend to use coordinated activities to maximise the visibility of specific content and call the attention of public and private entities (Schoch et al., 2022).

The scholarly literature describes these accounts that mimic the social behaviours of humans as social bots (Alothali et al., 2021; Khaund et al., 2022; Perna & Tagarelli, 2018a). However, bots are used interchangeably with fake (false) and spam accounts, which are not exactly the same. While fake (or false) accounts impersonate people on online social networks, spam accounts focus on disseminating unsolicited promotional content on a mass scale. Bots or automated accounts are mostly controlled by software to perform automatic interactions, such as posting messages, liking and sharing others' content. The main difference between automated and fake accounts is that the "former improves the metrics of itself while the latter improves the metrics of other users" and creates an unhealthy environment on online social networks (Akyon & Esat Kalfaoglu, 2019, p. 1). Some researchers use the term "inauthentic accounts" to refer to these collections of accounts (Yang & Menczer, 2022). The confusion is explained because these accounts often overlap, such as bots that automatically impersonate humans to post spam content.

Defining and distinguishing these account types also help to decide the proper interventions. Usually, fake and spam accounts violate platforms' policies, degrading the environment for users. Similarly, bots with malicious intent can disseminate false and misleading information to deceive users, exacerbate conflicts with manipulated opinions and disrupt communications (Yang & Menczer, 2022). However, some automated accounts, such as news bots, benefit online discussions (Lokot & Diakopoulos, 2016).

Twitter is the most news- and research-friendly of the online social platforms, as seen in the academic literature. As a result, most of the studies about coordinated campaigns were conducted using Twitter digital trace data (e.g., Fazil & Abulaish, 2020; Overbey et al., 2019; Sharma et al., 2021). Scholars are recently exploring other methods for tracking and analysing botnets on other platforms, such as Facebook. For example, Boshmaf et al. (2011a, 2011b) tested users' behaviour in response to a large-scale infiltration where social bots, showing that Facebook could not detect or stop extensive orchestrated campaigns. Similarly, Giglietto et al. (2020b) relied on URLs from Facebook posts to identify coordinated activities on political news stories published in the 2018 Italian general election and the 2019 Italian election for the European Parliament.

Scholars also employed a combination of computational methods to detect fake accounts on Facebook and Instagram, as previous requests for data from these platforms were available before Cambridge Analytical scandal (Akyon & Esat Kalfaoglu, 2019). Another study unveiled more than two million malicious accounts and 1156 large attack campaigns within one month on Facebook and Instagram (Cao et al., 2014). In a limited number of studies, researchers examined coordinated activities on other online social networks, such as VK and Reddit. In a cross-platform study, scholars analysed YouTube raids perpetrated by users of 4chan, a platform known for its controversies and alt-right communities. Despite these examples, few studies have explored these coordinated activities on other platforms, particularly messaging applications.

The Evolution: From Bots to Coordinated Actions

Throughout the years, digitally coordinated campaigns have been intended to undermine and disrupt public opinion, making scholars develop multiple methods to study them. While bot detection can seem a simple task for humans because we can observe emerging patterns or anomalies and evaluate conversational nuances such as sarcasm or persuasive language, machines do not have the same capacity (Khaund et al., 2018). Despite its limitation, detecting botnets' activity on Twitter has been successfully employed in the academic literature (e.g., Bastos & Mercea, 2018; Ferrara et al., 2016; Mendoza et al., 2021; Soto-Sanfiel et al., 2022) to identify coordinated activities on online social networks. For example, machine-learning techniques were applied to predict in real-time the type of account (human or bot) based on profile information (metadata) as features (Gilmary et al., 2023). According to these studies, bots spread rumours and false information, cyberbullying, spamming and manipulate the ecosystem of online social networks. Studies have also highlighted that bots share more news articles, fewer opinion tweets, no testimonial tweets and fewer conversational tweets than human users (Abokhodair et al., 2015).

However, the different number of bot detection techniques do not guarantee that all bot profiles are detected, as the methods are evolving, and malicious actors are creating orchestrated actions in online social networks that involve not only automated accounts. Some techniques are also not replicable, such as identifying false profiles during their creation on Facebook and Instagram, because these data are not publicly available and access is restricted to Meta employees, who conducted some studies (Akyon & Esat Kalfaoglu, 2019).

For this reason, Schuchard and Crooks (2021) proposed an ensemble bot detection coverage framework that harnesses the power of multiple detection sources to detect a wider variety of bots. They recognise the necessary efforts to incorporate numerous detection sources to account for the type of social bots operating in online social networks. Furthermore, this is important to

keep pace with the constant evolution of bot complexity, as malicious actors are incorporating improved or new techniques to overcome detection methods. Furthermore, scholars highlighted the importance of an explainable bot detection service, as AI-driven bot detection methods remain quite opaque and lack ethical responsibility, not contributing to the mitigation of coordinated campaigns (Kouvela et al., 2020).

Studies have also shown that bots participate in and contribute to online conversations in a manner that is distinguishable from human contributions. Using Benford's law tests for multiple user metrics, it was possible to identify that automatically controlled bots possibly disagree with it, while human-orchestrated bots follow a normal distribution. Similarly, social bots accounted for fewer than 1% of the total corpus of user contributors to online mass shooting conversations. Still, their significant prominence in networks could be recognised by centrality in these networks (Yuan et al., 2019). Some scholars relied on social network analysis methods to reach this conclusion. In fact, methods that rely on social connections and interactions between users by leveraging graph-based representation learning have been widely used to improve bot detection (Mendoza et al., 2021).

Community detection algorithms (Blondel et al., 2008) were used to get insight from the bot and human networks. Scholars observed that bots' communities are more hierarchical in structure. In other words, they have a central core of members who connect more strongly among themselves than the peripheral members, who are weakly connected (Abokhodair et al., 2015; Khaund et al., 2018). Similarly, human networks have more communities and tend to be smaller in size and denser than bot networks, that is, humans have more tightly knit and focused communities, while bots tend to be bigger. Consecutively, these connections have a weak sense of belongingness to a community (Khaund et al., 2018). Equally, graph-based unsupervised machine-learning methods were used to identify edge and node anomaly detection in social network data (Venkatesan & Prabhavathy, 2019).

Political astroturfing (centrally coordinated disinformation campaigns in which participants pretend to be ordinary users who act independently) is also a centrally coordinated disinformation campaign, as participants pretend to be ordinary users who act independently, aiming to reach and change the behaviour of as many regular users as possible. The campaign's success is contingent on a broad reach and an organic appearance of the coordinated activity, where central core members are highly influential in their networks. In contrast, grassroots movements exhibit some coordination but tend to be less synchronised in timing and content (Schoch et al., 2022). Although these coordinated activities tend to happen organically through cues sent by peers instead of centralised instructions, platforms have highlighted that it is hard for them to distinguish between coordinated campaigns from civil society organisations and malicious actors (Felipe, 2022). Therefore, graph features helped to detect orchestrated activities using network analysis approaches.

Similarly, experiments with social bots were promoted to quantify the infiltration effectiveness of different social coordinated strategies on online social networks (Boshmaf et al., 2011a, 2011b; Freitas et al., 2015). It is important to note that existing literature primarily focuses on bot detection and its roles in information campaigns rather than mapping coordinated actions. More recently, scholars have proposed an effective method to identify similar malicious activities on Facebook (e.g., Broniatowski, 2021; Giglietto et al., 2020b).

Topics that Emerged in Coordinated Campaigns

Prior study has shown that the health crisis and political events are periods when individuals are more likely to be exposed to mis- and disinformation, requiring fact-checking interventions (Ceron et al., 2021a, 2021b). Similarly, most topics used to study these orchestrated activities were related to elections and the COVID-19 pandemic. The US-focused campaigns were commonly found in these studies, restricting to a limited number of studies beyond this context. For example, a framework using three available bot detection sources was proposed to identify social bot activity within online social network interactions taking place during the 2018 US Midterm Election on Twitter. This framework aims to incorporate improved or new detection methods to keep pace with the constant evolution of bot complexity. Underlying socio-political processes behind the 2016 US Presidential Election, scholars used network science methods to study the social dynamics of automated accounts (Le et al., 2019). As a result, they could identify key groups associated with the US right wing promoting coordinated activities during the US election. Unexpectedly, a low number of automated accounts related to foreign intervention in the Trump-supporting group was detected.

Beyond the US, researchers analysed the dynamics of coordinated campaigns on Twitter during the 2018 government election in Rio de Janeiro, Brazil, using the Botometer API (Spiegel Rubin et al., 2021). In the Nigerian 2019 presidential elections, a lexicon-based public emotion mining and sentiment analysis was introduced to detect automated bots influencing the public's perception of the two major parties. Results indicated that these accounts created a higher positive and lower negative sentiment for the All Progressive Congress (APC) than the one observed with the People's Democratic Party (PDP) (Fagbola & Colin, 2019).

Besides these political studies, the COVID-19 pandemic boosted efforts to study health-related coordinated activities. The proliferation of misleading and false information surrounding COVID-19 by automated accounts coupled with human susceptibility to believing and sharing this content may well impact the course of the pandemic, as 66% of known bots were discussing pandemic themes (Himelein-Wachowiak et al., 2021). Conspiracy theories, such as "Film Your Hospital," aiming to show empty beds, enjoyed the promotion not only by verified users able to influence some Twitter users

but also by a small number of bots and deleted accounts within the network (Ahmed et al., 2020). The theme of COVID-19 vaccines was also used by large coordination networks involved in political astroturfing to deceive Twitter users about their efficacy (Jemielniak & Krempovych, 2021).

Studies have also considered disinformation campaigns on online social networks and how they hiddenly influence group behaviours (Sharma et al., 2021). In this disinformation warfare environment, it was possible to understand the role of state-sponsored trolls on Twitter (Vargas et al., 2020; Zannettou et al., 2019a, 2019b) and the use of political astroturfing (Keller et al., 2020) to influence and deceive users. Orchestrated activities also promote hate speech, although a prior study found the effects of automated bots sharing this type of content were insignificant (Beatty, 2020).

The polarisation of debates was also a topic that emerged from coordinated activities. An example is the 2019 Women's Strike conversation on Twitter, where automated accounts participated in the discussion using partisan hashtags and false information to promote polarisation (Calvo et al., 2021).

To a lesser extent, studies have shown that automated accounts promote discussions to artificially boost the visibility of specific content through commercial and quasi-commercial uses of bots. For example, Twitter bots were used to promote content from the audio-sharing platform SoundCloud (Bruns et al., 2018) or discussions about stocks traded in the leading US financial markets. Comprising accounts appear as untrustworthy and quite simplistic bots, speculative financial campaigns aimed at promoting low-value stocks by exploiting the popularity of high-value ones, likely aiming to fool automatic trading algorithms rather than human investors (Tardelli et al., 2020).

Tools and Methods Used to Detect Coordinated Campaigns

To detect these coordinated actions in online social networks, researchers relied on different methods and tools to assist them. Most of these studies used AI methods, publicly available datasets, third-party tools (e.g., Botometer/BotOrNot), and network analysis.

Publicly available datasets have been used to distinguish genuine users from bots (e.g., Cresci et al., 2017; Gong et al., 2018; Lee et al., 2013; Niranjan Koggalahewa et al., 2020). With these datasets, researchers combined deep learning techniques to distinguish tweets generated by legitimate users from those created by automated accounts (Ilias & Roussaki, 2021). These datasets were also used for the real-time detection of social bots on Twitter using machine-learning models (Alothali et al., 2021). Other studies focused on creating and keeping lists of potentially problematic sources, such as providing the URLs shared on Facebook by public groups, pages and verified profiles (Giglietto et al., 2020b).

The coordinated campaigns were also mapped using feature extraction for account-level data combined with deep learning for tweet-level classification (Ilias & Roussaki, 2021). In the same vein, scholars combined different

computational methods, such as Vector Support-Machine (SVM), logistic regression, decision tree, Naïve Bayes and K-Nearest Neighbours, to build a neural network–based classifier of bot accounts (Alothali et al., 2021).

In network analysis approaches, scholars looked at statistical features extracted from networks built based on these users and their interactions to detect coordinated activities. For example, researchers trained a binary classifier based on statistical components extracted from a time series of daily coordination networks on both Twitter disinformation campaigns and legitimate communities, allowing them to predict future disinformation-coordinated activity (Vargas et al., 2020).

On a small scale, publications also focused on the simulation of automated campaigns to study the strength of weak bots in promoting specific topics (Keijzer & Mäs, 2021) or measuring content coordination (Roussinov, 2018). Long short-term memory (LSTM) network was also used to determine whether an account is a bot or a human using a single tweet from that account (Chavoshi & Mueen, 2018). The probabilistic graphical model Markov Random Field (MRF) allowed scholars joint inference over-dependent random variables to make assumptions if a node is independent of its non-neighbouring nodes given its neighbours, helping to identify bot accounts (El-Mawass et al., 2018). A graph-based unsupervised learning method for edge and node anomaly detection was also used to detect irregular patterns in online social network activities, such as coordinated campaigns (Venkatesan & Prabhavathy, 2019).

A classification model from community-based features has also been used to examine coordinated activities. Using the node-level community structure from a weighted interaction graph of the social network, which represents the total number of messages, posts, etc., sent from the origin to the destination, it was possible to identify spam accounts (Bhat & Abulaish, 2013).

Another way used to detect coordinated activities is the reliance on detecting bots. As mentioned before, they are not necessarily harmful or spam accounts, but identifying this type of account help in this process. Several studies have relied on BotOrNot and Botometer to detect these bots (e.g., Ahmed et al., 2020; Bryden & Silverman, 2019; Furman & Tunç, 2020; Jemielniak & Krempovych, 2021; Spiegel Rubin et al., 2021).

Novel methods focused on detecting CIB in Facebook pages and groups based on a technique that identifies orchestrated actions based on a "near-simultaneous link sharing" activity (Broniatowski, 2021; Giglietto et al., 2020a).

Limitations of Such Approaches

While these different tools and methods provide a way to detect coordinated activities, they also come with several limitations. For example, using third-party tools, services or platforms, such as Botometer, entails certain risks, as they rely on "black box" functionality or end up discounting due to financial difficulties or business models in the long run (de-Lima-Santos et al., 2021). Equally

important, bot detection algorithms are not perfect. These methods report several false positives and negative results, which limit their effectiveness and put at risk their findings. A prior study has indicated that most accounts had 50% or lower bot scores, and many of the users' accounts exhibited bot-like behaviours due to their infrequency in posts or resharing content (Venkatesan & Prabhavathy, 2019). The major hurdle in social network anomaly detection is to identify irregular patterns in data that sometimes is not significantly different from regular patterns (Venkatesan & Prabhavathy, 2019).

Another important caveat of these tools is that they rely on access to API or publicly available datasets, which can be changed or restrained by platforms. For this reason, most research studying coordination mechanisms is confined to Twitter data, as the company provides more data than others. The recent announcement of Twitter ending its public API for researchers might change this scenario.

Furthermore, public availability datasets and computational algorithms need to keep pace with the constant evolution of bot complexity, as malicious actors are incorporating improved or new techniques to overcome detection methods. This also means it increases the complexity of classifiers and AI algorithms, requiring more computing power. In the same vein, several existing studies rely on lists of problematic content or news media sources compiled by fact-checkers or publicly available data. However, these lists "may quickly become obsolete, leading to unreliable estimates" (Giglietto et al., 2020b, p. 85).

Discussion and Conclusion

Overall, studies focused more on detecting bots, as elements that posted messages in an automated way and could be used to deceive people. However, there was no analysis that they could work in a coordinated manner to spread false or misleading content or hate speech. It is worth remembering that bots do not necessarily produce spam content on a large scale. Furthermore, the terminology concerning coordinated campaigns leads to confusion, as bots are used interchangeably with fake (false) and spam accounts. Accounts mimicking the social behaviours of humans are referred to as bots or automated accounts, which are controlled mainly by software (Alothali et al., 2021; Khaund et al., 2022; Perna & Tagarelli, 2018b). Fake or false accounts impersonate online users, while spam accounts spread unsolicited promotional content on a mass scale. Therefore, automated accounts improve their metrics, and fake ones help increase the metrics of other users, which could lead to an unhealthy environment on online social networks, depending on the users they boost (Akyon & Esat Kalfaoglu, 2019). The term "inauthentic accounts" refers to all these collections of accounts (Yang & Menczer, 2022), which can overlap each other.

We observed from the analysed works that the task of identifying bots, besides being difficult, is costly and becomes even more complex when it comes

to coordinated campaigns. In 2018, Facebook-owner, Meta, presented the concept of CIB to map automated accounts based on spam activities. Consequently, scholars have explored effective strategies to identify similar malicious actions by examining coordinated actions through "near-simultaneous link sharing" activity (Broniatowski, 2021; Giglietto et al., 2020a). However, social bots also rely on coordinated actions to help users consume quality information (Lokot & Diakopoulos, 2016).

Nevertheless, the evolution of the CIB shows that in addition to textual content, coordinated activities are evolving to audio-visual content, making detection even more complex. Future studies can explore orchestrated campaigns using multimedia content using advanced computational methods, such as computer vision and deep learning. Similarly, the rise of deepfake technology will undoubtedly become a more significant concern.

It is also important to note that humans can perform coordinated campaigns, hampering the ability of machines detected them. Since the posting time and content have no identifiable pattern, it is difficult, if not impossible, to detect these orchestrated actions. Upcoming research could explore new methods to identify these coordinated activities performed by humans and examine to what extent they pose risks for online social media users.

Lastly, the reliance on publicly available datasets might be an issue for coordinated mapping activities. Equally important is the support of public API, which is not a reality for many platforms. European regulation promises to change this reality, but scholars must rely on other data collection forms, such as data donation. Future studies should overcome these limitations.

Therefore, this study reveals the presence of coordinated activities on online social networks and how they can potentially undermine the public's trust and measures taken by spreading disinformation narratives on a large scale. By identifying these orchestrated campaigns in the literature, this study contributes to expanding the discussion of CIB or coordinated actions on online social networks, hoping it sheds light on new forms of tackling online mis- and disinformation activities and contributes to this theoretical discussion.

Note

1 This study was partially funded by the University of Amsterdam's RPA Human(e) AI and by AI4Media project under the European Union's Horizon 2020 research and innovation grant agreement No 951911.

References

Abokhodair, N., Yoo, D., & McDonald, D. W. (2015). Dissecting a social Botnet: Growth, content and influence in Twitter. *CSCW 2015 – Proceedings of the 2015 ACM International Conference on Computer-Supported Cooperative Work and Social Computing*, 839–851. https://doi.org/10.1145/2675133.2675208

Ahmed, W., Seguí, F. L., Vidal-Alaball, J., & Katz, M. S. (2020). COVID-19 and the "Film your Hospital" conspiracy theory: Social network analysis of twitter data. *Journal of Medical Internet Research*, *22*(10), e22374. https://doi.org/10.2196/22374

Akyon, F. C., & Esat Kalfaoglu, M. (2019). Instagram fake and automated account detection. *Proceedings – 2019 Innovations in Intelligent Systems and Applications Conference, ASYU 2019*, 1–7. https://doi.org/10.1109/ASYU48272.2019.8946437

Alothali, E., Alashwal, H., Salih, M., & Hayawi, K. (2021). Real time detection of social bots on Twitter using machine learning and Apache Kafka. *2021 5th Cyber Security in Networking Conference (CSNet)*, 98–102. https://doi.org/10.1109/CSNet52717.2021.9614282

Alothali, E., Hayawi, K., & Alashwal, H. (2021). Hybrid feature selection approach to identify optimal features of profile metadata to detect social bots in twitter. *Social Network Analysis and Mining*, *11*(1), 1–15. https://doi.org/10.1007/s13278-021-00786-4

Bastos, M., & Mercea, D. (2018). Parametrizing Brexit: Mapping twitter political space to parliamentary constituencies. *Information Communication and Society*, *21*(7), 921–939. https://doi.org/10.1080/1369118X.2018.1433224

Beatty, M. (2020). Graph-based methods to detect hate speech diffusion on Twitter. *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 502–506. https://doi.org/10.1109/ASONAM49781.2020.9381473

Bhat, S. Y., & Abulaish, M. (2013). Community-based features for identifying spammers in online social networks. *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 100–107. https://doi.org/10.1145/2492517.2492567

Bhimani, H., Mention, A. L., & Barlatier, P. J. (2019). Social media and innovation: A systematic literature review and future research directions, *Technological Forecasting and Social Change*, *144*, 251–269. https://doi.org/10.1016/j.techfore.2018.10.007

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), 10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Bode, L., & Vraga, E. (2021). The Swiss cheese model for mitigating online misinformation. *Bulletin of the Atomic Scientists*, *77*(3), 129–133. https://doi.org/10.1080/00963402.2021.1912170

Boshmaf, Y., Muslukhov, I., Beznosov, K., & Ripeanu, M. (2011a). The Socialbot Network: When bots socialize for fame and money. *ACM International Conference Proceeding Series*, 93–102. https://doi.org/10.1145/2076732.2076746

Boshmaf, Y., Muslukhov, I., Beznosov, K., & Ripeanu, M. (2011b). The Socialbot Network. *Proceedings of the 27th Annual Computer Security Applications Conference*, 93–102. https://doi.org/10.1145/2076732.2076746

Broniatowski, D. A. (2021). *Towards Statistical Foundations For Detecting Coordinated Inauthentic Behavior on Facebook*. https://www.facebook.com/communitystandards/inauthentic_behavior

Bruns, A., Moon, B., Münch, F. V., Wikström, P., Stieglitz, S., Brachten, F., & Ross, B. (2018). *Detecting Twitter Bots That Share SoundCloud Tracks*, 251–255. https://doi.org/10.1145/3217804.3217923

Bryden, J., & Silverman, E. (2019). Underlying socio-political processes behind the 2016 US election. *PLoS ONE*, *14*(4), 1–11. https://doi.org/10.1371/journal.pone.0214854

Calvo, D., Campos-Domínguez, E., & Simón-Astudillo, I. (2021). Towards a critical understanding of social networks for the feminist movement: Twitter and the Women's strike, *Tripodos*, *50*, 91–109. https://doi.org/10.51698/tripodos.2021.50p91-109

Cao, Q., Yang, X., Yu, J., & Palow, C. (2014). Uncovering large groups of active malicious accounts in online social networks. *Proceedings of the ACM Conference on Computer and Communications Security*, 477–488. https://doi.org/10.1145/2660267.2660269

Ceron, W., De-Lima-Santos, M.-F., & Quiles, M. G. (2021a). Fake news agenda in the era of COVID-19: Identifying trends through fact-checking content. *Online Social Networks and Media*, *21*, 100116. https://doi.org/10.1016/j.osnem.2020.100116

Ceron, W., Sanseverino, G. G., De-Lima-Santos, M.-F., & Quiles, M. G. (2021b). COVID-19 fake news diffusion across Latin America. *Social Network Analysis and Mining*, *11*(1), 47. https://doi.org/10.1007/s13278-021-00753-z

Chavoshi, N., & Mueen, A. (2018). Model bots, not humans on social media. *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 178–185. https://doi.org/10.1109/ASONAM.2018.8508279

Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., & Tesconi, M. (2017). The paradigm-shift of social spambots. *Proceedings of the 26th International Conference on World Wide Web Companion – WWW '17 Companion*, 963–972. https://doi.org/10.1145/3041021.3055135

de-Lima-Santos, M. F., Schapals, A. K., & Bruns, A. (2021). Out-of-the-box versus in-house tools: How are they affecting data journalism in Australia? *Media International Australia*, *181*(1), 152–166. https://doi.org/10.1177/1329878X20961569

El-Mawass, N., Honeine, P., & Vercouter, L. (2018). Supervised classification of social spammers using a similarity-based markov random field approach. In Proceedings of the 5th Multidisciplinary International Social Networks Conference (pp. 1–8).

Fagbola, T. M., & Colin, S. (2019). Lexicon-based bot-aware public emotion mining and sentiment analysis of the Nigerian 2019 presidential election on Twitter. *International Journal of Advanced Computer Science and Applications*, *10*(10), 329–336. https://doi.org/10.14569/IJACSA.2019.0101047

Farkas, J., & Schou, J. (2018). Fake news as a floating signifier: Hegemony, antagonism and the politics of falsehood. *Javnost*, *25*(3), 298–314. https://doi.org/10.1080/13183222.2018.1463047

Fazil, M., & Abulaish, M. (2020). A Socialbots analysis-driven graph-based approach for identifying coordinated campaigns in Twitter. *Journal of Intelligent and Fuzzy Systems*, *38*(3), 3301–3305. https://doi.org/10.3233/JIFS-182895

Felipe, M. (2022, July 12). *O que são os robôs que fizeram Elon Musk "desistir" da compra do Twitter?* *desinformante. https://desinformante.com.br/o-que-sao-os-robos-que-fizeram-elon-musk-desistir-da-compra-do-twitter/

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, *59*(7), 96–104. https://doi.org/10.1145/2818717

Freelon, D., & Wells, C. (2020). Disinformation as political communication. *Political Communication*, *37*(2), 145–156. https://doi.org/10.1080/10584609.2020.1723755

Freitas, C., Benevenuto, F., Ghosh, S., & Veloso, A. (2015). Reverse engineering Socialbot infiltration strategies in Twitter. *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2015*, 25–32. https://doi.org/10.1145/2808797.2809292

Furman, I., & Tunç, A. (2020). The end of the Habermassian ideal? Political communication on Twitter during the 2017 Turkish constitutional referendum. *Policy & Internet*, *12*(3), 311–331. https://doi.org/10.1002/poi3.218

Giglietto, F., Iannelli, L., Rossi, L., Valeriani, A., Righetti, N., Carabini, F., Marino, G., Usai, S., & Zurovac, E. (2018). Mapping Italian news media political coverage in the lead-up of 2018 general election. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3179930

Giglietto, F., Righetti, N., Rossi, L., & Marino, G. (2020a). It takes a village to manipulate the media: Coordinated link sharing behavior during 2018 and 2019 Italian elections. *Information Communication and Society*, *23*(6), 867–891. https://doi.org/10.1080/1369118X.2020.1739732

Giglietto, F., Righetti, N., Rossi, L., & Marino, G. (2020b). Coordinated link sharing behavior as a signal to surface sources of problematic information on Facebook. *ACM International Conference Proceeding Series*, *20*, 85–91. https://doi.org/10.1145/3400806.3400817

Gilmary, R., Venkatesan, A., & Vaiyapuri, G. (2023). Detection of automated behavior on twitter through approximate entropy and sample entropy. *Personal and Ubiquitous Computing*, *27*(1), 91–105. https://doi.org/10.1007/s00779-021-01647-9

Gleicher, N. (2018). *Coordinated inauthentic behavior explained*. Facebook Newsroom. https://about.fb.com/news/2018/12/inside-feed-coordinated-inauthentic-behavior/

Gong, Q., Chen, Y., He, X., Zhuang, Z., Wang, T., Huang, H., Wang, X., & Fu, X. (2018). deepScan: Exploiting deep learning for malicious account detection in location-based social networks. *IEEE Communications Magazine*, *56*(11), 21–27. https://doi.org/10.1109/MCOM.2018.1700575

Higgins, J. P. T., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., Savović, J., Schulz, K. F., Weeks, L., & Sterne, J. A. C. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ (Online)*, *343*(7829), d5928–d5928. https://doi.org/10.1136/bmj.d5928

Himelein-Wachowiak, M., Giorgi, S., Devoto, A., Rahman, M., Ungar, L., Schwartz, H. A., Epstein, D. H., Leggio, L., & Curtis, B. (2021). Bots and misinformation spread on social media: Implications for COVID-19. *Journal of Medical Internet Research*, *23*(5), e26933. https://doi.org/10.2196/26933

Hsu, T., & Thompson, S. A. (2023). *Disinformation researchers raise alarms about A.I. Chatbots*. The New York Times. https://www.nytimes.com/2023/02/08/technology/ai-chatbots-disinformation.html

Huijstee, M., Boheemen, P., Das, D., Nierling, L., Jahnel, J., Karaboga, M., Fatun, M., & Gerritsen, J. (2021). *Tackling deepfakes in European policy*. https://doi.org/10.2861/325063

Ilias, L., & Roussaki, I. (2021). Detecting malicious activity in Twitter using deep learning techniques. *Applied Soft Computing*, *107*, 107360. https://doi.org/10.1016/j.asoc.2021.107360

Jemielniak, D., & Krempovych, Y. (2021). An analysis of AstraZeneca COVID-19 vaccine misinformation and fear mongering on Twitter, *Public Health*, *200*, 4–6. https://doi.org/10.1016/j.puhe.2021.08.019

Keijzer, M. A., & Mäs, M. (2021). The strength of weak bots. *Online Social Networks and Media*, *21*, 100106. https://doi.org/10.1016/j.osnem.2020.100106

Keller, B. B., Schoch, F., Stier, D., & Yang, S. (2020). Political astroturfing on Twitter: How to coordinate a disinformation campaign. *Political Communication*, *37*(2), 256–280. https://doi.org/10.1080/10584609.2019.1661888

Khaund, T., Bandeli, K. K., Hussain, M. N., Obadimu, A., Al-Khateeb, S., & Agarwal, N. (2018). Analyzing social and communication network structures of social bots and humans. *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 794–797. https://doi.org/10.1109/ASONAM.2018.8508665

Khaund, T., Kirdemir, B., Agarwal, N., Liu, H., & Morstatter, F. (2022). Social bots and their coordination during online campaigns: A survey. *IEEE Transactions on Computational Social Systems*, *9*(2), 530–545. https://doi.org/10.1109/TCSS.2021.3103515

Kouvela, M., Dimitriadis, I., & Vakali, A. (2020). Bot-detective. *Proceedings of the 12th International Conference on Management of Digital EcoSystems*, 55–63. https://doi.org/10.1145/3415958.3433075

Le, H., Boynton, G. R., Shafiq, Z., & Srinivasan, P. (2019). A postmortem of suspended Twitter accounts in the 2016 U.S. presidential election. *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 258–265. https://doi.org/10.1145/3341161.3342878

Lee, K., Caverlee, J., Cheng, Z., & Sui, D. Z. (2013). Campaign extraction from social media. *ACM Transactions on Intelligent Systems and Technology*, *5*(1), 1–28. https://doi.org/10.1145/2542182.2542191

Lokot, T., & Diakopoulos, N. (2016). News bots: Automating news and information dissemination on Twitter. *Digital Journalism*, *4*(6), 682–699. https://doi.org/10.1080/21670811.2015.1081822

Malone, T. W. (1988). *What is coordination theory?* Science Foundation Coordination Theory Workshop. Sloan School of Management (No. 182).

Mendoza, M., Tesconi, M., & Cresci, S. (2021). Bots in social and interaction networks. *ACM Transactions on Information Systems*, *39*(1), 1–32. https://doi.org/10.1145/3419369

Mohamed Shaffril, H. A., Samsuddin, S. F., & Abu Samah, A. (2021). The ABC of systematic literature review: The basic methodological guidance for beginners. *Quality and Quantity*, *55*(4), 1319–1346. https://doi.org/10.1007/s11135-020-01059-6

Myllylahti, M. (2020). Paying attention to attention: A conceptual framework for studying news reader revenue models related to platforms. *Digital Journalism*, *8*(5), 567–575. https://doi.org/10.1080/21670811.2019.1691926

Ng, L. H. X., & Taeihagh, A. (2021). How does fake news spread? Understanding pathways of disinformation spread through APIs. *Policy & Internet*, *13*(4), 560–585. https://doi.org/10.1002/poi3.268

Niranjan Koggalahewa, D., Xu, Y., & Foo, E. (2020). Spam detection in social networks based on peer acceptance. *Proceedings of the Australasian Computer Science Week Multiconference*, 1–7. https://doi.org/10.1145/3373017.3373025

Overbey, L. A., Ek, B., Pinzhoffer, K., & Williams, B. (2019). Using common enemy graphs to identify communities of coordinated social media activity. In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics): Vol. 11549 LNCS*. Springer International Publishing. https://doi.org/10.1007/978-3-030-21741-9_10

Pariser, E. (2012). *The filter bubble: What the internet is hiding from you*. Penguin Books. https://books.google.com.br/books?id=Qn2ZnjzCE3gC

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, *592*(7855), 590–595. https://doi.org/10.1038/s41586-021-03344-2

Perna, D., & Tagarelli, A. (2018a). Learning to rank social bots. *HT 2018 – Proceedings of the 29th ACM Conference on Hypertext and Social Media*, 183–191. https://doi.org/10.1145/3209542.3209563

Perna, D., & Tagarelli, A. (2018b). Learning to rank social bots. *Proceedings of the 29th on Hypertext and Social Media*, 183–191. https://doi.org/10.1145/3209542.3209563

Piña-García, C. A., & Espinoza, A. (2022). Coordinated campaigns on Twitter during the coronavirus health crisis in Mexico. *Tapuya: Latin American Science, Technology and Society*. https://doi.org/10.1080/25729861.2022.2035935

Podsakoff, P. M., Mackenzie, S. B., Bachrach, D. G., & Podsakoff, N. P. (2005). The influence of management journals in the 1980s and 1990s. *Strategic Management Journal*, *26*(5), 473–488. https://doi.org/10.1002/smj.454

Roussinov, D. (2018). Towards measuring content coordination in microblogs. In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics): Vol. 10772 LNCS* (pp. 651–656). Springer. https://doi.org/10.1007/978-3-319-76941-7_58

Schoch, D., Keller, F. B., Stier, S., & Yang, J. (2022). Coordination patterns reveal online political astroturfing across the world. *Scientific Reports*, *12*(1), 1–10. https://doi.org/10.1038/s41598-022-08404-9

Schuchard, R. J., & Crooks, A. T. (2021). Insights into elections: An ensemble bot detection coverage framework applied to the 2018 U.S. Midterm elections. *PLoS One*, *16*(1), e0244309. https://doi.org/10.1371/journal.pone.0244309

Sharma, K., Ferrara, E., & Liu, Y. (2021). *Characterizing Online Engagement with Disinformation and Conspiracies in the 2020 U.S. Presidential Election*. http://arxiv.org/abs/2107.08319

Sharma, K., Zhang, Y., Ferrara, E., & Liu, Y. (2021). Identifying coordinated accounts on social media through hidden influence and group behaviours. *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, *1*(1), 1441–1451. https://doi.org/10.1145/3447548.3467391

Soto-Sanfiel, M. T., Ibiti, A., Machadom, M., Marín Ochoa, B. E., Mendoza Michilot, M., Rosell Arce, C. G., & Angulo-Brunet, A. (2022). In search of the global South: Assessing attitudes of Latin American journalists to artificial intelligence in journalism. *Journalism Studies 23*(10), 1197–1224. https://doi.org/10.1080/1461670X.2022.2075786

Spiegel Rubin, F., Luz de Almeida, Y., Alvim, A. C. D. F., Dias, V. F., & Santos, R. P. dos. (2021). Analysis of the first round of 2018 government election for the state of Rio de Janeiro based on Twitter. *XVII Brazilian Symposium on Information Systems*, 1–8. https://doi.org/10.1145/3466933.3466965

Sunstein, C. R. (2017). *#Republic: Divided democracy in the age of social media*. Princeton University Press. https://press.princeton.edu/books/hardcover/9780691175515/republic

Tardelli, S., Avvenuti, M., Tesconi, M., & Cresci, S. (2020). Characterizing social bots spreading financial disinformation. In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*, *12194 LNCS* (pp. 376–392). Springer. https://doi.org/10.1007/978-3-030-49570-1_26

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, *14*(3), 207–222. https://doi.org/10.1111/1467-8551.00375

Vargas, L., Emami, P., & Traynor, P. (2020). On the detection of disinformation campaign activity with network analysis. *Proceedings of the 2020 ACM SIGSAC Conference on Cloud Computing Security Workshop*, 133–146. https://doi.org/10.1145/3411495.3421363

Venkatesan, M., & Prabhavathy, P. (2019). Graph based unsupervised learning methods for edge and node anomaly detection in social network. *2019 IEEE 1st International Conference on Energy, Systems and Information Processing (ICESIP)*, 1–5. https://doi.org/10.1109/ICESIP46348.2019.8938364

Wardle, C., & Derakhshan, H. (2017). *Thinking about "information disorder": formats of misinformation, disinformation, and mal-information*. https://crosscheck

Weber, D., & Neumann, F. (2021). Amplifying influence through coordinated behaviour in social networks. *Social Network Analysis and Mining*, *11*(1), 111. https://doi.org/10.1007/s13278-021-00815-2

Yang, K. C., Pierri, F., Hui, P. M., Axelrod, D., Torres-Lugo, C., Bryden, J., & Menczer, F. (2021). The COVID-19 infodemic: Twitter versus Facebook. *Big Data and Society*, *8*(1), 1–12. https://doi.org/10.1177/20539517211013861

Yang, K.-C., & Menczer, F. (2022). *How many bots are on Twitter? The question is tough to answer — and misses the point*. Nieman Journalism Lab. https://www.niemanlab.org/2022/05/how-many-bots-are-on-twitter-the-question-is-tough-to-answer-and-misses-the-point/

Yuan, X., Schuchard, R. J., & Crooks, A. T. (2019). Examining emergent communities and social bots within the polarized online vaccination debate in twitter. *Social Media + Society*, *5*(3), 1–12. https://doi.org/10.1177/2056305119865465

Zannettou, S., Caulfield, T., Setzer, W., Sirivianos, M., Stringhini, G., & Blackburn, J. (2019a). Who let the trolls out? *Proceedings of the 10th ACM Conference on Web Science*, *3*, 353–362. https://doi.org/10.1145/3292522.3326016

Zannettou, S., Sirivianos, M., Blackburn, J., & Kourtellis, N. (2019b). The web of false information. *Journal of Data and Information Quality*, *11*(3), 1–37. https://doi.org/10.1145/3309699