From Fake News to Real Protests: WhatsApp's Role in Brazilian Political Coordination

Daniel Kansaon[1], Philipe de Freitas Melo[2], Savvas Zannettou[3], Fabricio Benevenuto[1]

[1] Universidade Federal de Minas Gerais, Brazil
[2] Universidade Federal de Viçosa, Brazil
[3] TU Delft, Netherlands
daniel.kansaon@dcc.ufmg.br, philipe.freitas@ufv.br, s.zannettou@tudelft.nl, fabricio@dcc.ufmg.br

Abstract

The growth of social networks has raised concerns about the misuse of these platforms by disinformation campaigns, social bots, and coordinated activities. Among these platforms, WhatsApp has become a focal point for this abuse, particularly in Brazil, one of the countries with the highest use of the platform. Despite acknowledging the presence of coordinated campaigns and implementing restrictions on the number of messages forwarded per user, the platform continues to be abused. Due to its private nature and the difficulty of collecting information, little is known about these campaigns and the messages they disseminate. Given this context, our study investigates the presence of coordinated activities on WhatsApp in Brazil, identifying their content and purpose, especially how these messages relate to recent Brazilian political events. To answer these questions, we analyzed 13 million messages from 1,444 political groups over seven months from July 2022 to January 2023. Using network analysis, our findings suggest a significant prevalence of coordinated activity in the propagation of news messages, 26% of which originate from misinformation sites. Furthermore, we found that images play a key role in coordinated activity, accounting for 15% of messages, which are also used to mislead. Finally, coordinated accounts were used to organize collective actions, including attacks and protests against election results.

1 Introduction

Today, upcoming elections in different countries are accompanied by serious concerns about election integrity. These concerns are mainly associated with the uncontrolled dissemination of misinformation on social media (Grinberg et al. 2019), rise of polarization (Conover et al. 2011) and radicalization (Ribeiro et al. 2020), the indiscriminate use of targeted advertising (Silva et al. 2020), social bots (Ferrara et al. 2016), and the increasingly personalized feed algorithms from social media platforms (Ribeiro et al. 2020). These concerns are more worrisome as different AI models become "off the shelf", making it easy to generate misinformation (Augenstein et al. 2024).

Since 2018, misinformation campaigns in Brazil have been taking place in a new and poorly understood digital space: messaging platforms such as WhatsApp (Benevenuto and Melo 2024a). To illustrate the extent of WhatsApp abuse in Brazil, an analysis of the most shared images in public WhatsApp groups collected in a period close to the 2018 presidential election found that 88% of them are false or misleading (Tardaguila, Benevenuto, and Ortellado 2018).

With millions of users worldwide, WhatsApp is especially popular in Brazil, where virtually everyone with a cell phone uses the app daily (Benevenuto and Melo 2024b). Many mobile data plans do not charge for the data used through WhatsApp in Brazil, i.e., zero-rating policy, making it an affordable alternative to SMS, voice, and video calls. A key feature of WhatsApp, often exploited by misinformation campaigns, is its public groups. These spaces typically connect users around specific interests and have become fertile ground for political activism, with many groups emerging to support specific candidates (Kansaon et al. 2024). Political campaigns have been orchestrated to flood public groups with content, which was then shared with the private part of the WhatsApp ecosystem (Resende et al. 2019).

After the 2018 presidential election in Brazil, WhatsApp acknowledged the presence of coordinated campaigns that spread massive amounts of messages during the election (Mello 2019). To mitigate the problem, they reduced its virality features by limiting how many times the content can be forwarded (Melo et al. 2019b). On the other hand, the Superior Electoral Court, responsible for Brazil's elections, criminalized the massive spread of political content through messaging applications.[1]

However, while these countermeasures are very welcome, they remain limited. First, bypassing the forwarding limit has proven to be relatively simple, rendering it ineffective (Melo et al. 2024). Second, WhatsApp has introduced new features that have increased virality and facilitated the widespread dissemination of content. For example, the Communities features allow users to manage and post to multiple groups simultaneously.[2] Finally, although coordinated campaigns to share political content on WhatsApp could be considered an electoral crime in Brazil, auditing any activity on WhatsApp is very difficult due to the closed nature of the application (Pasquetto et al. 2020).

In this study, we investigate the presence of coordinated campaigns on WhatsApp in Brazil. Specifically, we address

[1] https://folha.com/zdu068gh

[2] https://faq.whatsapp.com/495856382464992

the following research questions (RQs):

- **RQ1**: *Is there evidence of coordinated accounts involved in the propagation of messages on WhatsApp?*
- **RQ2**: *What is the content and purpose of the coordinated messages? How do they relate to recent political events in Brazil?*

To answer these questions, we collected a large set of messages from Brazilian political public groups, covering seven months from July 2022 to January 2023. This period includes significant events in Brazil, such as the 2022 presidential election, attacks on the electoral process, and riots that culminated in an attack on Brazil's federal government buildings.[3] Our dataset consists of 13 million messages, including text, images, and videos shared in 1,444 public groups. We employed a network-based method to identify accounts with synchronous activities, following the approach of (Pacheco et al. 2021). Our findings provide strong evidence of coordinated accounts on WhatsApp, not only to promote political content but also to orchestrate attacks on individuals and organize protests.

Main Findings Summarizing, our key findings are:

1. We observed the prevalence of coordinated activity in the dissemination of news content. Our analysis identified more than 1,575 coordinated users who systematically amplified more than 14,440 messages across various groups.
2. Coordinated accounts were primarily orchestrated to spread news. Nearly 70% of these coordinated messages are news-related, and approximately 26% are from misinformation websites.
3. Coordinated accounts notably amplified images, which are a key part of the content they share (15.54%), and also include misleading content.
4. Our analysis suggests that coordinated accounts played a key role in organizing major political events, including protests against election results and attacks on Supreme Court justices.

2 Related Work

With the growing influence of social media platforms and the impact of online speech and interactions on the real world, many studies have shown interest in studying coordinated online behavior in recent years. Coordination on social networks can influence online interactions and user perceptions, particularly in misinformation campaigns, in which coordinated accounts often work together to push false narratives (Keller et al. 2020).

Traditional methods for detecting online inappropriate behavior on social networks focus on identifying bots by analyzing individual user characteristics (Yang, Hui, and Menczer 2019; Yang et al. 2020). However, these approaches are challenging to apply when essential user data is unavailable, such as on WhatsApp, where users share information without public access to individual activity. Unlike traditional methods that estimate whether an account is a bot, the newer approach to identifying inauthentic coordinated accounts focuses on detecting coordinated actions at the group level by observing group users' patterns (Nizzoli et al. 2021; Pacheco et al. 2021). Current approaches focus on identifying similarities in the action sequences of two or more accounts, modeling user activities to build user similarity networks, and identifying coordinated user groups (Nizzoli et al. 2021; Pacheco et al. 2021).

(Nizzoli et al. 2021) proposed a network-based framework that identifies levels of similarity between coordinated accounts by connecting users who post similar content. This user similarity network is built by grouping coordinated actions. Furthermore, several studies incorporate temporal user similarity information to create the similarity network (Pacheco, Flammini, and Menczer 2020; Weber and Neumann 2021; Pacheco et al. 2021). (Pacheco et al. 2021) focuses on identifying coordinated accounts with synchronous activities, which means users who post messages simultaneously, using a threshold window that can be adjusted according to the context analyzed.

Based on the concept of group user activities to identify coordination, several studies have employed the similarity network to identify coordinated activities on different social networks, such as Facebook (Giglietto et al. 2020), Twitter (Pacheco, Flammini, and Menczer 2020; Vishnuprasad et al. 2024; Burghardt et al. 2024), and YouTube (Kirdemir and Adeliyi 2023). Other studies also focus on messaging platforms, such as Telegram (Venâncio et al. 2024) and WhatsApp (Nobre, Ferreira, and Almeida 2020, 2022). Although they are not structured like traditional social networks, public group spaces on these platforms connect users, forming a network ecosystem. To identify similarities between users, different metrics can be used, such as retweets (Pacheco, Flammini, and Menczer 2020), temporal and linguistic similarities (Burghardt et al. 2024), or even posting the same message (Venâncio et al. 2024).

Research Gap Although there are works studying coordination on WhatsApp (Nobre, Ferreira, and Almeida 2020, 2022), they provide a definition of coordination at the group level, using network structure and backbone extraction to identify coordination based on the similarity of the content posted. In contrast, our paper presents a different and more restrictive definition: rapid coordination, adapted from (Pacheco, Flammini, and Menczer 2020). This approach considers content similarity and synchronous posting behavior to identify coordinated accounts. This provides a completely different perspective on coordination on WhatsApp, which has not been explored before. Here, we focus on identifying consistent and synchronous coordination efforts on WhatsApp, leveraging the instantaneous nature of the platform to boost and amplify messages. Furthermore, to the best of our knowledge, this is the first large-scale study to explore rapid coordination on WhatsApp using a large dataset of more than 13 million messages collected over seven months, covering recent important events in Brazil. The large volume of data allows us to observe a different perspective by incorporating temporal similarity to identify synchronous coordination actions. By including a similarity of time posting, we can find more consistent cooperation

[3] https://en.wikipedia.org/wiki/2023_Brazilian_Congress_attack

between accounts (Giglietto et al. 2020). Furthermore, our work investigates different coordination formats, including text, images, and videos. Although textual content is prevalent, our study shows that coordinated actions take advantage of different media.

3 Methodology

We propose a comprehensive methodology that addresses the unique challenges posed by WhatsApp's architecture. This section describes our approach to collecting, processing, and analyzing to examine coordinated activity within Brazilian political public groups on WhatsApp. First, we detail our data collection strategy to obtain data from public WhatsApp groups, the approach to discovering and joining relevant groups, and the methods used to gather and manage data from these groups. Next, in the direction of investigating coordination, we structured our methodology as follows: first, we describe the network modeling used (Section 3.2) and the proposed strategy to investigate coordinated accounts (Section 3.3), addressing (RQ1). Next, we analyze the characteristics of coordinated messages (Section 4) and examine the topics discussed and their connection to political events (Section 5), addressing (RQ2).

3.1 Data Collection

WhatsApp presents challenges for data collection due to its encrypted and closed architecture, which limits external access to the (mis)information circulating through this network (Benevenuto and Melo 2024a; Melo et al. 2024). To address this limitation, we used a widely adopted methodology to gather data from publicly accessible WhatsApp groups (Melo et al. 2019a; Garimella and Tyson 2018; Resende et al. 2019; Bursztyn and Birnbaum 2020; Kansaon et al. 2024). Our methodology extracts data from WhatsApp publicly accessible groups, which are largely shared via invitation URLs on websites and other social media platforms in Brazil for political discussion over seven months, from July 2022 to January 2023. By collecting data from these open groups, we can capture large-scale messaging activity and also viral content shared within this platform during the Brazilian presidential election. This period includes important misinformation campaigns targeting the electoral process and protests against the election results.

For group discovery, we used a set of keywords related to Brazilian politics proposed by (Resende et al. 2019) and added terms related to relevant people and topics of the period analyzed, then we searched for WhatsApp groups on social media and online group repositories using these keywords (See Appendix A for more details about the keywords). After that, we manually joined each group found periodically, selecting only those related to political topics and suitable for our research. As a result, we included 1,444 public WhatsApp groups and gathered data about all messages sent within them. For each message, we extract (i) an anonymized user ID, (ii) group ID, (iii) timestamp, (iv) a label on whether the message was forwarded, (v) text, (vi) media type (e.g., images, audio, and videos, sticker) and (vii) if available, the attached multimedia files (e.g., images, audio, and videos). Data collection occurs at regular intervals. Since all real-time messages are stored locally on the smartphone, we periodically run processes to extract this data and download the media, typically every 7 days. During this process, we calculate metrics such as the total number of media appearances, repeated content, and other relevant details. The media is typically available for up to 15 days on WhatsApp, which is sufficient for our purposes, and we have not faced any difficulties retrieving the media. All multimedia and files shared on WhatsApp are assigned a unique identifier on the platform. Hence, it is not easy to associate multimedia files that share the same content. To address this, we used the MD5 hash values of the messages to deduplicate the dataset based on the unique set of hashes. Additionally, for images, we also calculate the perceptual hash (pHash), which enables us to determine how many times each piece of media has been shared.

In total, we collected 13,452,039 messages shared in 1,444 WhatsApp groups from 100 thousand users between July 2022 and January 2023. Our dataset is quite diverse and includes a lot of multimedia; 5.4M messages are text (3M unique messages), 3M messages are video messages (1.1M unique messages), 1.8M messages are image messages (1.1M unique messages), 1.2M messages are stickers (125K unique messages), 1M messages are links (584K unique messages), 679K messages are audio (512K unique messages), and 21K messages share documents (7K unique messages).

Limitations Despite the large-scale dataset collected, one of the main challenges is the inherent difficulty in determining the representativeness of this dataset, a common issue faced by studies focusing on messaging platforms such as WhatsApp. The groups we monitored may not fully capture the diversity of content and behavior found in private groups, where coordination and dissemination of information might differ significantly. Despite this limitation, we believe that our extensive data collection efforts enable us to characterize the coordinated activities of political groups on WhatsApp. Furthermore, WhatsApp's encryption and private nature limit our ability to track the origins of the messages and fully identify all the individuals coordinating these actions, which may be much larger than our results may capture. Nevertheless, we believe that these limitations do not undermine the findings of our work, which provide valuable contributions to understanding the dynamics of coordinated accounts on WhatsApp, since even small evidence of coordination highlights an important issue faced by the platform.

3.2 Rapid Spread Network Modeling

Coordinated activity is a well-known phenomenon in various online social networks, where users employ it for diverse purposes such as sharing beliefs, marketing, mobilizing people, shaping public opinion, and spreading misinformation (Nizzoli et al. 2021). On WhatsApp, the main messaging app in Brazil, particularly for political purposes, coordinated accounts can leverage the platform to increase engagement in specific activities. These groups of users often replicate similar behaviors over time, such as endorsing certain messages and amplifying specific content.

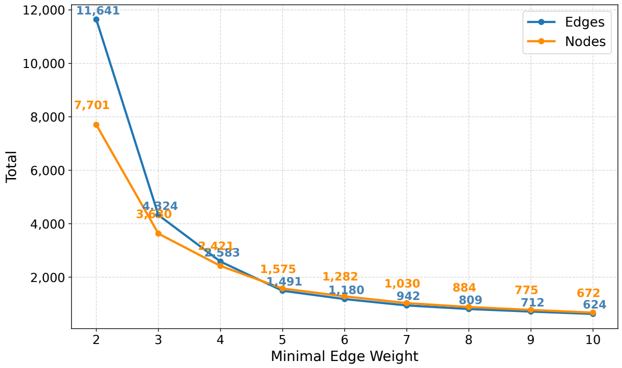


Figure 1: Coordinated users by alterations in parameters.

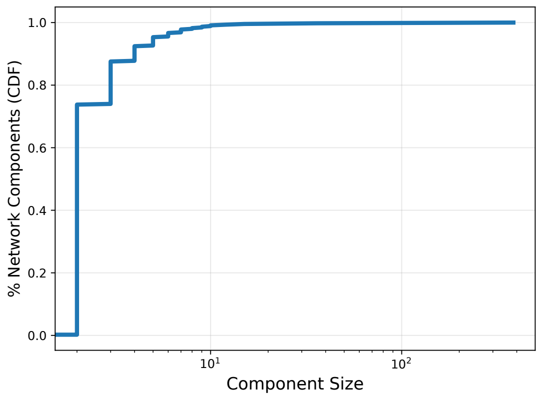


Figure 2: Component size distribution.

Due to the widespread use of WhatsApp for content dissemination, our goal is to identify networks of accounts that share the same content rapidly, simultaneously, and repeatedly in a coordinated way. To achieve this, we adapted the Rapid Retweet Network proposed by (Pacheco, Flammini, and Menczer 2020), which was originally applied to the Twitter ecosystem with a focus on retweets. This approach is particularly effective for identifying groups of accounts that consistently retweet the same source. On WhatsApp, there are no retweets like on Twitter. Instead, users typically receive content and share it in two main ways: directly forwarding the message to another group or copying and pasting the content into a different chat group.

We define the Rapid Spread Network based on the simultaneous sharing of similar messages within a specific time window. To analyze content similarity within WhatsApp data, we adopted the MD5 hash algorithm to detect identical messages and identify shares of the same content. This method assigns a unique identifier to each unique message, where two messages are considered identical only if they have the same hash. Any modification results in a different hash. With these definitions, we build a weighted network where users are connected if they post the same message in the same time window. In the network $G(V, E)$, a node $v \in V$ corresponds to users who post messages on WhatsApp. The undirected weighted edge $e = (v_i, v_j)$ is included in the users v_i and v_j if they posted the same message in the same time interval. The edge weight of w_{ij} represents the total number of messages they have shared in common during the time window.

On WhatsApp, the private nature of the platform prevents the identification of the source or promoters of a message. Consequently, the network analyzed in this study is composed of users who disseminate identical content simultaneously, highlighting the accounts and messages propagated concurrently across the network.

3.3 Identifying Coordinated Activity

Before applying the Rapid Spread Network to model the propagation of coordinated messages, we implemented a series of restrictive adjustments and parameter selections to ensure that we would only capture orchestrated actions by coordinated accounts. First, we selected a 60-second time window[4] to define rapid actions, allowing us to identify potential coordination by users acting simultaneously (Keller et al. 2020). To avoid misclassifying common messages frequently shared on WhatsApp (e.g., "good morning", "hello") as coordinated actions, we also filtered out short messages with less than two words. Given the message flow and the large period observed, we tested various threshold values, as shown in Figure 1. Using the elbow method, we determined the optimal threshold by identifying the inflection point on the curve. Consequently, we filtered out edges with weights below five to build the final graph. The resulting coordination network comprises 1,575 nodes and 1,491 edges, with an average degree of 1.89.

Although we are only examining a subset of the WhatsApp ecosystem, we can still observe a clear structure in the propagation of messages. One notable feature of this network is the presence of 450 unique components, which suggests that the network is not densely connected. This fragmentation is significant because it reveals the existence of many isolated pairs or small groups of accounts acting independently to coordinate messages. This high number of unique components suggests a decentralized coordination effort on WhatsApp, where many accounts operate separately to amplify their content rather than a single cohesive group controlling the flow of information. Figure 2 presents the cumulative distribution function (CDF) of component sizes, showing that the majority of coordinated components consist of two accounts (73.7%).

Furthermore, we identified a large connected component comprising 332 nodes (21% of the total network). This more connected group demonstrates that while much of the network consists of isolated actors, there is another aspect of coordination, with a substantial cluster of coordinated accounts operating together. This large component suggests a more organized structure within the WhatsApp ecosystem, in which many accounts are coordinated to propagate mes-

[4] We performed a sensitivity analysis with varying thresholds to determine whether 60 seconds is a suitable threshold. Please see Appendix B for more details.

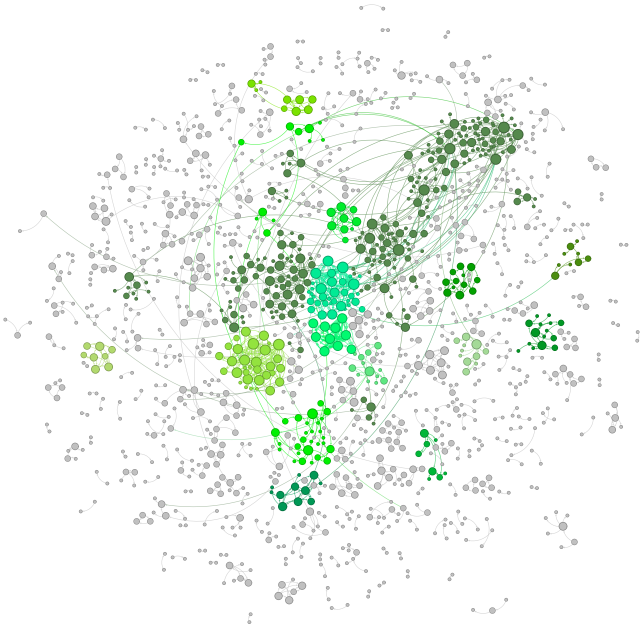


Figure 3: Rapid Coordination Network.

sages quickly and efficiently. Given WhatsApp's closed and encrypted architecture, it may be challenging for authentic users to differentiate this coordinated activity, making it even more concerning. Using the Louvain community detection algorithm (Blondel et al. 2008), we further identified some communities within the large connected component, highlighting the coordinated nature of these accounts, as shown in Figure 3. The presence of multiple communities suggests that coordination on WhatsApp can extend beyond pairs of synchronized accounts, allowing them to reach a wider audience. These communities could be used to amplify specific narratives or target particular themes within political discussions. We observed that the three largest identified communities contribute significantly to the flow of message propagation on WhatsApp. These communities consist of 301 accounts (19% of all coordinated accounts) and together posted 4,982 messages (34.5% of all coordinated messages). Coordinated accounts in these communities reached 664 groups (45.9% of all groups observed). This scenario helps us understand the impact of coordination and how the WhatsApp environment is conducive to coordinated actions, in which some users can affect a representative part of the group ecosystem.

After we identified the coordinated accounts, we analyzed their messages. According to our definition, messages are considered coordinated only if they are posted simultaneously by two or more coordinated accounts in the same time window. Considering these synchronous coordinated messages, we evaluated the portion of messages posted by the coordinated accounts within our dataset. We identified a concentration of coordinated activity, in which we observed that 80% of the coordinated messages are generated by 27.2% of the accounts, as shown in Figure 4. This suggests that, while many users are involved in sending messages, a relatively small subset of coordinated accounts is responsible for most of the messaging campaigns within the political public groups. This indicates a structured and orchestrated use of WhatsApp for coordinated political activity, with some actors playing a key role in shaping the network dynamics. However, this remains largely hidden from public view due to the platform's design.

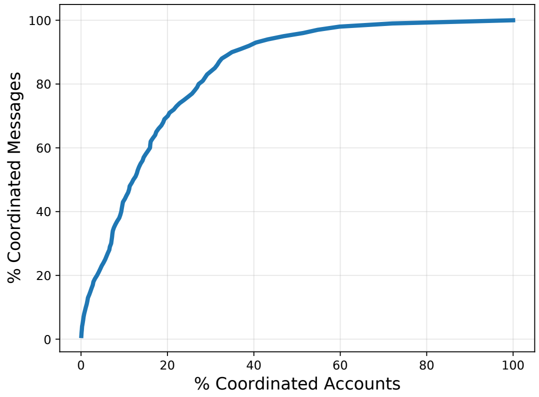


Figure 4: Coordinated messages by coordinated accounts.

Takeaways The main takeaways from this section are:

- Our methodology has identified more than 1,575 coordinated accounts actively disseminating political messages.
- There are decentralized coordination efforts in which accounts operate separately to amplify their messages, totaling 450 components.
- A small part of coordinated accounts (27.2%) are responsible for most of the flow of coordinated messages (80%).

4 Analyzing Coordinated Messages

To further delve into coordinated accounts, we examined their activities, focusing on understanding the messages they shared and their underlying motivations to address RQ2.

Once we identified the coordinated accounts, we analyzed their coordinated messages posted, totaling 14,414 messages. It is important to note that, according to our definition of coordination, a coordinated message is one posted simultaneously by two or more coordinated accounts.

To provide a more contextualized analysis, we created 35 random samples of non-coordinated messages, each matched in size to the coordinated messages. This allowed us to perform a comparative analysis to identify distinctive patterns between coordinated and non-coordinated messages, while ensuring the robustness of our findings with a 95% confidence interval. Similarly to rapid network construction, this sample only includes text messages with more than two words. Figure 5 compares the types of media found in these two groups of messages. Text messages are the most common form of communication in coordinated messages, comprising 70.24% (10,124 messages) of the total 14,414 coordinated messages, compared to 43.78% (±0.1702) in

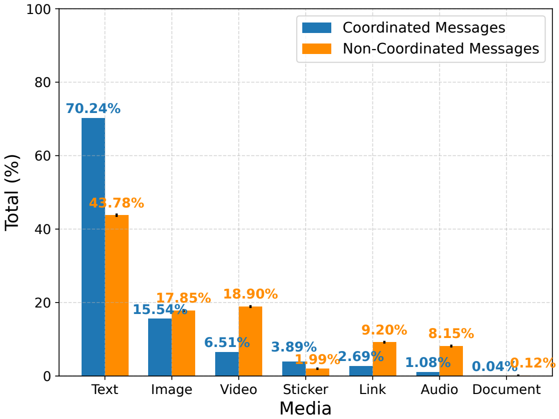


Figure 5: Messages media type differences from coordinated and non-coordinated messages.

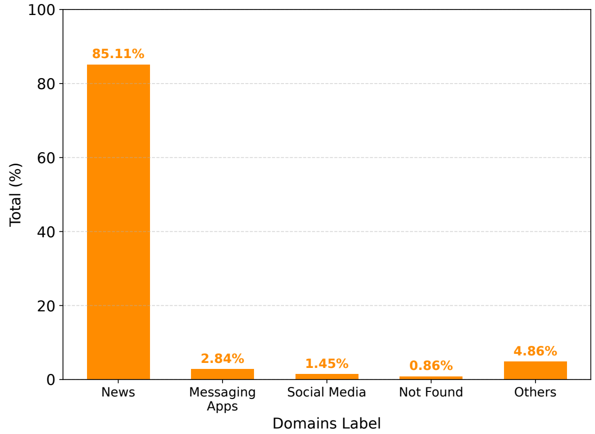


Figure 6: Category of URLs found in coordinated text messages.

non-coordinated. This significant difference highlights the efficiency of text messages, which are easily shared and forwarded without access to external media files or galleries. Images appeared in 15.54% of coordinated messages, aligning with 17.85% (±0.1145) in non-coordinated messages. Notably, videos, links, audio, and documents are rarely used in coordinated messages. Another notable observation is that stickers are more common in coordinated content. While stickers are generally used as a spontaneous form of communication, their use among coordinated users suggests that they can serve as a resource to a flooding attack in a group (Kansaon et al. 2024). A flooding attack occurs when one or more users overwhelm a group with a high volume of duplicate messages in a short period, intending to disrupt the flow of conversation or even crash the group, and usually use stickers (Kansaon et al. 2024). When we applied this definition to observe one of the most popular right-wing groups in our dataset, we identified a flooding attack in which three coordinated users sent over 1,200 duplicate sticker messages within seven minutes. These stickers were primarily composed of provocative content and political attacks. This suggests that, although stickers are less commonly used than text messages, coordinated users can significantly amplify the impact of a sticker flooding attack, making it more effective in achieving its disruptive goals.

When analyzing coordination efforts, it becomes clear that the main goal is to spread messages among as many groups as possible. Text messages are particularly well-suited for this purpose, as they are easy to share and read. Unlike other media formats that require downloads or additional actions, text messages allow for direct and easy communication. This simplicity makes them an ideal choice for disseminating coordinated content.

Additionally, we found that 58.9% of coordinated messages were not forwarded. This suggests that most coordinated message sharing is not done through WhatsApp's forwarding mechanism, indicating a deliberate effort to share messages directly, as this mechanism is not widely used. WhatsApp has introduced this feature to combat misinformation and virality by labeling popular forwarded messages with the “forwarded many times” tag (Melo et al. 2024). However, our findings suggest that this measure is ineffective in limiting the mass spread of coordinated accounts.

Characterizing Text Messages Looking at the structure of coordinated text messages, we note an average length of 17.93 words with a standard deviation of 24.19. In contrast, non-coordinated messages are longer, with an average of 24.68 words (±0.4659), but show a much higher standard deviation of 106.07 (±26.74), suggesting significant variation in length. While coordinated messages are generally shorter than the average, the low standard deviation suggests they have a more consistent length than non-coordinated messages, which show greater variability. This suggests that coordinated messages are not random but contain consistent information to engage users effectively.

Characterizing URLs Upon closer inspection of the 10,124 coordinated text messages, we observe that 97.31% contain embedded URLs, reflecting the widespread use of hyperlinks by coordinated accounts. These URLs are seamlessly integrated into the text, suggesting that these textual messages not only inform but also direct users to external web resources. Compared with the sample of non-coordinated texts, we find that only 16.91% (±0.0009) contain links, highlighting a distinct difference between coordinated and non-coordinated messages.

Further analysis of URLs within coordinated messages involved an extraction and manual labeling process. By parsing the coordinated text messages, we found 11,725 links, of which 10,269 were unique. We extracted the domain of each URL from the coordinated text messages, resulting in a subset of 116 unique domains. This indicates that coordinated messages disseminate content from a small number of websites. Here, we characterize the content by undertaking a qualitative analysis to better understand the nature of these domains. Initially, we compiled a list of all domains and created a codebook with preliminary codes, refining them iter-

Domain	Category	Total URLs
pensandodireita*	news	2,403
portaltocanews	news	2,119
redebrasilnews	news	682
gazetabrasil	news	432
whatsapp	messaging apps	293
macajubaacontece	news	291
terrabrasilnoticias*	news	274
brazilnewsinforma	news	263
portalcidade	news	232
direitaonline	news	203

Table 1: TOP-10 URLs domains found in coordinated text messages. Domains in bold are sites that employed misinformation strategies during the 2022 electoral campaign, as reported by Aos Fatos Fact Checker.

atively until no further changes were necessary.

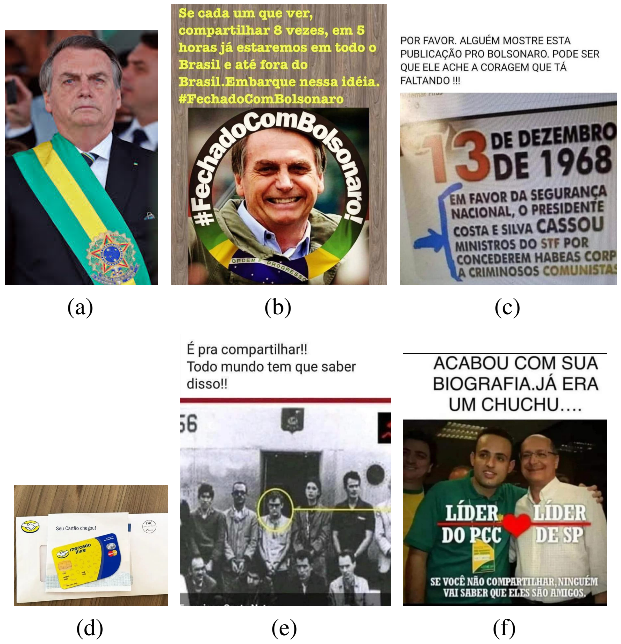
Our codebook consists of five codes:

- **News**: Websites that host structured news content.
- **Messaging Apps:** Invitations to Telegram and WhatsApp groups/channels.
- **Social Media:** Links to social media platforms such as Facebook, Twitter, Instagram, YouTube, and TikTok.
- **Not Found:** Domains that do not exist or are currently unavailable.
- **Others:** Links that lead to product sales, app stores, or banking services, as well as websites that show characteristics of spam or fraudulent activities.

After building the codebook, we applied it to categorize all domains identified in the coordinated messages. This labeling process allowed us to determine the thematic focus of each shared URL. Figure 6 shows the distribution of the coordinated link category. Upon analyzing the results, we observed that 85.11% of all URLs within coordinated messages lead to news websites. Furthermore, messaging apps account for 2.84%, which is particularly interesting because coordinated accounts leverage public platforms to invite others to more exclusive private communities through group invitations. Typically, they share a message that includes an invitation link to other groups.

The significant prevalence of coordinated messages containing links to news websites provides valuable insight into the strategies employed by coordinated users. This suggests a deliberate focus among coordinated users on disseminating news-related content within WhatsApp groups to engage users in propagating specific narratives. By including news links, coordinated users attempt to bring a sense of formality and credibility to their messages. Moreover, including news links serves as a mechanism to drive traffic to various news websites, thereby expanding the reach and influence of the disseminated information. Notably, websites hosting such coordinated news can generate revenue through advertisements, as shown by the Aos Fatos Fact-Checker.[5]

WhatsApp lacks mechanisms to verify information sources, and this is an ideal scenario in which coordinated users can exploit the platform's ecosystem to disseminate misinformation. This enables coordinated users to leverage WhatsApp as a tool for misinformation, potentially manipulating public opinion. Within political groups, individuals become deeply engaged with particular topics, facilitating the spread of news that aligns with specific narratives. As a result, these coordinated efforts can effectively spread messages that reinforce specific narratives in groups.


Se cada um que ver, compartilhar 8 vezes, em 5 horas já estaremos em todo o Brasil e até fora do Brasil.Embarque nessa idéia. #FechadoComBolsonaro
#FechadoComBolsonaro!
POR FAVOR. ALGUÉM MOSTRE ESTA PUBLICAÇÃO PRO BOLSONARO. PODE SER QUE ELE ACHE A CORAGEM QUE TÁ FALTANDO !!!
13 DE DEZEMBRO DE 1968
EM FAVOR DA SEGURANÇA NACIONAL, O PRESIDENTE COSTA E SILVA CASSOU MINISTROS DO STF POR CONCEDEREM HABEAS CORP A CRIMINOSOS COMUNISTAS
Seu Cartão chegou!
É pra compartilhar!! Todo mundo tem que saber disso!!
ACABOU COM SUA BIOGRAFIA.JÁ ERA UM CHUCHU....
LÍDER DO PCC
LÍDER DE SP
SE VOCÊ NÃO COMPARTILHAR, NINGUÉM VAI SABER QUE ELES SÃO AMIGOS.

(a) (b) (c)

(d) (e) (f)

Figure 7: TOP-6 coordinated images based on total shares.

Upon analyzing the most frequently shared domains by coordinated users, we found the ten most shared domains in the messages in Table 1. Fact-checking agencies revealed that *pensandodireita (1º)* and *terrabrasilnoticias (7º)* domains spread fake news and misinformation. We adopt the definition of misinformation provided by the Aos Fatos fact-checking, which highlights that these websites employed misinformation strategies in their news coverage during the 2022 electoral campaign. According to Aos Fatos, these two websites promote their news across messaging app platforms to attract users to their platforms. These websites typically use multiple ads that eventually engage users who click on these ads, thereby generating additional revenue for the website owner. This discovery sheds light on the deliberate strategies employed to exploit misinformation content for financial gain. Notably, we found that 26% of all links identified in coordinated text messages are from these two misinformation websites, highlighting the role of coordinated accounts in the dissemination of misinformation.

Characterizing Images Coordinated accounts also use image content to spread narratives, accounting for 15.54% of coordinated messages. Analyzing these images is crucial to understanding the similarities of media types shared in coordinated actions. Observing the ten most shared coordinated

[5] https://aosfatos.org/s/w6gbyzq/

messages, we identified that five of these images are related to politics and presidential elections, including promoting candidates or attacking opponents. These images have a different impact on the user's perception, and the visual content can condense more information into a small piece of content. This makes images particularly effective for spreading misinformation, often by presenting events out of context or making old events appear recent. We observed that two of the ten most shared images were misinformation. The figure 7(f), shared 396 times, is the sixth most shared coordinated image. The image is authentic, but the text puts it in the wrong context. Comprova's fact-checking team labeled this image misleading.[6] Another noteworthy example is the fifth most shared coordinated image, as shown in Figure 7(e), shared 427 times. This is an authentic picture, but the text puts this image in a different context, saying that one of the members present is a former Supreme Court justice. This image was classified as false by Aos Fatos Fact-checking organization.[7]

Furthermore, two of the analyzed images target the Supreme Federal Court and the electoral process, as shown in Figures 7(c) (shared 439 times) and 7(e). These two images aim to discredit the judicial decisions and accuse the court of political bias. This theme was particularly prominent in Brazil, as reflected in the identified topics presented in Table 2. A noteworthy aspect of the coordinated images is that some images encourage users to share the message, such as Figure 7(b), which has been shared 460 times. Additionally, three of the images are related to credit cards and financial topics. One example is Figure 7(d), which depicts a credit card and was shared 430 times.

Takeaways The main takeaways from this section are:

- Coordinated activities focus on news dissemination. 70.24% of coordinated messages consist of text, and 97.31% contain embedded links, with 85.11% of these links leading to political news.
- We found that 26% of all coordinated links are from news misinformation websites.
- Images are a key part of the content shared (15.54%) by coordinated accounts and also include misleading content. Notably, two of the ten most shared images were labeled as misleading content.

5 Topic Modeling and Case Studies

To further analyze the content of coordinated messages and address the (RQ2), we conducted a topic modeling analysis focused on identifying the connections of these messages with political events. We characterized the topics discussed in coordinated textual messages shared by the coordinated accounts. For this purpose, we employed BERTopic, a topic modeling technique that generates dense clusters to produce interpretable topics. This method uses vector representation (embeddings) and the concept of c-TF-IDF to create coherent topics, preserving key terms in the topic descriptions that enhance the clarity and interpretation (Grootendorst 2022).

We start by converting coordinated WhatsApp messages into vector embeddings using the PTT5 transformer language model, which was trained on Portuguese Wikipedia data and contains 200 million parameters (Carmo et al. 2020).[8] Next, we apply dimensionality reduction with the Uniform Manifold Approximation and Projection (UMAP) technique, which preserves both local and global structures of the embeddings. Then, we use a hierarchical clustering algorithm (HDBSCAN) to group the vector representations into clusters based on semantic similarities. Finally, we extract topics for each cluster using the Class Term Frequency-Inverse Document Frequency (c-TF-IDF), identifying the most relevant terms given all documents in a cluster. Then, we refined the approach to balance the number of topics with the size of our dataset. Following the recommendations in the BERTopic documentation, we set the number of topics to 15. To further improve the results, we applied the outlier reduction method to reassign these outlier documents to the appropriate topics. UMAP was configured with ten neighbors, ten components, and a minimum distance of 0.05 for effective dimensionality reduction. Finally, we configured HDBSCAN with a minimum cluster size of 50 and a minimal sample of 50, specifying the minimum number of messages a topic can represent and ensuring that only significant topics are considered for analysis. Additionally, the epsilon parameter was set to 0.3, defining the maximum distance for points to be considered neighbors.

Brazilian context Before analyzing the topics, it is important to understand the political scenario of Brazil during the period analyzed. In 2022, Brazil had a presidential election and, during this period, social networks and messaging apps were flooded with political discussions and campaigns. Furthermore, the analyzed period includes other important events, such as the intense protests marked by widespread fraud allegations about the results and the riots on January 8th. In addition, many protests were made against the decisions of the Supreme Federal Court and its ministers, which became a major topic in the news.

With that in mind, we can observe Table 2, which contains the discussed topics identified using the proposed framework. Initially, we can observe many political topics addressing different perspectives on the Brazilian elections. One major focus is the Supreme Court, which became the center of many protests, with several ministers' attacks (Topic 1). Additionally, many messages contain terms related to election fraud (Topic 5), as a large portion of the population refused to accept the results, claiming that it was rigged and asking for military intervention (Topic 15). We also observe discussions about journalists (Topic 6), political candidates (Topic 8), and political debate repercussions (Topic 10). Also, some topics reflect government arguments promoting the candidates' achievements (Topic 7), which relate to fuel and price reductions. Coordinated messages also include government assistance and subsidies for low-income individuals (Topic 14). Initially, we observed that the coordinated messages were related to the Brazilian political sce-

[6] https://projetocomprova.com.br/post/re_2B5W8XYjrLpY/

[7] https://aosfatos.org/s/r3i3wti/

[8] https://huggingface.co/unicamp-dl/ptt5-large-portuguese-vocab

Id	Label	Topic Terms
1	**Supreme Federal Court (42.9%)**	moraes, stf, pt, federal, minister, tse, round, whatsapp, police, pt supporter
2	**Credit Card (15%)**	thousand, card, credit, aid, cash, millions, bank, vacancies, request, receive
3	**Videos (8%)**	video, audio, activate, screen, husband, caught, wife, lover, videos, shows
4	**Social Networks (5.1%)**	telegram, whatsapp, twitter, network, social, facebook, instagram, follow, forget, follow us
5	**Election Fraud (4.6%)**	electoral court, electoral, crime, crimes, propaganda, fraud, operation, federal police, ballot boxes, elections
6	**Journalist Commentators (2.9%)**	shorts, amanda, klein, *toma, invertida, lapada*, guga, noblat, come through, live
7	**Economy and Fuel Price (3.5%)**	petrobras, price, gasoline, pf, seize, gas, reduction, tons, fuels, federal highway police (prf)
8	**Candidates (3.2%)**	candidate, candidates, presidency, candidacy, health, agenda, curses, covid-19, education, childish
9	**Health (2.4%)**	symptoms, cancer, know, hospital, disease, main, heart attack, treatment, signs, warning
10	**Political Debate (2.4%)**	debate, tv, criticize, band tv, knight, diego, globo tv, democracy, criticism, sbt tv
11	**Accident News (0.33%)**	death, dead, leaves, accident, found, dies, serious, deaths
12	**Pictures (1.8%)**	pictures, picture, images, body, globo tv, bikini, cameras, camera, happens, attention
13	**Money (1.8%)**	loan, aid, entrepreneurs, companies, payroll, beneficiaries, moraes, name, businessman, bank caixa
14	**Government Assistance (1.8%)**	family, families, scholarship, aid, receive, low, income, program, own
15	**Military (0.13%)**	military, military, police, defense, security, institutional security office, minister fachin, civil, organization, superior

Table 2: Discussion topics found in coordinated text messages. The topic terms are translated from the original Brazilian Portuguese. The percentages in the labels represent the proportion of coordinated text messages for each topic.

nario, which gave us an idea of the interest of coordinated accounts in reinforcing specific narratives.

Furthermore, we identified topics unrelated to politics, such as messages about credit cards, bank loans, and financial resources (Topics 2 and 13). There is also content that involves shared images, often featuring sexist themes (Topic 12). Videos are also popular, as seen in Topic 3, where many messages contain links to external websites hosting the videos. We also found health-related content that spreads tips on disease symptoms (Topic 9) and news about violence and accidents (Topic 11). As observed in domain analysis, many websites focus on increasing traffic and want to promote their content, often about curiosity or facts that intrigue people to know more, characteristics observed in these two identified topics.

The identified topics highlight the main discourses and narratives of the Brazilian political landscape during the period. WhatsApp plays a central role in Brazil's communication ecosystem, widely used for debates, gathering information, and tracking the repercussions of major events. Although WhatsApp is typically expected to be used to react and discuss real-world events, these coordinated actions suggest that, in some cases, it can be used as a strategy to organize, motivate, mobilize, and shape political narratives.

To better understand coordinated messages and their motivations, we focus on two key topics: the Supreme Federal Court (Topic 1) and Election Fraud (Topic 6). These topics were chosen because they were central to the Brazilian political discourse during the period covered by our dataset. The election is the most significant event, making these topics particularly relevant for analyzing how coordinated messages were used to influence the discussions. The Supreme Federal Court (Topic 1) became a focal point of political debates during the election, which was marked by widespread protests and many allegations of election fraud (Topic 6), making it a critical subject in public discourse.

Case Study 1: Supreme Federal Court (Topic 1) In this case, we identified a coordinated attack targeting the Brazilian Supreme Federal Court (STF), specifically aimed at doxxing the locations of the ministers. During the analyzed period captured by our dataset, the members of the Supreme Court became focal points of intense online discussions, which can be observed by Topic 1. To better understand this content, we analyzed the messages related to this topic and selected the most widely shared coordinated message about the ministers, which is shown as follows:

Message

"We have just discovered the hotel where the ministers of the Supreme Federal Court are staying in New York, please forward this to all Brazilians in the USA. xxxxW xxth St, New York, NY xxxxxx, United States" (translated and anonymized to not show address).

This message was shared 144 times within the entire dataset, with 29.2% of those shares coming from coordinated activity. The message reached 102 WhatsApp groups (7% of the total dataset), posted by 42 coordinated accounts.

Context In November, following the Brazilian presidential election, there were tensions surrounding the decisions of the Supreme Court. When the ministers traveled to New York for a conference on November 13th, their hotel location was maliciously doxxed online, inciting an attack. This led to a flood of messages on WhatsApp, encouraging people to gather and confront the ministers.

Analyzing the Impact The dissemination of this message had a significant real-world impact. Protesters quickly mobilized to the location, gathering outside the hotel on the night of November 13. The ministers faced harassment and confrontations when entering and leaving the hotel.[9] The rapid spread of this information was crucial in organizing these protests, demonstrating how coordinated actions on WhatsApp can escalate from the digital ecosystem to a physical response. The incident highlights the power of doxxing, coordinated by some users, to endanger public figures by inciting hostile actions within hours.

Case Study 2: Electoral Fraud (Topic 5) In this case, we analyzed the topic of Electoral Fraud. We searched for coordinated messages using the terms “fraud” and “ballot boxes”. From the top three most relevant messages based on the total number of shares, we found the following message:

> **Message**
>
> *“Our president said ALL PEACEFUL PROTESTS ARE WELCOME!!!! COME ON MY PEOPLE!! WE WILL NOT BACK DOWN! ALL THE RIGHT-WING IS GOING TO TAKE TO THE STREETS - THE ARMED FORCES ARE JUST WAITING TO REACH THE NUMBER TO HAVE THE NATIONAL AND INTERNATIONAL QUORUM THAT IS THE MASS OF THE POPULATION IN THE STREETS TO MEET THE DEMANDS OF THE PEOPLE”* (translated).

This message was shared 83 times during the analyzed period, and 24% of these shares were in coordinated activities. It reached 74 different groups, and two coordinated accounts were involved in the coordinated actions of this message.

Context After the election, former President Bolsonaro made his first public statement about the election. After that, messages like this one began to be shared, claiming that his pronouncement contained a subliminal message encouraging people to go to military posts and demand military intervention due to the alleged fraudulent election result.

Analyzing the Impact This coordinated message was widely shared after Bolsonaro’s speech. After that, the protests intensified.[10] Many people took to the streets in front of military posts, demanding military intervention and alleging fraud in the polls. This coordinated message reinforced and motivated users to continue protesting and taking specific actions that had a real impact on society. This kind of message needs to be spread quickly, and the rapid coordinated actions work perfectly in this context, reaching more people quickly.

By examining these coordinated examples, we can reinforce that WhatsApp is an extremely politically relevant tool in Brazil, and coordinated activities can have a much greater influence, expanding discourse and reinforcing narratives. In our context, we observed that coordinated actions impacted orchestrating real-world events, such as protests and specific mobilizations. These coordinated actions aim to reach as many people as possible quickly, which is evident in both cases analyzed.

Takeaways The main takeaways from this section are:

- The proposed topic analysis reveals that key Brazilian political events are highlighted in coordinated messages, particularly those that raise suspicions about electoral fraud and call for military intervention.
- We found evidence that the events discussed in the case studies (i.e., the mobilization of people against Supreme Court justices and protests over the election results) were also driven by coordinated accounts in WhatsApp groups.

6 Discussion & Conclusion

In this study, we provide valuable insights into the coordinated activities that drive message propagation on WhatsApp. By examining public political groups from July 2022 to January 2023 with a total of 13 million messages, we found a significant prevalence of coordinated accounts in disseminating messages on the platform. Our analysis reveals the presence of 1,575 coordinated accounts that work simultaneously to disseminate messages across multiple groups. These coordinated activities are focused on spreading news messages that can be easily shared across various groups. Our investigation showed that 26% of the links shared are from misinformation websites. This strategic news dissemination aims to engage audiences and promote specific political viewpoints. Furthermore, we observed that coordinated actions had a significant impact, as they were used to previously orchestrate protests and important political actions in the Brazilian political scenario.

Overall, this study sheds light on a frequent but little explored phenomenon on WhatsApp. While the influence of misinformation and messaging apps on society is well recognized, the influence of coordinated activities is quite new. Our research reveals compelling evidence of coordinated efforts to disseminate misinformation on the platform and mobilize people to specific events. Importantly, even though we can not access the entire WhatsApp network, we identified many coordinated accounts that are working on spreading messages through public groups. This suggests that the problem is likely to be even larger than observed. Even with WhatsApp recognizing the existence of coordinated campaigns in the last Brazilian elections (Mello 2019) and lim-

[9] https://www.cnnbrasil.com.br/politica/manifestantes-hostilizam-ministros-do-stf-na-porta-de-hotel-em-nova-york/

[10] https://www.reuters.com/world/americas/bolsonaro-backers-call-brazil-military-intervene-after-lula-victory-2022-11-02/

iting forwarding per user to control virality (Melo et al. 2019b), it was not enough to solve the problem, especially because coordinated accounts do not frequently use the forwarding tool to spread their content. By observing the real impact employed by coordinated users, more notable strategies are necessary to mitigate the problem of controlling the influence of coordinated accounts. In 2018, WhatsApp already banned hundreds of thousands of accounts detected as spammers in Brazil.[11] Banning is a strategy that temporarily mitigates the problem, but it needs to be aligned with other control policies to keep the WhatsApp environment healthier. Even because new features were introduced, facilitating massive spreading, such as increasing the number of participants per group and creating communities.

The methods to combat coordination misinformation should consider simultaneous posting patterns to mitigate coordination actions between multiple groups or accounts, since simultaneous posting activity is an important factor in the effectiveness of coordinated campaigns, such as protests or organized attacks. In this regard, effectively addressing coordinated misinformation is essential to foster collaboration between platforms and government authorities. This collaboration is particularly critical in high-impact contexts such as elections, where misinformation and coordinated accounts can directly and harmfully impact society. Increasing platform transparency is crucial. This should not only involve reporting on coordinated campaigns, but also include measures to limit the volume of messages during critical periods, improve content moderation algorithms, and detect and remove harmful and inauthentic activity.

Acknowledgements

This work was supported by grants from CAPES, CNPq and FAPEMIG.

References

Augenstein, I.; Baldwin, T.; Cha, M.; Chakraborty, T.; Ciampaglia, G. L.; Corney, D.; DiResta, R.; Ferrara, E.; Hale, S.; Halevy, A.; et al. 2024. Factuality challenges in the era of large language models and opportunities for fact-checking. *Nature Machine Intelligence*, 1–12.

Benevenuto, F.; and Melo, P. 2024a. Misinformation Campaigns through WhatsApp and Telegram in Presidential Elections in Brazil. *Commun. ACM*, 67(8): 72–77.

Benevenuto, F.; and Melo, P. 2024b. Misinformation Campaigns through WhatsApp and Telegram in Presidential Elections in Brazil. *Communications of the ACM*, 67(8): 72–77.

Blondel, V. D.; Guillaume, J.-L.; Lambiotte, R.; and Lefebvre, E. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10): P10008.

Burghardt, K.; Rao, A.; Chochlakis, G.; Sabyasachee, B.; Guo, S.; He, Z.; Rojecki, A.; Narayanan, S.; and Lerman, K. 2024. Socio-linguistic characteristics of coordinated inauthentic accounts. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 18, 164–176.

Bursztyn, V. S.; and Birnbaum, L. 2020. Thousands of small, constant rallies: a large-scale analysis of partisan WhatsApp groups. In *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining*, ASONAM '19, 484–488.

Carmo, D.; Piau, M.; Campiotti, I.; Nogueira, R.; and Lotufo, R. 2020. PTT5: Pretraining and validating the T5 model on Brazilian Portuguese data. arXiv:2008.09144.

Conover, M.; Ratkiewicz, J.; Francisco, M.; Gonçalves, B.; Menczer, F.; and Flammini, A. 2011. Political polarization on twitter. In *Proceedings of the international aaai conference on web and social media*, volume 5, 89–96.

Ferrara, E.; Varol, O.; Davis, C.; Menczer, F.; and Flammini, A. 2016. The rise of social bots. *Communications of the ACM*, 59(7): 96–104.

FORCE11. 2020. The FAIR Data principles. https://force11.org/info/the-fair-data-principles. Accessed: 2025-04-07.

Garimella, K.; and Tyson, G. 2018. WhatsApp, Doc? A First Look at WhatsApp Public Group Data. *Proceedings of the International AAAI Conference on Web and Social Media*, 12.

Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Iii, H. D.; and Crawford, K. 2021. Datasheets for datasets. *Communications of the ACM*, 64(12): 86–92.

Giglietto, F.; Righetti, N.; Rossi, L.; and Marino, G. 2020. It takes a village to manipulate the media: coordinated link sharing behavior during 2018 and 2019 Italian elections. *Information, Communication & Society*, 23(6): 867–891.

Grinberg, N.; Joseph, K.; Friedland, L.; Swire-Thompson, B.; and Lazer, D. 2019. Fake news on Twitter during the 2016 US presidential election. *Science*, 363(6425): 374–378.

Grootendorst, M. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv:2203.05794.

Kansaon, D.; Melo, P. F.; Zannettou, S.; Feldmann, A.; and Benevenuto, F. 2024. Strategies and Attacks of Digital Militias in WhatsApp Political Groups. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 18, 813–825.

Keller, F. B.; Schoch, D.; Stier, S.; and Yang, J. 2020. Political astroturfing on twitter: How to coordinate a disinformation campaign. *Political communication*, 37(2): 256–280.

Kirdemir, B.; and Adeliyi, O. 2023. Towards Characterizing Coordinated Inauthentic Behaviors on YouTube. In *The 2nd Workshop on Reducing Online Misinformation through Credible Information Retrieval (ROMCIR 2022).*

Mello, P. C. 2019. WhatsApp admite envio maciço ilegal de mensagens nas eleições de 2018. https://www1.folha.uol.com.br/internacional/en/brazil/2019/10/whatsapp-admits-to-illegal-mass-messaging-in-brazils-2018.shtml. Accessed: 2025-04-07.

[11] https://wapo.st/3ZudkSD

Melo, P.; Messias, J.; Resende, G.; Garimella, K.; Almeida, J.; and Benevenuto, F. 2019a. WhatsApp Monitor: A Fact-Checking System for WhatsApp. *Proceedings of the International AAAI Conference on Web and Social Media*, 13(01): 676–677.

Melo, P.; Vieira, C.; Garimella, K.; de Melo, P. O. V.; and Benevenuto, F. 2019b. Can WhatsApp Counter Misinformation by Limiting Message Forwarding? In *Proceedings of the International Conference on Complex Networks and their Applications*.

Melo, P. F.; Hoseini, M.; Zannettou, S.; and Benevenuto, F. 2024. Don't Break the Chain: Measuring Message Forwarding on WhatsApp. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 18 of *ICWSM'24*, 1054–1067.

Nizzoli, L.; Tardelli, S.; Avvenuti, M.; Cresci, S.; and Tesconi, M. 2021. Coordinated behavior on social media in 2019 UK general election. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, 443–454.

Nobre, G. P.; Ferreira, C. H.; and Almeida, J. M. 2022. A hierarchical network-oriented analysis of user participation in misinformation spread on WhatsApp. *Information Processing & Management*, 59(1): 102757.

Nobre, G. P.; Ferreira, C. H. G.; and Almeida, J. M. 2020. Beyond groups: Uncovering dynamic communities on the whatsapp network of information dissemination. In *Social Informatics: 12th International Conference, SocInfo 2020, Pisa, Italy, October 6–9, 2020, Proceedings 12*, 252–266.

Pacheco, D.; Flammini, A.; and Menczer, F. 2020. Unveiling coordinated groups behind white helmets disinformation. In *Companion proceedings of the web conference 2020*, 611–616.

Pacheco, D.; Hui, P.-M.; Torres-Lugo, C.; Truong, B. T.; Flammini, A.; and Menczer, F. 2021. Uncovering coordinated networks on social media: methods and case studies. In *Proceedings of the international AAAI conference on web and social media*, volume 15, 455–466.

Pasquetto, I. V.; Swire-Thompson, B.; Amazeen, M. A.; Benevenuto, F.; Brashier, N. M.; Bond, R. M.; Bozarth, L. C.; Budak, C.; Ecker, U. K.; Fazio, L. K.; et al. 2020. Tackling misinformation: What researchers could do with social media data. *The Harvard Kennedy School Misinformation Review*.

Resende, G.; Melo, P.; Sousa, H.; Messias, J.; Vasconcelos, M.; Almeida, J.; and Benevenuto, F. 2019. (Mis)Information Dissemination in WhatsApp: Gathering, Analyzing and Countermeasures. In *The Web Conference*, 818–828. ACM.

Ribeiro, M. H.; Ottoni, R.; West, R.; Almeida, V. A.; and Meira Jr, W. 2020. Auditing radicalization pathways on YouTube. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*, 131–141.

Silva, M.; Oliveira, L. S. d.; Andreou, A.; Melo, P. O. V. d.; Goga, O.; and Benevenuto, F. 2020. Facebook Ads Monitor: An Independent Auditing System for Political Ads on Facebook. In *Proceedings of The Web Conference (WWW'20)*, 224–234.

Tardaguila, C.; Benevenuto, F.; and Ortellado, P. 2018. Fake News Is Poisoning Brazilian Politics. WhatsApp Can Stop It. https://www.nytimes.com/2018/10/17/opinion/brazil-election-fake-news-whatsapp.html. Accessed: 2025-04-07.

Venâncio, O. R.; Ferreira, C. H.; Almeida, J. M.; and da Silva, A. P. C. 2024. Unraveling User Coordination on Telegram: A Comprehensive Analysis of Political Mobilization during the 2022 Brazilian Presidential Election. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 18, 1545–1556.

Vishnuprasad, P. S.; Nogara, G.; Cardoso, F.; Cresci, S.; Giordano, S.; and Luceri, L. 2024. Tracking fringe and coordinated activity on Twitter leading up to the US Capitol attack. In *Proceedings of the international AAAI conference on web and social media*, volume 18, 1557–1570.

Weber, D.; and Neumann, F. 2021. Amplifying influence through coordinated behaviour in social networks. *Social Network Analysis and Mining*, 11(1): 111.

Yang, K.-C.; Hui, P.-M.; and Menczer, F. 2019. Bot electioneering volume: Visualizing social bot activity during elections. In *Companion Proceedings of The 2019 World Wide Web Conference*, 214–217.

Yang, K.-C.; Varol, O.; Hui, P.-M.; and Menczer, F. 2020. Scalable and generalizable social bot detection through data selection. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 1096–1103.

Ethics Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes!

   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes, the claims in the abstract and introduction accurately reflect the paper's contribution and scope.

   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes, we state in the Introduction why our mixed-methods approach is suitable and appropriate for identifying and characterizing coordinated behavior on WhatsApp.

   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? No, because as described in Data Collection Section, we do not have access to representative samples from WhatsApp so we can not make any claims about the data used and its representativeness.

   (e) Did you describe the limitations of your work? Yes, the limitations of our work are mainly related to data collection (see Section Data Collection.

   (f) Did you discuss any potential negative societal impacts of your work? No, because we do not foresee any potential negative societal impact from this work.

(g) Did you discuss any potential misuse of your work? No, because we do not foresee any potential misuse of this work. Our research aims to raise awareness and inform the public and messaging platform operators about the existence of coordinated behaviour on WhatsApp.

(h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes, we describe measures we take to prevent or mitigate potential negative outcomes of our research in Ethics Sections, which includes a discussion about how we dealt with sensitive information.

(i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes, we have read the ethics review guidelines and ensured that our paper conforms to them.

2. Additionally, if your study involves hypotheses testing...

(a) Did you clearly state the assumptions underlying all theoretical results? NA

(b) Have you provided justifications for all theoretical results? NA

(c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? NA

(d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? NA

(e) Did you address potential biases or limitations in your theoretical framework? NA

(f) Have you related your theoretical results to the existing literature in social science? NA

(g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? NA

3. Additionally, if you are including theoretical proofs...

(a) Did you state the full set of assumptions of all theoretical results? NA

(b) Did you include complete proofs of all theoretical results? NA

4. Additionally, if you ran machine learning experiments...

(a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? NA

(b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? NA

(c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? NA

(d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? NA

(e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? NA

(f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? NA

5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...

(a) If your work uses existing assets, did you cite the creators? NA

(b) Did you mention the license of the assets? NA

(c) Did you include any new assets in the supplemental material or as a URL? NA

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? NA

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? NA

(f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see FORCE11 (2020))? NA

(g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see Gebru et al. (2021))? NA

6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...

(a) Did you include the full text of instructions given to participants and screenshots? NA

(b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA

(d) Did you discuss how data is stored, shared, and deidentified? NA

Ethics Statement

Handling WhatsApp data requires careful consideration of privacy and ethical guidelines, as messages may contain sensitive content. All data collected for this study came from public groups, with access granted through openly shared invitation links, ensuring that no private conversations or content were monitored. User anonymity was also preserved by not storing personally identifiable information, such as phone numbers or users' real names. Our analysis does not rely on this information.

Studying the dynamics of political discourse and misinformation on social media also requires careful attention. Our keyword list includes a balanced number of terms for different political alignments in Brazil to ensure greater diversity of groups and avoid methodological bias. Despite precautions, right-leaning groups are more prominent on WhatsApp in Brazil, which also affects our data. However, we believe that our methodology can capture an accurate perspective of the Brazilian political landscape on WhatsApp, which is inherently asymmetrical.

We do not intend to release the collected dataset, as the potential risks outweigh the possible benefits. The data may contain sensitive information, and we do not have a complete understanding of the entire dataset (e.g., the prevalence of hateful content remains uncertain). Given the controversial nature and the potential inclusion of harmful or extremist content in our dataset, we can not guarantee that all sensitive data has been anonymized, which could also violate the General Data Protection Regulation (GDPR) law. Furthermore, another reason is our concerns about the potential misuse of the dataset, including the incitement of violence and the spread of harmful ideologies.

A Dataset Collection Criteria and Keywords

A.1 Keywords

For group discovery, we used a set of keywords related to the Brazilian political scenario initially proposed by (Resende et al. 2019), and we included updated terms related to relevant people and topics of the period analyzed.

These keywords include terms related to relevant political figures during the period analyzed, such as *"Nikolas Ferreira"*, *"Padre Kelmon"*, *"Simone Tebet"*, and *"Alexandre de Moraes"*.

Moreover, we added keywords associated with the 2022 presidential election, including: *"Eleições 2022 (election 2022)"*, *"Fraude (fraud)"*, *"Supremo Tribunal Federal (Supreme Federal Court)"*, *"Tribunal Superior Eleitoral (Superior Electoral Court)"*, *"Urna Eletrônica (electronic ballot box)"*, and *"Ditadura Militar (military dictatorship)"*.

Additionally, we included keywords related to the COVID-19 pandemic, such as: *"Pandemia (pandemic)"*, *"Covid-19"*, *"Vacina (vaccine)"*, *"Coronavirus"*, *"Tratamento Precoce (early treatment)"*, and *"Ivermectina (ivermectin)"*.

A.2 Platforms Used for Group Discovery

The following sources and websites were searched using the keyword list to identify potential WhatsApp groups. We conducted searches on social networks, such as Twitter and Facebook, as well as online repositories indexed by Google, including *https://gruposwhats.app*, *https://www.gruposdewhatss.com.br*, and *https://gruposdezap.com*.

B Rapid Spread Network Threshold Sensitivity

To assess the sensitivity of the 60-second time window threshold in the Rapid Spread Network, we performed an additional analysis to evaluate the impact of varying time windows (30 and 90 seconds). Our goal was to determine whether the chosen time window significantly affects the network structure and to validate the applicability of the methodology to WhatsApp. We applied the methodology summarized in Section 3.2 and Section 3.3 across two different thresholds (30 and 90 seconds).

Creating the rapid spread network with a 60-second time window, we obtained a network with 1,575 nodes, 1,491 edges, and 14,414 coordinated messages (Section 4). With a 30-second time window, we have a network with 994 nodes, 918 edges, and 10,465 coordinated messages. For a 90-second time window, the network contains 2,038 nodes, 2,086 edges, and a total of 18,328 coordinated messages.

From these results, modifying the time window from 60 to 90 seconds leads to an increased number of nodes, edges, and coordinated messages. However, the core network structure remains unchanged, with over 70% of the nodes overlapping, indicating that the network is not highly sensitive to moderate changes in the time window. Coordinated users identified with a shorter time window remain part of the network structure when a longer threshold is applied. As the threshold rises, the network structure expands, but this does not necessarily lead to identifying more coordinated users. Instead, it may result in more coincidental users sharing the same message due to the larger time interval, highlighting the importance of finding a good balance. Similarly, using a very short time interval may cause many coordinated users to be missed. Based on these findings, the 60-second time window remains a well-supported and balanced choice, suggesting its applicability in the context of WhatsApp.