A Study of Misinformation in Audio Messages Shared in WhatsApp Groups

Alexandre Maros[1], Jussara M. Almeida[1], and Marisa Vasconcelos[2]

[1] Universidade Federal de Minas Gerais, Brazil
{alexandremaros,jussara}@dcc.ufmg.br
[2] IBM Research, Brazil
marisaav@br.ibm.com

Abstract. Recent studies have shown that group communication on WhatsApp plays a significant role to foster information dissemination at large, with evidence of its use for misinformation campaigns. We analyze more than 40K audio messages shared in over 364 publicly accessible groups in Brazil, covering six months of great social mobilization in the country. We identify the presence of misinformation in these audios by relying on previously checked facts. Our study focuses on content and propagation properties of audio misinformation, contrasting them with unchecked content as well as with prior findings of misinformation in other media types. We also rely on a set of volunteers to perform a qualitative analysis of the audios. We observed that audios with misinformation had a higher presence of negative emotions and also often used phrases in the future tense and talked directly to the listener. Moreover, audios with misinformation tend to spread quicker than unchecked content and last significantly longer in the network. The speaker's tone from the audios with misinformation was also considered less *friendly* and *natural* than the unchecked ones. Our study contributes to the literature by focusing on a media type that is gaining mainstream popularity recently, and, as we show here, is being used as vessel for misinformation spread.

Keywords: WhatsApp · Audio Messages · Misinformation.

1 Introduction

WhatsApp has become a major communication platform worldwide. In fact, as of January 2021, two billion users were accessing the messenger app on a monthly basis[3], surpassing, by far, the monthly usage of other popular platforms such as Facebook Messenger, Telegram, and Snapchat. The app stands out for offering a simple and easy-to-use set of features that allows anyone to quickly share texts, images, audios, videos, or files with individual users or several people at once, through the so-called group communication.

[3] www.statista.com/statistics/258749/most-popular-global-mobile-messenger-apps

The widespread use of WhatsApp motivated several studies on different aspects of the platform [7, 13, 15, 18, 20–22]. Most studies exploited the fact that, though private spaces by default, WhatsApp groups can be made publicly accessible as group managers share invitation links in public websites. By clicking on those links and joining these publicly accessible groups, researchers were able to gather data for further analysis. These studies showed that WhatsApp is not a mere communication tool but rather exhibits characteristics of social networks like Facebook, and Reddit, with the emergence of robust networks interconnecting users which facilitate the quick spread of information [18, 22]. With a particular focus on the misuse of the platform for spreading fake content, some authors analyzed content properties and general propagation dynamics of misinformation shared in WhatsApp groups, aiming at identifying distinctive properties of this type of content in image [22] and textual [21] messages.

However, online users in general, and WhatsApp users in particular [9], have been showing a growing interest in *audio* content. On one hand, neither text nor image messaging can fully convey the sender's tone, urgency, emotion, or purpose [23] as audio content can, and some prior studies relied on audio media to capture these peculiarities [8]. As such, audio conversations may lead to fewer misinterpretations than textual content. On the other, sharing audio content can be more convenient, especially for the sender. Unfortunately, the increase in popularity of audio messages on WhatsApp has been followed by reports on the use of this type of media as an effective vessel to spread misinformation on the platform[4]. Indeed, in a preliminary analysis of audio messages shared in publicly accessible WhatsApp groups in Brazil, we showed some initial evidence of the presence of audios with previously checked fake content [13].

We build on our prior work [13], focused mostly on developing a methodology to analyze audio content gathered from WhatsApp groups. Following the evidence raised in that work, we here delve into WhatsApp audio messages, offering what we believe to be *the first analysis of the spread of misinformation in audio messages*. Specifically, we aim to tackle the following research questions (RQs):

RQ1: What are the characteristics of audio messages with misinformation shared in publicly accessible WhatsApp groups in terms of content properties and propagation dynamics? How do they compare to prior findings of misinformation in other media types on the platform [21, 22]?

RQ2: How do the content and dynamics properties of audio messages carrying previously checked misinformation compare with the properties of the other audio messages?

To address these questions, we rely on a dataset obtained from [22], consisting of over 43 thousand messages collected from *publicly accessible* and *politically-oriented* WhatsApp groups in Brazil. Our study relies on qualitative and quantitative analyses to unveil content and propagation properties of misinformation in audio, comparing them against similar properties of unchecked audio messages as well as properties of misinformation in other media types reported in [21, 22].

[4] https://www.politico.com/news/2020/03/16/coronavirus-fake-news-pandemic-133447.

Our main findings reveal that audio communication is widely used in the 364 monitored WhatsApp groups, with more than 42 thousand audio messages across six months. Based on the misinformation detection, we marked over 120 unique audios that were shared more than 2,000 times across 260 groups during the monitored period. We observed that audios with misinformation appear in more groups and are shared by more users than their counterparts. Lastly, we noticed many particular characteristics that emerged more often in audios with misinformation, such as a call to action (actively asking the listener to take some action, such as share the audio) and being more related to negative emotions.

This paper is organized as follows. Section 2 presents related work. Section 3 describes our methodology to analyze audio content share on WhatsApp groups, notably how we identified audios containing misinformation. Our analyses on content properties, including the results of a qualitative investigation, are presented in Section 4, whereas results on propagation dynamics are discussed in Section 5. Conclusions and future work are offered in Section 6.

2 Related Work

Recent studies investigated the dissemination of misinformation on social media platforms. Since the 2016 U.S. presidential election, the spread of misinformation is increasing around the world. The so-called fake news may contribute to political polarization, decrease trust in public institutions, and lead people to have less faith in the political process. Social media bots on Twitter were observed during the 2016 U.S. Presidential election [3]. Out of almost 3 million distinct users involved in political discussions, 400 thousand were bots being responsible for 3.8 million tweets (one-fifth of all collected tweets). These numbers are worrisome since these bots can act in an orchestrated way to influence and promote discussion, impulsing content with misinformation, and influencing what is being discussed by real users [1]. Bots are not only targeting politics, but also several other areas, such as debates regarding vaccination campaigns [5], and are present in several social networks [10].

Recently, a few studies have looked into user behavior and content dissemination in WhatsApp groups. Garimella *et al.* [11, 14] proposed a general data collection methodology for collecting Whatsapp public groups and analyzed misinformation in images shared groups in India. Josemar *et al.* [7] analyzed political and non-political groups using cascade model as well as dissemination of misinformation. They observe that cascades with misinformation tend to be deeper, reach more users, and last longer in political groups than in non-political ones. Bursztyn *et al.* [6] focus on understanding the differences between right and left-wing Whatsapp groups during the 2018 Brazilian Presidential election . They found that right-wing groups are more abundant, tightly connected, and geographically distributed, while also sharing more multimedia messages. Melo *et al.* [15] evaluate the dynamics of the spread of misinformation in WhatsApp groups. Using an epidemiological model, the authors showed how the forwarding feature contributes to the content virality, and why system limitations are not effective to prevent a message to reach the entire network quickly. Resende *et al.* [22] studied the types of content shared in publicly accessible WhatsApp groups during two events in 2018 in Brazil. The authors proposed a method to identify misinformation in images shared across the groups. They found that images with previously checked misinformation tend to be reshared within shorter intervals and are more often shared first on WhatsApp and then on the Web. Later, the same authors extended their prior work by focusing on shared textual messages [21]. The analysis of psychological elements in the text showed a frequent presence of the insight category in messages with misinformation, often used in chain messages. In terms of propagation dynamics, textual messages are shared more times, by a larger number of users and in more groups. The ones containing misinformation tend to spread faster within particular groups, but take longer to propagate across different groups.

Lastly, our preliminary work [13] focused on studying **audio messages** in WhatsApp, going over a basic content analysis of all audio messages as well as some evidence on misinformation in audio messages. We build on this prior work and focus on identifying the differences of audio messages with misinformation versus unchecked content.

3 Methodology

In this section we describe the methodology we employed in our study, focusing particularly on the WhatsApp dataset used. We briefly review how we gathered and processed the dataset, following steps described in detail in [13]. We also present how we identified misinformation in the audio messages collected.

3.1 WhatsApp Dataset

This work relies on the same raw dataset collected in [21, 22], which contains messages collected from publicly accessible WhatsApp groups from 21st of May to 28th of October of 2018. Those studies focused on analyzing messages containing textual and image content only. More recently, we complemented those studies by looking into audio messages, offering a preliminary analysis of content properties, audio type (music vs. speech) and propagation dynamics [13].

To that end, we employed a multi-step methodology to process the audio content collected, consisting of (i) pre-processing, to guarantee that all audios are in the same format; (ii) similarity detection, to group audios that have similar content together; and (iii) speech recognition phase to transcribe the audios, allowing for the use of natural language processing tools to analyze their content. Very briefly, to identify and group audios with similar content, we employed the open-source library called Chromaprint[5], which processes and transforms the audio frequency in musical notes and uses this new representation to compare different files. Using the fingerprints produced by Chromprint, we performed a

[5] https://acoustid.org/chromaprint.

pairwise comparison of audios, grouping as "similar" (or near-duplicates), audio pairs for which Chromaprint returned a score of similarity above a given threshold, which was selected after a manual investigation of a sample of audios pairs. For each group of near-duplicates, we elected one audio as representative of the content. Moreover, to be able to analyze that content, we used Google's Speech-to-Text API[6] to produce a transcription of each (unique) audio content. This API returns a score of confidence on the transcription produced. Based on a manual evaluation, we selected a threshold of confidence of 0.8, only keeping transcriptions whose confidence exceeded this defined threshold. In this phase, we also filter audios that were not in Portuguese, as those yields a low confidence threshold since the API is set to Portuguese. We refer the reader to [13] to a detailed description of how these steps were performed.

Table 1. Dataset overview (* users and groups with at least one audio message).

	Truck Drivers' Strike	Election Campaign	Whole collected period
# Groups	117	330	364
# Users*	1,134	6,002	8,056
# Audio message	5,780	28,593	42,869
# Unique audios	1,450	8,505	16,503
# Transcripted Audios	987	5,913	11,700

Table 1 shows overall statistics about the dataset for the whole period of analysis, as well as for each of the two selected periods separately, the national truck drivers' strike (between May 21st and June 2nd) and the general election campaign in Brazil (from August 16th to October 28th). It shows the total number of audio messages shared, the total number of users who shared at least one audio, and groups where at least one audio was shared. Overall, we have more than 8 thousand different users who shared almost 43 thousand audio files in 364 different groups. The table also shows the total number of unique audio contents as well as the number of unique audios for which a transcription with enough confidence was obtained, identified following the methodology described above. The latter, corresponding to around 71% of all audios, was indeed the content used in our analysis. We note that the same audio content was shared 3-4 times on average. However, as we will see later, some audio contents were shared a much larger number of times.

3.2 Misinformation Detection

We expand the methodology developed in [13] by adding a fourth step: misinformation detection. Detecting misinformation is a challenging task. Prior efforts relied on various strategies, including the use of fuzzy analytic hierarchy process [2] and the detection of social bots as an initial step for computation fact-checking [16]. Another approach is by relying on fact-checking journalists and agencies,

[6] https://cloud.google.com/speech-to-text/.

which are experts in assessing the truth of a public claim by seeking reliable sources, analyzing facts, images, and videos as well as directly contacting those involved in these claims [12].

Following prior analyses of misinformation in textual and image content on WhatsApp [21, 22], we here chose to rely on fact-checking agencies to find misinformation in the content of the audios in our dataset. We used a dataset containing a list of fact-checked claims, made available by [22], gathered from 6 important fact-checking agencies sites in Brazil. We identified audios with misinformation by comparing each audio transcription in our dataset with a fact-checked claim marked as containing misinformation by at least one of the fact-checking agencies. Specifically, we first pre-processed all audio transcriptions and checked-as-fake claims by removing stopwords and using lemmatization. Next, we represented each transcription and checked-as-fake claim by a TF-IDF vector [17]. Then, we calculated the similarity of each audio transcription a with a checked-as-fake claim b by computing the cosine similarity between the corresponding TF-IDF vectors. We did that for each pair (a, b) of transcription a and checked-as-fake claim b.

We manually analyzed the 300 pairs of texts (audio transcription and checked-as-fake claims) with the highest cosine similarity. Our goal was to assess whether the audio transcription contained the same content as the previously checked fake claim. As per the manual analysis, we found that only 100 out of the 300 transcriptions analyzed indeed carried the same content as the claim they were matched to with the highest similarity. These audios were marked as containing misinformation[7], and are the focus of our analyses in the following sections. All other 200 audio messages, as well as all other audios with lower similarity compared to the collected claims, were marked as *unchecked*. We use the term unchecked to emphasize that all we can state is that they are not similar to any previously checked-as-fake claim collected. Thus, strictly speaking, they might or might not carry misinformation. However, we do expect that we were able to catch most audio messages containing misinformation in our dataset, especially those with greater impact on users, as they most probably were reported by at least one of the fact-checkers. As future work, we intend to explore other text similarity methods such as word embeddings.

4 Content Analysis

We now analyze the content of the audios shared, distinguishing between audios with misinformation and unchecked content. To that end, we focus on the audio transcriptions. We start by uncovering the main topics of discussion conveyed in each set of shared audios, and then we look into some psychological linguistic features extracted from their transcriptions. Next, we rely on volunteers to offer a qualitative analysis of the content of a sample of audios.

[7] These 100 audios represent distinct content that was shared during the period of analysis. As we will see, each content was indeed shared multiple times.

Table 2. Most representative words for each topic inferred by LDA method

Topic	Most representative words
1	Brazil, Country, Person, Brazilian, Politician, Year, PT, Family, Govern, Defend
2	Expensive, Talk, Stay, See, Understand, Marry, End, Impose, Woman, Nobody
3	Federal, Public, Congressperson, Million, Lula, Paulo, Money, Year, Candidate, Politician
4	Military, Brazil, Stop, Truck Driver, Army, World, Brazilian, Military intervention
5	Bolsonaro, Vote, Brazil, Haddad, PT, President, Jair, Election
6	God, Lord, Jesus, Life, Word, Day, Love, Heart, Father, Name
7	Guys, People, Talk, Understand, Stay, Do, Happen, Find
8	Day, Hour, Guys, City, Car, Night, Today, Come, Friend

4.1 Topic Analysis

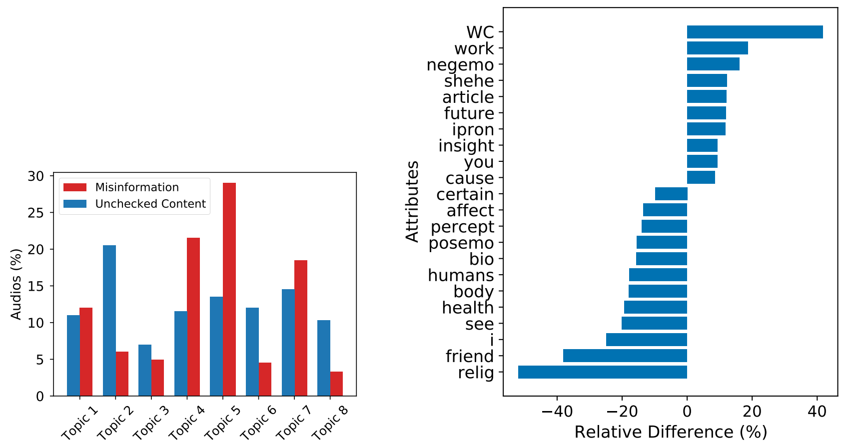
To infer the topics conveyed by the audio messages, we employed the Latent Dirichlet Allocation (LDA) algorithm [4] on our collection of audio transcriptions. LDA receives as input all audio transcriptions and the desired number of topics k, and it computes the topic distribution, which can be interpreted as k clusters, each one represented by a word distribution. Words with higher weights in this distribution are more representative of the given topic, thus for each input audio transcript, we can infer the most related topic.

We pre-processed all the transcriptions by removing punctuation marks and stopwords, lowercasing, and stemming all the words. As a next step, we used all the (pre-processed) transcriptions as input to the LDA model. We used the LDA implementation provided by Gensim[8], a Python library that implements LDA algorithm. Based on this, we obtain the words associated with the k topics learned by the model, and with those words, we can get a better understanding of what is discussed in each topic. To select the best number of topics k, we ran the algorithm varying the number of topics k from 2 to 20 and assessed the quality of the results, measuring the topic coherence c_v. We found the best topic coherence at $k = 8$ topics.

Table 2 presents the top-10 most representative words of each topic. Note that topics 1, 3, and 5 are closely related to politics since they are characterized by words such as “Campaign”, “Brazil”, “Mayor”, “Politician”, “PT” (an important political party in the country) and so on. Topic 4 is closely related to the truck drivers' strike event, identified by the words “Trucks”, “Truck Driver” and “Military Intervention” (a topic largely discussed during the truck driver's strike). Topic 6 contains mostly words related to religion, suggesting that many audios were recordings of members of Christian denominations members. Finally, topics 2, 7-8 are more loosely connected and encompass more general narratives.

To analyze the distribution of topics across different audio transcriptions, we first assigned to each transcription the most prevalent topic according to LDA results, i.e., the topic with the highest probability associated with the transcription. Figure 1a presents the distributions of topics across different transcriptions, separately considering audios with misinformation and unchecked ones. Note that 52% of audios with misinformation are characterized as containing content

[8] https://radimrehurek.com/gensim/.



(a) LDA topic distribution

(b) Relative difference in prevalence of LIWC attributes between misinformation and unchecked

Fig. 1. Content Properties of Audios with Misinformation and with Unchecked ones.

related to topics 4 and 5, which are the most strongly related topics to the social mobilization events that happened in the period – the truck drivers' strike and the elections. These two topics are characterized by words such as "Military", "Truck Driver", and "Election". Topic 7 is the third most predominant topic among audios containing misinformation: 18% of them are characterized by this topic which covers words such as "Guys" and "Do". Unchecked audios are more equally distributed across all topics. The topic that holds the largest fraction of audios with unchecked content is Topic 2, which is characterized by words like "Understand" and "Talk", with almost 21% of audios falling into this category.

4.2 Psychological Linguistic Features

We also analyzed the psycholinguistic properties of the audio transcriptions using the the 2015 Linguistic Inquiry and Word Count (LIWC) lexicon [19]. LIWC that categorizes words into linguistic style, affective and cognitive attributes. We ran each audio transcription through the Portuguese version of LIWC[9], computing the distributions of each LIWC attribute among audios that were marked as containing misinformation as well as the audios with unchecked content. For each LIWC attribute, we compared the two distributions to identify which attributes were significantly different across the two sets of audio messages. To compare each pair of distributions, we use the Kolmogorov-Smirnov test, selecting attributes that had a p-value < 0.05 as significantly different.

Aiming at contrasting the most common LIWC attributes in audio transcriptions classified as misinformation and transcriptions containing previously

[9] Provided by http://143.107.183.175:21380/portlex/index.php/pt/projetos/liwc.

unchecked content, we computed, for each LIWC attribute that was marked as being significantly different in the two types of messages, the ratio of the difference between the values of the attribute in audios with misinformation and audios with unchecked content to the value of the attribute in audios with unchecked content. This ratio implies how much more prevalent, percent-wise, the attribute is among audios with misinformation, compared to audios with unchecked content. Positive ratios imply that the attribute is more present in audios with misinformation, whereas negative ratios mean a greater prevalence among unchecked content.

Figure 1b shows the relative differences for attributes identified as significantly different across the two sets of audios. We note that audios with misinformation tend to be longer, with a higher word count (WC). Furthermore, messages with misinformation tend to be more related to work, characterized by words such as jobs and employment, have more negative emotions (e.g., "hate", "ugly"), use words from the third person singular, such as "she" and "he", carry phrases in the future tense and have words related to insights, such as "think" and "know". Moreover, audios with misinformation also tend to use words such as "you" or "your" (e.g., "it is your problem") and use words related to causation, such as "because" and "to that effect". In contrast, messages with previously unchecked content tend to carry more positive emotions, use the first person singular and cover words related to health, religion and friendship. These observations point towards a clear distinction in discourse in the misinformation audios. Interestingly, when comparing these results with those obtained for textual messages [21], we notice that in both audio and textual messages, there is the predominance of the insight attribute on misinformation. However, in textual messages, words such as "we" and "they" appear more frequently in misinformation, which is not the case here, with "you" being more frequent. Textual messages with misinformation also had a high presence of the sexual attribute, however we found no significant presence of this attribute in audio messages with misinformation. Lastly, textual messages also tend to be associated with the present, whereas audios are more often associated with the future tense.

4.3 Qualitative Analysis

To delve deeper into the content of the audios, we conducted a qualitative analysis of a sample of 100 audios by volunteers. The study is composed of two phases: an interview and a survey. The goal of the interview was to gather the perceptions of selected volunteers on the audios' content with a limited number of audios and use the insights to develop an online survey to reach a broader public.

First Phase: Interview We interviewed three volunteers separately, with each interview consisting of a one-hour online session, using a semi-structured format, with a defined list of questions to be answered by the volunteers. Each volunteer filled a consent form to allow the use of their responses in this study in an anonymous format. Each interview can be divided into two phases. The first phase consists of questions aimed at learning more about the volunteers, their participation in WhatsApp groups, and their perception of audio in general. During the second phase, the volunteers were asked to listen to four randomly selected audio files. After each audio, they were asked to answer questions regarding their perception of it. Out of the four audios, two contained misinformation, and two had unchecked content.

Regarding the volunteers' perception of the listened audios, we analyzed their answers separately for audios with misinformation and audios with unchecked content. Audios with misinformation were spotted easily as potential sources of misinformation, possibly due to the volunteers' close familiarity with studies on misinformation[10]. In some cases, they also reported finding a certain tone of artificiality in the tone of the speakers of these audios. The volunteers often pointed out that often the speaker of audios with misinformation tried to create a link with someone important as a strategy to bring credibility to the information being transmitted (e.g., the speaker was related to a famous newscaster) as well as trying to back their claims with sources. The volunteers also noted that audios with misinformation tried to engage more with the listener, often trying to create the illusion of familiarity and intimacy, referring to the listener by terms that relate to friendship or family. Overall, we noticed some peculiarities common in audios with misinformation, such as the feeling of uneasiness and the use of strategies to engage more with the public. With these peculiarities mapped out, we moved into the second part of the analysis to identify whether these characteristics were frequent.

Second Phase: Online Survey Based on the insights collected in the interview, we set up an online survey to gather more information on differences between audios with misinformation and unchecked content. We also wanted to check whether the previous remarks collected were also noticed by a larger group on a broad set of audios. The online form is composed of two sets of questions, one related to volunteers' demographics and the frequency of usage of audios in WhatsApp groups, and the other related to impressions they had after listening to the selected audio. We selected a random sample of 100 audios, 50 with unchecked content and 50 with misinformation content. Each audio was evaluated by 3 different people while each volunteer evaluated 5 different audios.

In total, 25 volunteers participated in the online survey. Starting with the responses to the first set of questions, the volunteers were, on average, 27 years old, 16 of them identified themselves as male and 9 as female. Moreover, the largest group they participated in varied from 25 to 256 members, with an average of 88, which coincides with previous observations that WhatsApp groups tend to connect a large group of people [7, 21]. In general, our volunteers reported that they receive audios more often than send them on WhatsApp.

For the second part of the survey, each answer was tied to a specific audio file, and we discuss them separately for audios with misinformation and with

[10] This potential bias of the volunteers is not problem in itself as the main focus of this phase was to raise the general impression that audios with misinformation evoked on the volunteers.

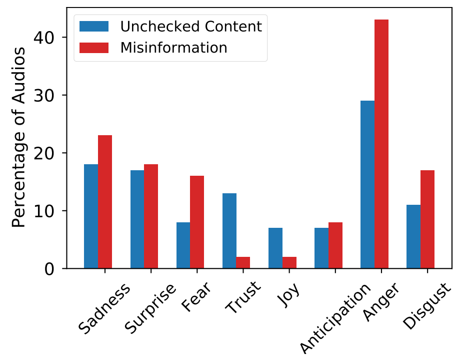


Fig. 2. Emotions felt by listeners of misinformation and unchecked content.

unchecked content. First we asked which emotion did the volunteer feel when listened to the audio. Figure 2 shows the percentages of audios for which different emotions were selected by the volunteers when listening to them. For each audio, we considered all the emotions selected by all volunteers. Thus, note that the percentages may exceed 100%. Note that volunteers felt more negative emotions when listening to audio messages with misinformation. This might be due to the higher presence of negative words, as we discussed in Section 4.2. Sadness, surprise, fear, disgust, and especially anger were most felt while listening to audios with misinformation, whereas trust and joy were most reported when listening to audios with unchecked content.

When asking whether did they think the audio contained false information, when presented audios with misinformation, our volunteers spotted them 76% of the time. When asked whether the audio had some form of data or source to back the information, 58% responded yes when presented an audio with misinformation and only 17% responded yes when presented an audio with unchecked content. However, when asked whether the provided source increased the credibility of the audio, only 24% of the volunteers said it did indeed increase the credibility. This reaffirms some points raised by the volunteers of the interview: many audios with misinformation try to back their history with some study or data, but they are often not reliable enough. This also links back to the greater prevalence of the *insight* attribute in audios with misinformation, as reported in Section 4.2. Words that characterize this attribute, such as “think”, “consider”, “know”, are often used to create a storyline. Overall, most volunteers said they would not share any of the audios with friends or family: indeed, for audios with misinformation, only 9% of the volunteers mentioned that they would share them, whereas, for audios with unchecked content, this fraction drops to 5%.

We then preceded to ask about the naturality, excitement, and friendliness of the audio. Each volunteer was able to select a number from zero to four for each of the three questions. For naturality, zero indicated “Very Artificial” whereas four indicated “Very Natural”. As for excitement and friendliness, zero indicated “Very Sad” and “Very Hostile”, and four indicated “Very Excited” and “Very

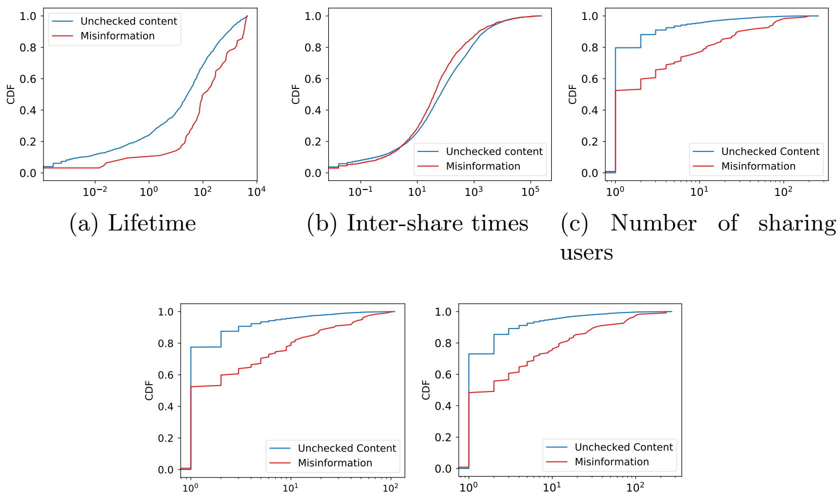
Friendly" respectively. We found a significant difference in the answers of volunteers regarding the *naturality* and *friendliness* of the speakers for audios with misinformation and with unchecked content, but no significant difference with respect to *excitement*. Volunteers reported that speakers of audios with misinformation tend to be more often less friendly (average score of 1.78) than speakers of audios with unchecked content (average score of 2.34), with a statistically significant difference according to a t-test (p-value ≤ 0.05). As to the naturality of the speaker, the gap is even larger: the average score was 1.65 for audios with misinformation and 2.56 for unchecked content (statistically significant difference with p-value ≤ 0.05). That is, speakers of audios with misinformation tend to more often give the impression of an artificial tone. Lastly, when asking whether the audio had any call to action, volunteers reported that they could often identify this characteristic in audios with misinformation. Indeed, some type of instruction to be executed by the listener (e.g., *share the audio in more groups*) was reported in 72% of the cases of audios with misinformation. For audios with unchecked content, this fraction falls to 32%.

In sum, the survey results suggest the following key observations. Audios with misinformation tend to make the listeners feel more negative emotions, such as sadness, fear, anger, and disgust, and more often mention some source to try to support their claims, although these sources were often seen as unreliable and, in many cases, did not make the information more believable. Moreover, the speaker's tone of audios with misinformation was considered less *friendly* and less *natural* than the audios with unchecked content. Finally, audios with misinformation more often resorted to some type of call to action, notably as a strategy to help spread the content.

5 Propagation Dynamics

In this section, we look into the propagation dynamics of audio messages, looking into the metrics lifetime and inter-share time. The former is the time interval between the first and the last times a particular audio content was shared in any monitored group, $t_n - t_1$, where n represents the number of times the audio was shared in any group, whereas the latter is the time interval between consecutive shares of the same content (regardless of the group in which it was shared), $t_2 - t_1, ..., t_n - t_{n-1}$. We also look into how many groups each audio message reaches and how many unique users share the same audio.

Figure 3a shows the distributions of lifetimes for audios with misinformation and unchecked content. As shown, audios with misinformation tend to last longer on the platform: 75% of audios with misinformation remained being (re-)shared by up to 31 days, whereas the same fraction of audios with unchecked content lasted at most 7 days. These numbers represent a significant increase compared to prior results on image messages. According to [22], no significant difference in the lifetimes of images with misinformation and unchecked content was observed, as roughly 70% of images in either category lasted up to 100 hours. With respect to textual messages, on the other hand, prior results [21] are somewhat similar



(a) Lifetime (b) Inter-share times (c) Number of sharing users

(d) Number of groups (e) Number of shares

Fig. 3. Propagation Dynamics of Misinformation and with Unchecked Audios.

to what was observed here for audios, though with longer lifetimes. For instance, 50% of textual messages with misinformation lasted up to 10 days in the system. In contrast, here we observe that the same fraction of audios with misinformation lasts for up to 6 days.

Figure 3b shows the distributions of inter-share times. We see only a small difference in the distributions for misinformation and unchecked content. Roughly speaking, around half of the audios with misinformation are re-shared within 40 minutes whereas the same fraction of audios with unchecked content are re-shared within 65 minutes. Thus, audios with misinformation tend to spread slightly more quickly than unchecked ones. This result is consistent with prior findings that misinformation in both image and textual content spreads faster. However, audios with misinformation tend to spread much more quickly than images with misinformation: according to [21], around 80% of the images with misinformation are re-shared within 100 minutes, but we found that only 65% of the audios with misinformation are re-shared with the same time interval.

We now turn to the analysis of the reach of the audio messages, in terms of number of users who shared them and number of groups where they were shared. Note that the latter gives an idea of the potential audience of these messages. Fig. 3c shows that audios with misinformation tend to be shared by a larger number of distinct users: around 80% of audios with misinformation are shared by at least 12 different users, while 80% of unchecked audio are shared by at most two people. Also, as shown in Fig. 3d, audios with misinformation tend to reach a much larger potential audience: 90% of audios with misinformation are shared in at least 27 different groups, while the same fraction of audios with unchecked content appears only in three groups. Ultimately, Fig. 3e shows that audios with misinformation tend to be shared a much larger number of times: 20% of them were shared more than 13 times, while 80% of the audios unchecked content had a maximum share count of only two.

These numbers show the greater potential of "viralization" that audios with misinformation have over general, unchecked audio. Indeed, audios with misinformation tend to often target topics that are incredibly relevant to the political scenario that they appear in, such as political candidates during the electoral period, or involving major opinions toward strikes, as seen in Section 4.1. They also bring many psychological attributes that catch people's attention and have a direct impact on our emotions, such as the use of negative words, or attributes regarding future, as seen in Section 4.2 and even in the response from the interviews conducted in Section 4.3. Finally, they include contents that make them potentially more engaging, such as "sources" that try to back their stories or employ strategies to engage the listener in actions (e.g., re-sharing), as seen in Section 4.3. These observed characteristics of audios with misinformation are consistent with prior findings for images and textual messages as well [21, 22] and may contribute to their attractiveness and virality. An avenue of future work is to explore the greater presence of these properties in the design of methods to detect misinformation and mitigate its harmful impact.

6 Conclusions and Future Work

Recent studies looked at the propagation of textual and image content in WhatsApp, but to our knowledge, no study focused on misinformation in audio content. In that context, our goal was to understand of how audio messages with misinformation are used in publicly accessible WhatsApp groups. We first focused on understanding the characteristics of these audio messages in terms of content properties and propagation dynamics, while also looking at the differences to prior findings for text and images.

Regarding the topics discussed in the audios, four were directly related to politics and had the largest fraction of misinformation. Other discussion topics were related to religion and chatter. Analyzing the psychological attributes, we identified a higher presence of words related to negative emotions and insight states in audios with misinformation. They also had more phrases in the future tense and referring directly to the listener using pronouns such as "you". Prior analyses on WhatsApp textual content found the frequent presence of terms that aggregate people, such as "we", and verbs often in the present tense. Thus, indicating different types of approaches depending on the type of media used.

We conducted a qualitative analysis to deepen our knowledge about the audio messages, gathering the perception of selected volunteers on the audio's content and potential feelings the speaker's voice triggered, analyzing audios with misinformation and with unchecked content separately. One key result from that analysis is that audios with misinformation tend to more often make the listener feel negative emotions, such as sadness and anger. Also, audios with misinformation often tried to back their claims by citing some sources, often perceived by the volunteers as unreliable. The speaker's tone from the audios with misinformation was considered less *friendly* and less *natural* than audios with unchecked content. Lastly, volunteers also noted that audios with misinformation carried some instruction for the listener, such as sharing the audio with other groups.

Finally, we looked into how these audios propagated in these groups. We observed that misinformation audios appear in more groups, are shared by more users, and have overall more shares than unchecked content. Many factors can explain it, such as being targeted for incredibly relevant topics to the political scenario that they appear in and having attributes that catch people's attention, and directly impacting the listeners' emotions.

A possible direction for the future consists of expanding our analysis to account for audios shared across many years, looking into how these properties behave across time, and possibly detecting seasonal events. We also would like to compare how the messages behave across the same topics of discussion. Lastly, another direction is to expand our misinformation detection pipeline to reliably and automatically detect audios with misinformation, thus expanding our current analysis to a larger quantity of audio files.

References

1. Allcott, H., Gentzkow, M.: Social media and fake news in the 2016 election. Journal of economic perspectives **31**(2), 211–36 (2017)
2. Baeth, M., Aktas, M.: Detecting misinformation in social networks using provenance data. In: SKG (2017)
3. Bessi, A., Ferrara, E.: Social bots distort the 2016 us presidential election online discussion. First Monday **21**(11-7) (2016)
4. Blei, D., Ng, A., Jordan, M.: Latent dirichlet allocation. JMLR **3**(Jan), 993–1022 (2003)
5. Broniatowski, D., Jamison, A., Qi, S., AlKulaib, L., Chen, T., Benton, A., Quinn, S., Dredze, M.: Weaponized health communication: Twitter bots and russian trolls amplify the vaccine debate. American journal of public health **108**(10) (2018)
6. Bursztyn, V., Birnbaum, L.: Thousands of small, constant rallies: A large-scale analysis of partisan whatsapp groups. In: ASONAM'19 (2019)
7. Caetano, J., Magno, G., Gonçalves, M., Almeida, J., Marques-Neto, H., Almeida, V.: Characterizing attention cascades in whatsapp groups. In: WebSci'19 (2019)
8. Cunningham, S., Ridley, H., Weinel, J., Picking, R.: Supervised machine learning for audio emotion recognition. Personal and Ubiquitous Computing (2020)
9. Fernando, G.: Why people are switching from texting to voice messages. www.news.com.au/technology/gadgets/mobile-phones/why-people-are-switching-from-texting-to-voice-messages/news-story/d36d6d80cc0c71da168b4e8ec96924e7 (2018), [accessed 02-23-2021]
10. Ferrara, E., Varol, O., Davis, C., Menczer, F., Flammini, A.: The rise of social bots. Commun. ACM **59**(7), 96–104 (Jun 2016)
11. Garimella, K., Eckles, D.: Images and misinformation in political groups: evidence from whatsapp in india. HKS Misinformation Review. (2020)
12. Graves, L.: Deciding what's true: Fact-checking journalism and the new ecology of news. Ph.D. thesis, Columbia University (2013)

13. Maros, A., Almeida, J., Benevenuto, F., Vasconcelos, M.: Analyzing the use of audio messages in whatsapp groups. In: TheWebConf (2020)
14. Melo, P., Messias, J., Resende, G., Garimella, K., Almeida, J., Benevenuto, F.: Whatsapp monitor: A fact-checking system for whatsapp. In: ICWSM'19 (2019)
15. Melo, P., Vieira, C., Garimella, K., Melo, P., Benevenuto, F.: Can whatsapp counter misinformation by limiting message forwarding? In: ICCA'19 (2019)
16. Menczer, F.: The spread of misinformation in social media. In: TheWebConf (2016)
17. Moon, A., Raju, T.: A survey on document clustering with similarity measures. IJERT **3**(11), 599–601 (2013)
18. Nobre, G., Ferreira, C., Almeida, J.: Beyond groups: Uncovering dynamic communities on the whatsapp network of information dissemination. In: SocInfo (2020)
19. Pennebaker, J., Boyd, R., Jordan, K., Blackburn, K.: The development and psychometric properties of liwc2015. Tech. rep. (2015)
20. Reis, J., Melo, P., Garimella, K., Benevenuto, F.: Can whatsapp benefit from debunked fact-checked stories to reduce misinformation? (2020)
21. Resende, G., Melo, P., Reis, J., Vasconcelos, M., Almeida, J., Benevenuto, F.: Analyzing textual (mis)information shared in whatsapp groups. In: WebSci (2019)
22. Resende, G., Melo, P., Sousa, H., Messias, J., Vasconcelos, M., Almeida, J., Benevenuto, F.: (mis)information dissemination in whatsapp: Gathering, analyzing and countermeasures. In: TheWebConf (2019)
23. Sherman, L., Michikyan, M., Greenfield, P.: The effects of text, audio, video, and in-person communication on bonding between friends. Cyberpsychology **7**(2) (2013)