Analyzing the Use of Audio Messages in WhatsApp Groups

Conference Paper · April 2020

DOI: 10.1145/3366423.3380070

Jussara Almeida
Federal University of Minas Gerais

Fabrício Benevenuto
Federal University of Minas Gerais

Marisa Vasconcelos
IBM Research, Sao Paulo, Brazil

Analyzing the Use of Audio Messages in WhatsApp Groups

Alexandre Maros
alexandremaros@dcc.ufmg.br
UFMG, Brazil

Jussara Almeida
jussara@dcc.ufmg.br
UFMG, Brazil

Fabrício Benevenuto
fabricio@dcc.ufmg.br
UFMG, Brazil

Marisa Vasconcelos
marisaav@br.ibm.com
IBM Research, Brazil

ABSTRACT

WhatsApp is a free messaging app with more than one billion active monthly users which has become one of the main communication platforms in many countries, including Saudi Arabia, Germany, and Brazil. In addition to allowing the direct exchange of messages among pairs of users, the app also enables group conversations, where multiple people can interact with one another. A number of recent studies have shown that WhatsApp groups play an important role as an information dissemination platform, especially during important social mobilization events. In this paper, we build upon those prior efforts by taking a first look into the use of *audio* messages in WhatsApp groups, a type of content that is becoming increasingly important in the platform. We present a methodology to analyze audio messages shared in WhatsApp groups, characterizing content properties (e.g, topics and language characteristics), their propagation dynamics and the impact of different types of audios (e.g., speech versus music) on such dynamics.

CCS CONCEPTS

• **Human-centered computing → Social media**; • **Applied computing → Sociology**.

KEYWORDS

WhatsApp, Audio Messages, Content Propagation

ACM Reference Format:
Alexandre Maros, Jussara Almeida, Fabrício Benevenuto, and Marisa Vasconcelos. 2020. Analyzing the Use of Audio Messages in WhatsApp Groups. In *Proceedings of The Web Conference 2020 (WWW '20), April 20–24, 2020, Taipei, Taiwan.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3366423.3380070

1 INTRODUCTION

WhatsApp is a free world-wide messaging app which currently has more than 1.5 billion active monthly users [14], and has become one of the main communication platforms in many countries, including Saudi Arabia, Malaysia, Germany, and Brazil. The numbers in Brazil are especially expressive: approximately 65% of the country population (209 million people) have access to and use WhatsApp as a communication platform [17].

WhatsApp has some key features that make it stand out. Firstly, the platform offers a simple and easy-to-use set of features that allows anyone to quickly share texts, pictures, audios, videos or any kind of file with individual users or several people at once, through the so-called groups. These groups are limited to 256 people and the access is controlled by the group administrators. Secondly, the contents shared on WhatsApp are end-to-end encrypted which means that the content is only encrypted or decrypted in the phones of those involved in the communication, and technically cannot be sent by anyone else. This feature is especially interesting when compared to other popular messaging apps, such as Telegram and Facebook's Messenger, which do not have end-to-end encryption enabled by default. Finally, there are tools for quickly spreading information in the network, such as broadcasting, which allows users to send a message to 256 contacts or groups at once, or forward content to 5 different people or groups[1].

A number of recent studies have analyzed content dissemination in WhatsApp groups, focusing on groups whose administrators chose to share invitation links online thus effectively turning the access to the group conversations public to anyone with those links [7, 8, 10, 16, 20, 21]. These prior studies focused mostly on image and textual content, characterizing content properties as well as propagation dynamics and offering quantitative evidence of the use of the platform to spread misinformation (e.g., fake news) [8, 20, 21].

Yet, neither text nor image messaging can fully convey sender's tone, urgency, emotion or purpose as audio content can. Moreover, audio conversations tend to lead to less misinterpretations than textual content. Indeed, it has been reported that the use of voice messages on WhatsApp has rapidly increased recently: over 200 million voice messages are sent by WhatsApp app every day in some regions[2]. Thus we here take a first look into audio content shared on WhatsApp groups, tackling the following questions:

RQ1: What are the characteristics of audio messages shared in WhatsApp groups in terms of content properties and propagation dynamics? How do they relate to prior findings for other types of content (text and images)?

RQ2: What are the introspect properties of audio content (e.g., gender of speaker, music versus speech content) and how these properties correlate with propagation dynamics?

To address these questions, we use the dataset collected in [20, 21] which consists of messages collected from *publicly accessible* and *political-oriented* WhatsApp groups during two major social

WWW '20, April 20–24, 2020, Taipei, Taiwan

ACM ISBN 978-1-4503-7023-3/20/04.
https://doi.org/10.1145/3366423.3380070

[1] Before 2019, the forward limit was 20

[2] https://www.news.com.au/technology/gadgets/mobile-phones/why-people-are-switching-from-texting-to-voice-messages/news-story/d36d6d80cc0c71da168b4e8ec96924e7

events in Brazil. However, unlike those prior studies which focused on textual and image content, we here restrict our analysis to *audio* content.

Our analysis confirmed that audio is an important source of communication in WhatsApp, having more than 20 thousand audio messages being shared in 330 groups in a span of a few days. Audio messages are shared, on average, 3-4 times and last 1-2 days in the network. To understand what was being said, we automatically transcribed the audios and analyzed the different psychological features between audios that were more shared than others while also doing a topic analysis of the conversations. We also categorized audios between music and speech and observed that music is overall more shared than speech. A second classification was on the gender of the primary speaker but found no significant difference in the number of shares between these two groups. Finally, we briefly discuss some evidences of misinformation in audio messages and show that these audios have a higher spread than audios who were unchecked. We also plan to release our dataset in due time with the audio files alongside their transcriptions.

2 RELATED WORK

A number of studies have analyzed information dissemination in different online platforms (e.g., Twitter, Facebook, Reddit) [3, 12, 13] and its influence on major real life events such as electoral campaigns and social mobilization [1, 6]. Some of them focused particularly on the propagation of misinformation [13, 20], often supported by automated bots [3].

More recently, a number of studies have analyzed user behavior and content dissemination in publicly accessible WhatsApp groups. Melo *et al.* proposed a general data collection methodology [16] while Seufert *et al.* analyzed the implications of group conversations on mobile network traffic [22]. Caetano *et al.*, in turn, proposed a hierarchical methodology to analyze user interactions [8]. Resende *et al.* analyzed the content properties and propagation dynamics of images [20] and textual messages [21] shared on WhatsApp political oriented publicly accessible groups during two major events in Brazil. The authors revealed the presence of misinformation being disseminated in those groups, and concluded the messages with misinformation, both in image and textual content, tend to be shared more times and more quickly than other messages

In [8], the authors also looked into differences between misinformation and other content shared in WhatsApp groups with the goal of modeling collective user attention. To that end, they used the concept of *cascades*. They analyzed the structural and dynamic properties of cascades in political and non-political groups and analyzed the impact of misinformation on such properties.

In [7], the authors analyze differences between WhatsApp groups that are primarily left-wing and right-wing. The authors found that right-wing groups, such as how they are more tightly connected and geographically distributed, while also sharing more multimedia messages. Finally, Melo *et al* explores the the speed with which information is spread WhatsApp groups considering an epidemiological model [10]. Subsequently, the work identified parameters that could control, difficult or slow down the propagation of misinformation, such as to how many people can a person forward a single message.

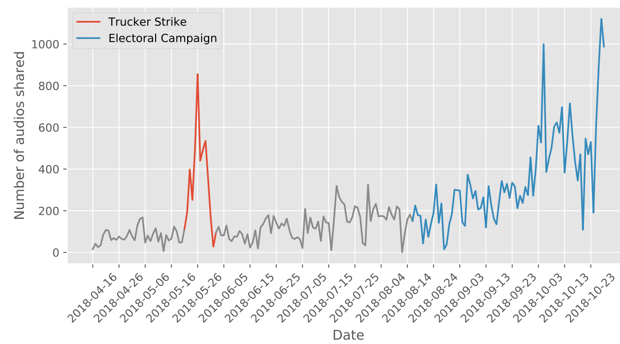


Figure 1: Number of audios shared in the monitored period

3 METHODOLOGY

In this section we describe the dataset used in this work as well as the methods used to process it and analyze it.

3.1 WhatsApp Dataset

We used the original dataset collected in [20, 21], which contains all messages shared in hundreds of political oriented publicly accessible WhatsApp groups during 6 months in Brazil: from April 16th to October 28th 2018[3]. But unlike those studies, which focused on image and text messages, we here constrain our analyses to audio data.

Figure 1 shows the number of audio messages shared on a daily basis during the period of collection. There is a peak of activity in May (in red), which coincides with a national truck drivers' strike (between May 21st and June 2nd) that generated a lot of social mobilization in the country. We also note a steady increase in the volume of audios shared by the end of the collection (marked in blue), which coincides with the period of the general election campaign in Brazil (from August 16th to October 28th). Thus, we focus our analyses only on audio content shared during these periods.

Period	Type	Quantity
Trucker Drivers' Strike	# Groups*	117
	# Users*	1,134
	# Audio messages	5,780
	# Unique Audios	1,450
Election Campaign	# Groups*	330
	# Users*	6,002
	# Audio messages	28,593
	# Unique Audios	8,505

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

Table 1 shows overall statistics about the data for the two selected periods. It shows the total number of audio messages shared as well as the total numbers of users who shared at least one audio and groups where at least one audio was shared. In general, compared

[3] We thank the authors for sharing the data with us.

to the total numbers reported in [20], roughly 32% and 21% of all users active in the monitored groups during the electoral campaign and truck drivers' strike periods, respectively, shared at least one audio message. The fraction of monitored groups with at least one audio content also increased from 83%, during the truck drivers' strike to 90% during the election period. These numbers illustrate the increasing user participation in sharing audio content within WhatsApp groups.

The table also shows the total number of unique audio contents, which refer to potentially different audio files that convey the same content. We note that each audio content was shared 3-4 times on average, although, as we will see later, some audio contents were shared a large number of times. We describe how we identified different audios with similar content next.

3.2 Grouping Audios with Similar Content

The identification of audios with similar contents cannot be performed by simply comparing whether both files are identical (on a bit level) as there might be many variations of the same audio. For example, different compression methods could be used, one audio file could be cut slightly shorter, the same audio could be recorded by different devices, etc. Thus, to identify and group audios with the same content, we employed a fingerprinting method. Specifically, we used an open source library called Chromaprint[4], which processes and transforms the audio frequency in musical notes and uses this new representation to compare different files and is used in many other studies, such as [2, 4, 15, 19]. The method returns a score between 0 and 1, where 0 indicates completely different audios, and 1 being a perfect match.

We only compared pairs of audios with durations that differ by no more than two seconds from each other as larger differences were assumed to reflect different contents. We manually analyzed a sample of 300 pairs of audios (i.e., listened to both audios and decided whether they had the same content), finding that all comparisons for which Chromaprint returned a score of 0.9 or above indeed consisted of the same content. Lower values otherwise resulted in some false positives. Thus, we selected 0.9 as a threshold and used it to group the audios with similar content.

3.3 Audio Content Analysis

In order to be able to analyze the properties of the audio contents, we relied on two strategies. First, we relied on 20 volunteers to manually annotate a sample of audios. Specifically we randomly selected 100 audios from the truck drivers' strike period, 100 audios from the electoral period and 100 audios from the top 500 audios most shared in our dataset, adding up 300 different audios. We asked the group of volunteers to categorize each audio into eight categories to better understand the information they conveyed. For each sampled audio, we required three annotations from three different volunteers. Finally, the volunteers were instructed to select all categories that fit its content. For comparison purposes, we adopted the same eight categories used in [20], however, we listened to a great share of the audios to see whether the categories still made sense for this media format and whether there were additional categories missing. We were satisfied with the original categories, which are:

- **Opinion**: a content expressing the speaker's opinion;
- **News**: information about an event, quoting or referencing a newspaper, magazine or news portal;
- **Politics**: information related to a candidate or party to publicize or praise some political subject;
- **Advertising**: commercials or ads related to a product, venue or service;
- **Satire**: Humorous content about current events or people;
- **Activism**: content encouraging or mentioning social movements, protests or other events
- **Inappropriate**: Hate speech, pornography;
- **Others**: content does not fit any other category;

We measured the inter-annotator agreement using Cohen's kappa coefficient (κ) [9]. The average Cohen's kappa considering all categories is $\kappa = 0.49$, which indicates a moderate agreement. Considering each category individually, the one that had the lowest agreement is *Inappropriate* with $\kappa = 0.18$ (therefore it was filtered out), while the category with the highest agreement is *Politics*, with $\kappa = 0.78$.

As a second strategy, we also performed a speech to text transcription of the audio messages in our dataset. To that end, we employed an automatic translation method, the Google Cloud's Speech-to-Text[5] tool. Like Chromaprint, the Google Speech's API also returns a score between 0 and 1 which reflects the confidence of the model on the transcription produced. We did validate the tool by asking a number of volunteers to judge the transcriptions performed on a sample of 300 audios. Specifically, each volunteer was asked to first listen to the audio and then respond whether the transcription correctly reflected the audio content. Each volunteer responded using a 0-4 Likert scale, where 0 indicates a complete disagreement and 4 indicates completely agreement. Each audio was judged by 3 volunteers.

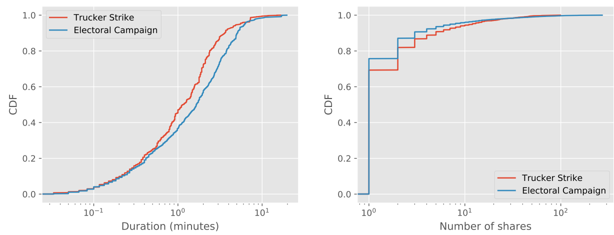
We found a strong correlation between the transcription scores and the average responses of the volunteers, with a pearson correlation coefficient equal to 0.86. Indeed, we found that all audios with transcription scores above 0.8 received on average 3.6 points by the volunteers, suggesting high transcription quality. In turn, audios with scores below 0.8 received on average 1.7 points, suggesting very poor transcriptions. Interestingly we found that almost all audio messages containing music content fell in the poor transcription category. Thus, in our analysis of the textual properties of the audio messages, we only used audios for which the transcription score was 0.8 or above.

4 GENERAL CHARACTERISTICS

In this section, we discuss general characteristics of the audio messages shared in the monitored groups. We start by discussing the duration of the audio messages. Figure 2 shows, for both analyzed periods, the cumulative distributions (CDF) of the durations for audios with up to 20 minutes (only 139 unique audios are longer than 20 minutes). The average duration is around 2 minutes for both periods, though audios shared during the election period tend to be somewhat longer: around 20% of the audios shared during that

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

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



(a) CDF of audio duration for audios under 20 minutes (b) CDF of number of audio shares

Figure 2: CDF of Audio duration and number of shares

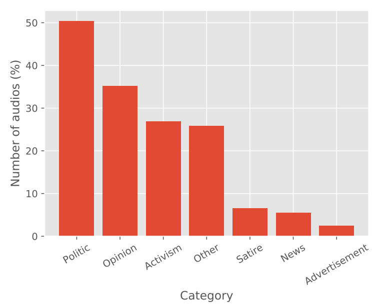
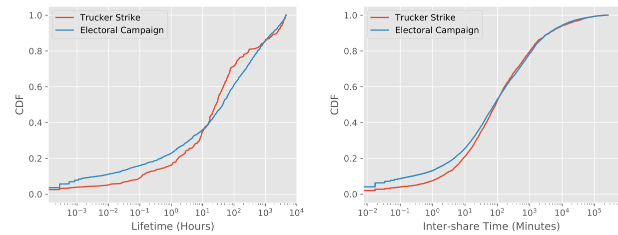


Figure 3: Distribution of audios across categories.

period have more than 3,5 minutes (versus 2.5 minutes during the strike). Figure 2 also shows the CDFs of the total number of times each audio message (i.e., all audios grouped as similar content) was shared in all monitored groups in both periods analyzed. As shown, some audios have a large reach: for instance, 10% of the audios where shared more than 10 times during the election campaign and the audio that appeared most times had 270 shares. We computed the Pearson and Spearman coefficient between the number of shares and the duration of the audio but found no noticeable correlation.

Next, we discuss the contents of the audio messages by looking into the manual categorization performed by the volunteers. Figure 3 shows how the 300 sampled audios are distributed across the eight categories. Note that the sum exceeds 100% as an audio may have been associated to more than one category at once. The main categories of audios are *Politics*, followed by *Opinion* and *Activism*. Most audios that contain opinions also belong to the politics category. Audios in the *Others* category mostly relate to religious content, specific events, or unrelated chatter. *Satire*, *News* and *Advertisement* categories appeared in less than 10% of audios. Compared to a similar categorization of image messages reported in [20], we observe a much larger presence of personal opinions and activism related content among the audio messages but a less frequent use of this type of media to spread satirical content. These results illustrate important differences in how different media types are used to disseminate content in WhatsApp.

We also analyze the propagation dynamics of audio messages. Figure 4 shows the distributions of lifetimes and inter-share times.



(a) Lifetimes (b) Inter-share times

Figure 4: Distributions of lifetimes and inter-share times of audio messages.

The former is the time interval between the first and the last times a particular audio content was shared in any monitored group, whereas the latter is the time interval between consecutive shares of the same content (irrespective of the group), during both periods.

As shown in the figure 4-a), 50% of the audios stopped being reshared after only one day since the first appearance during the truck drivers' strike. During the election campaign, half of the audios remained being reshared for up to two days. Moreover, according to Figure 4-b), roughly 60% of the audios are re-shared within 3 hours and 20% are re-shared within 6 minutes, in both periods. These numbers are significantly different from those previously reported for textual messages [21]: audio messages tend to spread more slowly (longer inter-share times) but also remain for shorter periods in the system (shorter lifetimes). Such differences may simply reflect the greater effort required to listen to an audio message (compared to reading a text). In any case it is interesting to note that a fraction of the audios remained in the system for quite some time: the lifetimes exceed 10 days for roughly 20% of the audios.

5 CONTENT PROPERTIES

In this section, we explore the properties of audio messages shared in the groups. For each audio message, we analyze the type of audio (e.g., speech vs. music) and using the textual transcription, the topics and psychological features of the ones related to speech.

5.1 Speech vs Music Classification

Using a convolutional neural network described in [11], we classify Whatsapp audio messages into three categories: speech, music and random noises/inactivity periods [6]. For audio messages classified as speech audio, we also extract the gender of the speaker. Audios messages that contained both male and female speakers were classified as the predominant gender (e.g., who had spoken more time).

To evaluate the model efficiency in our data set, since the model is trained in French and we are using it in Portuguese, we asked the participants that manually classified the 300 audios (described in 4) to also answer the question of whether the audio was Speech (and if it was speech, which gender was predominant) or Music. That way, we had a test set that consisted of 300 random audios from our data set. We ran the classification model on this test set

[6]There was none in our data set, all audios fell under speech or music categories

Type	Class	Gender	%
Truck Drivers' Strike	Speech	Male	75.8%
		Female	15.8%
	Music	-	8.4%
Election Campaign	Speech	Male	65.1%
		Female	18.0%
	Music	-	16.9%

Table 2: Distribution of Speech and Musical messages

and achieved an F1 score of 0.97, a recall of 0.97 and a precision of 0.98, meaning that the model worked well in our audio messages.

Table 2 shows the percentage of speech and music messages in the dataset. In both periods, we have a predominance of male speakers, reaching almost 76% in the Trucker Strike period while female speakers are between 16-18% in both periods. The music messages doubled during the Electoral Campaign period, mainly due to the many massive dissemination of "Electoral jingles", which is a common campaign method in Brazil to promote politicians during their campaign.

Table 3 shows average values for three distributions for the two audio categories: unique groups per message (how many unique groups an audio message appeared), unique users per message (how many unique users shared the audio) and the total number of shares per message. In all three distributions, we notice that the music audios spread more (t-test *p-value* < 0.05 for all analysis comparisons), appearing in more groups (43% more), being shared by more users (28% more) and having an average share count greater than the speech audios (30% more). Most of the music audios contained some type of political content. The top ten most shared musical audios, that had more than 80 (the first had 223 shares) appearances in the network, were political propaganda for the presidential candidate Jair Bolsonaro.

Type	Analysis	Mean ± SE
Speech	Unique groups per message	3.39 ± 0.08
	Unique users per message	3.01 ± 0.13
	Shares per message	2.53 ± 0.15
Music	Unique groups per message	4.88 ± 0.28
	Unique users per message	3.88 ± 0.45
	Shares per message	3.29 ± 0.57

Table 3: Speech vs. music spreading.

We also explored how Male versus Female audio speakers behaved in the network when looking at the three measures (unique groups, unique users and shares per message). Audio messages in which the gender was primarily female had a slightly higher average for all the three measures (e.g, 3.81 vs 3.29 unique groups per message). However, we found no evidence of these distributions being significantly different from each other (by calculating t-test and Kolmogorv-Smirnov p-value tests), which suggests that gender does not seem to influence popularity or dissemination speech messages in Whatsapp.

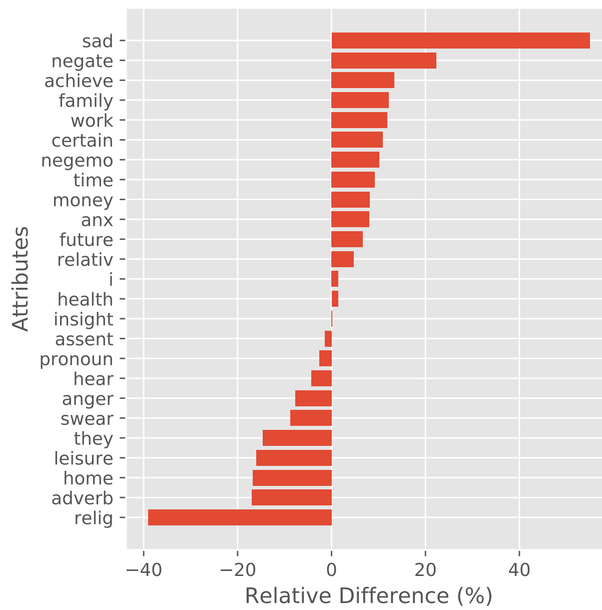


Figure 5: Relative difference between audio messages with more than 20 shares vs. with one share only.

5.2 Psychological Linguistic Features

To extract the distribution of psychological linguistic features from the audio transcriptions, we used the 2015 LIWC lexicon[18]. We use the Portuguese dictionary which has categories that map the presence of some words to cognitive and emotional components present in oral and written language samples.

We ran the LIWC analysis on each transcription and calculated the relative differences between each feature from audio messages that were shared more than 20 times compared to audios that were shared a single time. Figure 5 shows the relative differences for attributes were significantly different (p-value < 0.05) in terms of their distribution. A positive difference means that messages with more than 20 shares had more of that attribute than the ones that were shared only once and vice-versa.

From Figure 5, we note that the attributes with a greater presence in the most popular shared audio messages were related to sad emotions, negations (negate), needs, achievement, family, work, time, money, anxiety, and future. This can be explained by the political context when and where (political groups) that these audios were shared in both periods. We selected examples of some audio transcriptions shown in Table 4 with the respective words associate with the LIWC category in bold. Less popular audios have more attributes related to anger, swear and religion.

5.3 Topic Analysis

Finally, we further characterized the audio transcriptions in terms of the distribution of the topics they talk about. We use LDA (Latent Dirichlet Allocation)[5] model to infer the topics in our set of audio messages and run it varying the number of topics (k) from 2 to 20. We found the optimal LDA topic Coherence with $k = 8$. From

Attribute	Example
sad	I'm not saying that anyone who is listening to this audio votes in this or in that candidate, what **I regret** is that we, as believers, are constantly going to church and prayers, asking God to have **mercy of this country**
	[...] Brazil's democracy suffered **another blow**, and at this time, the **blow was almost fatal**, the presidential candidate fighting corruption put himself in front of a knife, a **sad and cowardly blow** [...]
work	[...] Lindbergh Farias receives an absurd amount of money each month and **is roaming in Curitiba instead of working at the Congress** [...]
	[...] I saw bandits being victimized and the **working citizen** arrested, held hostage to violence, I saw the schools violate the innocence of our children [...]
negate	**I don't know what is happening with Brazil**, with the working class, the people suddenly created a hate against PT, against the red flag, against the working class [...]

Table 4: Examples of transcriptions (Translated from Portuguese)

Topic	Most representative words
1	Bolsonaro, Vote, Go, Brazil, Haddad, 17, PT, Jair, Candidate, Election, Urn, Turn, Ok, President, Win, Research, Ciro, Lula, Campaign
2	Brasil, Country, Brazilian, Politic, Rule, Lula, PT, Be, Military, Money, Comunist, Defend, Family, Left, Arm, President, Nation, Corruption
3	City, Deputy, 13, São, Paulo, Rio, Mayor, State, Marry, Federal, Friend, Apply, Minas, Region, Amazonas
4	Go, People, Guys, Stop, Trucks, World, Car, Ask, Truck drivers, Audio, Go out, Streets, Want, Do
5	God, Jesus, Life, Lord, Word, Love, Be, Heart, Father, Bless, Church, Say, Brother, Christ, Man
6	Guys, Speak, Understand, Expensive, Find, Make, See, People, Broke, Want, Thing
7	Be, Be able, Have, Do, Exist, Get, Important, Situation, Have, Some, Life, Say, Fact, Duty, Law, Need, Feel, Information
8	Expensive, See, Speak, Stay, Marry, End, Woman, Order, Kill, Man, Kid, Child, Leave

Table 5: Topics distribution inferred by LDA method

Table 5, we can see that topics 1 to 3 are closely related to politics since they are characterized by words such as "Campaign", "Brazil", "Mayor" and so on. Topic 4 is closely related to the Trucker Strike event, identified by the words ' "Trucks", "Truck Drivers" and "Go Out" (as in go out to protest). Topic 5 contained mostly words related to religion suggesting that many audios were recordings of Christian denominations members. Finally, topics 6 and 7 are more loosely connect and encompass more general narratives.

5.4 Evidence of Misinformation

We briefly analyzed the presence of misinformation in the audio messages. To do so, we used the same methodology presented in [20]. Firstly, we calculate the cosine similarity based on the TF-IDF vectors of the audio transcription and a fact-checked news from the dataset collected in [20]. After that, we manually analyzed the 300 more similar comparisons (the comparisons that had the highest cosine similarity) and confirmed if the audio contained misinformation or not. Out of the 300 audios, 100 of them were marked as containing fake news. Comparatively, the audios with misinformation reached, on average, 376% more groups (2.51 versus 9.46 unique groups) and were shared by a number of users that was 480% higher than the messages that were marked as unchecked (2.95 versus 13.51 unique users).

6 CONCLUSIONS AND FUTURE WORK

We presented in this paper the first step in understanding audio messages dynamics in WhatsApp groups. To do so, we collected a dataset of audio messages from publicly accessible groups and analyzed their content and propagation dynamics on two major events in Brazil, the Trucker Strike, and the Presidential Electoral campaign.

We confirmed that audio messages are an important information vehicle type in WhatsApp, having more than 20,000 audio messages being shared in 330 groups in a span of a few days. The messages are shared 3-4 times on average and last 1-2 days in the network. When comparing this behavior with text messages, we could see that audios messages tend to spread more slowly and also remain for shorter periods in the systems, which could reflect the greater effort required to listen to an audio message when compared to reading a text. By automatically transcribing the audio messages, we analyzed the different psychological features between the most popular/shared audios, finding out that audio messages with a more sad connotation and closely related to family, work, time and money are the most shared. We also did a topic analysis of the conversations to understand the major topics discussed and found evidence of having audios containing misinformation. Finally, we observed that music audios were the most shared type of audio in the groups and we did not found a correlation between the gender of the primary speaker and their propagation dynamics.

As future work, we intend to analyze misinformation dissemination as found in images and text messages shared on Whatsapp [20, 21]. We will investigate which specific audio and text properties turn messages containing misinformation more viral.

ACKNOWLEDGMENTS

This work was partially supported by grants CNPq and Fapemig.

REFERENCES

[1] Rubens Achilles Ana Letícia Loubak. 2019. WhatsApp admite envio ilegal de mensagens em massa nas Eleições 2018. https://www.techtudo.com.br/noticias/2019/10/whatsapp-admite-envio-ilegal-de-mensagens-nas-eleicoes-2018.ghtml. Acessed in 2019-10-12.

[2] M. Bartsch and G. Wakefield. 2005. Audio thumbnailing of popular music using chroma-based representations. *IEEE Transactions on Multimedia* 7, 1 (Feb 2005), 96–104.

[3] Alessandro Bessi and Emilio Ferrara. 2016. Social bots distort the 2016 US Presidential election online discussion. *First Monday* 21, 11-7 (2016).

[4] R. Bhatia, S. Srivastava, V. Bhatia, and M. Singh. 2018. Analysis of Audio Features for Music Representation. In *2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*. 261–266. https://doi.org/10.1109/ICRITO.2018.8748783

[5] D. Blei, A. Ng, and M. Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.

[6] David A Broniatowski, Amelia M Jamison, SiHua Qi, Lulwah AlKulaib, Tao Chen, Adrian Benton, Sandra C Quinn, and Mark Dredze. 2018. Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. *American journal of public health* 108, 10 (2018), 1378–1384.

[7] Victor S Bursztyn and Larry Birnbaum. 2019. Thousands of Small, Constant Rallies: A Large-Scale Analysis of Partisan WhatsApp Groups. (2019).

[8] Josemar Alves Caetano, Gabriel Magno, Marcos André Gonçalves, Jussara M. Almeida, Humberto Torres Marques-Neto, and Virgílio A. F. Almeida. 2019. Characterizing Attention Cascades in WhatsApp Groups. *CoRR* abs/1905.00825 (2019). arXiv:1905.00825 http://arxiv.org/abs/1905.00825

[9] J. Carletta. 1996. Assessing agreement on classification tasks: the kappa statistic. *Computational linguistics* 22, 2 (1996), 249–254.

[10] Philipe de Freitas Melo, Carolina Coimbra Vieira, Kiran Garimella, Pedro O. S. Vaz de Melo, and Fabrício Benevenuto. 2019. Can WhatsApp Counter Misinformation by Limiting Message Forwarding? arXiv:cs.CY/1909.08740

[11] D. Doukhan, J. Carrive, F. Vallet, A. Larcher, and S. Meignier. 2018. An Open-Source Speaker Gender Detection Framework for Monitoring Gender Equality. In *IEEE ICASSP*.

[12] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2016. The Rise of Social Bots. *Commun. ACM* 59, 7 (June 2016), 96–104. https://doi.org/10.1145/2818717

[13] Adam Fourney, Miklos Z Racz, Gireeja Ranade, Markus Mobius, and Eric Horvitz. 2017. Geographic and temporal trends in fake news consumption during the 2016 US Presidential Election. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, 2071–2074.

[14] M. Iqbal. 2019. WhatsApp Revenue and Usage Statistics. https://www.businessofapps.com/data/whatsapp-statistics/. Accessed in 2019-09-01.

[15] D. Jang, C. D. Yoo, S. Lee, S. Kim, and T. Kalker. 2009. Pairwise Boosted Audio Fingerprint. *IEEE Transactions on Information Forensics and Security* 4, 4 (Dec 2009), 995–1004.

[16] Philipe Melo, Johnnatan Messias, Gustavo Resende, Kiran Garimella, Jussara Almeida, and Fabrício Benevenuto. 2019. WhatsApp Monitor: A Fact-Checking System for WhatsApp. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 13. 676–677.

[17] F. Oliveira. 2019. Facebook chega a 127 milhões de usuários mensais no Brasil. https://www1.folha.uol.com.br/tec/2018/07/facebook-chega-a-127-milhoes-de-usuarios-mensais-no-brasil.shtml. Accessed in 2019-09-01.

[18] J. Pennebaker, R. Boyd, K. Jordan, and K. Blackburn. 2015. *The development and psychometric properties of LIWC2015*. Technical Report.

[19] A. Porter. 2013. *Evaluating musical fingerprinting systems*. Ph.D. Dissertation. McGill University Libraries.

[20] G. Resende, P. Melo, J. Reis, M. Vasconcelos, J. Almeida, and F. Benevenuto. 2019. Analyzing Textual (Mis)Information Shared in WhatsApp Groups. In *WebSci '19*.

[21] G. Resende, P. Melo, H. Sousa, J. Messias, M. Vasconcelos, J. Almeida, and F. Benevenuto. 2019. (Mis)Information Dissemination in WhatsApp: Gathering, Analyzing and Countermeasures. In *WWW'19*.

[22] Michael Seufert, Tobias Hoßfeld, Anika Schwind, Valentin Burger, and Phuoc Tran-Gia. 2016. Group-based communication in WhatsApp. In *2016 IFIP networking conference (IFIP networking) and workshops*. IEEE, 536–541.