WhatsApp Monitor: A Fact-Checking System for WhatsApp

Philipe Melo,[1] **Johnnatan Messias,**[2] **Gustavo Resende,**[1]
Kiran Garimella,[3] **Jussara Almeida,**[1] **Farbrício Benevenuto**[1]
[1]Universidade Federal de Minas Gerais, [2]MPI-SWS, [3]MIT,
philipe@dcc.ufmg.br, johnme@mpi-sws.org, gustavo.jota@dcc.ufmg.br
garimell@mit.edu, jussara@dcc.ufmg.br, fabricio@dcc.ufmg.br

Abstract

WhatsApp is the most popular communication application in many developing countries such as Brazil, India, and Mexico, where many people use it as an interface to the web. Due to its encrypted and peer-to-peer nature feature, it is hard for researchers to study which content people share through WhatsApp at scale. In this demo paper, we propose *WhatsApp Monitor* (`http://www.whatsapp-monitor.dcc.ufmg.br/`), a web-based system that helps researchers and journalists explore the nature of content shared on WhatsApp public groups from two different contexts: Brazil and India. Our tool monitors multiple content categories such as images, videos, audio, and textual messages posted on a set of WhatsApp groups and displays the most shared content per day. Our tool has been used for monitoring content during the 2018 Brazilian general election and was one of the major sources for estimating the spread of misinformation and helping fact-checking efforts.

1 Introduction

The ever-increasing availability of cheap smartphones has generated a massive surge of new users having Internet access for the first time. This trend is especially higher in developing countries, where hundreds of millions of people have started using the internet recently, primarily through mobile devices. WhatsApp benefited from this trend and it has been heavily used in developing countries since. For instance, in Brazil, around 48% of the population uses WhatsApp (Resende et al. 2019) while in India it is even higher, 94% of all Android devices have the app installed (Garimella and Tyson 2018). For most of these users, WhatsApp is the main source of information, which leans towards multimedia based messages to news consumption in local languages.

Various established actors, including political parties, have taken notice of this wave of new users and started to engage and reach a new potential base through WhatsApp. For example, in a 2017 survey, around one-sixth of WhatsApp users in India said that they were members of a group created by a political leader or by the political party.[1]

WhatsApp is also being massively used for election campaigning.[2] For instance, in 2018, Brazil has gone through its general election process to elect a new president, deputies, and governors. (Resende et al. 2019) study the use of WhatsApp during the election campaign and show that WhatsApp as one of the sources of misinformation, and its potential to spread misleading information. WhatsApp was also a protagonist in some major events around the world, such as in the Brazilian truck drivers' strike[3], in the general elections in Brazil[4], and in the spread of fake news with rumors about child kidnappings in India[5] and in Mexico.[6]

WhatsApp is a closed ecosystem due to its encrypted and peer-to-peer nature, and hence it is difficult for researchers, journalists, and law enforcement to study how information spreads in the platform. In this demonstration, we take first steps towards understanding WhatsApp conversations by developing the *WhatsApp Monitor* (`www.whatsapp-monitor.dcc.ufmg.br/`). Our system could be useful to help researchers and journalists to explore the nature of content shared on WhatsApp public groups at scale. WhatsApp groups are limited in size to 256 people and they can be private, requiring an administrator to add new users, or public, being freely accessible to anyone with a invitation link. We focus only on publicly accessible groups in two use cases: (i) the usage of WhatsApp during the 2018 Brazilian general election, where our tool was a primary source for fact-checking efforts on WhatsApp; and (ii) India, in which we also deployed the system for a large number of groups and where there will also be elections in 2019.

The purpose of the *WhatsApp Monitor* is to inform and anticipate communicators about the type of information shared on public groups. The idea is to share to a restrict set of researches and journalists or fact-checking agencies the access to the system to perform the fact-checking on the information shared on those groups.

[1]`www.livemint.com/Technology/O6DLmIibCCV5luEG9XuJWL/How-widespread-is-WhatsApps-usage-in-India.html`

[2]`http://nyti.ms/2L3AV3M`

[3]`https://bbc.in/2FqPeQF`

[4]`https://brasil.elpais.com/especiais/2018/eleicoes-brasil/conversacoes-whatsapp/`

[5]`https://bbc.in/2M9XKDT`

[6]`https://bbc.in/2QMSEyt`

2 Data Collection

Our system uses data collected from picked WhatsApp public groups, discussing political topics. These groups are operated by individuals affiliated with political parties, or local community leaders. Given a set of groups, we use a strategy developed by (Garimella and Tyson 2018) to automate the joining of these groups. After that, we daily download all data shared within each group and persist them in a database ranked by the total number of shares for each item. From each message, we extract its data attributes such as the group name where the message was posted, group ID, user ID, and timestamp. For the multimedia messages (i.e., audios and videos), we also download their respective files.

To identify duplicated images and count their popularity, we use the Perceptual Hashing (pHash) algorithm [7] to calculate a fingerprint for every image as did in (Resende et al. 2019). This allows us to group a set of the same content as they have the same hash value. For audio and video, we compare their checksum codes, since people tend to only forward those instead of changing their content. Thus, the checksum for duplicates remain the same. We also analyze similarity for identical URLs and use the Jaccard index to compare text messages.

2.1 Privacy Concerns

Our system gathers a considerable amount of data from many WhatsApp groups. To ensure the privacy of users, we do not share or disclose any Personally Identifiable Information (PII) such as cellphone numbers. We use them only to measure aggregate statistics (e.g. posts per user and number of unique users per group).

To avoid any misuse of even aggregate information, we limit the access of our system to a restricted number of journalists and researchers, through a login account. Moreover, they are also informed about the data limitations and the potential bias present on our system. Since we only specifically use publicly available WhatsApp groups, our data collection does not violate WhatsApp terms of service.

3 Description of the Demo

We provide an online system where users can oversee the daily trends shared on the WhatsApp public groups for a particular country (e.g. Brazil and India) or domain (e.g. politics and news). Our system ingests various types of available media (text, URL, audio, image, and video) from the groups being daily monitored. Then, we rank the most popular pieces of media content related to the political and news topics.

Currently, our system replicates two distinct instances: a Brazilian and an Indian version. Once an instance is chosen, and the user logs in, the user is taken to a dashboard where she can navigate between dates and observe the most shared multimedia content in our monitored groups for a given date. Five types of content were identified and persisted: images, videos, audio messages, external links, and text messages (longer than 200 characters). Furthermore, our system daily displays the content divided by media type and shows them

---

[7]https://www.phash.org/

ranked by number of shares. This allows journalists to daily get an idea about critical content shared in WhatsApp that may be worth being fact-checked. Figure 1 depicts a screenshot of our system once a user logs in.
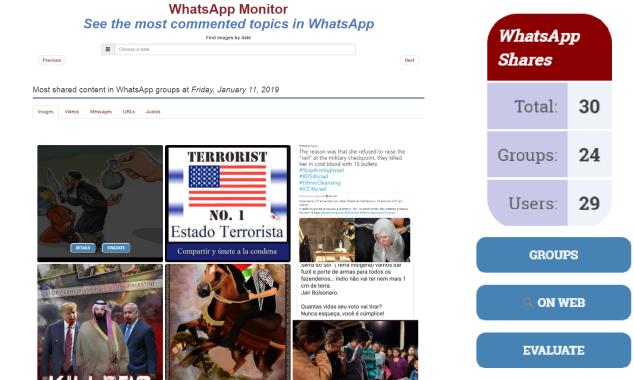

Figure 1: A screenshot of the WhatsApp Monitor interface showing main page and details of an image.

Finally, we provide tools to enable fact-checkers to verify information quickly for each media content. For instance, by clicking on an image, one can see the search results on Google for that image, allowing the track of external sources of the content and giving clues on how it is shared across different web platforms. In addition, by clicking on "Details", it shows the number of shares and names of the groups in which the media appears, to help them identify some context associated with the content.

4 Real World Impact

During the recent Brazilian elections, we gave access to the system to more than a hundred journalists and to three fact-checking agencies who explicitly mentioned our system as a data source. More importantly, dozens of pieces of news have referred to our system or used our its data during the Brazilian elections[8], suggesting it was useful to better understand the political campaigns and discussions within WhatsApp. We hope our system can be useful in other countries in which WhatsApp is popular and help fact-checking agencies to fight misinformation.

References

[Garimella and Tyson 2018] Garimella, K., and Tyson, G. 2018. Whatapp Doc? A First Look at Whatsapp Public Group Data. In *Proc. of Int'l Conf. on Web and Social Media (ICWSM)*.

[Resende et al. 2019] 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 *Proc. of the The Web Conference (WWW'19)*.

---

[8]https://bbc.in/2QFdhMR