BATMAN: A Big Data Platform for Misinformation Monitoring

Ivandro Claudino, Luciano Galic, Wellington Franco, Thiago Gadelha, José Maria Monteiro
and Javam Machado
Universidade Federal do Ceará, Brazil

Keywords: Big Data, Misinformation Monitoring, Platform, Social Networks.

Abstract: The large-scale dissemination of misinformation through social media has become a critical issue, harming social stability, democracy, and public health. The WhatsApp instant messaging application is very popular in Brazil, with more than 165 million users. On the other hand, in just one year, the proportion of smartphones with Telegram installed grew in Brazil from 45% to 60% in 2022. If on one hand, these platforms offer security and privacy to its users, on other hand they are spaces with little or no moderation. Consequently,they have been used to spread misinformation. In this context, we present BATMAN, a Big Data Platform for Misinformation Monitoring, a real-time platform for finding, gathering, analyzing, and visualizing misinformation in social networks, in particular, in instant message applications such as WhatsApp and Telegram. To evaluate the proposed platform, we used it to build two different messages datasets, concerning the Brazilian general elections campaign in 2022, obtained from public chat groups on WhatsApp and Telegram, respectively.

1 INTRODUCTION

Lately, the popularity of instant messaging applications, such as WhatsApp and Telegram, has contributed to the spread of misinformation. Through these systems, misinformation can deceive thousands of people in seconds and cause significant harm to individuals or society. Such platforms allow content to be spread without editorial judgment. In this context, misinformation has been used to change political scenarios, to spread ineffective treatments, and even to cause deaths (Martins et al., 2022; Martins et al., 2021a; Silva and Benevenuto, 2021).

The WhatsApp instant messaging application is very popular in Brazil, with more than 165 million users in about 214 million people. In Brazil, more than 95% of smatphone users use WhatsApp daily and 48% of the population use WhatsApp to get, share and discuss news (de Sá et al., 2021; Newman et al., 2021). On the other hand, in just one year, the proportion of smartphones with Telegram installed grew in Brazil from 45% to 60% in 2022. The popularity of these platforms is due to the versatility and ease of use. They make it possible to instantly share different media types, such as images, audios, and videos. Besides, they provide a significant feature: the public chat groups. These public groups are accessible through invitation links and, usually, they have specific topics for discussion, such as politics and education. Both WhatsApp and Telegram, allow users to join or even share their public groups to simultaneously connect to hundreds of people at once, and quickly receive and share digital content.

In this context, monitoring the content that circulates in public chat groups is a fundamental task to understand the misinformation spreading and get insights to address this problem. However, collecting a database of messages already in circulation in chat public groups is a challenging task. To fill this gap, we built the BATMAN, a Big Data Platform for Misinformation Monitoring, which supports finding, gathering, analyzing, and visualizing misinformation in different social networks, in particular, in instant message applications such as WhatsApp and Telegram. To evaluate the proposed platform, we used it to build two different datasets, concerning the Brazilian general elections campaign in 2022, obtained from public chat groups on WhatsApp and Telegram, respectively.

The remainder of this paper is organized as follows. Section 2 presents the main related work. Section 3 describes the BATMAN platform. Section 4 details a case study performed to evaluate the proposed platform. Conclusions and future work are presented in Section 5.

Claudino, I., Galic, L., Franco, W., Gadelha, T., Monteiro, J. and Machado, J.
BATMAN: A Big Data Platform for Misinformation Monitoring.
DOI: 10.5220/0011995500003467
In *Proceedings of the 25th International Conference on Enterprise Information Systems (ICEIS 2023) - Volume 1*, pages 237-246
ISBN: 978-989-758-648-4; ISSN: 2184-4992

2 RELATED WORK

Despite the scientific community's efforts, there is still a need for monitoring and identifying misinformation in WhatsApp and Telegram messages, mainly in Portuguese. The paper presented in (Garimella and Tyson, 2018) is a seminal work in collecting and analyzing WhatsApp' messages. The authors built a dataset by crawling 178 public groups, containing 45K users and 454K messages, from different countries and languages, such as India, Pakistan, Russia, Brazil, and Colombia. In the study presented in (Machado et al., 2019), the authors collected and analyzed 298,892 WhatsApp' messages, from 130 public groups, in the period leading up to the two rounds of the 2018 Brazilian presidential elections. In (Resende et al., 2019), the authors analyzed different aspects of WhatsApp messages from public political-oriented groups. The messages were collected during major social events in Brazil: a national truck drivers' strike and the Brazilian presidential campaign. The authors analyzed the types of content shared within such groups and the network structures that emerge from user interactions.

In (Resende et al., 2018), the authors presented a system for gathering, analyzing, and visualize public groups in WhatsApp. Besides describing their methodology, the authors also provide a brief characterization of the 169.154 messages shared by 6,314 users in 127 public groups to help journalists and researchers understand the repercussion of events related to the 2018 Brazilian elections. In (de Sá et al., 2021), the authors presented the Digital Lighthouse, an entire platform for finding, gathering, analyzing, and visualize public groups in WhatsApp. In (Cabral et al., 2021) the author built a large-scale, labeled, anonymized, and public dataset formed by WhatsApp messages in Portuguese (PT-BR), concerning the Brazilian general elections campaign in 2018, collected from public chat groups, using the platform proposed by (de Sá et al., 2021). Then, the authors conducted a series of classification experiments using combinations of Bag-Of-Words features and classical machine learning methods, resulting in a total of 108 experiments, in order to build a specif MID for WhatsApp messages. Their best results achieved a F1-score of 0.733, which served as a baseline for other work. As a practical result of this work, the authors built and deployed a Misinformation Detector, which receives a text as input and returns as output the probability that the text contains some misinformation [1].

[1]Currently, the Misinformation Detector can be accessed through the link https://faroldigital.info/classifier/misinformation-text

In (Martins et al., 2021a), the authors presented a large-scale, labeled, and public data set of WhatsApp messages in Brazilian Portuguese about coronavirus pandemic, called COVID-19.BR, which was collected from public chat groups, using the platform proposed by (de Sá et al., 2021). In that work, they conduct a series of classification experiments using nine different machine learning methods to build an efficient misinformation classifier for WhatsApp messages. The best result reached by (Martins et al., 2021a) had an F1 score of 0.778, considering the full corpus of COVID-19.BR dataset. In (Martins et al., 2021b) the authors detailed the dataset proposed in (Martins et al., 2021a) and presented a case study exploring data visualization concepts to represent information graphically, highlighting patterns and trends in data and achieving new insights concerning COVID-19 misinformation on WhatsApp. In (Martins et al., 2021c), the authors proposed a new approach to misinformation detection, called MIDeepBR, based on BiLSTM neural networks, BERT Embeddings, pooling operations and attention mechanisms. MIDeepBR can automatically detect misinformation in PT-BR WhatsApp messages. Their best results achieved an F1 score of 0.834. In (Martins et al., 2022), the authors explored a posthoc interpretability method called LIME to explain the predictions of misinformation detection approaches. Besides, they applied a textual analysis tool called LIWC to analyze WhatsApp messages' linguistic characteristics and identify psychological aspects present in misinformation and non-misinformation messages. The results indicated that it is feasible to understand relevant aspects of the MID model's predictions and find patterns on WhatsApp messages about COVID19.

In (Ng and Loke, 2021), the authors analyzed a Singapore-based COVID-19 Telegram group with more than 10000 participants focusing on five dimensions: participation, sentiment, negative emotions, topics, and message types. In (Júnior et al., 2022a; Júnior et al., 2022b), the authors presented the "Telegram Monitor", a web-based system that monitors the political debate in this platform and enables the analysis of the most shared content in multiple channels and public groups. In (Paschalides et al., 2020), the authors presented MANDOLA, a big-data processing system that monitors, detects, visualizes, and reports the spread and penetration of online hate-related speech using big-data approaches. MANDOLA consists of six individual components that intercommunicate to consume, process, store, and visualize statistical information regarding hate speech spread online.

3 THE BATMAN PLATFORM

This section will present the main components of the BATMAN (Big dATa platforM for misinformAtion moNitoring) platform, a real-time platform for finding, gathering, analyzing, and visualizing misinformation in social networks, in particular, in instant message applications such as WhatsApp and Telegram, based on Big Data technologies.

The proposed platform architecture comprises six layers, as illustrated in Figure 1. Next, we will discuss in detail each one of these layers and its components.

3.1 Data Collector Layer

The main goal of this layer is make possible gathering data from different social networks using a common interface. In this layer, the components are called "connectors", which are applications that collect data from a particular social network. The connectors run in Docker containers and they are independent of each other. So, a failure in a certain connector does not affect the others. They collect messages (text) and media (image, audio and video) shared in a specific social network. A connector converts each captured message to the JSON (JavaScript Object Notation) format and send it to the Message Broker (Redis) on the process layer. We chose JSON because it is a language-independent, standard format for storing and exchanging data. Besides, a connector sends each captured media file to the File Server component on the Persistence Layer. In order to avoid storing duplicate media files, we apply the MD5 hash algorithm on the file content and generate a a unique identifier, which is used as the name of the media file. Thus, we avoid wasting disk space, as well as making it possible to aggregate similar content and quantify how many times each one was shared by the users, with the purpose of understand the popularity of each content. Currently, the BATMAN platform has two different connectors running: one for WhatsApp and other to Telegram. Listings 1 and 2 illustrate examples of JSON "files" caught by WhatsApp and Telegram connectors, respectively.

3.2 Process Layer

The main purpose of this layer is ensure a common interface for receiving the messages collect by connectors. This layer has two components: Message Broker (Redis) and ETL Application.

The Message Broker is a software that make it possible that applications, systems and services communicate with each other and share information. It is responsible for validating, storing, routing and delivering messages to the appropriate destinations. It acts as an intermediary between other applications, allowing senders (connectors, in this case) to issue messages without knowing where the receivers (ETL Application, in this case) are, whether they are active or not, or how many of them there are. This facilitates the decoupling of processes and services within the proposed architecture. The Message Broker allows reliable storage and ensures message delivery. It has a set of message queues, which store and sort messages until consuming applications can process them. Furthermore, it ensures that each queued message is consumed only once. To implement the Message Broker, we use Redis, which is an in-memory, key-value, open source, versatile and easy-to-use storage system. In addition, it provides high performance, persistence and data replication.

The ETL Application is responsible for the message processing, which includes different tasks, such as: parsing, anonymization, user geolocation discovery (only for WhatsApp messages), misinformation detection and sentiment analysis. Many of these tasks use the services of the Data Processing API on the Data Processing Layer. We took into consideration privacy issues by anonymizing users' names and cell phone numbers. For this, we create an anonymous and unique ID for each user by using an MD5 hash function on their phone number. Similarly, we create an anonymous alias for each group. After a message processing, the ETL Application component sends the resulting data to a Relational Database Server (PostgreSQL) and a Search Engine (Elasticsearch), both on the Persistence Layer.

3.3 Data Processing Layer

The main goal of this layer is to provide a set of services to support the data processing, through a standardized API (Application Programming Interface).This API integrates several independent components, which will be detailed next. The Machine Learning Models component includes a service to compute the probability of a text message received as input to contain misinformation. The Geographic Component provides a service to discover the geographic location (DDD and DDI) of a WhatsApp user. The Natural Language Processing Component has a service to compute the sentiment (polarity) of a text message received as input. The Image Processing component is under development and will provides services for extract text from image files collected from the connectors, for find similar images and for identify objects in an image.

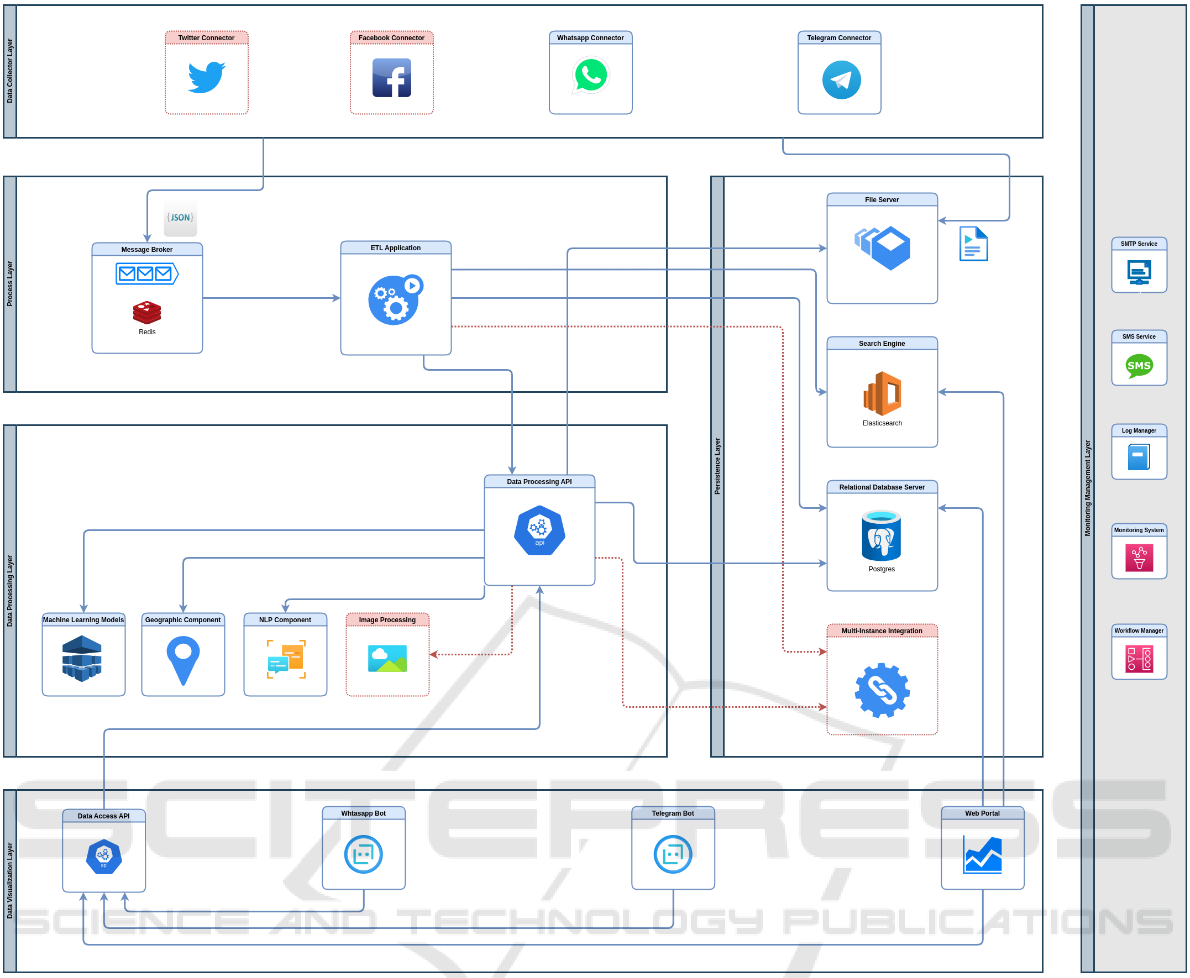

Figure 1: The BATMAN Platform Architecture.

3.4 Persistence Layer

The main purpose of this layer is to provide support for storing and querying data. This layer has four components: File Server, Search Engine, Relational Database Server and Multi-instance Integration. Next, we will describe each one of them.

The File Server component is responsible for store the media files (audios, images and videos) captured by the connectors in a persistent and safe manner. The Search Engine component aims to provide textual queries directly on the captured messages. For this, it uses Elasticsearch, a search engine based on the Lucene library that provides a distributed, multitenant-capable full-text search engine with an HTTP web interface and schema-free JSON documents. The Relational Database Server supports storing and querying data on the traditional flat model. Thereunto, it uses PostgreSQL, a free and open-source relational database management system (RDBMS) emphasizing extensibility and SQL compliance. Figure 2 illustrates the PostgreSQL database schema, that is, its tables and columns. It is important to highlight that the audios, images, and videos are stored by the File Server. The PostgreSQL database stores only the path to these files. The Multi-instance Integration component is under development and will provides support to integrate and communicate several distributed instances of the BATMAN platform.

3.5 Data Visualization Layer

The main goal of this layer is to support visualization and consuming of the data previously stored and processed. This layer includes four components: Data Access API, Web Portal, WhatsApp Bot and Telegram Bot. Next, we will describe each one of these components.

The Data Access API has a set of services to access: i) processed data stored on the relational model

Listing 1: An Example of a JSON Caught from WhatsApp.


```
{
  "id_message": "EE6CF9B6E75AE22708BDDE4B17548D6D",
  "messenger": "whatsapp",
  "message_type": "DocumentoComLegenda",
  "id_persona": "XXXXXXXXXXXX.0:12@s.whatsapp.net",
  "date_message": "2022-09-17 01:41:05 +0000 UTC",
  "text_content": "",
  "id_member": "XXXXXXXXXXXXX@s.whatsapp.net",
  "id_group": "YYXXZZZZ09789-1513745318@g.us",
  "media": "833f14a3bcbbb7f5c4d356cfe1ab19fa.pdf",
  "media_name": "Paginas do Lulaflix,site com as falcatruas
  de Lulla que a Justica mandou tirar do ar-Consegui salvar todas as
      paginas.pdf",
  "media_type": "application/pdf",
  "media_url": "",
  "media_md5": "833f14a3bcbbb7f5c4d356cfe1ab19fa",
  "display_name": "",
  "address_message": "",
  "latitude_message": 0,
  "longitude_message": 0,
  "contacts_message": null
}
```


(PostgreSQL), ii) processed data stored on text format (Elasticsearch) and iii) media files previously collected by the connectors. Thus, it uses the Data Processing API on the Data Processing Layer. Today, there is a great need for displaying massive amounts of data in a way that is easily accessible and understandable. In this context, data visualization is a way to represent information graphically, highlighting patterns and trends in data and helping to achieve news insights. It enables the data exploration via the manipulation of charts and images. More specifically, it enables users to analyze the data by interacting directly with a visual representation of it. In this work, the Web Portal component is a web application developed using Python programming language and Django 3 framework, which explores relational (from PostgreSQL) and textual (from Elasticsearch) data. The last two components of the Data Visualization Layer, WhatsApp Bot and Telegram Bot, are proactive chatbots, which automatically detects and alerts the presence of misinformation in social chats. Initially, they need to be added to a certain group. Then it will automatically monitor and analyze the content that travels in the group. Finally, if they detect that a certain content has a high probability of containing misinformation, an alert message is sent to the group.

3.6 Monitoring Management Layer

The main purpose of this layer is monitoring the operation of the BATMAN platform as a whole and alerts a human administrator by email and SMS (short message system) in case of failures. Besides, this layer maintains a set of logs, which can be used to audit, troubleshooting and repairs.

4 CASE STUDY

To evaluate the BATMAN platform, we performed an exploratory case study using two different datasets, covering the Brazilian general elections campaign in 2022, collected by WhatsApp and Telegram, respectively. Next, we will describe these two datasets in detail.

- Brazilian general elections on WhatsApp: This dataset contains 798,882 messages, obtained from 17,717 users (cell phone chips), which participated of 331 WhatsApp public groups, in the period from August to November 2022.
- Brazilian general elections on Telegram: This dataset contains 561,449 messages, obtained from 14,866 users, which participated of 180 Telegram public groups, in the period from September to November 2022.

Listing 2: An Example of a JSON Caught from Telegram.

```
{
  "id_message": "1049819",
  "messenger": "telegram",
  "message_type": "Url",
  "id_persona": #########,
  "date_message": "2022-09-28T16:18:09+00:00",
  "text_content": "Banco Mundial confirma fala de Paulo Guedes: PIB do
      Brasil tende a crescer mais que o da China \u261b https://
      terrabrasilnoticias.com/2022/09/banco-mundial-confirma-falaa-de-paulo
      -guedes-pib-do-brasil-tende-a-crescer-mais-que-o-da-china/\n\nBom dia
       \ud83d\udd25\ud83d\udd25\n\n@FYIBRASIL",
  "id_member_telegram": #########,
  "id_group": #########,
  "media": "ca9a6c25f29526930d60852a74ab6940 (1).jpg",
  "media_name": "",
  "media_type": "url",
  "media_url": "https://terrabrasilnoticias.com/2022/09/banco-mundial-
      confirma-falaa-de-paulo-guedes-pib-do-brasil-tende-a-crescer-mais-que
      -o-da-china/",
  "media_md5": "ca9a6c25f29526930d60852a74ab6940",
  "display_name": "Banco Mundial confirma fala de Paulo Guedes: PIB do
      Brasil tende a crescer mais que o da China - Terra Brasil Not\
      u00edcias\nCompartilhe: A nova previs\u00e3o do Banco Mundial \u00e9
      de que o PIB (Produto Interno Bruto)",
  "address_message": "",
  "latitude_message": 0,
  "longitude_message": 0,
  "contacts_message": null
}
```


Using the Web Portal component of the the Data Visualization layer from the BATMAN Platform, the user can choose a specific dataset or all data from all datasets to build a set of visualizations.

4.1 Messages Characterization

In general, messages created to spread misinformation include a URL, often from a little-known website or blog, to give it credibility. Thus, the presence of a URL can be a criterion for selecting messages to be analyzed by fact-checkers. We observed that a significant proportion of the caught Telegram (29.85%) and WhatsApp (20.16%) messages contains some URL.

Currently, audios, images, and videos are commonly used to spread misinformation. Therefore, the messages associated with these files are potential candidates to undergo a verification process. We observed that a significant proportion of the caught Telegram (63.32%) and WhatsApp (62.60%) messages contains some media file.

Figure 3 shows the distribution messages sending time by the day hours on Telegram. As we can imagine, the peak of sending messages occurs at the time reserved for lunch (between 12 and 15 hours) and in the early evening, just after work hours.

Figure 4 shows the distribution messages sending time by day on Telegram. As we can imagine, the peak of sending messages occurs on October 2nd (date of the first round of elections) and October 30th (date of the second round of elections).

4.2 Geographic Distribution

In the 2020 Brazilian elections, some cell phone chips from foreign countries were used in the electoral advertisement. Thus, monitor the messages sent by these chips is an important task to identify misinformation spreading. We observed that 1,83% of WhatsApp messages were sent by cell phone chips of foreign countries.

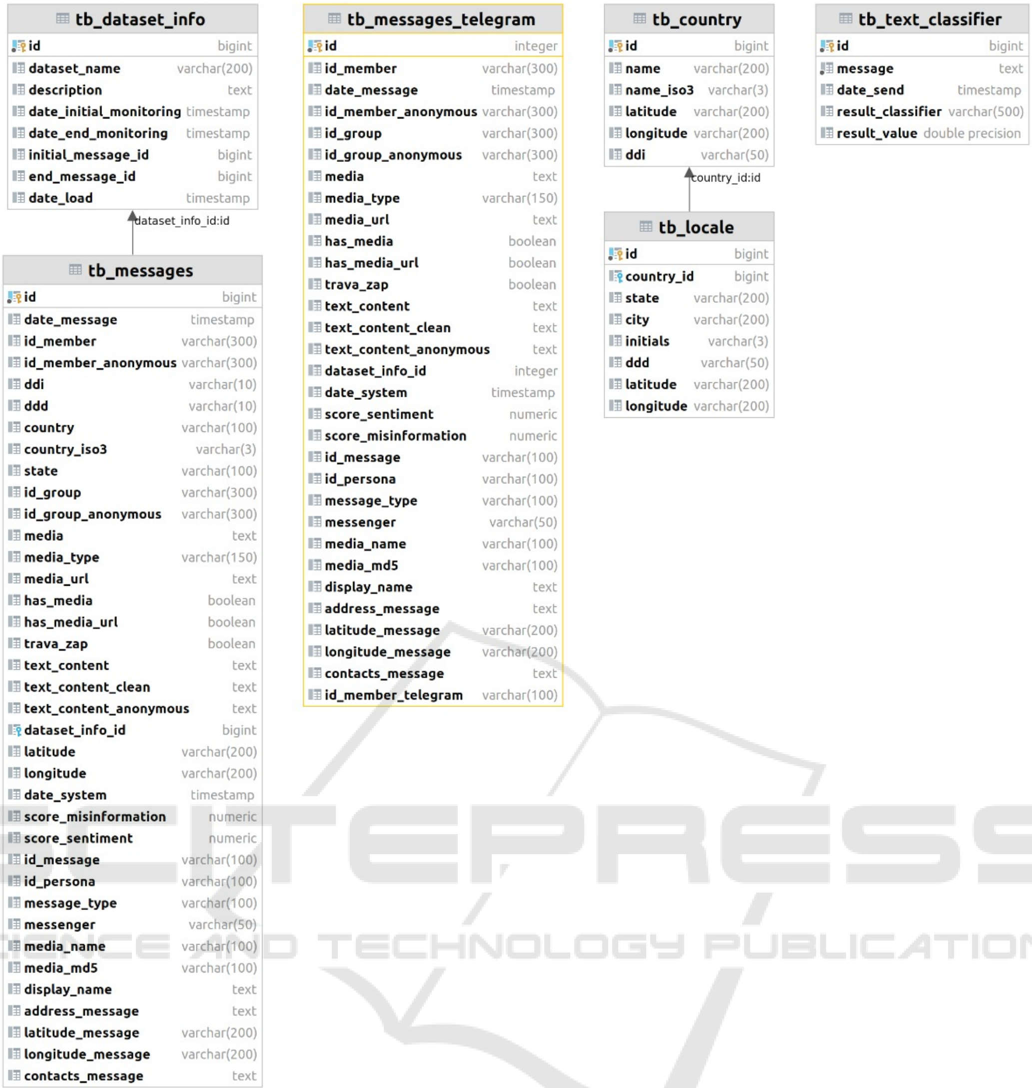

Figure 2: The PostgreSQL Database Schema.

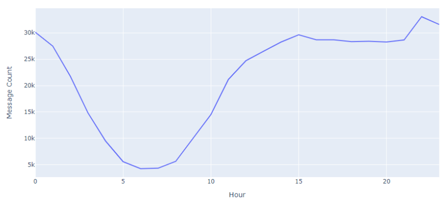

Figure 3: Number of Messages by Hour.

Another relevant aspect to observe in the monitored groups is the geographic location of users (cell phone chips), both Brazilians and foreigners, besides these users' activity level. Figure 5 shows the Brazilian states with more quantity of messages on WhatsApp. As might be expected, the most populous states have the most significant amount of messages sent.

Figure 6 illustrates the Brazilian states with more

Figure 4: Number of Messages by day.

users' on WhatsApp. The most populous states have the most significant amount of users.

However, when analyzing the states with more messages per user (Figure 7), we can observe that not so populous states such as Paraiba, Alagoas, and Distrito Federal, have the most active users.

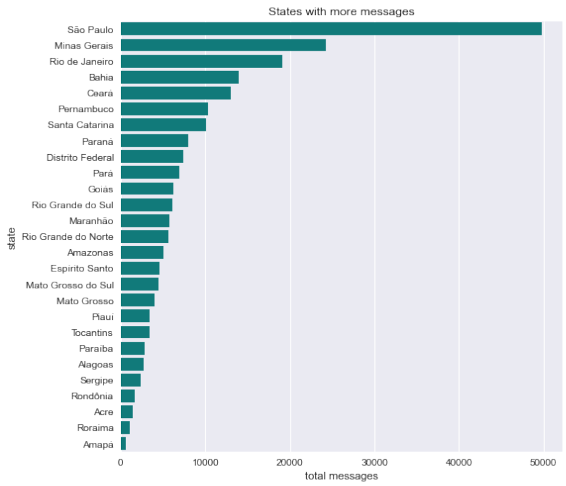

Figure 5: States with more Messages.

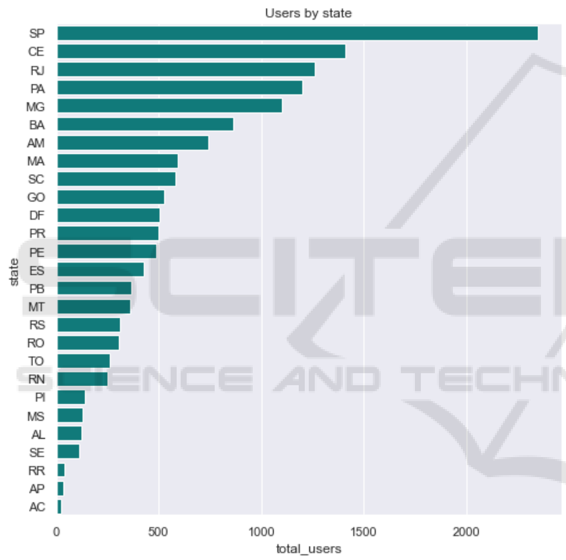

Figure 6: States with more Users.

4.3 Vocabulary Characterization

Another aspect that needs to be analyzed is the characteristics of the vocabulary used in the text messages, since there is a strong relationship between the used vocabulary and the social network, in this case, WhatsApp. Figures 8 and 9 show the number of messages by the number of words contained in the message, for the WhatsApp and Telegram datasets, respectively. As we can note, there are few messages with a large number of words and a high number of messages with few words.

Figures 10 and 11 show the word cloud highlighting the most popular words on WhatsApp and Telegram, respectively.

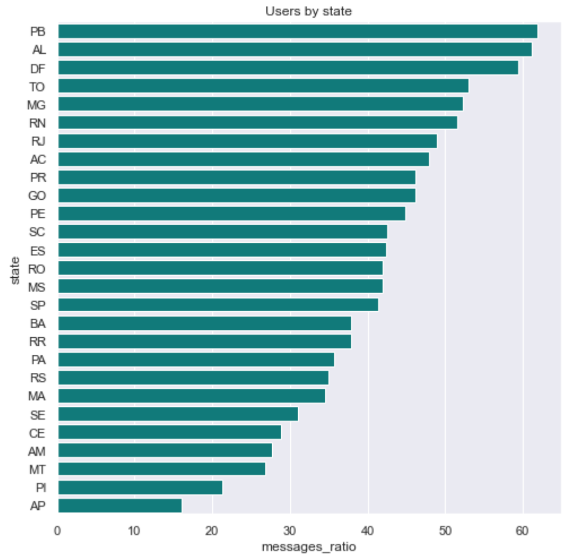

Figure 7: States with more Messages per User.

4.4 Misinformation Analysis

The last aspect to be explored using the Web Portal component on the Data Visualization layer of the BATMAN Platform is the misinformation analysis. In this context, various information about messages and users are explored to identify text messages containing misinformation and super-spreaders, that is, the users that most spread misinformation.

Table 1 contains the five most shared messages on WhatsApp. The “Sharings” column indicates how many times the message was shared. It is important to highlight that all the five most shared messages contain misinformation.

Finally, we can query the URLs most used in the messages. Tables 2 and 3 contain the five most shared URLs together with the number of messages that refers each URL, on WhatsApp and Telegram, respectively.

5 CONCLUSIONS

The fast spread of misinformation through social networks, such as WhatsApp and Telegram, messages poses a significant social problem. In this work, we presented BATMAN, a platform for finding, gathering, analyzing, and visualizing misinformation in social networks. To evaluate our methodology, we built two different datasets. Besides, we presented a case study using the proposed platform. We hope that our platform can help journalists and researchers to understand the misinformation propagation in Brazil.

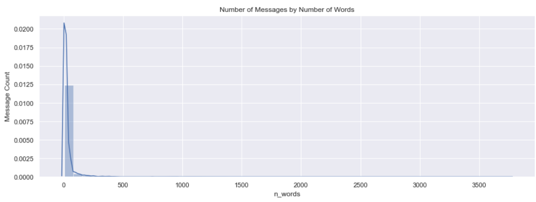

Figure 8: Number of Messages by the Number of Words in the Message on WhatsApp.

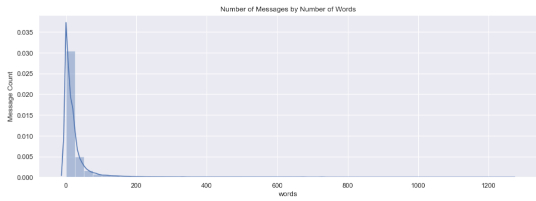

Figure 9: Number of Messages by the Number of Words in the Message on Telegram.

Figure 10: Word Cloud on WhatsApp.

Figure 11: Word Cloud on Telegram.

Table 1: Most Shared Messages on WhatsApp.

Sharings	Text
275	Pesquisa eleitoral para presidente da República em 2022. Vote https://financas-web.club/enquete/
190	Lula ou Bolsonaro? Quem está ganhando? Descubra a verdade e vote https://eleicoes2022.info/atualizada
173	Pesquisa eleitoral para presidente da República em 2022. Vote https://www.creditoparacartao.com/enquete/
79	AO VIVO: NOVA ENQUETE OFICIAL para Presidente da Republica 2022: Entre e Vote ==¿¿ https://enqueteeleicoes2022.com.br/
77	Atenção, Brasil! Chegou a hora de colocarmos fim aos abusos praticados pelo ministro Alexandre de Moraes e mostrarmos aos demais ministros do STF que o povo não aceitará mais qualquer excesso praticado por eles. Como qualquer outro membro do Estado, eles devem agir nos limites da lei e de sua competência. Para isso, vamos fazer circular essa petição pelo IMPEACHMENT de Alexandre de Moraes. O Brasil acordou e vamos mostrar que supremo é o POVO! Assine e compartilhe! Pra frente, Brasil! Vamos mostrar nossa força! https://peticaopublica.me/impeachment-alexandre-de-moraes-2021/

Table 2: Most Shared URLs on WhatsApp.

Qtd	Site
4212	https://h3r0.link/9XLudJ8VU12CYzpJ7
3583	https://api.whatsapp.com/sendphone=CCDDDXXXXXXX %20interesse%20
2473	https://t.me/+SLVkezIiNKkkH4sy
761	https://t.me/apostagem
622	https://www.youtube.com/channel/UCou3uZZFuvu5oB_E7BOXf6Q

Table 3: Most Shared URLs on Telegram

Qtd	Site
1221	https://t.me/canalselvabrasiloficial
1188	https://youtu.be/qbTzhB0akt8
1049	https://youtu.be/zDuOoyhyN-4
659	https://youtu.be/4DHk9KZ01HM
593	https://youtu.be/x2uiakywcrI

REFERENCES

Cabral, L., Monteiro, J. M., da Silva, J. W. F., Mattos, C. L. C., and Mourão, P. J. C. (2021). Fakewhastapp.br: NLP and machine learning techniques for misinformation detection in brazilian portuguese whatsapp messages. In Filipe, J., Smialek, M., Brodsky, A., and Hammoudi, S., editors, *Proceedings of the 23rd International Conference on Enterprise Information Systems, ICEIS 2021, Online Streaming, April 26-28, 2021, Volume 1*, pages 63–74. SCITEPRESS.

de Sá, I. C., Monteiro, J. M., da Silva, J. W. F., Medeiros, L. M., Mourão, P. J. C., and da Cunha, L. C. C. (2021). Digital lighthouse: A platform for monitoring public groups in whatsapp. In Filipe, J., Smialek, M., Brodsky, A., and Hammoudi, S., editors, *Proceedings of the 23rd International Conference on Enterprise Information Systems, ICEIS 2021, Online Streaming, April 26-28, 2021, Volume 1*, pages 297–304. SCITEPRESS.

Garimella, K. and Tyson, G. (2018). Whatsapp, doc? a first look at whatsapp public group data. *arXiv preprint arXiv:1804.01473*.

Júnior, M., Melo, P. F., Kansaon, D., Mafra, V., Sá, K., and Benevenuto, F. (2022a). Telegram monitor: Monitoring brazilian political groups and channels on telegram. *CoRR*, abs/2202.04737.

Júnior, M., Melo, P. F., Kansaon, D., Mafra, V., Sá, K., and Benevenuto, F. (2022b). Telegram monitor: Monitoring brazilian political groups and channels on telegram. In Bellogín, A., Boratto, L., and Cena, F., editors, *HT '22: 33rd ACM Conference on Hypertext and Social Media, Barcelona, Spain, 28 June 2022- 1 July 2022*, pages 228–231. ACM.

Machado, C., Kira, B., Narayanan, V., Kollanyi, B., and Howard, P. (2019). A study of misinformation in whatsapp groups with a focus on the brazilian presidential elections. WWW '19, page 1013–1019, New York, NY, USA. Association for Computing Machinery.

Martins, A. D. F., Cabral, L., Mourão, P. J. C., Monteiro, J. M., and Machado, J. (2021a). Detection of misinformation about covid-19 in brazilian portuguese whatsapp messages. In *International Conference on Applications of Natural Language to Information Systems*, pages 199–206. Springer.

Martins, A. D. F., da Cunha, L. C. C., Mourão, P. J. C., de Sá, I. C., Monteiro, J. M., and de Castro Machado, J. (2021b). Covid19.br: A dataset of misinformation about covid-19 in brazilian portuguese whatsapp messages. In *III Dataset Showcase Workshop, DSW 2021, Rio de Janeiro, RJ, Brazil, October 4-8, 2021 (To appear)*. SBC.

Martins, A. D. F., da Cunha, L. C. C., Mourão, P. J. C., Monteiro, J. M., and de Castro Machado, J. (2021c). Detection of misinformation about covid-19 in brazilian portuguese whatsapp messages using deep learning. In *XXXVI Simpósio Brasileiro de Banco de Dados, SBBD 2021, Rio de Janeiro, RJ, Brazil, October 4-8, 2021 (To appear)*. SBC.

Martins, A. D. F., Monteiro, J. M., and Machado, J. C. (2022). Understanding misinformation about COVID-19 in whatsapp messages. In Chiusano, S., Cerquitelli, T., Wrembel, R., Nørvåg, K., Catania, B., Vargas-Solar, G., and Zumpano, E., editors, *New Trends in Database and Information Systems - ADBIS 2022 Short Papers, Doctoral Consortium and Workshops: DOING, K-GALS, MADEISD, MegaData, SWODCH, Turin, Italy, September 5-8, 2022, Proceedings*, volume 1652 of *Communications in Computer and Information Science*, pages 14–23. Springer.

Newman, N., Fletcher, R., Schultz, A., Andi, S., and Nielsen, R. K. (2021). Reuters institute digital news report 2021. Dados sobre a utilização de redes sociais no Brasil.

Ng, L. H. X. and Loke, J. Y. (2021). Analyzing public opinion and misinformation in a covid-19 telegram group chat. *IEEE Internet Computing*, 25(2):84–91.

Paschalides, D., Stephanidis, D., Andreou, A., Orphanou, K., Pallis, G., Dikaiakos, M. D., and Markatos, E. (2020). Mandola: A big-data processing and visualization platform for monitoring and detecting online hate speech. *ACM Trans. Internet Technol.*, 20(2).

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.

Resende, G., Messias, J., Silva, M., Almeida, J., Vasconcelos, M., and Benevenuto, F. (2018). A system for monitoring public political groups in whatsapp. In *Proceedings of the 24th Brazilian Symposium on Multimedia and the Web*, WebMedia '18, page 387–390, New York, NY, USA. Association for Computing Machinery.

Silva, M. and Benevenuto, F. (2021). COVID-19 ads as political weapon. In Hung, C., Hong, J., Bechini, A., and Song, E., editors, *SAC '21: The 36th ACM/SIGAPP Symposium on Applied Computing, Virtual Event, Republic of Korea, March 22-26, 2021*, pages 1705–1710. ACM.