9^{th} International Conference on Computational Social Science IC^2S^2
July 17–20, 2023, Copenhagen, Denmark

Analyzing misinformation claims during the 2022 Brazilian general election on WhatsApp, Twitter, and Kwai

Keywords: Brazil 2022, elections, misinformation, claim matching, NLP

Extended Abstract

This study analyzes misinformation from WhatsApp, Twitter, and Kwai during the 2022 Brazilian general election. Given the democratic importance of accurate information during elections, five fact-checking organizations collaborated to identify and respond to misinformation via WhatsApp tiplines and power a fact-checking feature within a chatbot operated by Brazil's election authority, the TSE. WhatsApp is installed on over 99% of smartphones in Brazil, and the TSE chatbot was used by millions of citizens in the run-up to the elections. During the same period, we collected social media data from Twitter and Kwai, a popular video sharing app similar to TikTok. In this paper, we use claim matching and fact-checks from three Brazilian fact-checking organizations to examine the overlap in claims between platforms and across different WhatsApp misinformation tiplines. We perform a descriptive analysis examining the formats (image, video, audio, text) and content themes of popular misinformation claims.

Overall, we find the unique characteristics of each platform (formats, text length limits, etc.) influence the specific content circulating on the platforms. These differences also complicate any large-scale, quantitative comparison, and our paper suggests areas for further algorithmic development. While there is overlap across the platforms and the heavy-tailed patterns often observed in collective human behavior can be observed here, we also find clear differences between platforms. There is clear separation, for instance, between the videos we capture from WhatsApp and Kwai. The nature of each interface also affects its use: the WhatsApp tiplines operated by fact-checkers captured many long forwarded messages, whereas the bot operated by Brazil's election authority, the TSE, received many shorter, search-query-like messages.

Data & Methods. Prior quantitative research on misinformation or public opinion using WhatsApp data in Brazil usually relies on the collection of messages exchanged in large groups that have invite links publicly available on the Internet [2, 4]. In contrast to this, we use a crowdsourcing approach working with fact-checking organizations running misinformation tiplines. Tiplines are accounts on WhatsApp to which users can send possible misinformation content and questions and in exchange receive fact-checks and trusted information [1]. We analyze anonymized data from three fact-checking organizations operating such tiplines in Brazil during the elections. We also include anonymous content sent to these fact-checking organizations via the fact-checking feature on the chatbot operated by the TSE and heavily promoted by WhatsApp. All data is anonymous: phone numbers were replaced with random ids and names and other metadata beyond the timestamp were not included in the data made available to us for analysis.

We also collect data from Kwai via a third-party API finding approximately 600 thousand posts and from Twitter with elevated access to the streaming API collecting approximately 54 million tweets. We represented text content from WhatsApp, Kwai, and Twitter as dense vectors using a MPNet language model trained to produce semantic sentence embeddings.[1] This model produces similar embeddings (i.e., vectors) for content with similar meanings even if the content uses different words [3].

We used TMK video embeddings to compare the videos received on WhatsApp tiplines and

[1] `https://huggingface.co/sentence-transformers/paraphrase-multilingual-mpnet-base-v2`

the TSE bot directly to Kwai videos.[2] TMK is a C++ library, but we develop and release open-source Python bindings to make it easier to calculate TMK embeddings using Python.

Results. We find that there are examples of common misinformation claims across the three platforms, but also that the unique characteristics of each platform make a large-scale quantitative comparison difficult. WhatsApp messages submitted to the misinformation tiplines operated by fact-checkers are often an order of magnitude larger (up to 4,000 characters) than the maximum size allowed for tweets.

The most successful, large-scale comparison we can conduct is between the videos on Kwai and WhatsApp using TMK video fingerprinting. In that comparison, we find that there is little overlap between the videos from the two sources, as shown in Figure 1c.

Our comparative analysis of WhatsApp tiplines revealed that most of the users submit only one message, although a smaller number of 'power-users' submit many messages. This mirrors the heavy-tailed patterns found for content creation on other platforms. Of the users who submit multiple messages, most interact exclusively with one tipline. Given the end-to-end encryption of WhatsApp, it's impossible to know how representative any collection of messages is. Our analysis suggests that new claims are constantly circulating: each surge in use of the tiplines resulted in new content (new clusters over time shown in Figure 1a). Similarly, it appears that the tiplines are not reaching a saturation point: as the number of users interacting with a tipline increases so too does the amount of new, unseen content.

Overall, overlap with the fact-checking misinformation tiplines was low, but there is a weak, but positive and statistically significant correlation between the number of appearances in both the tiplines and the TSE bot (see Figure 1b).

Conclusions. On the methodological front, our research highlights the challenges of matching data across social media platforms with different format constraints such as between long WhatsApp messages, short tweets, and Twitter threads of unlimited length. Our results also show the promise of crowdsourcing approaches to identifying misinformation on end-to-end encrypted messaging platforms. Partnering with fact-checking organizations to better understand the content circulating on such platforms and to develop tools to assist fact-checkers to respond is an important area for further research. Finally, our results highlight the potential of Kwai as a data source for further research. Altogether, we hope that our research encourages further analysis into new social media platforms, particularly those used in the Global South, as this direction is likely to bring the methodological development needed for this research, along with novel insights about how public opinion is formed and mediated through online platforms.

References

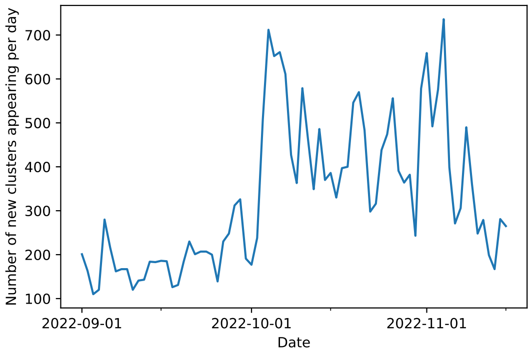
[1] Kazemi, A. et al. Tiplines to uncover misinformation on encrypted platforms: A case study of the 2019 Indian general election on WhatsApp. *Harvard Kennedy School (HKS) Misinformation Review*, 2022.

[2] Machado, C. et al. A Study of Misinformation in WhatsApp groups with a focus on the Brazilian Presidential Elections. In *Companion Proceedings of The 2019 World Wide Web Conference*, pages 1013–1019, San Francisco USA, May 2019. ACM.

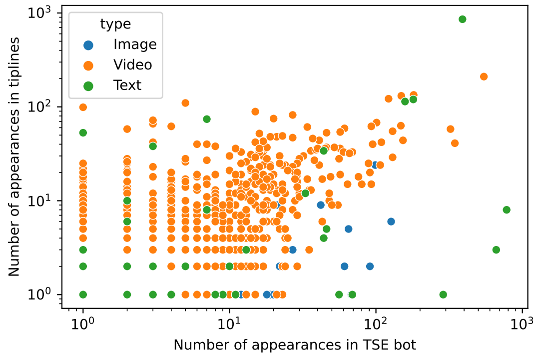
[3] N. Reimers and I. Gurevych. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *EMNLP-IJCNLP 2019*, pages 3982–3992, Hong Kong, China, November 2019. Association for Computational Linguistics.

[4] Resende, G. et al. (Mis)Information Dissemination in WhatsApp: Gathering, Analyzing and Countermeasures. In *The World Wide Web Conference*, WWW '19, pages 818–828, New York, NY, USA, May 2019. Association for Computing Machinery.

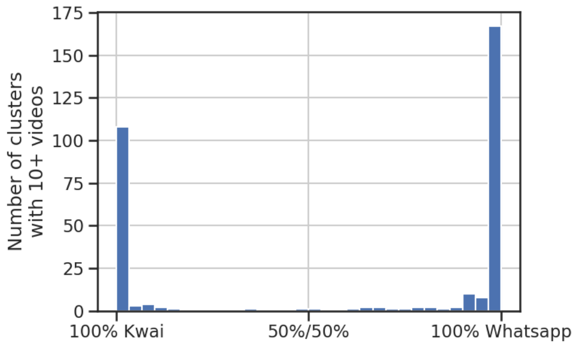
[2] `https://github.com/facebook/ThreatExchange/blob/main/hashing/hashing.pdf`



(a) New tipline clusters



(b) TSE vs. tiplines



(c) Kwai vs. WhatsApp

Figure 1: (a) Number of new clusters appearing per day in the fact-checkers' misinformation tiplines. (b) Overlap between content submitted to the tiplines and TSE is low, but there is a weak, positive correlation. (c) Histogram of the percentage of Kwai vs. WhatsApp videos per cluster, for the 320 clusters with 10 or more videos.