

Title: Methods appendix for “Brazilian Capitol attack: The interaction between Bolsonaro’s supporters’ content, WhatsApp, Twitter, and news media”

Authors: Joao V. S. Ozawa (1), Josephine Lukito (1), Felipe Bailez (2), Luis G. P. Fakhouri (2)

Date: April 9th, 2024

Note: The material contained herein is supplementary to the article named in the title and published in the Harvard Kennedy School (HKS) Misinformation Review.

Appendix B: Methods

This section explains in more detail how we conducted our data analysis. First, we will provide details about our topic modeling analysis. We decided to use 15 categories in our topic modeling analysis after a saturation process. We first ran the analysis with 10 categories and found several different topics. Then, we ran the analysis with 15 categories, and several of them were repeated. Therefore, we concluded that 15 categories were enough to represent the topics of the collected data.

We chose the names for each category according to a qualitative assessment. One of our researchers is a native Portuguese speaker and analyzed which themes were represented by each topic that the automated analysis identified. We will describe in detail the words that were decisive for the decision of each theme. We found five themes: *infiltrators*, *religion*, *armed forces*, *left-wing*, and *nationalism*. *Infiltrators* refers to the word “infiltrados” (in English, “infiltrated”), *religion* refers to the word “Deus” (“God”), *armed forces* refers to the words “exército” (“army”) and “quartéis” (“headquarters”). *Left-wing* refers to the words “esquerda” (“left”), “contra ladrão” (“against the thieves”). *Nationalism* refers to the word “brasileiros” (“Brazilians”), “país” (“country”), “patriotas” (“patriots”), “verde amarelo” (“green yellow,” which are the colors of the Brazilian flag), and links to Twitter profiles of politicians who are Bolsonaro’s allies. We recognize that this analysis has limitations as the decision of each theme depends on a limited number of words, and different interpretations of each theme could be possible. Future studies could benefit from our themes as a starting point and develop deeper content analysis through methods such as Critical Discourse Analysis.

We also recognize that our categories characterize diverse themes, such a broad social aspect as *religion*, a context-specific aspect such as *infiltrators*, and an institution such as the *armed forces*. However, those themes were automatically generated through a topic modeling analysis of words that are commonly mentioned together. Furthermore, our qualitative assessment found that these categories adequately reflect the narratives spread by Bolsonaro’s supporters on WhatsApp during the insurrection.

Samples

We collected the daily volume of messages with the keywords “#patriotas” and “patriotas” ($N = 16,238$) on WhatsApp between September 1, 2022, and January 31, 2023, and we collected all Twitter posts using the hashtag #patriotas, except for retweets ($N = 1,410,897$), during the same time period. Given the symbolic meaning of the term that we described in the main manuscript, we considered those messages to be content related to propaganda. However, we did not assess the proportion of false information within this content. Future research could benefit from investigating this content in more depth.

Palver provided data from exclusively public WhatsApp groups. Any person can join these groups, though the groups are limited to a maximum of 1,024 people. Once a person joins the group, the content is accessible. Due to privacy concerns, Palver does not provide information on the origin of each individual message that the platform monitors. The platform presents the overall volume of daily messages, and the user can see the content of each message, but the source remains anonymous. Relatedly, the platform does not inform which or how many groups spread specific messages, or how many people are in each of the groups. In other words, we do not know how much our sample represents within the overall number of groups that Palver monitors. This is a limitation of our study, though we believe that our sample can bring important insights given the scarcity of available data on WhatsApp.