

## **“Thou Shalt Not Take the Lord’s Name in Vain”: A Methodological Proposal to Identify Religious Hate Content on Digital Social Networks**

LUIZ ROGÉRIO LOPES-SILVA  
RODRIGO EDUARDO BOTELHO-FRANCISCO  
PAULO SERGIO DA CONCEIÇÃO MOREIRA  
ANDRÉ JOSÉ RIBEIRO GUIMARÃES  
Federal University of Paraná, Brazil

This study explores biblical terms in contexts of hate speech dissemination on digital social networks and proposes a method to filter hateful terms on Facebook. The objective is to identify and index words and expressions of religious intolerance and violence in Portuguese. Specifically, it aims to identify the biblical terms used in the context of the dissemination of hateful content on Facebook and, based on the results, build a block list that expands the analytical capacity of response, investigation, and timely intervention methods to abusive, fundamentalist, and extremist content. The methodological approach is based on text mining and content analysis procedures, combined with the application of Zipf’s law, with the adoption of Goffman’s transition point (T). As a result, we collected 4,214,699 comments from the official page of a neo-Pentecostal pastor who served as federal deputy between 2013 and 2016, recognized for intolerant and radical statements against nonpractitioners of his religion, non-Christians, agnostics, atheists, and secularists. In addition, the method determines 108 terms distributed in 3,614 comments. The content analysis verifies 25 terms used in the context of hate and systematizes them about the forms of speech (intolerant, fundamentalist, extremist, curse, or praise), as well as different recipients of the message found. The findings support the validation of the methodological proposal and provide the creation of a controlled vocabulary that organizes and monitors the hateful comments, the speech form, and the message recipient.

*Keywords: religious intolerance, hate speech, digital social networks, text mining, content analysis*

---

Luiz Rogério Lopes-Silva: [luiz.lobes@ufpr.br](mailto:luiz.lobes@ufpr.br)  
Rodrigo Eduardo Botelho-Francisco: [rodrigobotelho@ufpr.br](mailto:rodrigobotelho@ufpr.br)  
Paulo Sergio da Conceição Moreira: [paulomoreira@ufpr.br](mailto:paulomoreira@ufpr.br)  
André José Ribeiro Guimarães: [andrejose@ufpr.br](mailto:andrejose@ufpr.br)  
Date submitted: 2021-04-09

Copyright © 2022 (Luiz Rogério Lopes-Silva, Rodrigo Eduardo Botelho-Francisco, Paulo Sergio da Conceição Moreira, and André José Ribeiro Guimarães). Licensed under the Creative Commons Attribution Non-commercial No Derivatives (by-nc-nd). Available at <http://ijoc.org>.

Online religious practices and their implications in political and social contexts, including a disdainful and extremist discourse against members of other religions, LGBTQIA+<sup>1</sup> communities, women, and other subordinate groups, have been the objects of research in several knowledge areas. Part of such studies focuses on five major world religions—Buddhism, Christianity, Hinduism, Islamism, and Judaism—and aims to understand how technological mediation increases, restrains, or modifies the relationship among religious, nonreligious, and historically marginalized groups.

Some of the main *loci* where the manifestations of hatred occur on the Internet are digital platforms such as Facebook, Twitter, and Youtube (Blaya, 2019; Silva, Botelho-Francisco, Oliveira, & Ramos, 2019). In these spaces, the so-called echo chambers limit exposure to different perspectives and favor the construction of groups of users with similar ideas, framing, and reinforcing a shared narrative (Cinelli, Morales, Galeazzi, Quattrociochi, & Starnini, 2021). As a result, platforms play instrumental roles in spreading hatred and translating discourse into action (Cohen-Almagor, 2009; Schäfer, Leivas, & Santos, 2015).

Applied research has been undertaken to understand the radical behavior in digital platforms, mainly in the global south, in countries with solid religious adherence. Works such as those developed by Lee (2019), Samaratunge and Hattotuwa (2014), and Tsuria and Yadlin-Segal (2021) report the way the Internet may be a space for radicalization, especially from radical Muslims and Christians. These researchers also debate how digital platforms such as Facebook have increased the extremist discourse against minority groups in Myanmar and Sri Lanka.

In the Latin American context, religious hate discourse has been strategically used in authoritarian and conservative regimes to obtain power. For instance, in Brazil, a country of Christian majority (Datafolha, 2021), political and religious leaderships use the platforms to spread their doctrines, reach new audiences, and raise public agenda within the ideals of Christian morality, such as the criminalization of abortion and intervention and psychological assistance to homosexuals (Cunha, Lopes, & Lui, 2017; Lionço, 2017; Natividade & Nagamine, 2016).

In such works, we also observe that religious freedom has been used as an excuse to disseminate hate content, using words and expressions taken from the sacred Bible to justify intolerance and curse the recipients. This fact occurs without sufficient intervention in moderating this kind of content by the platforms.

Because of this problem, this article proposes a methodological approach to identify and index words and expressions of religious intolerance and violence in Portuguese on Facebook, using text mining, bibliometric laws, and content analysis (CA). Specifically, it aims to identify the biblical terms used in the context of the dissemination of hateful content on the platform and, based on the results, build a list of terms that expands the analytical capacity of the response, investigation, and intervention methods appropriate to abusive, fundamentalist, and extremist content.

---

<sup>1</sup> Acronym for Lesbian, Gay, Bisexual, Transgender, Transvestite, Transsexual, Queer, Intersex, Asexual, and Pansexual.

The study considers that the textual content that expresses intolerant behavior in a network is characterized by some type of word, radical and/or n-gram (Wang & Manning, 2012). It also uses Zipf's law (Araújo, 2006) to determine terms that would not be commonly identified and indexed in high-frequency dictionaries and block lists.

The construction of the methodological approach is based on the use of a database extracted from the official fan page of neo-Pentecostal pastor Marco Feliciano, accused of racist, homophobic, and religious intolerance statements. This pastor has also been a federal deputy since 2011 and has a strong presence on digital platforms, reaching more than 4 million followers on Facebook alone. He is a public figure of significant influence in the media and who, together with his colleagues from the Evangelical Front, promotes a "pro-life" and "pro-family" discourse, disregarding rights acquired by women, gays, Blacks, indigenous people, and adherents of African-based religions.

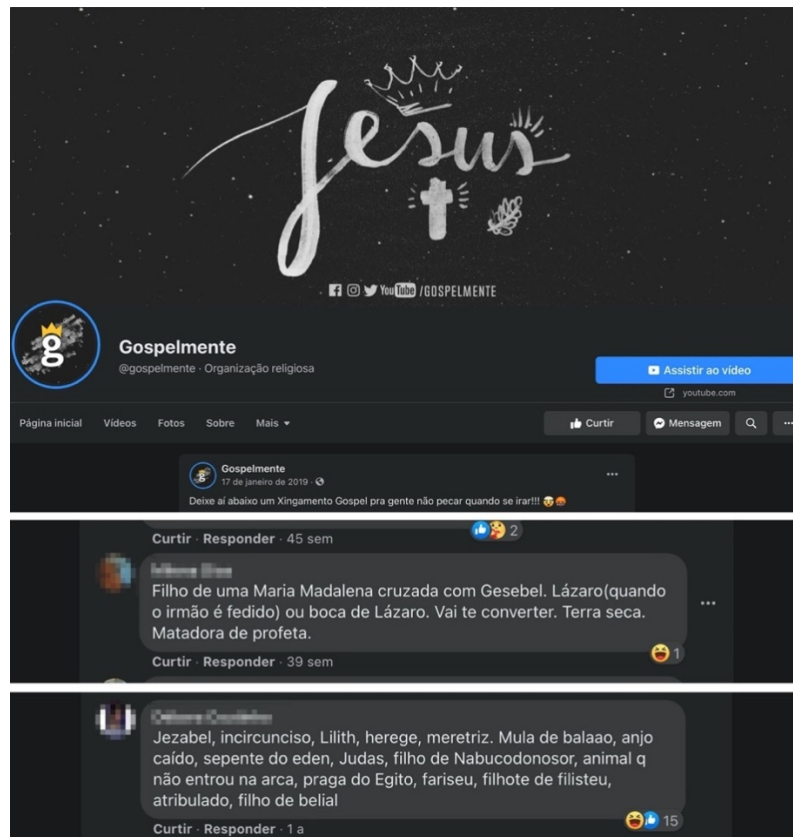
We crossed the comments collected on Feliciano's fan page with texts from the Catholic Holy Bible to identify common terms. From there, the terms were frequently identified at the transition point (T) of Goffman (Guedes & Santos, 2013; Santos, 2009) and through content analysis (Neuendorf, 2017), the retrieved terms were categorized as part or not as a hate sentence. Finally, we organized these terms in a list that contains the term, the whole comment, the recipient of the speech, and the form of hate speech.

In summary, the work concludes with the systematization of 25 terms used in the context of hatred, organized, and systematized into four categories: "intolerant," "fundamentalist," "extremist," and "curse." These terms do not constitute hate speech on a first impression and in isolation. For this reason, it probably would not be noticed through content moderation techniques that use block lists or other alternatives that rely on computational resources and languages that support one another in term dictionaries. However, by the methodology used, it is possible to observe them being used for this purpose. Furthermore, the results show how promising a methodological triangulation based on text mining, bibliometric laws, and content analysis can be for retrieving this type of information.

For discussion of the proposal, this article was organized into five sections: the first focused on defining what is religious hate speech online; the second presents the context where this practice takes place—digital platforms—and the challenges of content moderation in these spaces; the third aimed at presenting and discussing in detail the proposed and tested methodological approach; the fourth with the results; and the fifth with the discussion about the study.

### **Online Religious Hate Speech**

In 2019, a publication on the Gospelmente (n.d.) page on Facebook read as follows: "Leave a Gospel Cursing below, so we do not sin when we get angry!!! 🤔🙄." The said page has more than 2,750,000 followers, and the post features 305 comments containing biblical expressions used as curses and insults.



**Figure 1. Post Gospelmente (n.d).**

Based on this example, we can observe how these manifestations of intolerance, prejudice, and discrimination may be hidden in biblical expressions that are not recognized as hate discourse. Expressions such as “Belial’s daughter,” “Amalekite,” and “Pharaoh’s horse” refer to ungodly, perverse, impure, bitchy, useless, dishonest, despicable, and evil people. Religious people use such terms to offend, oppress, and disqualify people or groups because of their behavior, lifestyle, beliefs, and convictions. Among the victims of this abusive discourse are believers and nonbelievers, Jews and Semites, women, the LGBTQIA+ community, Blacks, indigenous, immigrants, political opponents, obese, elderly, and people with physical and mental disabilities.

Like other types of abusive speech, religious hate speech can be neutral without overtly hostile or degrading language (Blaya, 2019; Roberts, 2019). Religious intolerance can be camouflaged in words and expressions that, at first glance, do not appear to be inappropriate content, including the use of revealed texts to justify discrimination and foment violence (Britt, 2010). They can also be linked “to the secular nature of the State, the rights of the LGBTQIA+ population and religious education in schools” (Filla & Fantinni, 2016, p. 201), in addition to serving to recruit people for hate crimes and terrorism (Budke, 2016; Nobata, Tetreault, Thomas, Mehdad, & Chang, 2016).

Part of these discourses is based on the idea that "one religion is superior to others or the only holder of absolute truth" (Santos, Simões, & Salaroli, 2017, p. 363). It can also be recognized in discrimination, prejudice, and violence concerning those considered dissidents, heterodox, agnostics, seculars, and atheists (Tamayo, 2019). Furthermore, demonstrations can have fundamentalist and extremist characteristics, including disseminating content that threatens or promotes actions against (secular) civil rights, insulting, or expressing contempt based on religious arguments.

The concept of fundamentalism emerged at the end of the 19th century at the Niagara Biblical Conference. In this time, contrary to modern theology, a group defended the Christian Bible as a foundation for understanding the world. Since then, this reductionist vision, which is not exclusive to Christians but from several monotheistic religions, has hindered points of contact and dialogue with other religious manifestations, cultural dimensions, and worldviews (Tamayo, 2019; Vieira, 2018).

Tamayo (2019) associates fundamentalism with fanaticism, especially regarding the literal way the religious interpret revealed texts (such as the Qur'an, Bible, Sunnah or Hadith, Torah, Tri-Pitakas, Vedas) and the dogmatic definitions of their religions. In such a way, the author points out that groups such as ultra-Orthodox Jews, fundamentalist Muslims, and traditionalist Christians do not consider the historical context for interpreting the holy books. Moreover, this stand attributes meanings that often segregate and discriminate against groups that do not share the same faith.

In addition to interfering with the right of belief, groups of believers of these religions, in an organized or spontaneous way, also hinder the whole exercise of other fundamental rights, especially of women, LGBTQIA+ people, indigenous people, believers of African-based religions, and progressive popular movements in general (Vieira, 2018). Furthermore, fundamentalists stress-sensitive issues such as legalization of abortion, same-sex marriage, female emancipation, and fostering ethnic-religious prejudice, addressed to religions considered "primitive" (Buddhism) or animist (Shintoism; Budke, 2016; Nobata et al., 2016).

Religious extremism, in turn, can be identified as beliefs and actions that support violence to achieve religious goals, which in some cases can be manifested in acts of terrorism (Coleman & Bartoli, 2014; United Nations Educational, Scientific and Cultural Organization [UNESCO], 2017). Coleman and Bartoli (2014) consider extremism as activities, beliefs, actions, or strategies that go out of the ordinary and break with established consensus. In this same sense, Cunha (2012) argues that extremism is about "practical measures" to punish those who disagree with interpreting sacred texts.

Another aspect of religious hate speech is the use of curse words, that is, excerpts from sacred texts considered as "strong words" but serve to "curse" people or groups of people with protected characteristics (e.g., Blacks, women, gays, Jews). For Britt (2010), when a person uses these words to cause or invoke harm to someone, by supernatural means or to belittle or deliver the target (person or group) to a stigmatized condition, the "supernatural" dimension of the discourse is revealed of hatred tied to religious history and curse traditions.

Only to have an idea of what is happening, one of the Internet content moderators states that in the book by Roberts (2019), part of the demeaning comments, even those that do not use bad words or

swearing, use the scripture and religion as weapons. The moderator disagrees with the passive-aggressive way other people use Bible terms or religious doctrines to attack other people and point out expressions such as "You are going to hell" or "You are an abomination" recurrently in social networks. Consequently, it is necessary to read the Bible to understand some contexts.

As can be seen, it is a complex, multifaceted phenomenon that needs careful analysis. In this research, the identification and analysis of religious hate speech considered all content that aims to discriminate, offend, reduce, or encourage violence about those considered dissidents, heterodox, agnostics, seculars, or atheists. In addition, content can be manifested in intolerance, fundamentalism, extremism, and the curse. The categories of analysis used will be explained in the method procedures section.

### **Moderation of Fundamental and Extremist Content on Social Media Platforms**

Most digital platforms, including Facebook, do not tolerate hate speech and have increased their investments in more effective solutions for moderating this type of content (Blaya, 2019; Silva & Sampaio, 2018). Despite advances, companies assume the complexity of identifying and removing hate speech (Blaya, 2019). All moderation techniques face the challenge of overcoming aspects such as platform heterogeneity, the fragmented nature of user-generated content, multiple languages, in addition to security issues and the speed with which content circulates (Ahmad, 2016; Blaya, 2019; Gillespie, 2018; Silva & Sampaio, 2018).

On the one hand, content moderation processes remain opaque, and hate discourse policies in companies do not prioritize healthy interactions and the protection of users. North American Media Consortium points out the negligence of Facebook with abusive content in its platforms and the safety of its users (Cantú, 2021; Wall Street Journal staff, 2021). Former employees have confirmed that Facebook values growth and profits instead of fighting hate discourse, misinformation, and other threats to the public. The story also points out that moderation in hate discourses and misinformation in countries out of the English-speaking world is also poor. Artificial intelligence in these platforms is scarcely adjusted to local languages, and the company invests very little in human moderators fluent in such languages and dialects.

On the other hand, Facebook has opposed the allegations and has shown developments as to the exclusion of hate content. Among the methods used are: (a) social moderation—in which the interactor himself identifies and signals what he understands as profane or inappropriate (Gillespie, 2018); (b) automatic detection procedures, including natural language learning and Web filtering—filtering words or expressions through block lists (Ma, Saul, Savage, & Voelker, 2011; Sood, Antin, & Churchill, 2012); and (c) human moderation (Gillespie, 2018).

Interactor flagging moderation (also called user reporting) uses the platforms' interactor base to help identify offensive content and behavior (Gillespie, 2018). Interactors become regulators through mechanisms that allow "flagging" content that violates the platforms' community policies and guidelines. Each of these mechanisms allows users to organize the platform's content and its reviewing, rating, and presentation to others.

The alternatives for automatic detection of inappropriate content, on the other hand, aim at the problem of scale (a massive amount of content to be analyzed) and the distance from any human bias and

subjectivity (Gillespie, 2018). Fortuna and Nunes (2018), in a systematic review of the literature on hate speech detection, identified several automated approaches, including text mining, distance metrics, bag-of-words (BoW), n-grams, word embeddings, and deep learning. The authors also identified that the main approaches are to create machine learning models for hate speech classification and that the most common algorithms used are: (a) support vector machines (SVM), (b) random forest, and (c) decision trees.

Among the automatic techniques is Web filtering—an approach used by people and companies to determine if the data received is harmful to the network or includes any intellectual property (Banday & Shah, 2010). The technique uses a firewall to check for the presence of a set of predefined settings, blocking harmful and objectionable data. The filter can be implemented nationwide between the Internet backbone and the country's network; at the organizational level, between the organizational network and the Internet gateway; on a server, where the filter is installed on the ISP gateway to provide filtered content to the interactors; at an individual level, when installed on the computer or local workstation; or through third parties, when companies/applications use SOCs (security operation centers) for filtering.

Web filtering uses classifiers that contain block lists formed by terms, expressions, and links identified as inappropriate content or interactors. This type of moderation with lexical, syntactic patterns has a high sensitivity (Kwok & Wang, 2013) and a lower cost than other existing classifiers (Ma et al., 2011). Therefore, they are essential tools for the prevention of harassment (Geiger, 2016; Jhaver, Ghoshal, Bruckman, & Gilbert, 2018), for the maintenance of safe spaces (Clark-Parsons, 2018), and even to preserve the mental health of interactors (Wheatley & Vatnoey, 2020).

On their part, block lists, since they understand Internet communication as unimodal, use the written text and do not consider images or language syntax, for instance. However, unfortunately, this technique typically has high false-positive rates because of the ambiguity of the words in a single text. Besides, they cannot track down offensive content in real time nor block interagent offenders (Hashmi, Ikram, & Kaafar, 2019).

Other limitations are not identifying spelling errors (intentional or not), the absence of analysis of the context/nature of profanity, and the rapid change of speech (Sood et al., 2012). The resource can also cause structural overblocking, which is verified when an entire domain/profile/page is blocked instead of blocking a specific publication, causing unintended technical interruptions to innocent parties (Hashmi et al., 2019).

In any case, block lists are practical tools to combat harassment and inappropriate language, either by restoring control of interactions or improving the interactor's experience (Golbeck et al., 2017; Jhaver et al., 2018; Maréchal, 2016). Further, block lists can even reach the "creation of collective meaning" (Geiger, 2016, p. 789) with shared values, identities, and ideas about what constitutes harassment.

### **Methodological Proposal**

The methodological proposal of this study tests the use of text mining, bibliometric laws, and CA procedures to identify religious hate speech in texts produced and disseminated on Facebook. These texts

contain terms that are also present in the Christian Holy Bible. The steps of data collection, treatment, analysis, and indexing considered that some religious groups use inflamed words to cause public indignation and promote their agendas (Britt, 2010). Moreover, these isolated terms cannot reveal hate discourse and, thus, they would not be present in block lists and term dictionaries used for moderation.

We adopted text mining procedures because of the need to identify and extract information from unstructured texts (Westergaard, Stærfeldt, Tønsberg, Jensen, & Brunak, 2018). In this sense, the procedures used in this study correspond to the stage called preprocessing (Kobayashi et al., 2017). Thus, the study follows these steps: (a) data collection, (b) data cleaning and transformation, and (c) selection of terms using bibliometric laws.

In addition, we used content analysis, following the model proposed by Silva and Sampaio (2018), to identify which of the selected terms were inserted in the context of hate speech. Content analysis emphasized the validity, replicability, and reliability of the coding (Neuendorf, 2017). Furthermore, from an open science perspective, the coded data are available in the Scientific Database of the Federal University of Paraná (BDC/UFPR).<sup>2</sup>

Both approaches used methods and tools deemed pertinent for this research goal, as seen below.

### ***Data Collection***

Data collection took place on two fronts: Facebook and the Christian Bible. Regarding Facebook, the platform's public API was used through the discontinued application Netvizz.<sup>3</sup> The extraction took place from the fan page of pastor and federal deputy Marco Antônio Feliciano, parliamentarian recognized for statements that associated biblical concepts of "spiritual curse" to the African people and their descendants and insults to the LGBTQIA+ community and other historically marginalized minorities.<sup>4</sup>

We collected 4,214,699 comments from May 2013 to December 31, 2016. The time analyzed covers the mandate of the parliamentarian as chairman of the Commission on Human Rights and Minorities of the Chamber of Deputies (2013–2014), the 2014 election, and the first two years of his second term as federal deputy (2015–2016).

For data collection in the Christian Bible, the first step was to list which version of the Holy Scriptures would be analyzed since there are differences between Catholic and Neo-Pentecostal Bibles and differences in the translations of the texts. The Catholic canon (list) has 73 books, while the "Evangelical

---

<sup>2</sup> Available at: <https://bdc.c3sl.ufpr.br/>

<sup>3</sup> Netvizz is a tool developed in the Digital Methods Initiative (DMI) by Bernhard Rieder. Its function is to collect different data types from Facebook, mainly from Pages, Groups, and Events. For example, extracting textual content from posts and comments and likes, comments, shares, and reactions data is possible. The data were extracted in 2016 when the platform's API allowed access to the application.

<sup>4</sup> Read more at: <https://noticias.uol.com.br/politica/ultimas-noticias/2011/03/31/deputado-federal-diz-no-twitter-que-africanos-descendem-de-ancestral-amaldicoado.html>



Bible” is composed of 66 books, divided into Old and New Testaments. Therefore, the “Catholic Bible” has seven more books than the neo-Pentecostals, which are the books of Tobias, Judith, Wisdom, Baruch, Ecclesiasticus (Siracid or Sirac), I Maccabees, and II Maccabees, in addition to some parts of the books of Esther and of Daniel. In this context, the Catholic version of the Pastoral Bible was chosen (*Nova Bíblia Pastoral*, 2021), considering that the canon includes the “Evangelical Bible” books. Another point that weighed on the decision is that most of the Brazilian population declares itself to be a Catholic Christian (Datafolha, 2021).

### ***Data Cleaning and Transformation***

Data cleaning and transformation procedures were performed in the comments extracted from Facebook and the Bible file. Both converted to plain text to process the data and find the most recurrent nouns and adjectives present in both databases. The proposal understands that verbs, pronouns, numerals, and adverbs are associated with nouns and adjectives. Therefore, by identifying the noun or adjective of “hate” in the sentence, it is also possible to identify the other grammatical classes that constitute hate speech. For this, tools developed with the Python programming language, version 3.6, were used.

The first procedure consisted of “cleaning the data” to remove unnecessary noise and characters for the defined scope. In this step, empty records, URLs, punctuation elements, and special characters like accents and “ç” have been removed or replaced by the equivalent, without special characters. Then, we changed the uppercase characters to lowercase to synchronize the terms. Thus, we avoided words like “Civilização” and “civilizacao” being categorized as different, even they have the same meaning.

Next, we removed a set of words considered irrelevant (*stop words*) for this study. Five groups of *stop words* were defined:

1. The Natural Language Toolkit (NLTK) package list of stop words.
2. The PreText tool list of stop words.
3. A list of the 5,000 most used infinitive verbs in the Portuguese language.
4. A list of proper names.
5. Adverbs and superlatives.

The first group of stop words removed contained the words present in the NLTK package.<sup>5</sup> However, we noted that terms such as “below,” “however,” and “where,” for example, had high frequencies of occurrence in the corpus and were not included in the removal list. Therefore, we added the words listed in the PreText software (Matsubara, Martins, & Monard, 2003) to the list of stop words as it contains many Portuguese terms than the NLTK package.

Then, we removed the 5,000 most used verbs in Portuguese and their respective conjugations.<sup>6</sup> At this stage, we built a Python script that allowed the collection of verbs directly from the website

---

<sup>5</sup> Available at: <https://www.nltk.org/>.

<sup>6</sup> Available at: <https://www.conjugacao.com.br/verbos-populares/>.

*conjugacao.com.br* and the removal of terms in the two databases (Comments on Facebook and Pastoral Bible). Another script removed the most registered proper names in Brazil between 2015 and 2019 according to the Civil Registry Transparency portal,<sup>7</sup> which amounted to 91 names removed. Adverbs were another group of stop words taken from the corpus—words with the suffix “*mente*”—and their superlatives—words with the suffixes “*íssimo*,” “*ísimos*,” “*íssima*,” and “*íssimas*.”

Finally, we tokenized the sentences (Jiang & Zhai, 2007), that is, the words in the comments were separated according to the blank space between them. Then, we calculated the frequency of occurrence of the terms, disregarding terms with fewer than three characters. The total of terms remaining was 412,441 from the Facebook comments and 12,742 from the biblical text.

### **Selection of Terms**

Based on the most frequent terms collected from the comments and the Bible, we intersected the two lists to reveal the terms used in the comments that were also present in the Bible. This fusion identified 9,505 terms that were then analyzed based on Zipf's law and Goffman's transition point ( $T$ ; Guedes & Santos, 2013; Santos, 2009).

Zipf's law analyzes the frequency of occurrence of words in a given text. Although it is referred to as a single law, it has two variations: (1) the first represents high-frequency words in a text; and (2) the second, also known as Zipf-Booth's law, is applied to the representation of words with a low frequency of occurrence (Guedes & Santos, 2013; Santos, 2009). Goffman's point ( $T$ ) is a way to “enrich” this law, allowing the identification of a region composed of terms with high semantic content (Guedes & Santos, 2013; Santos, 2009).

To apply Zipf's law, we grouped the terms in the comments by frequency; the resulting list was sorted in decreasing order of frequency ( $f$ ), and a serial order ( $r$ ) was assigned to represent the ranking of this frequency. Finally, we multiplied these elements to obtain the constant of the first of Zipf's laws ( $r \times f = c$ ), and to calculate Goffman's point ( $T$ ), we used Equation (1):

$$n = \frac{-1 + \sqrt{1 + 8 * 681}}{2} = 36,41 \quad (1)$$

where  $n$  represents the point ( $T$ ); 8 symbolizes a constant and 681 corresponds with the number of terms with frequency of occurrence equal to one.

---

<sup>7</sup> Available at: <https://transparencia.registrocivil.org.br/inicio>.

By means of Equation (1), we found that the point ( $T$ ) occurs at frequency 36.41. By rounding this value, the zone with high semantic value is found in terms whose frequency of occurrence is equal to 36, which amounted to 44 terms. To obtain many terms for the analysis, we considered increasing this zone for frequencies 35 (33 terms) and 37 (31 terms), totaling 108 terms.

Table 1 shows the list of terms considered as having high semantic value and observed in the light of the CA.

**Table 1. List of Terms With Frequency 35, 36, and 37.**

Freq.	Terms
35	<i>abolition / welcome / archer / boats / deer / understood / completed / contiguous / delicious / devastating / sweetness / weakened / that / emerald / excessive / fertile / grandeur / horrified / harmless / insatiable / unreal / donkeys / light / mantles / messianic / buttocks / redeemed / divided / plundered / solemn / drums / theoretical / threshold</i>
36	<i>abysses / beaten / amplitude / aruma / tormented / atria / caleb / carcass / caramel / closed / tight / combatants / combined / comparable / conversions / corrected / cosmic / cook / deposits / day laborer / dismissed / divorces / eber / flood / scorpion / thin / immature / indulgence / inflamed / wide / marine / dipped / nocturnal / occupations / wetland / begging / dripping / precise / recall / remembrance / remote / uphill / transmitted / brought</i>
37	<i>lit / suitable / alexandrine / camels / duress / colorful / consecutive / troubled / misfit / devastating / debtors / devoured / dubious / escape / eliminated / fertile / outsider / habitat / imminent / infamous / insane / insulting / unpalatable / vegetables / magdiel / moved / muscles / disturbed / spared / insured / dark</i>

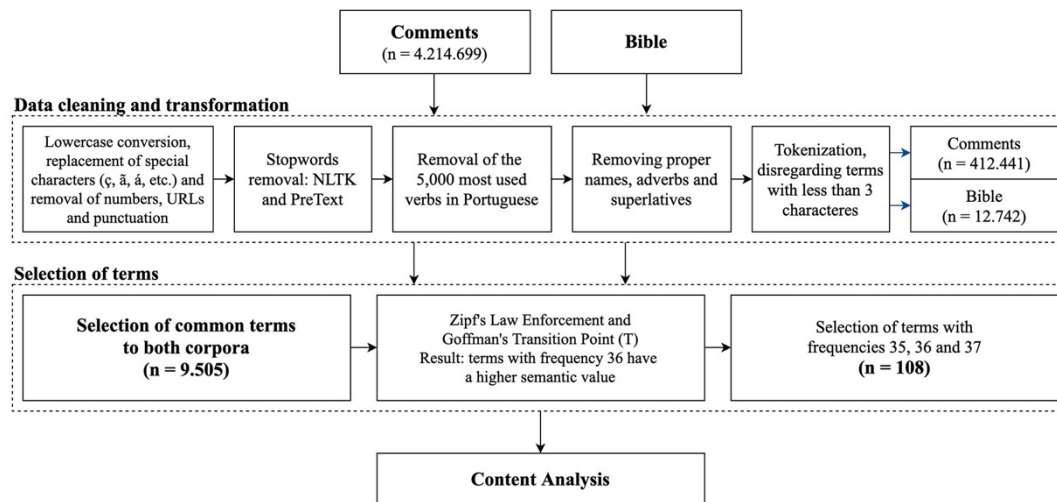
To elaborate the frequency rankings, we considered the frequency of occurrence of the terms found in the comments to avoid analyzing terms with high frequencies of occurrence in the Bible and low frequencies in the comments.

### **Content Analysis**

Content analysis used the categories of Silva and Sampaio (2018) to determine whether the term was within the context of hate speech or not. The 3,614 comments were analyzed considering the type, manifestation, and form of hate speech. We also used an analysis category that considers: (a) intolerant—the text that intentionally offends or insults one or more members of the community; (b) fundamentalist—when based on a literal and decontextualized view of the Bible; (c) extremist—when it values, encourages, or supports violence on account of a cult or creed; and (d) curse—when based on words that cause or invoke damage to someone by supernatural means and depreciate or surrender the target to a stigmatized condition.

The unit of analysis was the textual content of the comments collected containing one of the terms with frequency 35, 36, and 37. We disregarded duplicate comments and those exceeding 3,000 characters

as they were biblical quotes *ipsis litteris* or texts replicated from other sources. Figure 2 shows the synthesis of the methodological guidelines carried out in this study.



**Figure 2. Flowchart of the methodology.**

Content analysis (Neuendorf, 2017) was done manually by two previously trained codifiers until they reached Kappa coefficient reliability of 0.646. The Kappa concordance coefficient may vary from 0 (zero) to 1. The closer to 1 is its value, the greater the indication of a concordance between the codifiers. In this case, the value 0.646 indicates a strong (significant) concordance.

## Results

Twenty-five of the 108 terms were used in hate contexts. For our analysis, this result was organized in an analytical table containing the term, the comment, the form of speech, and the addressee, as shown in Table 2. The systematization shows the forms of speech used as the CA categories (intolerant, fundamentalist, extremist, curse, or praise) and the different recipients of the message. The complete table is in Appendix I (English)<sup>8</sup> and II (Portuguese).<sup>9</sup> We have translated terms and comments into English to facilitate reader understanding. However, we also recommend reading comments in their structure in Portuguese. Furthermore, the original text extracted and analyzed makes more evident the aspects that characterize it as hate speech in the Brazilian context and contain linguistic aspects that deserve attention during the analysis, such as the presence of figures of speech, abbreviations, spelling, and grammatical errors.

<sup>8</sup> Available at: <https://bdc.c3sl.ufpr.br/bitstream/handle/123456789/109/apendiceI.pdf>

<sup>9</sup> Available at: <https://bdc.c3sl.ufpr.br/bitstream/handle/123456789/109/apendiceII.pdf>

**Table 2. Biblical Terms Used in Religious Hate Context.**

Term	Comment	Form of Speech	Recipient
ESCAPE	I tell you only one thing Lucio Walker E, you will know the truth, and the truth will set you free. John 8:32 the path of truth and life and Jesus, whether you accept it or not, he is unique and true .... And everyone who sin here will be judged and from that, there is no escape whether you accept it or not #Ijustsayit	Curse: they use the term as a threat— “there is no escape”— against those who do not profess the same faith.	Not identified
ELIMINATED	They are true apatrid chameleons contaminated by the malignant cancer that settled in our country!!! They have to be eliminated.	Extremist: they incite extinction, violence.	Not identified
MOVED	Pastor Marcos Feliciano is not concerned with these threats from people moved by Satan. Jesus is with you this is more important all evil will fall to the ground.	Fundamentalist: They use the biblical expression “moved by Satan” to justify any kind of contrary thinking or persecution of the pastor.	Not identified
TIGHT	It must be difficult for Feliciano to live as many others, hiding, not being able to admit to their identity, having to lead a life of lies, to straighten his hair and wear tight pants swearing he is not gay. He heals so many people and does not heal himself. You do not represent anyone, Mr. Feliciano, because whoever follows you is alienated.	Intolerant: They make generalizations and assumptions regarding the deputy’s way of dressing—“tight pants.”	Deputy Marco Feliciano
	It is true . . . but one thing is certain, God wants us all! Whether she is wearing makeup or not, she is serving God. Now if she was wearing clothes that	Intolerant: They make generalizations and assumptions about the way women	Women

	were too tight, she would be scandalizing the Holy name.	dress—"tight clothes."	
DRUMS	The truth is that some spirits try to overthrow Pastor Marcos, you are ridiculous, you have no morals, much less intellect to do that, go beat your drums	Intolerant: They consider other religious practices as inferior—"go beat your drums."	Spiritist Community
DIVISION	Jesus is just a division of the bipolar and outsourced company that they call God.	Intolerant: They associate religion with animals/sub-humans.	Christians

*Note.* From *Discurso de ódio no Facebook: A construção da incivildade e do desrespeito nas fan-pages dos parlamentares Jair Bolsonaro, Marco Feliciano e Rogério Peninha* (Lopes-Silva, 2017; <https://bdc.c3sl.ufpr.br/handle/123456789/110>).

The 25 terms found in the context of hatred were: flogged, tormented, camels, tight, colored, misfits, devastating, escape, eliminated, outsider, habitat, infamies, foolish, unpalatable, donkeys, light, dipped, moved, buttocks, beggar, disturbed, divided, drums, dark, threshold. If we observed in a first impression a good part of the terms in isolation, they do not constitute hate speech. However, they are used for this purpose—here, the need to observe the n-grams of the sentence is demonstrated, the context in which the interactant's manifestations and intention occur; otherwise, the content will likely go unnoticed in content moderation techniques that use block lists. For example, the term "habitat," as observed by the following comment (Marco Feliciano, n.d.):

Paola, I thought you were a woman, lol . . . Paola, here is the thing, I don't know what goes on in the mind of a man who chooses to be gay, heartbreak, family disappointment, friendships, all this has an influence, right? As to those who say they have been born like this, which I do not believe, they were exposed to a female habitat, which ended up creating doubt in the child and the female side prevailed because it was in greater number than the male, then masculinity seeks to emerge and suddenly that child looks for the male habitat, and at some point, the traces of the female habitat will appear and detergent (sic) that young person will be feeling sexual desires for a friend or cousin, and so on. There is a way to avoid this, not to give in to carnal desires and seek help.

The term habitat in isolation does not qualify as hate speech, but when the n-gram "female" is added, the expression "female habitat" takes the form of intolerant speech in the sentence, as it considers sexual orientation because of a dysfunctional environment. The expression shows the relationship with the feminine as a determining factor for the individual's sexual orientation; the LGBTQIA+ community is targeted in this message.

In another example—"Globo's artists are all disturbed because they only get involved in macumba" (Marco Feliciano, n.d.)—the term "disturbed" alone is an offense, but it doesn't qualify as hate speech. At the end, the sentence reads "get involved in macumba," an African-Brazilian religious ritual. Such

association between “being disturbed” and “getting involved in macumba” qualifies as religious hate speech given that it expresses contempt for another religious practice and is a judgmental statement about the behavior of a group of people, TV Globo’s artists, as misguided and immoral.

These examples show the difficulty in identifying what qualifies as hate speech and what is closer to other abusive content such as aggressive, crude, and offensive language. On the other hand, they provide an important characteristic of hate content, which is the recurrence of associations of terms, regardless of the distance between them in the sentence. In this sense, filtering and classification options for suspicious messages need attention throughout the sentence structure and in terms of the relationships between terms.

Regarding the form of speech, the comments of the 25 terms in the AC were categorized into four of the five aforementioned categories; the praise category was not found in the content analyzed. Regarding frequency, the intolerant category prevails; it was verified in 36 comments. The extremist category was verified in two; and curse and fundamentalist, both in only one comment each. The analysis was careful in preserving the indications of the codebook by Silva and Sampaio (2018) so as not to confuse hate speech with a simple divergence of opinions or other forms of intolerance, mainly because the corpus was collected from a politician fan page, where the divergence of ideas is expected.

The content analysis also revealed other 242 terms and expressions used as hate speech but were not among those identified in Goffman’s transition point (7). The terms will integrate an ontology of hate speech in digital social networks in Brazil and are not included in this work.

The results confirm the diverse and multifaceted nature of religious intolerance on the Internet and how the content is fragmented, with spelling errors (intentional or not) or typos. In a sentence such as “Those unhappy pastors will suffer at the threshold!!!” (Marco Feliciano, n.d.), we assume there is a typo [three S’s in a row, in Portuguese], while in the sentence below, the original Portuguese words “*estasiado*” [ecstatic] and “*faser*” [to organize, in this case] are simply misspelled.

The day you come here to our church in Alvorada you will be ecstatic [*estasiado*] with our praise. Alvorada needs your preaching pastor Marco Feliciano. Alvorada is a city deeply immersed in macumba and violence. If it wasn’t for the churches of Jesus and (sic) the city would be in hell. You could mobilize a union of churches and organize [*faser*] an event in the fire department. We are waiting for you, pastor. Peace of the Lord. (Marco Feliciano, n.d.)

Many of the comments produced and disseminated by believers and nonbelievers with intolerant and violent speech were based on biblical texts. Therefore, the stage of intersecting the content extracted from Facebook and that of the Pastoral Bible proved relevant for term mining. The stage implementation recognizes the importance of the language common to the group (in this case, the Christian community that follows the pastor on Facebook). From this common vocabulary among members of the community, meaning is attributed to the speech. It is likely that without this crossing, the terms identified in Goffman’s transition point (7) would not necessarily be found in a context of hatred.

It was not possible to identify the recipients of the messages in 14 comments, which may indicate that the content is part of a counterhate narrative targeting another interactant with a diverging perspective, or it may be a reply to the pastor's post. It would be necessary to analyze the publication in the administrator's page to better understand the context in which the content was produced and to whom it is intended. On the other hand, the LGBTQIA+ community appears as the most frequent recipient, which was verified in eight comments. The other recipients can be divided among the class of pastors and Deputy Marco Feliciano (Pastor, four; Marco Feliciano, two); religion (religious communities, two; Muslims, one; Islam, one; Christians, one; Spiritist community, one); artists (artistic class, one; Globo's artists, one); minorities (Black man, one; women, one); personalities (Jean Wyllis, one; Graciete, one; Felipe Silva, one).

As to the hate content moderation point of view, we observe that when political/religious leaders use digital platforms aiming to legislate in the name of God, the risk that religious content becomes misinformation and hate discourse increases. Furthermore, several comments identified as hate discourse were based on publications in which minister Feliciano discusses identity and human rights issues under the fundamentalist religious discourse perspective. Thus, we can notice the need for such platforms to reconsider their hate discourse policies to offer a quicker and more efficient treatment. It is especially true when hate discourse is encouraged by public people such as politicians, artists, and religious people with a broad scope, influence, and prestige.

### **Discussion**

This article explored religious hate speech produced on digital social networks, especially on Facebook, using a biblical term or expression to attack people or groups with protected characteristics, including professed non-Christians, agnostics, atheists, and secularists. Fundamentalist language is among the forms of hate speech identified; it resorts to literal interpretation of sacred texts, extremist manifestations inciting violent acts and actions, words and expressions that evoke curses, and words of praise for those who practice discrimination or fight equality of rights for minority groups. The study also presented moderation strategies and techniques used by platforms to identify and remove this type of content, including automatic detection methods such as Web filtering.

As a methodological proposal, we used the triangulation of text mining and content analysis procedures and the Zipf bibliometric law, aiming to expand the number of terms in the block lists used by the platforms. The findings revealed terms used in a nonevident way to justify or camouflage the intention to offend, diminish, or harm a person or group of people with protected characteristics.

The results revealed 25 terms that apparently or alone are not considered hate speech, but that within the religious context of the pastor's Facebook page is intended to harm others. The terms were organized to contribute to hate speech studies by expanding the block lists already used in other systems of inappropriate content detection used by the platforms.

Further contributions of the methodological proposal are (a) improvement of the preprocessing step, adding new stop words in Portuguese; (b) the intersection of the terms mined on the DSN with terms from other texts, considering substitution according to the origin of the database; for example, if the page to be analyzed



belongs to a personality in the legal system, the text used for intersection may be extracted from legal vocabularies that extend to several legal expressions; and (c) the use of Zipf's law and Goffman's transition point (T) to identify highly semantic content, which in the case of this study was focused on nouns and adjectives.

The comment files used utf-8 encoding, which hindered the text mining process. Certain records required "manual encoding" since special characters such as "@" and diacritics ("à," "é") hinder their correct representations in other types of encoding (ASCII, ANSI, etc.). The presence of grammatical errors such as "favos" [instead of *favor*], "ourives" [*horríveis*, horrible], and "umidade" [*humildade*, humility] was also problematic because the correct interpretations of their meanings were possible only within the content analysis.

Regarding this study's limitations, we point to the difficulty in carrying out preprocessing procedures in a Portuguese-speaking corpus. Mainly concerning the lemmatization steps—remove plural and consider verbs in the infinitive—and stemming—reduce words inflected to the root. The lack of these procedures resulted in creating different functions to guarantee the quality of the analyzed content. Furthermore, the computational performance was also a limitation because of the number of analyzed records, causing the preprocessing to take approximately 15 hours of computer work.

In any case, the study showed promise for analyzing suspicious messages, apparently neutral, but hiding hate speech. The findings demonstrate the methodological proposal as feasible and indicate improvements that need to be made but that do not compromise the validity of the research. The methodological proposal helps investigate and prevent online hate speech in human and automated moderation procedures, and it is also beneficial in intolerance education initiatives. The cross between comments published on the platform and a glossary of terms common to an online religious community optimizes identification. For example, in a context where Islam prevails, cross-referencing the comments with the Qur'an will likely lessen the mining effort and, with the use of Goffman's transition point (T), will reach more quickly in terms that have a high semantic load of hate. Another advantage of this methodology is to make the terms and their context explicit, which can support the education of civility and respect insofar as people may not use that term or offer new meaning. Other approaches using databases from other platforms can contribute to the evolution and refinement of the model. As future works, the possibilities of using the results for automatic grouping are listed; database validation; and creating templates for automatic classification of hate speech.

## References

- Ahmad, A. N. (2016). *Masculinity, sexuality and illegal migration: Human smuggling from Pakistan to Europe*. Abingdon, UK: Routledge.
- Araújo, C. A. (2006, December 10). Bibliometria: Evolução histórica e questões atuais [Bibliometrics: Historical evolution and current issues]. *Em Questão*, 12(1), 11–32. Retrieved from <https://seer.ufrgs.br/EmQuestao/article/view/16>

- Banday, M. T., & Shah, N. A. (2010). A concise study of web filtering. *AIS Electronic Library (AISEL)*, 10(31). Retrieved from <http://sprouts.aisnet.org/10-31>
- Blaya, C. (2019). Cyberhate: A review and content analysis of intervention strategies. *Aggression and Violent Behavior*, 45(March–April 2019), 163–172. doi:10.1016/j.avb.2018.05.006
- Britt, B. M. (2010). Curses left and right: Hate speech and biblical tradition. *Journal of the American Academy of Religion*, 78(3), 633–661. doi:10.1093/jaarel/lfq020
- Budke, S. (2016). Mídia & religião: Conflitos e oportunidades do diálogo inter-religioso no ciberespaço [Media & religion: Conflicts and opportunities for interreligious dialogue in cyberspace]. *Anais do Congresso Estadual de Teologia*. São Leopoldo, Brazil: EST.
- Cantú, E. (2021, October 28). The Facebook Papers and their fallout. *The New York Times*. Retrieved from <https://www.nytimes.com/2021/10/25/business/facebook-papers-takeaways.html>
- Cinelli, M., Morales, G. F. M., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9), 1–9. doi:10.1073/pnas.2023301118
- Clark-Parsons, R. (2018). Building a digital Girl Army: The cultivation of feminist safe spaces online. *New Media & Society*, 20(6), 2125–2144. doi:10.1177/1461444817731919
- Cohen-Almagor, R. (2009). Belgian Euthanasia Law: A critical analysis. *Journal of Medical Ethics*, 35(7), 436–439. doi:10.1136/jme.2008.026799
- Coleman, P. T., & Bartoli, A. (2014). *Addressing extremism* [White paper]. International Center for Cooperation and Conflict Resolution, Columbia University. Retrieved from [https://www.cve-kenya.org/media/library/Coleman\\_\\_Bartoli\\_2002\\_Addresssing\\_Extremism.pdf](https://www.cve-kenya.org/media/library/Coleman__Bartoli_2002_Addresssing_Extremism.pdf)
- Cunha, C. V. D. (2012). Conflitos religiosos e a construção do respeito à diversidade: Breve histórico e iniciativas recentes [Religious conflicts and the construction of respect for diversity: A brief history and recent initiatives]. In A. Libonatti & D. G. K. Kitta (Orgs.), *Comunicação e transformação social* [Social communication and transformation] (pp. 95–122). Caxias do Sul, Brazil: Unisinos.
- Cunha, C. V. D., Lopes, P. V. L., & Lui, J. (2017). Religião e política: Medos sociais, extremismo religioso e as eleições de 2014 [Religion and politics: Social fears, religious extremism and the 2014 elections]. Rio de Janeiro, Brazil: Fundação Heinrich Böll.
- Datafolha. (2021). Grau de confiança nas instituições [Degree of trust in institutions]. Retrieved from <http://media.folha.uol.com.br/datafolha/2021/09/24/avali24968insti94782congress8472.pdf>

- Filla, M. G., & Fantini, J. A. (2016). A construção mutual de discursos intolerantes: Ateus, agnósticos e religiosos [The mutual construction of intolerant discourses: Atheists, agnostics and religious]. *Memorandum: Memória e História em Psicologia*, 30, 199–223.
- Fortuna, P., & Nunes, S. (2018). A survey on automatic detection of hate speech in text. *ACM Computing Surveys (CSUR)*, 51(4), 1–30. doi:10.1145/3232676
- Geiger, R. S. (2016). Bot-based collective blocklists in Twitter: The counterpublic moderation of harassment in a networked public space. *Information, Communication & Society*, 19(6), 787–803. doi:10.1080/1369118X.2016.1153700
- Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. London, UK: Yale University Press.
- Golbeck, J., Gnanasekaran, R. K., Gunasekaran, R. R., Hoffman, K. M., Hottle, J., Jienjiltert, V., . . . Gergory, Q. (2017). A large labeled corpus for online harassment research. In *Proceedings of the 2017 ACM on Web Science Conference—WebSci '17* (pp. 229–233). New York, NY: ACM.
- Gospelmente. (n.d). *Home* [Facebook page]. Facebook. Retrieved January 17, 2019 from <https://www.facebook.com/gospelmente/photos/a.279205512183446/2027742767329703/>
- Guedes, V. L. S., & Santos, M. J. V. C. (2013). Recorrência de nominalizações deverbais em resumos de cartas científicas em língua Portuguesa e a indexação temática [Recurrence of declarative nominalizations in abstracts of scientific letters in Portuguese and thematic indexing]. *Linguística*, 29(1), 37–57. Retrieved from [http://www.scielo.edu.uy/scielo.php?pid=S2079-312X2013000100003&script=sci\\_arttext&tlng=en](http://www.scielo.edu.uy/scielo.php?pid=S2079-312X2013000100003&script=sci_arttext&tlng=en)
- Hashmi, S. S., Ikram, M., & Kaafar, M. A. (2019). A longitudinal analysis of online ad-blocking blacklists. *IEEE 44th LCN Symposium on Emerging Topics in Networking (LCN Symposium)*, 1, 158–165, doi:10.1109/LCNSymposium47956.2019.9000671
- Jhaver, S., Ghoshal, S., Bruckman, A., & Gilbert, E. (2018). Online harassment and content moderation: The case of blocklists. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 25(2), 1–33. doi:10.1145/3185593
- Jiang, J., & Zhai, C. (2007). An empirical study of tokenization strategies for biomedical information retrieval. *Information Retrieval*, 10(4–5), 341–363. doi:10.1007/s10791-007-9027-7
- Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihók, G., & Den Hartog, D. N. (2017). Text mining in organizational research. *Organizational Research Methods*, 21(3), 733–765. doi:10.1177/1094428117722619

- Kwok, I., & Wang, Y. (2013). Locate the hate: Detecting tweets against Blacks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 27(1), 1621–1622.
- Lee, R. (2019). Extreme speech| Extreme speech in Myanmar: The role of state media in the Rohingya forced migration crisis. *International Journal of Communication*, 13, 3203–3224.
- Lionço, T. (2017). Psychology, democracy and laicity in times of religious fundamentalism in Brazil. *Psicologia: Ciência e Profissão*, 37(Special), 208–233. doi:10.1590/1982-3703160002017
- Lopes-Silva, L. R. (2017, February 2). *Discurso de ódio no Facebook: A construção da incivilidade e do desrespeito nas fan-pages dos parlamentares Jair Bolsonaro, Marco Feliciano e Rogério Peninha* [Hate speech on Facebook: The construction of incivility and disrespect in the fan-pages of parliamentarians Jair Bolsonaro, Marco Feliciano and Rogério Peninha]. InfoMedia. Retrieved from <https://bdc.c3sl.ufpr.br/handle/123456789/110>
- Ma, J., Saul, L. K., Savage, S., & Voelker, G. M. (2011). Learning to detect malicious URLs. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3), 1–24. doi:10.1145/1961189.1961202
- Marco Feliciano. (n.d.). In *Facebook* [Facebook page]. Retrieved February 16, 2017 from <http://www.facebook.com/PastorMarcoFeliciano>
- Maréchal, N. (2016). When bots tweet: Toward a normative framework for bots on social networking sites. *International Journal of Communication*, 10, 5022–5031.
- Matsubara, E. T., Martins, C. A., & Monard, M. C. (2003). Pretext: Uma ferramenta para pré-processamento de textos utilizando a abordagem bag-of-words [Pretext: A tool for preprocessing texts using the bag-of-words approach]. *Technical Report*, 209(4), 10–11. Retrieved from [https://repositorio.usp.br/directbitstream/1bd81b2c-853b-44f8-8a9e-ae2fd4bb0dab/BIBLIOTECA\\_113\\_RT\\_209.pdf](https://repositorio.usp.br/directbitstream/1bd81b2c-853b-44f8-8a9e-ae2fd4bb0dab/BIBLIOTECA_113_RT_209.pdf)
- Natividade, M., & Nagamine, R (2016). Entre hospitalidade e hostilidade [Between hospitality and hostility]. In M. Natividade (Ed.), *Margens da política: Estado, direitos sexuais e religiões* [Margins of politics: State, sexual rights and religions] (pp. 187–241). Rio de Janeiro, Brazil: Garamond.
- Neuendorf, K. A. (2017). *The content analysis guidebook*. Newbury Park, CA: SAGE Publications.
- Nobata, C., Tetreault, J., Thomas, A., Mehdad, Y., & Chang, Y. (2016). Abusive language detection in online user content. In *25th International Conference on World Wide Web* (pp. 145–153). Montréal, Canada: ACM.
- Nova Bíblia Pastoral*. (2021). Nova bíblia pastoral [New pastoral bible]. São Paulo, Brazil: Paulus Editora.

- Roberts, S. T. (2019). *Behind the screen: Content moderation in the shadows of social media*. London, UK: Yale University Press.
- Samaratunge, S., & Hattotuwa, S. (2014). *Liking violence: A study of hate speech on Facebook in Sri Lanka*. Colombo, Sri Lanka: Center for Policy Alternatives.
- Santos, A. D. S. M., Simões, M., & Salaroli, T. P. P. (2017). O retrato da intolerância religiosa no Brasil e os meios de combatê-la [The portrait of religious intolerance in Brazil and the means to combat it]. *UNITAS-Revista Eletrônica de Teologia e Ciências das Religiões*, 5(2), 411–430. doi:10.35521/unitas.v5i2.570
- Santos, M. J. V. C. (2009, December 22). Correspondência científica de Bertha Lutz: Um estudo de aplicação da lei de Zipf e ponto de transição de Goffman em um arquivo pessoal [Bertha Lutz's scientific correspondence: A study of Zipf's law enforcement and Goffman's transition point in a personal archive]. *PontodeAcesso*, 3(3), 317–326. Retrieved from <https://periodicos.ufba.br/index.php/revistaici/article/view/3628>
- Schäfer, G., Leivas, P. G. C., & Santos, R. H. D. (2015). Discurso de ódio: Da abordagem conceitual ao discurso parlamentar [Hate speech: From the conceptual approach to parliamentary discourse]. *Revista de Informação Legislativa-RIL*, 52, 143–158. Retrieved from [https://www12.senado.leg.br/ril/edicoes/52/207/ril\\_v52\\_n207\\_p143](https://www12.senado.leg.br/ril/edicoes/52/207/ril_v52_n207_p143)
- Silva, L. R. L., Botelho-Francisco, R. E., Oliveira, A. A. de, & Pontes, V. R. (2019). A gestão do discurso de ódio nas plataformas de redes sociais digitais: Um comparativo entre Facebook, Twitter e YouTube [The management of hate speech on the platforms of digital social networks: A comparison between Facebook, Twitter and YouTube]. *Revista Ibero-Americana de Ciência da Informação*, 12(2), 470–492. doi:10.26512/rici.v12.n2.2019.22025
- Silva, L. R. L., & Sampaio, R. C. (2018). Impeachment, Facebook e discurso de ódio: A incivilidade e o desrespeito nas fanpages das senadoras da república [Impeachment, Facebook and hate speech: Incivility and disrespect in the fanpages of senators of the republic]. *Esferas*, 1(10), 95–107. dx.doi:10.31501/esf.v1i10.9334
- Sood, S. O., Antin, J., & Churchill, E. (2012). Using crowdsourcing to improve profanity detection [White paper]. In *Association for the Advancement of Artificial Intelligence* (pp. 69–74). Palo Alto, CA: AAAI.
- Tamayo, J. J. (2019). Apologia do diálogo perante os fundamentalismos. In B. Sousa Santos & B. S. Martins (Eds.), *O pluriverso dos direitos humanos: A diversidade das lutas pela dignidade* [The pluriverse of human rights: The diversity of struggles for dignity]. Belo Horizonte, Brazil: Autêntica.

- Tsuria, R., & Yadlin-Segal, A. (2021). Digital religion and global media: Flows, communities, and radicalizations. In S. J. A. Ward (Ed.), *Handbook of global media ethics* (pp. 157–175). New York, NY: Springer.
- United Nations for Education, Scientific and Culture Organization (UNESCO). (2017). *Preventing violent extremism through education: A guide for policy makers*. Paris, France: Author.
- Vieira, H. (2018). Fundamentalismo e extremismo não esgotam experiência do sagrado nas religiões [Fundamentalism and extremism do not exhaust the experience of the sacred in religions]. In E. S. Gallego (Ed.), *O ódio como política: A reinvenção das direitas no Brasil* [Hate as politics: The reinvention of the rights in Brazil] (pp. 206–220). São Paulo, Brazil: Boitempo Editorial.
- Wall Street Journal staff. (2021, September 29). Facebook's documents about Instagram and teens. *The Wall Street Journal*. Retrieved from <https://www.wsj.com/articles/facebook-documents-instagram-teens-11632953840?page=1>
- Wang, S. I., & Manning, C. D. (2012). Baselines and bigrams: Simple, good sentiment and topic classification. In *50th Annual Meeting of the Association for Computational Linguistics* (pp. 90–94). Jeju Island, Korea: ACL.
- Westergaard, D., Stærfeldt, H. H., Tønsberg, C., Jensen, L. J., & Brunak, S. (2018). A comprehensive and quantitative comparison of text-mining in 15 million full-text articles versus their corresponding abstracts. *PLoS Computational Biology*, *14*(2), 1–16. doi:10.1371/journal.pcbi.1005962
- Wheatley, D., & Vatnoey, E. (2020). Understanding attitudes towards social media segregation: Spatial metaphors in the discussion of Twitter blocklists. *Information, Communication & Society*, *25*(1), 1–16. doi:10.1080/1369118X.2020.1749696