UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Exatas
Programa de Pós-Graduação em Ciência da Computação

Samuel da Silva Guimarães

Characterizing Reactions and Comments Associated with News on Facebook

Belo Horizonte
2020

Samuel da Silva Guimarães

Characterizing Reactions and Comments Associated with News on Facebook

Final Version

Thesis presented to the Graduate Program in Computer Science of the Federal University of Minas Gerais in partial fulfillment of the requirements for the degree of Master in Computer Science.

Advisor: Fabrício Benevenuto de Souza

Belo Horizonte
2020

Guimarães, Samuel da Silva.

G963c Caracterizando reações e comentários associados a notícias no facebook [recurso eletrônico] / Samuel da Silva Guimarães. – 2020.
1 recurso online (98 f. il, color.): pdf.il.il.

Orientador: Fabrício Benevenuto de Souza.

Dissertação (mestrado) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Ciência da Computação.
Referências: f. 77-83.

1. Computação – Teses. 2. Mídia – Aspectos políticos – Teses. 3. Mídia social – Teses. 4. Desinformação – Teses. 5. Computação social – Teses. 6. Aprendizado de máquina– Teses. I. Souza, Fabrício Benevenuto de. II. Universidade Federal de Minas Gerais; Instituto de Ciências Exatas, Departamento de Ciência da Computação. III. Título.

CDU 519.6*04(043)

Ficha catalográfica elaborada pela bibliotecária Belkiz Inez Rezende Costa
CRB 6/1510 Universidade Federal de Minas Gerais - ICEx

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

FOLHA DE APROVAÇÃO

Caracterizando Reações e Comentários Associados a Notícias no Facebook

SAMUEL DA SILVA GUIMARÃES

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:

Prof. Fabrício Benevenuto de Souza - Orientador
Departamento de Ciência da Computação - UFMG

Prof. Eduardo Freire Nakamura
Instituto de Computação - UFAM

Prof. Pedro Olmo Stancioli Vaz de Melo
Departamento de Ciência da Computação - UFMG

Prof. Fabrício Murai Ferreira
Departamento de Ciência da Computação - UFMG

Belo Horizonte, 11 de Dezembro de 2020.

To all who helped me, and to all the people who can be helped by this work.

Acknowledgments

To start, I am more than required to thank my parents, who gave me life and every opportunity to improve myself. I also have appreciated every help, advice, and support from everyone in our Locus Lab. In particular, my favorite researchers Júlio Reis, the two, differently spelled, Filipe Ribeiro and Philipe Mello, besides Lucas Lima, and of course my fellow master student Daniel Kansaon.

Outside of this amazing research team, I have also shared my grievances and relied a lot on Keyla Rocha, Vanessa Vasconcelos, Matthias Novaes, Jucélia Nascimento, and Simone the cat, all who have my deepest gratitude.

And to the person whom, without him, even this short text would not be here, my advisor Fabrício Benevenuto, my gratefulness to you is very high. You lead me through failures and successes, which eventually brought me to become a better and more accomplished professional than ever before.

Lastly, I would like to thank the other professors and staff from the Federal University of Minas Gerais, particularly from the Department of Computer Science, for keeping everything working despite these troubling times and being always there for me too. To anyone also reading this and that might help me in the future, thank you very much.

Resumo

À medida que os usuários da Internet adotam cada vez mais sites de mídia social como forma de se informar sobre as notícias, eles encontram um número cada vez maior de opções de veículos de mídia para sua dieta de notícias. Esse número cresceu, em particular, neste ambiente onde qualquer pessoa pode se cadastrar como uma página do Facebook que se apresenta como uma fonte de notícias. Hoje encontramos milhares de páginas do Facebook categorizadas como algum tipo de página de notícias. Essa situação impulsionou o chamado jornalismo independente, também conhecido como mídia alternativa. No entanto, entre esses veículos, podemos ter desde um jornalista independente cuidadoso apresentando objetivamente informações confiáveis até um grupo político que age de má-fé para mudar a opinião pública e promover suas ideias e objetivos. As ações deles podem até provocar diferentes tipos de respostas odiosas por parte de seu público, incluindo conteúdo ofensivo e discurso de ódio. Esse ódio então pode se espalhar para muitas outras seções de comentários em várias páginas. Portanto, identificar e caracterizar todas as páginas de notícias que desempenham um papel vital na disseminação da informação é essencial para a compreensão desse ecossistema de jornalismo e mídia em um país. O objetivo desta dissertação é fornecer um diagnóstico detalhado das notícias e opiniões políticas, ambas compartilhadas dentro do Facebook, especialmente focado no impacto das mídias alternativas. Para isso, apresentamos as seguintes contribuições: 1) criação e validação de uma metodologia para identificar e medir o viés político das páginas do Facebook para um determinado país, e 2) uma caracterização aprofundada do viés político, demografia do público, reações nas postagens e toxicidade dos comentários de uma grande amostra dos três tipos de páginas citados: mídia tradicional, mídia alternativa e figuras públicas, nas páginas brasileiras do Facebook.

Palavras-chave: Mídia alternativa, Viés político, Conteúdo tóxico, Ecossistema de notícias, Computação social, Aprendizado de máquina.

Abstract

As Internet users increasingly adopt social media sites as a way to learn about the news, they encounter an increasing number of media outlet choices for their news diet. That number has grown, in particular, in this environment where anyone can register as a Facebook page claiming to be a news source. Today we find thousands of Facebook pages categorized as some form of news media outlet. This situation boosted the so-called independent journalism, also known as alternative news media. However, among these outlets, we can have from a diligent independent journalist objectively presenting reliable information to a political group acting in bad faith to shape public opinion and promote their ideas and goals. Their actions can even provoke different kinds of hateful responses from their audience, including offensive content and hate speech. This hatred can then spread to many other comment sections on multiple pages. Therefore, identifying and characterizing all the news pages that play a vital role in the dissemination of information is essential for understanding this ecosystem of journalism and media in a country. The goal of this dissertation is to provide a detailed diagnostic of news stories and political opinions shared inside Facebook, primarily focused on the impact of alternative media. To accomplish this, we present the following contributions: 1) creation and validation of a methodology to identify and measure the political bias of Facebook pages for a given country, and 2) an in-depth characterization of the political bias, public demographics, reactions in posts and comment toxicity of a large sample of the three actors cited: mainstream media, alternative media, and public figures, in Brazilian Facebook pages.

Keywords: Alternative media, Political bias, Toxic content, News ecosystem, Social computing, Machine learning.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

Today, social media sites like Facebook and Twitter are popular destinations for users to find, share, and discuss real-time news about the world around them. Recent surveys estimate that 68% of U.S. adults [55] and 66% of Brazilians [44] consume news primarily from social media sites. Furthermore, since 2018 social media sites have passed print newspapers as a news source for North Americans [54]. Meanwhile, this surpassing happened in Brazil since 2014 [44].

These social media sites brought new features that helped the promotion of this kind of online environment as a place for people to share news. For example, anyone can register as a news producer on a social media site, creating a Facebook page claiming to be a news outlet. This situation creates a shift in how news is consumed and produced, lowering the barrier to entry, and therefore promoting independent journalism, with a good example being the citizen journalism from the Arab spring [52]. That shift was indirectly measured by a recent work that counted 20,448 self-reported pages of U.S. news outlets located on Facebook [47].

This independent journalism, also called **alternative news media**, still generates some debates about firm definitions, at times challenging the definition of journalism [43]. One article conceptualized key dimensions where this journalism differs from the so-called **mainstream media** as *the producers*, *the content*, *the media organizations* formed, and *the media systems* where it lives [30]. A unique type of these news *producers* are **public figures** and political entities that can also replace traditional news, similar to U.S. President Donald Trump's Twitter use. Alternative media are gaining considerable space and power in the last years, sometimes using **political bias** as fuel [33, 44].

Therefore, understanding how each one of these types of pages is crucial to comprehending the news ecosystem on social media for different reasons. Firstly, alternative media pages can influence public opinion by composing the news diet of a large following. Serious independent journalists can create these pages, but political powers can act in bad faith and use outlets they conceive to shape public discourse to disseminate their talking points or attack political enemies [29, 33]. These circumstances make the mapping of these pages and the discovery of their political stance crucial to a more transparent environment.

Based on that conclusion, our main contribution is presenting a methodology capable of identifying Facebook pages for a given country on the three types cited, mainstream media, alternative media, and public figures, and also measure the political bias of these pages. We additionally characterize the demographics of the pages' audience, presenting an overview of the media ecosystem.

Beyond that, social media sites also offer a novel dissemination strategy, in which users help to share news pieces in an attempt to influence their friends [4, 67]. Besides sharing, these sites provide ways for users to interact and discuss with each other through comments sections, and among people that comment on news, a survey found that 77.9% of them comment on social media [59]. Among these messages, a significant proportion of toxic comments flood these news posts [60, 28], similarly to comments on news websites, which are mostly negative independently of the headline [45].

Our second contribution is an extensive diagnostic about the toxicity in the comments associated with the news posts shared on Facebook. Besides messages, Facebook post reactions, like Haha and Angry, can also express the negativity and anger from the users, being more descriptive than regular likes. Considering that, we also analyze these reactions for each post, and compare them to the results of toxic comments, presenting the possible correlations.

As a case study of our methodology's use, we analyze the ecosystem of Facebook in Brazil, which has a history of interaction between alternative media and politics.

In resume, this dissertation aims at filling this research gap, presenting a twofold contribution: 1) creation and validation of a methodology to identify and measure the political bias of Facebook pages for a given country, and 2) an in-depth characterization of the ideology, public demographics, reactions in posts and comment toxicity of a large sample of the three actors cited: mainstream media, alternative media, and public figures, in Brazilian Facebook pages. In the next section, we present the main motivations of our work.

1.1 Motivation

Understanding alternative media pages and their audience's interactions between themselves and with the content can help us find trends in how they affect different communities and assess their societal impacts. Despite their importance, studying alternative media on social media platforms is still challenging as it requires extensive manual efforts in identifying them in the first place. For example, while Facebook is the most used platform for news reading in Brazil, studying its ecosystem is limited, with case studies of existing groups predominating [41, 44]. Facebook is, in fact, the largest social network worldwide, with 2.50 billion monthly active users as of December 2019 [18], and with 67% of U.S. Facebook users getting their news there [55]. For Brazil, this statistic is 54% [44]. However, despite the importance of this part of our current news ecosystem, we still know little about the distribution of these comments, especially in Brazil. In this country, several alternative media outlets have emerged during both left and right-wing governments, mostly mixing activism and reporting. Three of the four last presidents, former Presidents Luis Inácio Lula da Silva and Dilma Rousseff, and current President Jair Bolsonaro, had some media problems. More recently, in 2018, Facebook dismantled what they called a coordinated web of disinformation, composed of 196 pages and 87 personal profiles [27]. The owners of these fake accounts hid the nature and origin of their publications, intending to propagate division and disinformation. In the same year, Facebook discovered that a Brazilian marketing group, Raposo Fernandes Associados (RFA), with another 68 pages and 43 accounts associated with it, was violating the social media network's misrepresentation and spam policies [63], with more user engagement than some celebrities [64]. As the number and influence of these alternative media outlets grow, questions about their political accountability and compromise with the truth emerge.

1.2 Goals

The goal of this project is to provide a diagnostic about news stories and political opinions, both shared inside Facebook, especially looking into mainstream media, alternative media, and public figures' impact, bias, and toxicity present in their publications. In particular, we want to provide answers to the following questions:

1. *What is the distribution of the mainstream media, alternative media, and public figures pages on Facebook?*

2. *How the characteristics of the pages from Alternative Media compare to the pages of Mainstream Media and Public Figures?*

3. *Which factors affect the proportion of toxic comments on some of these pages?*

4. *What are the common characteristics of these toxic comments?*

1.3 Contributions

With our goals in mind, this dissertation presents the following five tasks which compose our contributions:

1. Identify a sizable list of Facebook pages related to news and politics, classifying them as one of the three studied page types;
2. Characterize Public Figures, Mainstream Media and Alternative News Media Pages on Facebook with respect to their audience and bias;
3. Measure the toxicity of a large sample of Facebook comments from those pages;
4. Create profiles to those pages considering the reactions to a large group of Facebook posts from them;
5. Analyze the comments' relation to the other information from the pages and posts that they respond.

The execution of the first task of our work was gradual. Starting from the strategy from a recent related work [47], we first exploit the Facebook Marketing API to find pages based on a list of interests related to politics. Then, we search pages based on a list of media pages from previous work that studied the division of media based on political sides. Finally, we followed the same process with the Marketing API, but now considering a bigger set of types of pages.

Next, for the second task, we adapted previous work by using the Marketing API to gather both demographic information and indirect data of the audience shared by different pages and interests related to them. This data allowed us to create a method to calculate bias based on a previous Twitter focused methodology of using graphs and label propagation.

For the third task, we mainly relied on Google's Perspective API to label the posts and comments after a simple cleaning of the text. Then, we analyzed the reliability of this method, statistically confirming that the concept of toxicity understood by human volunteers was how the API measured it.

In the fourth task, with the data from all posts from the most comprehensive set of pages we could use, we created reaction profiles from the clustering of the posts by the similarity of the proportion of each possible reaction. With these profiles, we can measure how these reactions can show how differently the audience of various page types react to their content and how this other possible form of negative interaction affect the toxicity of the comments.

Finally, for our fifth task, we gather all the information produced in the other tasks and analyzing the context, we were able to draw some statistically significant conclusions on what factors correlate with the toxicity in the comment section of news media on Facebook. The results presented for this dissertation are part of the following two papers, in chronological order of publication:

- **Guimarães, S. S.**, Reis, J. C. S., Ribeiro, F. N., and Benevenuto, F. (2020). Characterizing Toxicity on Facebook Comments in Brazil. In Proceedings of the 26th Brazilian Symposium on Multimedia and the Web (WebMedia);

- **Guimarães, S. S.**, Reis, J. C. S., Lima, L., Ribeiro, F. N., Vasconcelos, M., An, J., Kwak, H., and Benevenuto, F. (2020). Identifying and Characterizing Alternative News Media on Facebook. In Proceedings of the 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).

1.4 Dissertation Organization

The rest of the dissertation is organized as follows.

- **Chapter 2 - Background.** Chapter 2 presents the main concepts related to this work, which includes the definition of alternative media the overlapping descriptions of online hate speech, online harassment, cyberbullying, toxicity and incivility, and finally, explain the research of the bias present on online news.

- **Chapter 3 - Related Work.** Chapter 3 presents a brief review of the many methods and datasets in the literature for political bias measurement. Additionally, the chapter also describes in more depth the main related works that were crucial to help the creation of our methodology.

- **Chapter 4 - Methodology and Datasets.** Chapter 4 presents the methodology used to accomplish our four tasks described in the previous section. In particular, it describes how we use two Facebook APIs to find an extensive list of Facebook pages, how we identify them as public figures, alternative, or mainstream media, along with how our new methodology calculates political bias based on Facebook data. We introduce the list of Brazilian Facebook pages found by each iteration of our methodology, pointing out the differences and showing the details of our filtering of the pages to reach the final datasets. Besides that, we also detail the steps we took to gather the comments and audience demographics from the pages we found, along with the statistical methodology taken to analyze the demographic difference.

- **Chapter 5 - Characterizing Alternative Media.** In Chapter 5, we characterize the age, gender, relationship status, and education of the audiences of the different types of pages, also discussing the results of our method to measure political bias and the distribution of the bias in those pages.

- **Chapter 6 - Toxicity and Reactions in Alternative Media.** Inside Chapter 6, we present the results from collecting Facebook posts and comments from the pages we found, and then the results of measuring the toxicity of those using the Perspective API. We show how the toxic content in the post and comments relate, also comparing how this content is affected by the characteristics of the pages from where they originated. Along with that analysis, we present the reactions from the users to different posts on different pages, showing how the toxicity from the page's posts can generate a response from the users.

- **Chapter 7 - Concluding Discussion and Future Work.** Finally, Chapter 7 concludes the dissertation by discussing the limitations of this work, how these results can be used, and also the possible directions of future work.

Chapter 2

Background

2.1 Alternative media

As stated before, with the low barrier to entry of news outlets present in social media, independent journalism becomes more popular and gains more followers. This increase in influence creates a shift in how news is consumed and produced. The news' process that started as mediated and somewhat centralized, with professional journalists working for media companies, changed to more dispersed and direct, more personal and audience-centered, possibly with one individual alone able to create and share their opinionated news stories.

This **alternative news media** does not have a clear definition. But, as [26] puts it: "alternative media processes and products have been described as inhabiting - indeed, as being inseparable from - an alternative or plebian public sphere". In other words, alternative media is or at least tries to appear as a more organic and less corporate type of media, in contrast to the traditional journalism of the so-called **mainstream media**. This *alternative* term can be understood as an alternative way that this new type of media operates.

Some recent examples of the impact of alternative media are the role of social media in the Arab spring [52], immigration-critical alternative media gaining reaching one-tenth of Sweden [29] and the use of Twitter by the U.S. President, Donald Trump [36]. Recently, the impact of the alternative media has been examined [33, 29]. The ecosystem of alternative right-wing outlets in the USA [33] creates an asymmetry in the perception of world affairs, being particularly relevant in the U.S. election of 2016. During this election, Russian troll farms spread propaganda on social media through fake news and fear-mongering, resulting in a lot of hate on the platform.

In Sweden and Germany, right-wing movements share the anti-system rhetoric with the alternative media to make their audiences participate in voting. In those cases, Immigration-critical alternative media gained a significant reach, and a word for the traditional media was coined in German, "Lügenpresse," or the Lying Press [29].

These examples across the countries show that politicians and political movements purposely use alternative media to amplify their message, influencing their followers to distrust the established media landscape, and trust the new form of journalism. Sometimes these alternative media outlets return significant media support to the politicians.

Looking to measure the impact of this phenomenon in Brazil, we looked for a working definition of alternative media and divided public figures pages as a separated division. To accomplish this classification of the pages between public figures, mainstream media, or alternative media, we use the definition proposed by [30] cited before, especially applying the dimensions of *producers* and *media organizations*. Therefore, we consider a page as alternative media if it does not represent an outlet registered in any official press organizations. That is, it is alternative media if it cannot be confirmed as a mainstream outlet or public figure. Finally, we characterize these pages by generating an ideological bias score based on graph-based semi-supervised learning.

2.2 Toxicity and other related metrics

In our analysis of this situation, we also investigate the impact of the pages of news outlets and politics-related pages in the propagation of hatred and insults on Facebook. Because this problem manifests itself in many degrees, including incivility, offensive content, toxicity, and hate speech, we now present the background of this research area and explain the diverse definitions of each type of hatred, and its relations.

Knowing that some countries now have laws intended to reduce the amount of hate speech in online media, several studies created after that started to provide a better understanding of this content on the Internet.

Davidson et al. [13] build a dataset primarily focused in the distinction between hateful and offensive language. The main difference of hate speech, according to the authors is the derogatory aspect of using obscene language. The authors define hate speech as “language that is used to expresses hatred towards a targeted group or is intended to be derogatory, to humiliate, or to insult the members of the group”. To properly separate the definitions they give, they used three labels: hateful, offensive and non-offensive. They built the corpora by using the Twitter API to search for tweets containing terms of a keyword list of hate terms.

Golbeck et al. [25] has a specific division for hate speech, labelling a tweet as hate speech if it “express hate or extreme bias to a particular group”. The hate can be about the religion, race, gender, sexual orientation, among others. In general, the targeted groups are defined by their *inherent attributes*, not by their *actions or thoughts*.

They propose many labels on their work besides hate speech, including direct harassment, threats, and potentially offensive. However, the authors also have created a version of their data with only the labels of harassment or non-harassment.

This vision of hate speech that focus on the existence of a target is also shared by [57] and [40]. In both cases, the authors provide a deeper understanding of the hateful messages exchanged in social networks, studying the most common hate speech targets in these systems. Meanwhile, [53] created a taxonomy for hate speech, dividing hate into accusations, promotion of violence, humiliation, and swearing. Then, they did some experiments using it to see how different features and different machine learning algorithms can influence the results of the hatefulness classification of text.

Founta et al. [21] follows some of the aspects of the other previous definitions, being influenced especially by [13], and specifies that hate speech is "Language used to express hatred towards a targeted individual or group, or is intended to be derogatory, to humiliate, or to insult the members of the group, on the basis of attributes such as race, religion, ethnic origin, sexual orientation, disability, or gender". They used CrowdFlower to label instances in a large set of labels, which are Offensive, Abusive, Hateful, Aggressive, Cyberbullying, Spam, or Normal.

Fortuna and Nunes [20] noted that these studies used similar, but not identical, descriptions of what is hateful and how it is different from offensive content. In its survey, they found that the definition generally covers the presence of some aspects:

1. Hate speech has specific targets;
2. Hate speech is to incite violence or hate;
3. Hate speech is to attack or diminish;
4. Humour can be used to some degree.

Not surprisingly, many recent research efforts have attempted to operationalize the concept of hate speech, and other overlapping behaviors, by defining them in terms of measurable factors to be able to identify and counter them. The key challenge for that is that, even in our society, there is not a unanimously accepted definition of hate speech. So, the strategy to operationalize hate speech has been usually a data-driven approach in which humans are recruited to label a predefined set of messages as hate speech or non-hate speech. An effort by [50] showed that even human annotators provided with hate speech definitions do not lead to highly reliable corpora, indicating that this task is difficult even for humans.

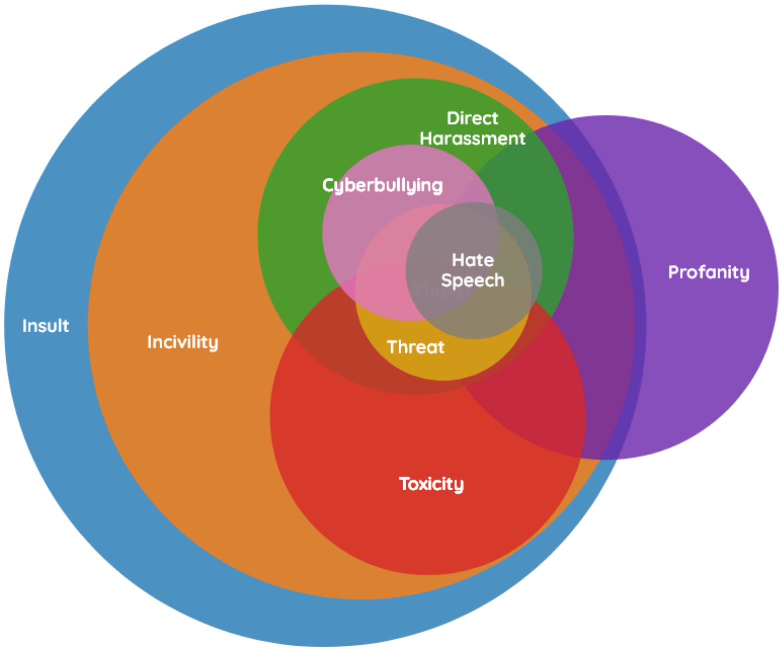
Evaluating this various researches, we see that in an attempt to detect the very harmful problem of hate speech, they defined a series of slightly less harmful, or more specific, forms of hatred.

Concept	Description
Cyberbullying	Hostile online behavior performed individually or by a group, repeatedly, with the use of force, threats, or coercion to abuse, embarrass, intimidate, or aggressively dominate others [21].
Direct Harassment	Language directed at a specific person or group designed to upset, intimidate, or threat the target [25].
Hate Speech	Public speech that expresses hate or encourages violence towards a person or group based on something such as race, religion, sex, or sexual orientation [8].
Incivility	Mild uncivil messages are impolite or hostile language containing expressions describing or attempting to offend others, but without any explicit name-calling or curse words. Extreme uncivil messages uses profane language with the explicit intention to threat or attack someone [60].
Insult	Insulting, inflammatory, or negative comment towards a person or a group [31].
Profanity	Swear words, curse words, or other obscene or profane language [31]. Offensive or obscene word or phrase [7].
Threat	Describes an intention to inflict pain, injury, or violence against an individual or group [31]. The language used has the intention to create a unsafe feeling, or make the target fearful [21].
Toxic language	Rude, disrespectful, or unreasonable comment that is likely to make people leave a discussion [31].

Table 2.1: Comparison between definitions of concepts related to toxicity

More recently, the Jigsaw Group launched the Perspective API [31]. This API uses machine learning models to score the perceived impact a comment might have on a conversation, creating a toxicity index to measure how toxic a message can be perceived by a user. The concept of toxicity presented has an overlap with many of the reviewed subdivisions but focuses on the fact that these types of messages disrupt the online environment, possibly ending the interactions. Inspired by a table included in [20], we grouped the main definitions of the research we analyzed in Table 2.1. Using these descriptions, we also created a Venn diagram of our understanding of how these many concepts overlap, shown in Figure 2.1.

Figure 2.1: Venn Diagram of the concepts related to toxicity



Source: The author.

Based on this, we believe that toxicity includes the most relevant types of hatred found. As hate detection is an open problem, and the hate subdivisions are still not clear enough to produce very consistent results between them, in this work [66], we used the Perspective API, one of the models that are considered state-of-the-art.

2.3 Media Bias and Political Bias

Previous work in the area of journalism has discussed different kinds of biases present in the media, with diverse definitions for media bias and its more specific forms.

Mullainathan and Shleifer [42] defined a type of bias they called editorial slant as “the quantity and tone of a newspaper’s candidate coverage as influenced by its editorial position”, and with that definition measured the impact of the media on voters. Meanwhile, D’Alessio and Allen [12] have identified three main media biases: coverage, selection, and statement bias. First, the coverage bias concerns how much attention is paid to a news story by a vehicle. The selection bias represents which stories are presented and which are ignored by journalistic sources. Finally, the statement bias is related to how a fact is reported, possibly adding negative or positive connotations to a news story, creating a more opinionated report. In this sense, the editorial slant of the first work is related to the coverage bias present by this second group of authors.

Entman [17] presents similar divisions of distortion, content, and decision-making bias, which are, respectively, presenting distortions of reality, favoring one political side, and the motivation of journalist who produces biased content. In this case, they are related, in order, to statement, coverage, and selection bias. In a definition more focused on intention, [15] divided media bias into two, with ideological bias attempting to transform the opinions of the reader into a specific set of beliefs, and the second bias, called spin, focusing on simply creating a story that is provoking, sometimes regardless of realism.

Looking at all of these definitions, we see that, in the area of journalism, there are works that subdivide bias by the way it affects the news articles, while others focus on what were the objectives of the insertion of the biased perspectives. As intention is difficult to measure, for our purposes, it is not a useful definition for us to focus on the objectives of bias. However, to evaluate the diverse ways that news outlets transform the content, we need to analyze the content produced. We expect that how much data is examined might interfere with the accuracy of the calculated bias. Processing more information means that it is more computationally costly.

Because of that, our approach is closer to the work of [42], which tries to quantify the way the content was altered, and more importantly, chose to investigate the effects on the people who use the different news outlets in their news diet. Therefore, our strategy is to measure what we will call political bias, which is closer to the editorial slant or ideological bias cited, by analyzing the politics of the audience that each outlet acquired. This strategy of measuring the ideology of the public of news media was already used before in different ways, especially on researches in social computing [38, 3, 47].

Chapter 3

Related Work

We here review related work that was useful for this dissertation along four distinct dimensions: (i) identifying online news pages, (ii) polarization and relationship graphs, (iii) political bias measurement studies, (iv) the use of toxicity in news comments and (v) analysis of posts reactions.

3.1 Identifying online news pages

Some studies working to find online news pages found alternative news outlets by screening the most popular links on Facebook groups and pages [3], while others search for alternative media shared narratives [58]. Also notably, [47] used recommendations from Facebook Marketing API to create a snowball process collecting all of these recommendations for U.S. pages. We have a similar approach using a different tool from the Facebook Ads platform.

3.2 Polarization and relationship graphs

Just as important, researches considering political polarization and relationship graphs mostly focus on Twitter [37, 58]. Conover et al. [10] is one example, having compared different methods for measuring the political alignment of Twitter users, including text, hashtags, and label propagation analysis using both mentions and retweets graphs. This retweet graph-based approach inspired part of our graph approach.

3.3 Political bias measurement studies

For assessing ideological bias for news outlets, we introduce four studies that were compared with our method. First, the Pew Research Center [38] analyzed the audiences from 36 news outlets by interviewing 2,901 people and asking them what media they know, which one they read, their political self-identification, and their trust in the media. This analysis allowed them to make a diagnostic about how different political leanings affected the perception of the news.

Budak et al. [6] used content analysis to identify the overall ideological bias of 15 major U.S. news outlets by compiling 803,146 published stories over an entire year, and later using 749 human judges to classify 10,502 specific political articles. The overall leaning was measured using these articles, and the results showed little difference in the coverage by outlets of different bias, except in scandals.

Bakshy et al. [3] also used Facebook data to calculate the ideology from 500 websites with links shared on the platform. It used the ideological affiliations from 10.1 million users that declared their bias to classify 226,000 URLs from an initial seven million shared by them over six months. They showed that the content on social media could cross ideological lines and reach people from the opposite perspective.

Finally, Ribeiro et al. [47] used the Facebook Marketing API[1] to get information on the proportion of users identified within different parts of the political spectrum, then calculated a bias score for 20,448 American media outlets. They provide a demographic analysis of the U.S. audience, especially using the demographic division of the users in Very Conservative, Conservative, Moderate, Liberal, and Very Liberal, to generate the bias score. However, their approach cannot be exploited for countries other than the U.S. because the political leaning of the pages' audience is not readily available.

3.4 Toxicity in news comments

In the topic of hateful and toxic speech, some of the recent researches used a grading of how much hate there is on online messages. Others have discussed how to detect the hate, and a few analyzed the general picture, which can be getting worse due to the political polarization reaching an apparent peak [22].

[1] developers.facebook.com/docs/marketing-apis

Between works measuring comments, [60] found that impoliteness and incivility were more prevalent on comments on Facebook pages from conservative and local news sites, with around 20% to 40% of user comments on those news stories consisting of uncivil comments. Similarly, [45] showed that comment sections from influential newspapers are becoming home for hostility and trolls by collecting messages from news websites and using sentiment analysis to confirm this trend. We also focus on comments from news pages but complement our study with a comparison to other relevant page categories, such as politicians and political activists.

Few articles try to create or use databases of Facebook news comments or post reactions like we intend to use. Khan and Chang [34] built a dataset from Facebook data. In their work, the Amazon Facebook page had its posts collected during five years, using the Graph API. After feature extraction on the information, they used different neural networks to predict the number of distinct interactions the posts would receive, based on the post content, the form of content (video, link, picture, and text), and some time-related information. Kolhatkar and Taboada [35] created a dataset focused on the notion of the constructiveness of news comments, evaluating the result with a deep-learning approach. Their concept of constructiveness was that: "Constructive comments intend to create a civil dialogue through remarks that are relevant to the article and not intended to merely provoke an emotional response. They are typically targeted at specific points and supported by appropriate evidence." With the dataset, the relationship of the toxicity with constructiveness was analyzed using the Perspective API. For Brazilian data, [14] created the dataset *OffComBr* using one of the biggest news websites in Brazil, called G1. Their focus was on creating a database that could be used in machine learning, labeling 1,250 comments from the website commenting section.

3.5 Analysis of posts reactions

Meanwhile, dealing with the reactions to the posts, a few articles were also relevant to our project. First, Tian et al. [62] analyses 21,000 Facebook posts from 15 media pages from four different countries: UK, USA, France, and Germany. After collecting approximately 57 million reactions and 8 million comments, they test the relation of emojis in these comments to reaction profiles of pages, using emoji sentiment scores from a previous work to calculate the sentiment of the message. They created reaction profiles of the pages using the K-means clustering algorithm after transforming the raw number of each reaction to a proportion of the total of all the posts received.

Between their results, they conclude that Facebook comments, and especially for us, the reactions to posts may work as an indication of user emotional state and possible attitudes. Because of that, we use the same strategy to compare these reactions to the toxicity score of our comments.

Another related work that was directly interested in confirming with users that their use of emoticons and reactions was able to express the emotions they desired was [24]. The authors invited several Facebook users to participate in the study and respond to a questionnaire. This questionnaire showed 24 posts prejudged as the most relevant to their inquiries and asked users to indicate which Facebook reaction they would give each one and the other two questions to show how they feel about the publication. They then use the clustering algorithm Expectation-Maximization to group the responses and analyze the relationship between the real feelings from the users and used reaction, showing a correspondence between the two.

From a different perspective, some articles deal with the problem of prediction involving reactions. Moers et al. [39] tries to create and evaluate alternative methods for predicting these reactions to posts on supermarket chains pages. Indeed, their final model was able to predict the reaction distribution with a Mean Square Error of 0.135, using their dataset of 8,103 posts on the customer service page of 12 US/UK big supermarket chains. In their analysis, they found a problem in using Facebook like that was initially the only reaction, with the early results showing that results become increasingly meaningless when including likes, due to its large predominance in the data. A similar analysis lead us to remove the likes in our study too.

Finally, Basile et al. [5] also tried to predict results using the data from Facebook reactions, using them as proxies for predicting news controversy. They say these controversies are "situations where, even after lengthy interactions, opinions of the involved participants tend to remain unchanged and become more and more polarized towards extreme values". Based on this definition, they hypothesize that a news story has a higher controversy if it has a large number of reactions divided into two or more of the different emotions represented, measuring this by the use of the entropy of these reactions. In their results, they use a simple type of regression to prove the ability of predicting the controversy of a text by using the cited definition. Although they used a small and local dataset, predominantly with content from Italy, we also used this concept of controversy to compare the reactions to the toxicity received by posts.

As the above studies show, calculating the ideology of pages by both the audience or relations between them is useful and accurate. We followed this trend, proposing a new method that can be extrapolated to any country with sufficient adherence to Facebook. A few articles analyzed the Brazilian media ecosystem with minor conclusions about the news consumption in the country [44, 49, 41].

We also see that calculating toxicity based online data is becoming accurate and useful. On the line of these studies, we conduct a large-scale analysis in the Brazilian context, and we focus on Facebook, the largest social media platform used for news sharing. However, to the best of our knowledge, there is no previous work that compared alternative media, mainstream media, and public figures in the current context.

Chapter 4

Methodology and Datasets

In this chapter, we briefly describe the methodology adopted for the analysis, including our strategy to select and group news outlets and political pages on Facebook, along with how we collected reactions and inferred toxicity of comments and posts associated with them. We also present the different datasets we used for each task, showing some details relevant to their use and contrasting them between themselves, besides discussing their limitations.

4.1 Methodology

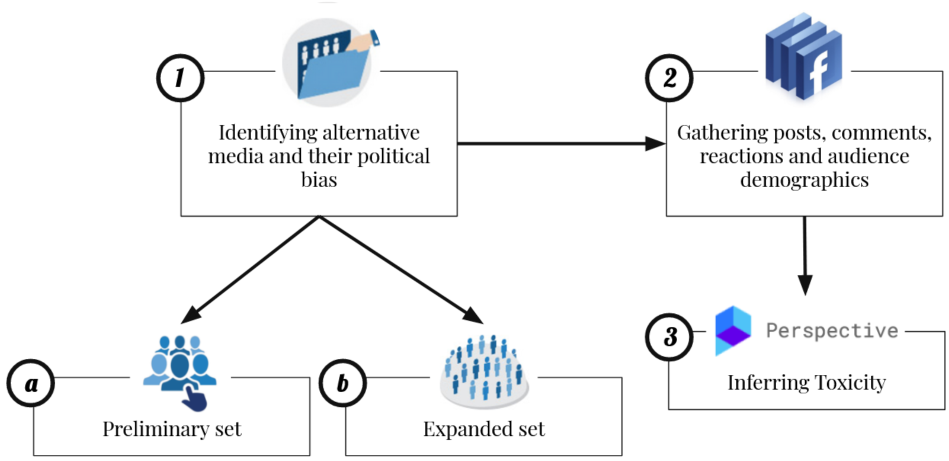
First, we present our methodology. To illustrate the general steps and decisions we executed, Figure 4.1 displays a step-by-step diagram of our actions.

We start finding the pages to study, which we separate into a preliminary dataset, and an extended dataset, following the steps of the development of our strategy to encounter the outlets. Next, we explain how we gathered the different types of data we use. Finally, we give more specific details of how we infer the toxicity from that data.

4.1.1 Identifying alternative media and their political bias (Preliminary Approach)

The first step in our methodology is selecting the News and Politicians Facebook pages to monitor. We start from a list of 22 Facebook pages introduced in [41], which includes Brazilian news outlets from mainstream and alternative media outlets as well as other meta-information such as their political affiliation and reach.

Figure 4.1: The steps of our methodology



Source: The author.

As there were only 14 pages still active, we used the Facebook Audience Insights Tool[1] to expand the initial set.

This tool helps advertisers to refine the audience they want to show an ad, by defining a set of attributes such as age, location, gender, and *interests*. These *interests* are a comprehensive set of topics that Facebook infers, and that represents an individual user concerns and topics that are likely to engage him, including public figures, politicians, political parties, types of food, restaurants, or activities. One of the Audience Insights functions helps to specialize in the desired public by suggesting related topics. So given an input Interest, the tool appoint related Facebook Pages with a similar audience ('Page Likes' menu).

Thus, for the first iteration of our method to search and identify pages, we choose four interests related to the Brazilian politics scenario as seeds and manually searched for related pages. For that, we used the following:

1. **Jair Messias Bolsonaro**, the current right-wing President;
2. **Lula**, the left-wing Former President;
3. **The Social Liberal Party (PSL)**, a Brazilian right-wing political party[2];
4. **The Worker's Party (PT)**, a Brazilian left-wing political party.

From the suggested pages matching to each of the four interests, we included those from the following 13 relevant categories[3]: Public Figures, Politicians, Government

[1]https://www.facebook.com/ads/audience-insights/.
[2]This was Bolsonaro's Party during the search, but he latter left the party.
[3]When a page is created, a pre-defined category can be assigned to that page.

Officials, Authors, Political Organizations, Political Parties, Media, News & Media Websites, Media/News Companies, Broadcasting & Media Production Companies, Magazines, Journalists, TV Programs (News related) and Newspapers.

As a result, we ended up with 63 Brazilian News and Political Facebook pages from various categories. But, we mainly grouped them into two categories: Public figures and Media, which have five subdivisions. These five sub-groups are:

1. Public Figures:
   a) **Right-Wing Figures**;
   b) **Centrist Figures**;
   c) **Left-Wing Figures**.
2. Media:
   a) **Mainstream Media**;
   b) **Alternative Media**.

Since our list includes not only politicians but also political activists, we use the generic term ‘public figure’ instead of ‘politicians’ in our study.

For public figures, we use their claimed political leanings for assigning the subgroup, and we ignore pages if they do not clearly state their political leanings. But in this initial approach, we did not classify media pages by their political bias, only as mainstream media or alternative media, because we still did not have a political bias measuring methodology.

For media, we then classified media as **alternative media** if it was not registered in any of these institutions or related to registered companies, and **mainstream media** if they were. We use data from Brazilian official press organizations like the National Association of Journals (ANJ), National Association of Magazine Editors (ANER), or National Agency of Telecommunications (ANATEL). Alternative Media tends to present online-only, or sometimes they only have Facebook Pages and publish only on Facebook.

4.1.2 Identifying alternative media and their political bias (Snowball Process)

One main challenge we faced throughout our work is the lack of ground-truth political bias for Brazilian media. As in the preliminary version of the method to identify pages, our principal ground-truth group of pages was from the list of 22 Facebook pages introduced in [41].

Inspired by the method proposed by [47], we make use of the tools from the Facebook Ads platform to classify the political leaning of a given Facebook page. First, we use the Facebook Marketing API that allows the creation and management of ads on Facebook by specifying the target audience with attributes such as age, location, gender, and *interests*. Besides that, we also again use the Audience Insights Tool, developing a new method to find pages.

By using these two tools from the Ads platform, we iteratively collect a list of relevant pages as follows:

1. Compile a small number of "seed" pages that were, preferentially, manually curated;
2. Get the associated interests for each page;
3. Use these returned interests to find related pages;
4. Go back to (2) until no new page is suggested.

In step 3, we consider pages that are one of same 13 relevant categories from the previous version.

This new proposed method assesses the political bias of a Facebook page by utilizing audience interaction information. However, note that existing methods are limited to U.S. pages [47], as they collected the available bias information directly from the demographic information in the API. In our proposed methodology, we infer the political leaning from an initial list of pages with self-reported ideology, using the interests graph to propagate the political stance to related pages. Whit this, we expand the possibility of bias measuring to other countries, like Brazil.

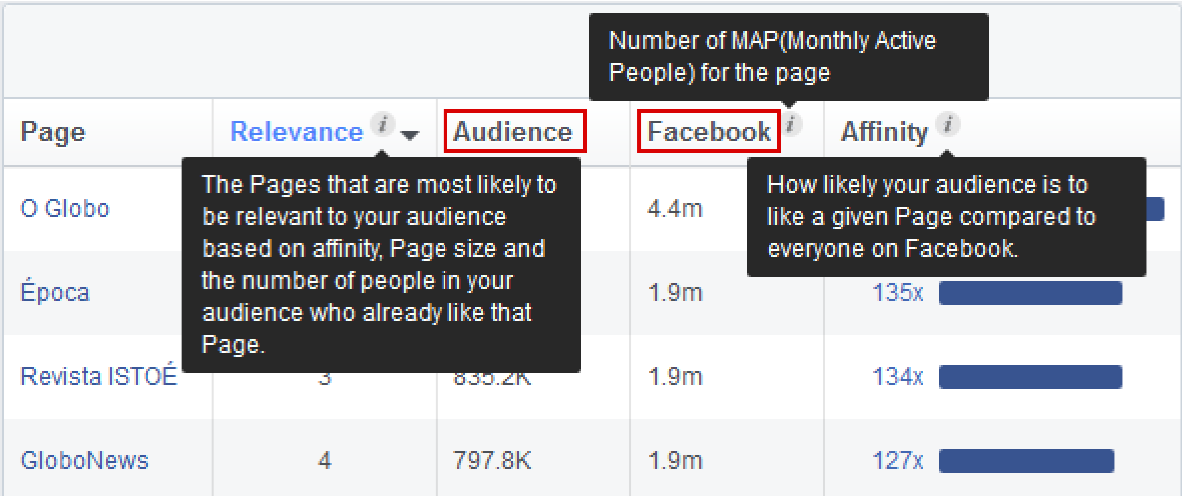
To demonstrate how differently we use the interests, we present a hypothetical scenario. For example, consider two pages with associated interests, a and b. Given an interest in page a, the Audience Insights provides a list of associated pages, including b. For each related page, the tool provides three metrics: Monthly Active People (MAP), Audience, and Affinity score, shown in Figure 4.2. In our example, MAP is the number of monthly active users of page b. The audience is the number of users who are active on page b, given the interest in page a. Then, the affinity score measures how likely a user that is interested in a is to like page b compared to a random user.

While the affinity score allows us to compare related pages found using the same interest in the snowball process, it is not straightforward to compare pages found by searching different interests. We thus propose a new normalized affinity score, $\mathcal{A}$, between pages (e.g., for pages a and b), that is calculated based on the MAP and Audience as:

$$\mathcal{A}_{a,b} = \left(\frac{Audience_{a,b} + Audience_{b,a}}{MAP_a + MAP_b}\right)$$

Basically, $\mathcal{A}$ is the sum of the number of people who are interested in one page and like the other, and vice-versa, divided by the sum of the number of active users of both pages.

Figure 4.2: Facebook Audience Insights tool.



Source: The author.

It is important to note that all the audience of one interest is not equal to the MAP of the related page. For instance, a person may like a fan page of one celebrity but not the official page. That person is still counted as interested in that celebrity.

With this new affinity score, we compute the political leaning. For that, we construct a graph using the score and use a semi-supervised learning (SSL) method to propagate the ideological bias of some known pages to all others in the following steps:

1. For each page found as interest, we calculate our new affinity score $\mathcal{A}$;
2. We create an undirected weighted graph whose nodes are pages and edges are established when one page was found as related to the other on the Audience Insights, with edge weight as the complement of affinity: $w(u, v) = 1 - \mathcal{A}_{u,v}$;
3. We apply the Floyd-Warshall algorithm [19] to find the distance between all pairs of nodes of that graph;
4. We verify which pages from the selected 13 categories can be identified as right-wing or left-wing and then label them as such;
5. We use one graph-based SSL method to classify the remaining pages as left or right, passing the graph as a parameter;
6. We define the ideological leaning as the probability of a page being classified as right-wing minus the likelihood of the page being left-wing, giving a skew between -1 (left) to 1 (right). As we use cross-validation, we take the average of this bias on all folds.

For step (5), we experiment with three existing graph-based SSL algorithms. They are classic label propagation (LP) [68], label propagation with smooth function classes (Smooth LP) [69], and spectral graph transducer (SGT) [32].

As the baseline method, we use the K-nearest neighbors algorithm (KNN) [11] using only the known part of the graph in supervised learning. We perform 10-fold cross-validation and report the area under the ROC curve (AUC) for all instances, using this metric and cross-validation scheme to perform a grid search on the best hyper-parameters. In the next chapter, we present the results of this assessment, showing which algorithm was finally chosen to be used in step (5).

4.1.3 Gathering posts, comments, reactions and demographics

Once we obtain the list of news and politics Facebook pages, we use the Facebook Graph API to collect Facebook posts and comments on the pages of our list. The data from these posts and comments included textual content, the number of likes, published date, and mentions of other Facebook users in the text. We also collect all the replies to comments. Using the API, we found posts ranging from mid-2018 to end-2019 in the pages we selected but, we focused on the posts from the period of October 27th, 2019 to November 16th, 2019 to compose our dataset, circa a week prior and after Lula was released, capturing all posts available in the API.

Page Reaction Profiles

As a complement to the toxicity measure, we also examine the reactions given to the posts by the users. To analyze these reactions, for each publication we collect, individually and grouped by page, we had each of these reactions out of five possible (Love, Haha, Wow, Sad and Angry) normalized by their proportion in the total, creating a profile by page. With these profiles, we did a clustering of page profiles to analyze possible patterns between the political bias and page types. We used the K-means algorithm for this clustering, following tests with the algorithms used in previous work. We analyze these reactions ignoring Likes because it is the immense majority of all interactions and ignoring Care because it was not available in 2019 [39]. All the data was collected in the Facebook CrowdTangle platform [4], searching in the same period and pages we analyze for the toxicity research, but also expanding the data gathering to a longer period and to data from more pages to further improve the characterization of the page types. In total, we gathered all reactions from 2,006,359 posts published in 767 pages from 79 categories over six months, from May 15th, 2019 to November 30th, 2019.

[4] https://www.crowdtangle.com

Reaction Entropy as Controversy measure

Following one previous work [5], we measure the entropy of the reactions as a proxy to how controversial one post is. In this metric, we use the concept of entropy from information theory, with the formula being:

$$H(X) = -\sum_{i=1}^{n} P(x_i) \times \log(P(x_i))$$

For each post, we calculate this metric from the proportions of each one of the five relevant reactions.

For pages and page types, we calculate the average of this metric to compare the influence of these divisions.

Gathering Demographic Information

After assessing the information about the political bias of the pages, we collect the demographic information of the audiences for each relevant interest. For that, we use Facebook Marketing API which gives demographic data about the target audience for ads. That tool is normally used as analytics tool for announcers understand their potential audience. Specifically, the Marketing API gives demographic data of the audiences for target audience analysis of ads. If we decide a target public for a possible ad and pass a combination of properties that the audience should have, such as *interests*, Facebook estimates the size of that targeted public.

Using these estimates, we collect data from four demographic dimensions: Gender, Age Group, Relationship Status, and Education. Previous works have used this same information for relevant work. These include inferring useful health patterns in user's offline life [2], comparing census data to Facebook to evaluate the usefulness of possible estimates deduced from that [46], approximate election polling [48], besides some analysis of cultural differences [65].

Analyzing Demographics of Facebook Brazilian Pages

We further characterize the audience of the pages we collected by measuring how different a page's audience is from the general Brazilian Facebook audience in their demographic composition. To that end, we calculate the Mean Absolute Deviation (MAD) for each demographic dimension. We formally define the deviation, Δ_a^p, of Facebook page p for one demographic attribute a as follows: $\Delta_a^p = \pi_a^p - \pi_a^{\mathrm{B}}$ where π_a^p is the proportion of users who are the audience of Facebook page p and have the demographic attribute

a, and π_a^{B} is the proportion of all Brazilian users who has the demographic attribute a. Intuitively, Δ_a^p measures how one demographic group is under-represented (< 0) or over-represented (> 0) in an audience of certain page p compared to the average.

For example, $\Delta_{female}^{PublicFigureA} = \pi_{female}^{PublicFigureA} - \pi_{female}^{\mathrm{B}} = 0.9135 - 0.5396 = 0.3739$, which indicates females are 37.39% over-represented in the audience of "Public Figure A" page. From this deviation, we calculate the MAD for each demographic dimension by page as:

$$MAD_d^p = \frac{\left(\sum_{a_i \in d} |\Delta_{a_i}^p|\right)}{||d||}$$

where d is one demographic dimension (i.e., gender), and $||d||$ is the number of possible values of that dimension (i.e., 2 for male and female). In contrast to Δ_a^p, MAD_d^p does not differentiate under- from over-representation but assess the magnitude of the total shift from the baseline.

4.1.4 Inferring Toxicity

We use Google's Perspective API[5] to infer the toxicity of the posts and comments on the Facebook pages. There are various models provided by the Perspective API, which Table 4.1 describes.

In all cases, given a text, the models return a probability of it being toxic or an attack, which we call their scores. When this text is confusing or misspelled, the model returns no score. The API did not estimate the toxicity for 8.17% of the posts and 9.13% of the comments in our dataset. We used all models to measure the toxicity of posts and comments, but we found that the results have a high correlation to each other (Pearson's correlation coefficient is higher than 0.8 ($p < 0.05$) for all pairs). Thus, in this work, we present the results using the toxicity model. Table 4.2 shows example comments with their corresponding toxicity scores.

Potential Limitations

There are a few limitations of our toxicity data, discussed next.

Accuracy of the inference of toxicity by Perspective API models for Portuguese. Measuring the toxicity of a text is still a research topic that is in development. The Perspective API represents one of the first "off-the-shelf" tools available, and there are no current studies available about its accuracy in the Portuguese language.

[5] https://www.perspectiveapi.com

Metric	Description
TOXICITY	Rude, disrespectful, or unreasonable comment that is likely to make people leave a discussion
SEVERE_TOXICITY	A version of Toxicity which is less sensitive to comments that include positive uses of curse words
IDENTITY_ATTACK	Negative or hateful comments targeting someone because of their identity
INSULT	Insulting, inflammatory, or negative comment towards a person or a group
PROFANITY	Swear words, curse words, or other obscene or profane language
THREAT	Describes an intention to inflict pain, injury, or violence against an individual or group

Table 4.1: Overview of metrics from perspective API.

Comment (Translated to English)	**Comment (Original in Portuguese)**	**Toxicity Score**
This M*fo gets nothing more, he is a f*got	F*P não ganha mais nada, é vi*do	0.905
How about ... killing Lula and all the supreme court bad guys	tal ... matar-mos Lula e tds os bandidos do STF	0.884
They will all go to jail	Vão tudo pra cadeia	0.506
U know he doesn't like the poor	sabe ele não gosta de pobre	0.673
Guuuuuuys ... nobody does anything to stop this scum?	Genteeeee... ninguém faz nada pra cessar essa corja?	0.369
Don't let them set Barabbas out again. Can't make this mistake again ...	Não deixem soltar Barrabás novamente Não podem cometerem este erro de novo ...	0.199
of course it is ... he wants it to remain state owned to get his hands on the money that comes in .. very simple	claro que ta ...ele quer que continue sendo estatal pra meter a mao na grana que entra ..muito simples	0.128

Table 4.2: Examples of comments and their toxicity score.

Therefore, to estimate its accuracy, we manually labeled a sample of the data and measured the Fleiss's Kappa coefficient of agreement [9] for our labeling and our agreement with the results provided by Perspective API.

Three volunteers were instructed on the definition o toxicity given by the API, along with an explanation from the official website that showed the intended use of the tool to flag toxic content, where the volunteers should reflect if they would flag that content in the same situation. Then, they are instructed to label $2,000$ comments, randomly selected, as toxic or non-toxic, reaching a kappa coefficient of 0.41, with a confidence interval of [0.38-0.43]. To arrive at a final verdict on the labeling, we adopted a majority vote strategy, where if a message was vote by at least two people as toxic, it was marked as such. Finally, to compare it to the classification given by the Perspective API model, we tested different cutoffs based on previous works and a live demo on the official website. After testing, we decided to use the value of **0.8** as the cutoff. The kappa coefficient between the human label and the API was 0.52, with a confidence interval of [0.48-0.56]. Likewise, using translation and then using the Perspective API in English gives a kappa of 0.44, with a confidence interval of [0.40-0.49], which makes the use of the Portuguese version better than using what is available with translation. These results also show that measuring the toxicity of texts is a difficult test, and the Perspective API can be as good as a Human.

When looking into where the API did not agree with the volunteers, the metric was much more likely to diverge from the humans, labeling as toxic, than labeling non-toxic, with an accuracy of 0.9045 and f1-score of 0.5727. Therefore, we expect to find a slightly higher amount of toxic messages compare to human judgment.

Toxic comments detection by Facebook itself. Another limitation of our dataset is related to Facebook actively trying to diminish the toxic environment that can occur on comment threads. As collecting all the data took a few days, we expect our data do not represent the situation of the comments as they were posted, considering that Facebook has deleted part of the most toxic ones. After measuring how many comments were deleted two months after our initial search on a random sample, we found that near 1% of them got deleted.

Even with those limitations, we believe that our dataset can provide relevant insights on toxicity in Facebook comments. In the following section, we present and discuss the main results from characterizing toxicity in Facebook comments.

4.2 Datasets

4.2.1 Preliminary dataset

At the start of this work, we had a preliminary approach to search and identify pages, which focused on expanding a small list of pages. The primary goal we had was creating a slightly larger group of the main pages about news and politics on Facebook.

Our initial set was the 22 Facebook pages introduced in [41], and, having applied our process of searching for related pages using Audience Insights, we ended up with 63 Brazilian News and Political Facebook pages from various categories.

We then divide those 63 pages into the following nine categories: 16 Media/News Companies, 15 Politicians, 11 Public Figures, 9 News & Media Websites, 5 Magazines, 4 Broadcasting & Media Production Company, 1 Author, 1 Media, and 1 Political Organization.

The Graph API we used to collect the comments for our analysis of toxic messages had some limitations. This API limits the number of posts visible in a period, especially for posts older than one week. Because of that, we used the pages found in this preliminary approach as the dataset for measuring toxicity. After trying to collect data in two other political events throughout 2019 and failing to gather a large sample because of this limitation, we finally used this consolidated list to collect comments exactly as the release of Former President Lula was happening.

As stated in the previous chapter, we use the categories of the pages to group them on more broad types of pages, focusing on the division of mainstream media and alternative media, with public figures as a special kind of alternative media. Then, from the claimed political leanings of public figures, we also subdivided them to examine how political affiliation affected toxicity.

After collecting all the data available in the Graph API during the event that we decided to analyze, we gathered the dataset presented in Table 4.3.

Category	**Pages**	**Posts**	**Comms**	**Page Likes**	**Followers**	**Talking About**
Right-wing Figures	10	875	1,308,573	27,533,991	30,224,588	2,468,575
Centrist Figures	4	125	65,793	4,646,812	4,618,715	55,567
Left-wing Figures	12	1,100	547,581	13,024,855	13,907,604	1,194,221
All Public Figures	*26*	*2,100*	*1,921,947*	*45,205,658*	*48,750,907*	*3,718,363*
Mainstream Media	15	6,725	1,404,119	63,962,840	64,071,959	7,644,183
Alternative Media	22	7,000	769,853	12,443,604	12,484,364	3,151,038
Media	*37*	*13,725*	*2,173,972*	*76,406,444*	*76,556,323*	*10,795,221*
Total	**63**	**15,825**	**4,095,919**	**121,612,102**	**125,307,230**	**14,513,584**

Table 4.3: Overview of the collected data used to analyze toxicity

As we can see, the majority of the pages found are from public figures, followed by alternative media and then mainstream media. However, even that mainstream media are a smaller group, they have an aggregated popularity more substantial than the other two, with public figures, especially right-wing public figures, coming in second. This result shows that more people interact with mainstream media, which was expected, and also shows the divide between the groups that the next datasets used also confirm.

4.2.2 Extended dataset

Subsequently, going beyond the preliminary version of the method to identify pages, our second version used a snowball process to collect more pages. For the dataset of Brazilian pages, we used the same list of 21 Facebook pages introduced in [41] as our starting point, similar to our initial method. In total, we found 156 pages, with 36 public figures and political entities, which are identifiable as left-wing (19) or right-wing (17) pages. Later, we use this set of 36 pages in our method to calculate the political bias of all pages. Beyond this division, we also separate the pages by the same three types we used in the preliminary approach. Table 4.4 shows the distribution of this dataset in these types.

Type	**Pages**	**Avg. Audience Size**	**Std.**	**Interests by Category**
Alternative Media	31	7,031,467.74	12,385,139.80	News & Media Website (8), Media/News Company (6), Political Party (5), Broadcasting & Media Production Company (2), Local Business (2), TV Show (2), Website (2), Newspaper (1), Magazine (1), Business Service (1), Society & Culture Website (1)
Mainstream Media	59	5,678,779.66	10,381,588.20	Media/News Company (16), TV Show (15), News & Media Website (8), Magazine (7), TV Channel (4), Local Business (4), TV Network (3), Newspaper (2)
Public Figures	64	2,808,625	4,688,947.03	Politician (28), Public Figure (21), Journalist (6), Author (4), Artist (2), Arts & Entertainment (1), Writer (1), Book (1)

Table 4.4: Overview of the interest related to the Brazilian pages with calculated political bias

In contrast to the previous dataset of Brazilian pages, as we now want to create a list of pages that can be used in our method for ideological bias, we only keep the pages that have interests. Therefore, while the previous set had a larger proportion of alternative media, here mainstream media dominates it, as it was more likely to have a related interest. Meanwhile, public figures still are the majority of pages, which is comprehensible considering that, only for the current elected officials from federal and state level, Brazil has 1.572 congressmen, all of which could have Facebook pages.

As we were dealing with interests, instead of the likes of posts and other metrics we had, we primarily compare the reach of the pages we collected by the audience of their interest. As the audience is not cumulative in the same way the previous metrics were, we now compare averages. In this case, the alternative media we found to have a much higher popularity, even higher than mainstream media. But, this comparison with the mainstream is not entirely fair at this point, as one of the alternative media pages is from a large website from an internet provider that aggregates news from other outlets, not producing content itself, and which alone had an audience of circa 6.500.000 users.

Looking into the categories these pages have, we can see that in contrast with the initial method, we were able to find a more diverse set of pages. Especially for alternative media, we included political party public pages, pages categorized as local business, and even some self-declared TV shows. Mainstream media also increased in variety, with TV Channels, TV Networks, and also regional news outlets categorized as local businesses. We expect that this final list of Brazilian pages represents a sample of the most relevant pages on Facebook and that the later characterization of these outlets and public figures in the next chapters will help fill the gap in the research of the Brazilian media landscape, notably including alternative media.

4.2.3 Extended dataset in our reaction analysis

Finally, to complement the analysis of the toxicity present in the content, we also do a more extensive examination of the users' reaction to the posts. As we do not need to analyze posts only from the pages with complete audience information to examine these reactions to content, we can also add the pages found during our snowball process without a direct interest. For these pages, we can also measure political bias, further using our methodology to find ideology to present useful data on political leaning. Besides that, we also use the data to create page reaction profiles that can help us compare the data from toxicity in the comments to the possible negative reactions given to a post.

To create this dataset, we also add the pages from the snowball process that appeared as related pages but did not have interests. Table 4.5 shows the results of our search. Alternative media again becomes the majority, reinforcing our impressions from the last section. As stated before, when our list of pages depends too much on interests, the number of alternative media pages found becomes the smallest of all three types. But in this set, where we do not limit the pages and use our snowball process, we have alternative media as the most substantial set.

Type	Pages	Posts	Avg. Likes	Avg. Comments	Avg. Reactions	Total Reactions	Categories
Alter-native Media	332	802,956	662.697	159.783	263.912	211,909,710	News & Media Website (110), Media/News Company (100), Magazines (24), Political Organization (20), Political Party (8), Broadcasting & Media Production Company (7), Community (6), Newspaper (6), Personal Blog (5), Public Figure (5), TV Channel (3), Arts & Entertainment (3), Local Business (3), Religious Organization (3), TV Show (3), Entertainment Website (3), Video Creator (2), Radio Station (2), Political Ideology (2), Jornalist (2), Community Organization (2), Government Organization (2), Fan Page (2), Media Agency (1), Political Candidate (1), Cause (1), Editorial/Opinion (1), Writer (1), Nonprofit Organization (1), Science Website (1), Education Website (1), Digital Creator (1)
Mains-tream Media	180	991,374	679.336	199.984	285.152	282,692,338	TV Channel (46), Media/News Company (41), News & Media Website (32), TV Show (27), Magazines (14), TV Network (7), Newspaper (6), Radio Station (4), Local Business (2), Broadcasting & Media Production Company (1)
Public Figures	255	212,029	2,990.927	576.097	697.831	147,960,457	Public Figure (144), Politician (60), Jornalist (16), Writer (15), Author (11), Artist (2), Comedian (2), Actor (1), Media/News Company (1), Local Business (1), TV Show (1), Government Official (1)

Table 4.5: Overview of the pages with reactions collected

Going into more details of our data, we can see that this group of pages repeats some trends we saw before. Alternative media still receives the least comments and likes on average, while the public figures create the least amount of posts. However, what differentiates this set of pages from other ones is the popularity of public figures and the share of publications. While public figures only were more popular than alternative media in other datasets, in this case, we have their pages as the most popular, with the highest average of all metrics of response from their audience. Also, interestingly, we found that in the number of shares, alternative media surpasses mainstream media.

Complementing the conclusions from our first dataset and our second set from Brazilian pages, we can aggregate our results in a more clear picture. Mainstream media have a larger audience, more page likes, and page followers than public figures and alternative media. Meanwhile, the posts produced by both types of media pages do not receive the same amount of attention as public figures' posts, probably due to the more constant stream of content from media. The larger average of shares per publication of alternative media compared to the mainstream might indicate posts that are polemical or that the endorsement that the audience of this type of media commonly shows is the sharing of the content, displaying a more direct engagement. Meanwhile, this same engagement becomes even more pronounced in the public figures audience, and, although it is the smallest group, it is the most reactive.

Lastly, the categories of the pages found by this last use of our snowball method show that when we deal with a larger sample of pages, these categories from Facebook are not as informative. Especially when dealing with alternative media, some of these categories that might appear to indicate that a page is from a person, and therefore classified as a public figure, the page has no relation to a specific politician or media person and is just a news sharing page.

Chapter 5

Characterizing Alternative Media

In this chapter, we present the results of the four semi-supervised learning algorithms, compare our methodology for quantifying the ideological leaning of Facebook pages with other previous methods, and show the estimated political bias of Facebook pages in our data collection. After that, we also examine the demographics of their audiences in the context of the ecosystem on Brazilian political Facebook pages.

5.1 Dataset used for validation

As stated before, in the verification of our methodology to calculate political bias, we needed to use a dataset of U.S. news outlets to compare our results to previous work [38, 3, 6, 47].

To accomplish that, first, we compiled a list of seed pages for step (1) of our snowball method from Chapter 4. We use the list of 15 news outlets created in [6] as our starting list. After ten iterations of the snowball described, we found 832 pages that had an interest in Facebook. Using their political self-identification, we also divided a group of 136 public figures and political entities from these pages almost evenly into 65 left-wing and 71 right-wing pages to use in the training part of our method. Table 5.1 shows the categories of these pages and the intercession with our data to the datasets from previous work we want to compare to it.

The dataset created by our snowball process of U.S. pages has more pages in common with the more extensive datasets, but contain more of the pages of the smaller datasets, proportionally. This fact shows one limitation of our approach, which is the use of interests to find new pages. Compared to the most comprehensive list of pages we have from [47], we were only capable of finding 302 of the 20,448 pages with our method.

The most probable reason for this is that this dataset includes a high percentage of smaller alternative media that do not have interests related to them. It is also possible that some large regional or national mainstream media has an interest, with a smaller

Dataset	Total Size	Intercession with our entire data	Categories
Mitchell [38]	32	24 (75.00%)	Media/News Company (10), Public Figure (3), Newspaper (2), News & Media Website (2), Local Business (2), TV Show (2), Broadcasting & Media Production Company (1), TV Network (1), Website (1)
Bakshy et al. [3]	500	111 (22.20%)	Media/News Company (30), News & Media Website (29), Newspaper (11), Broadcasting & Media Production Company (9), Magazine (8), Website (7), TV Show (5), Local Business (3), Public Figure (2), Nonprofit Organization (2), TV Network (2), Arts & Entertainment (1), Media (1), News Personality (1)
Budak et al. [6]	15	14 (93.33%)	Media/News Company (7), Newspaper (2), Website (2), Broadcasting & Media Production Company (1), Local Business (1), News & Media Website (1)
Ribeiro et al. [47]	20,448	302 (1.48%)	Media/News Company (75), News & Media Website (72), Newspaper (40), TV Show (39), Broadcasting & Media Production Company (25), Magazine (19), Nonprofit Organization (10), Website (9), Local Business (3), Public Figure (2), Political Organization (2), TV Network (2), Arts & Entertainment (1), Media (1), Musician/Band (1), News Personality (1)
Our extra data	527	527 (100.00%)	Public Figure (109), Politician (102), Nonprofit Organization (81), Journalist (43), Media/News Company (29), Political Organization (28), Author (24), News Personality (24), Government Official (15), News & Media Website (12), Political Candidate (10), Political Party (9), TV Show (9), Magazine (7), Local Business (6), Broadcasting & Media Production Company (4), Newspaper (4), Media (3), Cause (2), Political Ideology (2), Athlete (1), TV Channel (1), Entrepreneur (1), Interest (1)

Table 5.1: Overview of dataset used to validate our political bias measuring method

local news outlet affiliated to not receiving one, as the topic of the larger outlet covers them. One evidence that seems to support this possibility is the interaction of this broader dataset with the others. All the pages analyzed by Bakshy et al. [3] and Budak et al. [6] also existed in the data from Ribeiro et al. [47], and only three pages from Mitchell [38] were not present. These three pages were two public figures, Rush Limbaugh and Sean Hannity, and one news outlet, Buzzfeed News, related to a broader news and entertainment outlet already included in this dataset, Buzzfeed. Therefore, as Ribeiro et al. [47] presented such an extensive dataset, it is the most important in our comparison.

Another observation we can make is that public figures are barely present in previous work, with most of the ones encountered by our methodology only found by us. Table 5.1 shows them in the extra data that did not have an intercession with any previous dataset. We use the self-identification of public figures to measure the political bias of all pages. Consequently, we can test our results for them using just their declared ideology. Therefore, the absence in the other works does not impact our analysis. Nonetheless, it is still important to note that the proportion of alternative media is higher in the more extensive datasets, with the smallest ones almost entirely composed of mainstream media outlets.

5.2 Comparing Graph-based SSL Algorithms

To identify the best graph-based SSL algorithm to use in step (5) of our methodology, detailed in section 4.1.2, we compare well-known graph-based SSL methods in the task of classifying Facebook pages as either left or right, which we will call the classification step. Later, we use the best algorithms to calculate an actual political bias score using the ideological leaning of U.S. pages calculated from the average outcome of the ten folds of the supervised learning and comparing our results to the other four related works [47, 3, 38, 6].

From the 832 pages, we found in the previous section, which had an interest in Facebook, we reserved ten test sets of 83 pages for each fold and proceeded with a 10-fold cross-validation. For the hyper-parameters of these algorithms, we used a grid search over the parameters' space of each one of them, using this cross-validation scheme to test the results. For all algorithms, except SGT, we used the scikit-learn implementation, with the code for SGT acquired from the original author's website [1]. We found the best results when using 50 neighbors for KNN, the RBF kernel for both LP and Smooth LP, and the SGT with its c hyper-parameter equal to 1000.

The σ used in the kernel for the LP method was equal to 0.35, while the same parameter was 0.55 for the Smooth LP method. The best RBF kernel found was not the standard implementation available on the scikit-learn library, but our version that used our affinity score directly instead of calculating the euclidean distance on it. Table 5.2 shows the results of each tested model for both training and test sets.

[1] http://sgt.joachims.org/

	AUC (Train)	AUC (Test)
LP	0.9546 [0.9414-0.9679]	0.8440 [0.8091-0.8790]
Smooth LP	0.9509 [0.9298-0.9719]	0.8926 [0.8718-0.9133]
SGT	0.9615 [0.9462-0.9768]	**0.9482** [0.9290-0.9674]
KNN	**1.0000** [1.0000-1.0000]	0.9122 [0.8806-0.9437]

Table 5.2: AUC scores for different SSL methods with 95% confidence intervals.

We find that SGT has the best result on average for the test set, surpassing the KNN baseline. Smooth LP came in second, being statistically equivalent to the baseline, while LP was the worst. In the training set, KNN was better than all others. However, as the training set for KNN was only composed of the labeled data, it had an easier task when learning the classes on that set, making it a less meaningful comparison than the test data. That is, as the KNN did not have to try to generalize the classification using the training data without labels. Therefore, maybe allowing some error for the training data, it could just learn the correct class for all data points in that partition.

5.3 Comparing our Proposed Method with Previous Work

We now take the three algorithms that were satisfactory in the previous section (SGT, Smooth LP, and the KNN Baseline) to compare the results of using them as step (5) of our methodology with four well-known datasets of political bias.

5.3.1 Comparing Algorithms for the Complete Task

To compare the results from the graph-based SSL methods with the four datasets created by other methods from relevant related work, we use the Pearson correlation coefficients, shown in Table 5.3[2]. We observe that almost all these methods had statistically equivalent results, with statistically significant differences only on [47] data. In this case, Smooth LP and KNN have the highest correlations, and SGT is worse than all other options. As KNN and Smooth LP were also satisfactory in the classification step, the SGT

[2]The confidence intervals were calculated using Fisher z-transformation as presente on: http://onlinestatbook.com/2/estimation/correlation_ci.html

advantage becomes less relevant as it only won in that step alone. Additionally, considering that the KNN uses only the labeled data, we can deem the **Smooth LP** the best method, as it uses semi-supervised learning, training with most of the graph. In other words, as the amount of labeled data affect the overall result, in cases where labeling is difficult, we expect the semi-supervised approach to be better.

	Smooth LP	SGT	KNN
Mitchell [38]	**0.7642** [0.5216-0.8925]	**0.8193** [0.6212-0.9190]	**0.8091** [0.6022-0.9141]
Bakshy et al. [3]	**0.8353** [0.7686-0.8841]	**0.8483** [0.7863-0.8934]	**0.8204** [0.7485-0.8733]
Budak et al. [6]	**0.6267** [0.1440-0.8685]	**0.6616** [0.2019-0.8824]	**0.6414** [0.1681-0.8744]
Ribeiro et al. [47]	**0.8225** [0.7823-0.8559]	0.6266 [0.5528-0.6906]	**0.8263** [0.7868-0.8590]

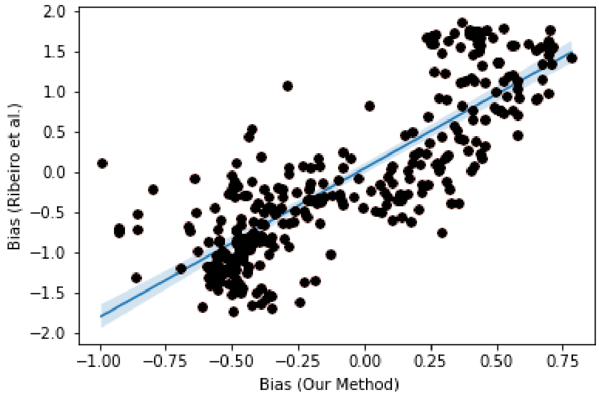
Table 5.3: Pearson's r for each combination of political leaning baseline and graph-based SSL method.

5.3.2 Comparing our Method to other Methodology

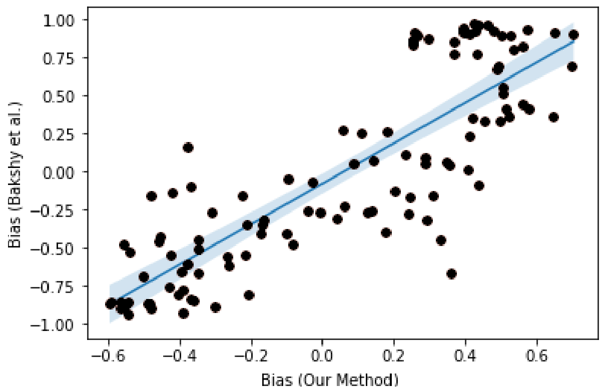
After establishing the best algorithm, we now compare the results of our entire methodology with the results of the four relevant methods from related work. Figures 5.1 and 5.2 depict how similar our bias scores computed with the Smooth LP algorithm are to the ground truth datasets.

We see that the results of our method had a high correlation with the results of those three audience-based datasets (Person's $r = 0.8$ on average). Notably, the sets that also used Facebook data [47, 3] had narrower confidence intervals. Meanwhile, [38] had a slightly wider confidence interval, with the lower bound of its correlation being as low as 0.5216, probably because audience bias was assessed using a survey instead of Facebook data, being less comparable to our strategy. Following that trend, the dataset that measured the political bias of news stories by showing them to Amazon Mechanical Turk human judges instead of using Facebook [6] had the broadest confidence intervals and the worst correlation with our method. This result happened even after using their list of pages for our starting list in step (1) of our methodology. This problem was equally detrimental for all algorithms. We theorize that the number of outlets, together with the content-based scores from human labeling, generated this lower performance. Unfortunately, the paper does not provide any rater reliability metric (e.g., Kappa score), which makes it harder to analyze other possible causes of that discrepancy.

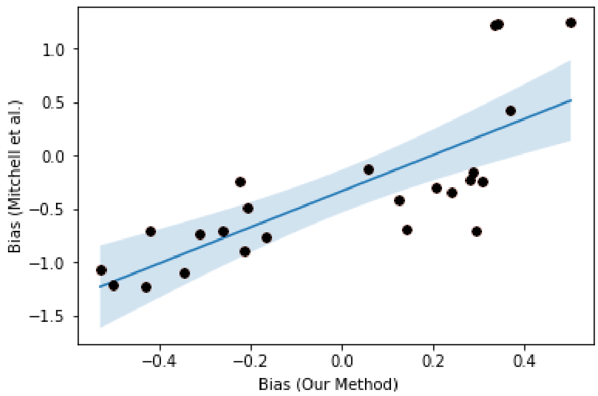
Figure 5.1: Comparison of our method to the baseline political leanings from methods based on audience metrics.



(a) Ribeiro et al. [47]



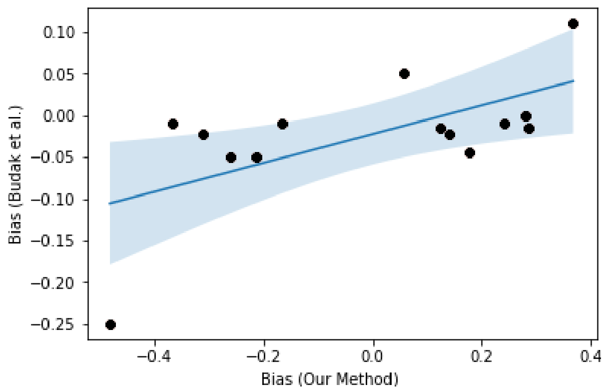
(b) Bakshy et al. [3]



(c) Mitchell [38]

Source: The author.

Figure 5.2: Comparison of our method to the baseline political leanings from Budak et al. [6], based on media content.



Source: The author.

Nonetheless, the fact that our method performed well with data from audience analysis is an indication of how it is reliable compared to other similar methodologies. These methodologies have a margin of error, something also found in our method, possibly estimating a bias that is slightly off. However, because these related works have presented reasonable results and our methodology is comparable to them, we effectively built another functional alternative.

5.4 The Brazilian Alternative News Landscape

5.4.1 Our Brazilian Dataset

With our method validated, we now apply the same methodology to analyze Brazilian pages.

We found 156 Brazilian pages with 36 public figures and political entities, which are identifiable as left-wing (19) or right-wing (17) pages. We used these 36 pages as our labeled data. Using the Smooth LP algorithm as the step (5) of our method, described in Section 4.1.2, the test set AUC and its 95% confidence interval was 0.9875 [0.9686-1.0]. With this classification, we correctly identified all the ideological leanings of the "seed" pages from the original article [41].

5.4.2 Calculating Political Polarization of Brazilian Pages

To further analyze our measured bias, we divided the range of [-1,1] of our score into three parts to represent **Left**, **Center**, and **Right** political leanings. To accomplish this, we used the standard deviation (δ) of the bias from the ten folds of the cross-validation, assigning data to their political positions based on the sign of average score adding and subtracting δ. If it is negative in both cases, we considered it **Left**. If it stays positive, we labeled it **Right**. It is **Center** otherwise. To better understand how the alternative news media differ from other types of news media and answer the **first research question**, we also label the collected pages by three types: public figure, mainstream media, and alternative news media. We grouped all politicians and public figures in the **public figure** category, and we classified the Journals, Websites, TV, Radio, and Magazines as **mainstream media** if they had a registry in any Brazilian official press organization[3]. If there was no registry, we considered them as **alternative news media**.

[3]We use data from the National Association of Journals (ANJ), the National Association of Magazine Editors (ANER), and the National Agency of Telecommunications (ANATEL).

Table 5.4 shows the distribution of pages by their political bias and types, helping us answer the **first research question**. We see that most alternative media outlets are classified as left-wing, while mainstream media outlets are primarily right-wing.

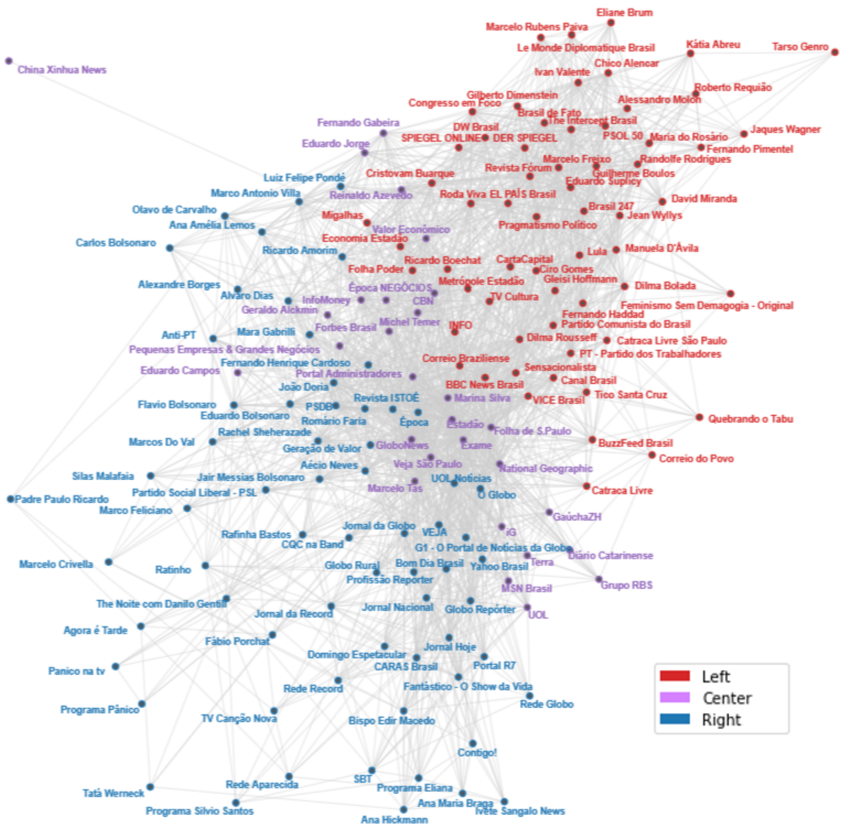
A possible explanation for this polarization is that our method collects pages that also have *interests* in the Facebook Ads platform. As pages are added as *interests* based on user interactions, most alternative media we found are from the time of the previous left-wing governments. Meanwhile, some big mainstream outlets have more center-right positions, as they exist since the right-wing Brazilian military government, and are pro-business [23].

From the visualization of the graph of Brazilian Facebook pages constructed by our method (Figure 5.3), we also notice that the right-wing pages are less connected, and politicians like *Marcelo Crivella* and *Maria do Rosário* are more isolated from the opposite political side. This isolation is useful for reinforcing the bias in our method.

	Left	Center	Right	**Total by Type**
Alt. Media	17	5	6	28
Main. Media	14	16	29	59
Pub. Figures	30	8	31	69
Total by Bias	61	29	66	**All Pages:156**

Table 5.4: Overview of Brazilian Facebook pages data.

Figure 5.3: Graph of the all pages.

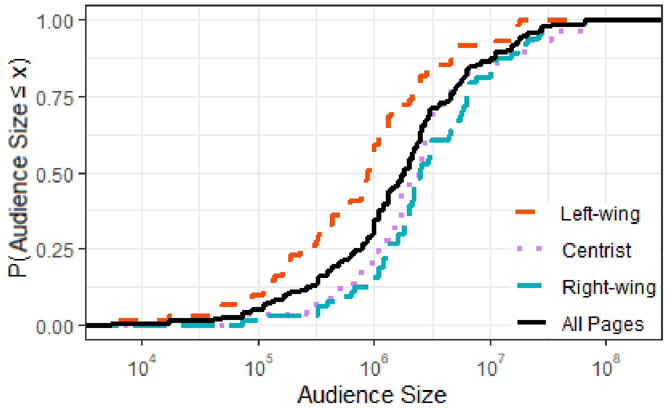


Source: The author.

5.4.3 Size of the Audience across Political Leaning

Here we examine the size of the audiences of each page. Figure 5.4 depicts the cumulative distribution function (CDF) of audience size from all the pages and considering the political leanings.

Figure 5.4: Distribution of the audience size for each page.



Source: The author.

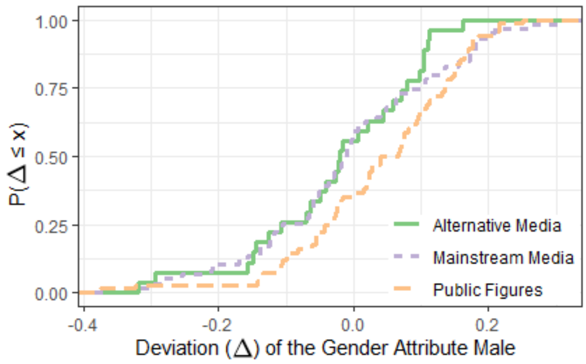
As we can see, 50% of all pages have an audience of up to 3 million users, while circa of 12.5% of them had an audience of more than 10 million users. Grouping by ideology and testing statistically significant differences using the Mann-Whitney U test, we see that left-wing pages had significantly smaller audiences compared to both right-wing pages (U = 2417, $p < 0.005$) and centrist pages (U = 513.5, $p < 0.005$). Yet, when we consider the two types of media and public figures, there were no significant deviations from the distribution of the audience compared to all pages.

5.4.4 Contrasting demographic dimensions

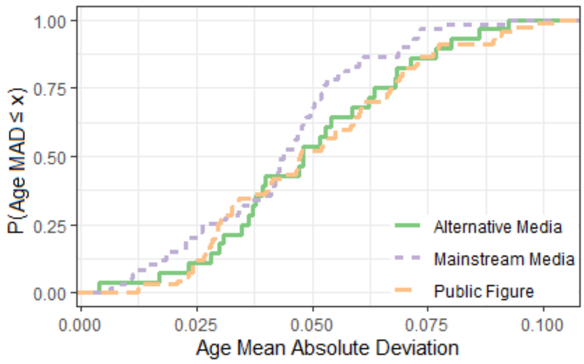
Going beyond the audience size, we use the MAD of each demographic dimension to compare the audience of different media types, in a first step to answer the **second research question**. Here the MAD represents how far the audience of one page differs from the average Brazilian Facebook users. Then, to confirm which demographic attributes were under- or over-represented, we used the attribute deviation (Δ).

In these demographics, the dimension of *gender* had the highest values of MAD. 30% of all pages had an over-representation of men in their audience. In those, there was a 10% higher proportion of men than the average.

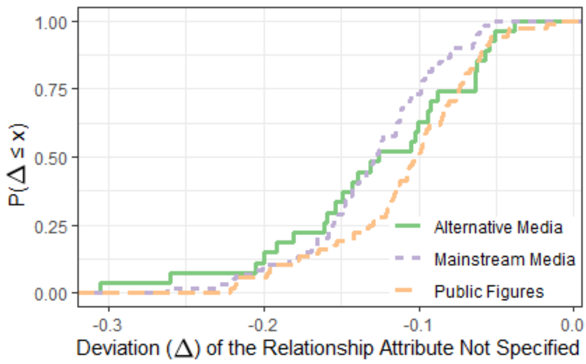
Figure 5.5: Cumulative Distribution Function (CDF) of the deviation (Δ) and MAD score for demographic attributes and dimensions



(a) Δ_{male}



(b) MAD_{age}



(c) $\Delta_{not\ specified\ (relationship)}$

Source: The author.

However, the over-representation was not uniform, with **public figures** having MAD scores significantly higher than the other categories (U = 3661, $p < 0.05$). It means that even with various pages having some over-representation, the deviation in this category was the highest; in this case, **public figures were especially more followed by men**, as shown in Figure 5.5a.

Another demographic dimension where pages of public figures had significantly higher values of MAD was *age*. For this characteristic, **the audience of these pages contained more people over the age of 45**, with the MAD score being statistically higher than of **mainstream media** (U = 1633, $p < 0.05$), shown in Figure 5.5b.

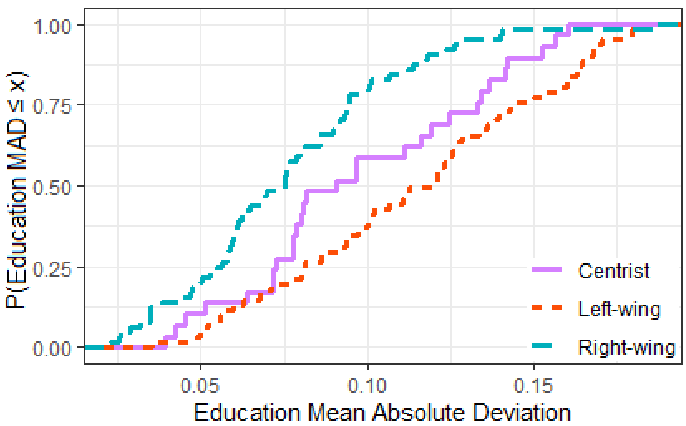
In contrast, **alternative media** had this metric varying between the values of the other two groups, having no significant difference with them. [41] found a similar trend with partisan **alternative media having more interactions from users that were men and older than 41 years old compared to traditional media**. This shift from users interacting with alternative news media pages to following public figures relates possibly to the deactivation of some pages from their study since 2017.

One demographic dimension that we also analyze, and is not present in that work, is the *relationship status*. We divide this dimension in the Facebook status of *Single*, *In a Relationship*, *Married*, *Not specified*, and **Others**.

The *Others* value aggregates relationship types that were a minority in our data, like *Divorced* or *In an Open Relationship*. For all pages in our data, the proportion of people that had some relation, in particular more serious relationships like marriage, was higher than the average. But, this deviation is not uniform again, and when we examine the values of MAD, we see the inverse of the *age* dimension. **Mainstream media had a significantly higher shift from the average compared to public figures** (U = 2471, $p < 0.05$). That is most evident when we look at people without any status, counted as *Not Specified*. 45.55% of all Brazilians on Facebook does not have any relationship status. Figure 5.5c shows how the three types of pages deviate from this average. For 72.5% of all Mainstream media pages, less than 35% of their audience do not have any status, while only 47% of public figures page have this proportion of users with this *Not Specified* status.

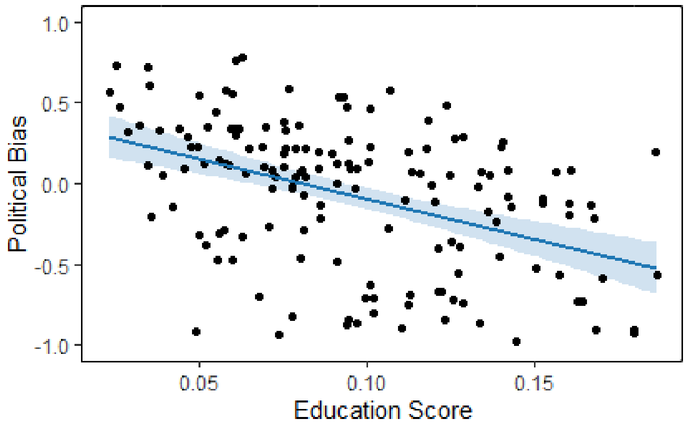
Lastly, we look at the demographics of *education* in the pages we collected. In our analysis, we notice that all pages tend to have an audience with an over-representation of people with higher education (e.g. college and grad school) and an under-representation of people with only more basic education (e.g. high school and professional degrees). Particularly, **alternative news media** audiences had the highest MAD scores in this dimension, statistically higher than both mainstream media (U = 542, $p < 0.05$) and public figures (U = 1156, $p < 0.05$). However, as we saw in Section 5.4.2 that alternative media is mostly left-wing in our data, we also examine the correlation between education and political leaning. Figure 5.6 depicts a clear difference in the MAD score, with a significant ($p < 0.0005$) relation of right-wing < centrist < left-wing. To test this trend, we also calculated the Pearson correlation coefficient (ρ) between the deviation and the political bias, as shown in Figure 5.7.

Figure 5.6: Cummulative Distribution Function (CDF) for education MAD score grouped by poltiical leaning



Source: The author.

Figure 5.7: Correlation of education score and political bias.



Source: The author.

We found ρ equal to -0.4456 with a confidence interval of [-0.5640, -0.3092], showing that a more left-wing bias (more negative) had a higher significant correlation with education. These results align with the findings of [56] that cultural and lifestyle differences affected by political bias.

To recapitulate, in our analysis, we found **an older and more male audience for public figures**, confirming some similar trends from other works [41], now also adding an examination of relationship status. For all types, **the people with some relation were more prevalent than those without any**. Their percentage was higher than the average, especially with committed relationships like marriage. But that was more accentuated for mainstream media, which had a significantly higher variation from the mean, compared to public figures.

Between the two types of media, **alternative media had a public that was composed of users that were more commonly men and older than 41 years old** compared to traditional media.

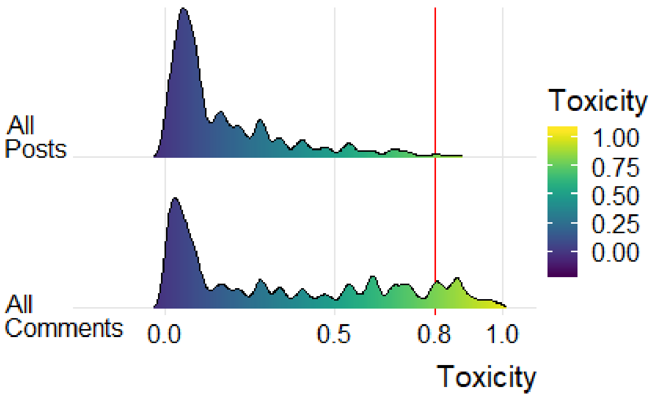
Lastly, we also found a correlation between politics and how much education people reportedly received, which showed **pages further to the left having an audience that self-described as more educated**.

Chapter 6

Toxicity and Reactions in Alternative Media

In this chapter, we aim at analyzing the toxicity level of comments and posts to check the extent to which the toxicity is correlated with different categories and ideological leaning.

Figure 6.1: Distribution of toxicity for all comments and posts.



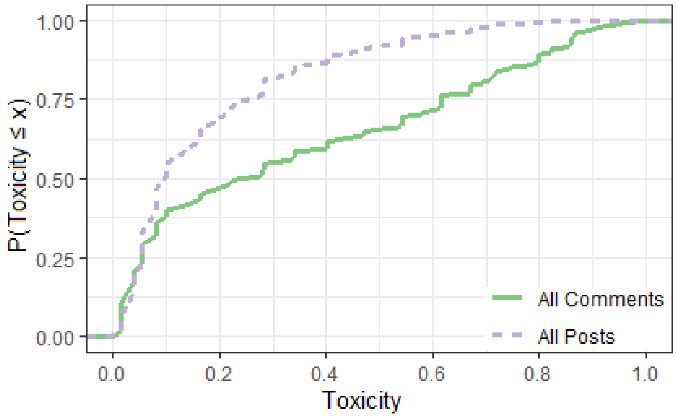
Source: The author.

Figure 6.1 shows the distribution of the toxicity scores for all comments and posts. In this work, we consider a comment or post to be toxic if the toxicity score is above 0.8, following our validation of the toxicity measure and the previous work of [16]. Using this threshold, we find that 13.27% of comments and 0.84% of posts were considered toxic. While these percentages of toxic messages and posts may appear low, the level of toxicity varies between pages by a factor of two to ten compared to average values. Figure 6.2 shows the CDF of the same distribution as Figure 6.1.

We can see that 50% of all publications have less than 0.1 toxicity and that in the range between 0.1 and 0.9 toxicity, the distribution of comments is more skewed to higher values than the posts.

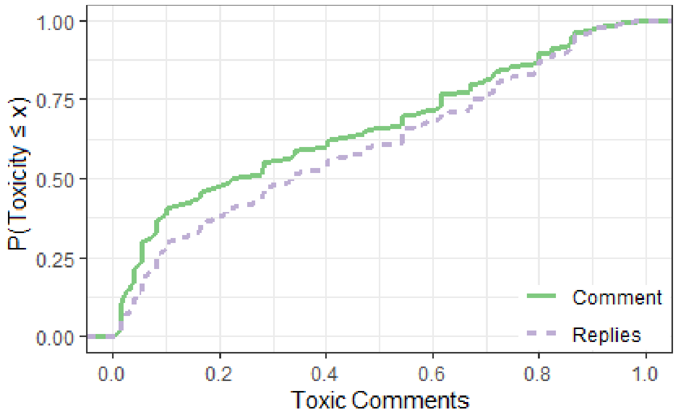
When verifying how much this distribution varies between pages, we found that 20% of all pages are responsible for 60% of toxic comments and 56% of all toxic posts. A large number of these comments on a page may indicate that a group of angry users is attacking the page or that there is a fight occurring between users, created by discussion in the comment section. But, it can also indicate that a page is inviting those messages.

Figure 6.2: Cumulative Distribution Function (CDF) of the amount of toxic comments and posts.



Source: The author.

Figure 6.3: Cumulative Distribution Function (CDF) of toxicity of comments compared to their responses.

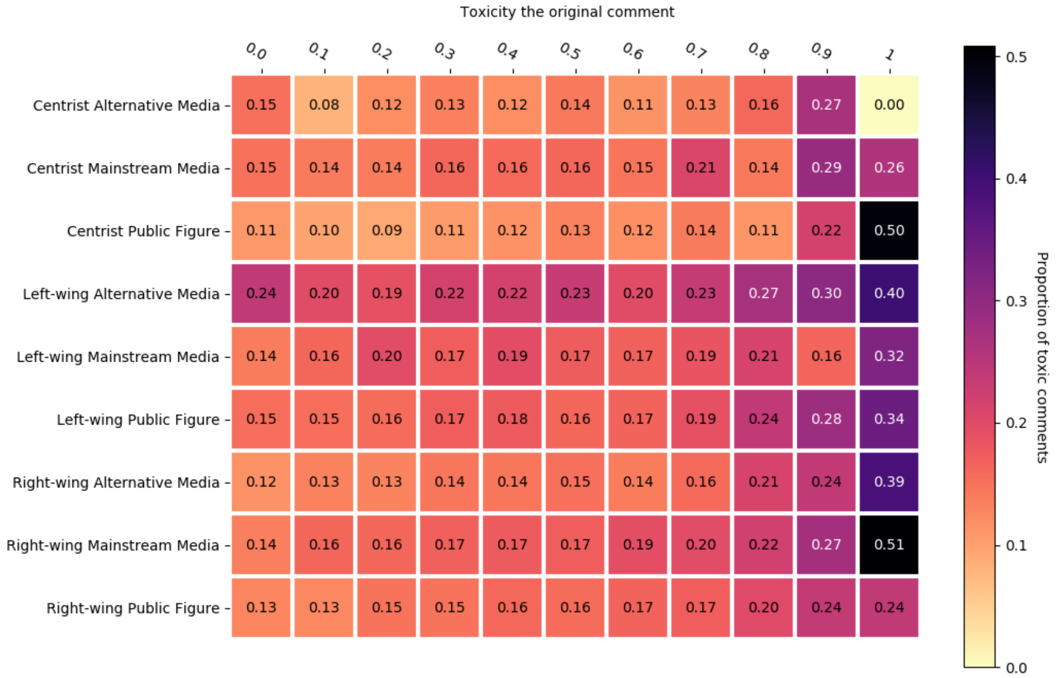


Source: The author.

We test that by using the Chi-squared test between the toxic posts and posts with the proportion of toxic comments above the average of our data. We then found that the toxicity of the publication and the percentage of these messages are not independent (χ^2=36.867, $p < 0.005$). Furthermore, using the Mann-Whitney U test, we find **the responses to a comment tend to be more toxic than a normal comment to a post** (U $> 3.6 \times 10^{11}$, $p < 0.005$). In Figure 6.3, more details of this comparison show that even though there is little difference in the percentage at the cutoff point, as the test shows, there is a greater probability of the toxicity of the responses being greater.

To better examine this result, we look into the toxic responses aggregated by the value of metric for the original comment and the type of page where the reply is present. Figure 6.4 shows the average percentage of toxic response a comment receives, given the page type and its toxicity level of the original message. In general, we can see that comments with toxicity above 0.8 provoke more messages above this limit. Yet, this phenomenon is not uniform in all page types. The proportion of toxic comments received by a message with maximum toxicity in a page from a right-wing public figure is equal

Figure 6.4: Heatmap of proportion of toxic reply comments by page type and original comment toxicity.



Source: The author.

to the percentage for one with minimum toxicity in a left-wing alternative media page.

These analyses already answer some factors that influence the proportion of toxic comments, which is the question asked as **the third research question**.

6.1 Toxicity in Brazilian Pages

In this section, we examine how various characteristics of the pages are related to the level of toxicity in the comment sections, further answering the **second research question**. To analyze the toxicity of several pages, initially, we calculated the proportions of the toxic messages commented in each of the posts.

We first investigate how the political leaning of public figures (left, center, or right-leaning) and the type of media (mainstream or alternative) of pages are relating to the level of toxicity in their comment sections.

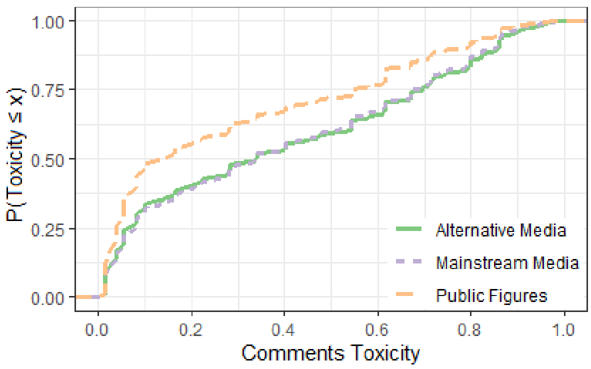
For each sub-group, we aggregate all the comments of the corresponding pages. Table 6.1 shows that the average toxicity of these comments for public figures (0.2886) is slightly lower than for media (0.3906), but that the political position of public figures does not significantly affect such an average.

Category	**Toxic Comms**	**Toxic Posts**	**Avg.**	**Std.**	**Mean Sq. Error**	**En-tropy**
Right-wing Figures	9.50%	2.29%	0.2855	0.1028	0.0111	0.3141
Centrist Figures	9.09%	0.00%	0.3135	0.1273	0.0180	0.3045
Left-wing Figures	9.82%	1.00%	0.2929	0.1237	0.0169	0.3212
Public Figures	*9.58%*	*1.48%*	*0.2886*	*0.1168*	*0.0146*	*0.3158*
Mainstream Media	16.19%	0.22%	0.3903	0.1633	0.0323	0.4428
Alternative Media	17.15%	1.24%	0.3912	0.1945	0.0430	0.4583
Media	*16.53%*	*0,74%*	*0.3906*	*0.1799*	*0.0377*	*0.4483*
Total	**13.27%**	**0.84%**	**0.3427**	**0.1728**	**0.0346**	**0.3915**

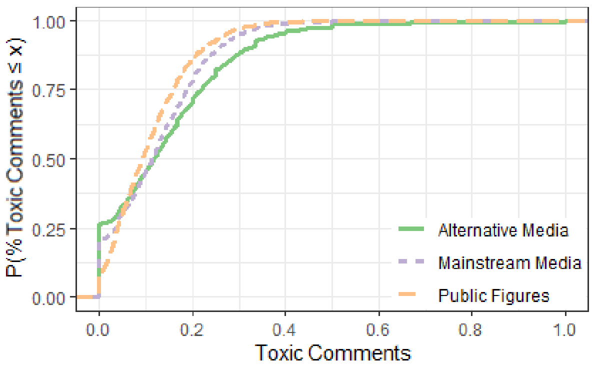
Table 6.1: Overview of toxicity metrics calculated only on the preliminary dataset

Figure 6.5a illustrates this difference, showing that it extends to the entire distribution, with toxicity being statistically lower for public figures, with significance according to the Mann-Whitney U test (U $> 3.52 \times 10^{12}$, $p < 0.005$). Figure 6.5b similarly shows that in terms of the proportion of toxic comments (that is, comments whose toxicity is higher than 0.8), the averages of the percentage of these comments in publications of public figures are lower than that of publications in the media with averages of 9.58% and 16.53%, respectively.

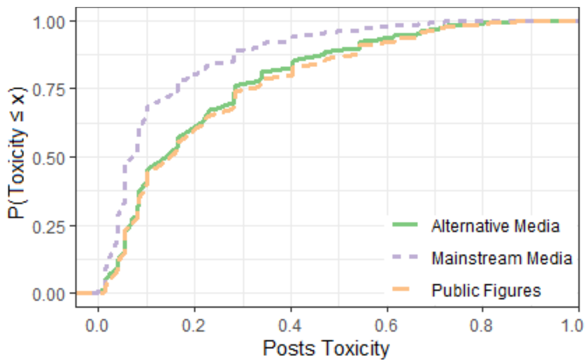
Figure 6.5: Cumulative Distribution Function (CDF) of toxicity in comments and posts by page type.



(a) Comments toxicity.



(b) Toxic comments.



(c) Posts toxicity.

Source: The author.

Using the Mann-Whitney U test to assess the difference in these proportions, a significant effect was seen on the type of pages (U $> 1.17 \times 10^7$, $p < 0.005$).

This result showed that publications in the media receive statistically more toxicity than publications by public figures. One possible reason is that people may consider the pages of public figures as homogeneous political discussion spaces where they support the corresponding person.

Meanwhile, they might recognize the media pages as cross-cutting political discussion spaces where people with various political leanings come together and argue or discuss [1]. In the first case, when talking to people who think the same way, there may be less toxicity or the second heterogeneous environment could trigger more toxicity. Figure 6.5c presents a sign that the environment itself may be responsible for the difference. In it, we see that the traditional media has a distribution of the toxicity of the posts with less toxicity, showing that the higher proportion of toxic comments does not come from many posts with toxic text. Among the media pages, all but one have more than 10% of toxic comments, while half of the pages of public figures do not show such a level of toxicity in their messages. The pages by Lula, Michel Temer, and Jair Bolsonaro, who are the former and current presidents of Brazil, are the first, second, and fifth pages by the lowest proportions of the toxic comments in their pages.

With Lula and Bolsonaro being the most polarizing figures of recent years, it is unexpected that they have such low values. Besides the possibility of being homogeneous political discussion spaces, another possible explanation is that, as politicians, their accounts may be maintained by professionals. As the online presence is remarkably crucial for public figures, their assistants could flag the toxic comments and try to remove them. In contrast, news media might not have the same amount of effort, and, in some cases, they might even benefit if people come to their page fight about current events, especially alternative media. Then, their effort to flag comments and moderate them might be of a lesser extent.

6.2 Toxicity in Posts

Next, we examine the proportion of toxic comments to select the worst posts and also analyze to see if it was related to politics. To filter posts with few comments, we calculate the average number of received messages, ignoring publications below the average, as they can have only one of these toxic comments, and that 100% proportion would be irrelevant. The results are in Table 6.2.

When considering posts with above the average number of comments, right-wing

Page	Post Content	Post Toxicity	Total Comms	Toxic%
Jornal da Cidade Online	Politics	0.8652	1730	49.94%
Jornal da Cidade Online	Politics and Bolsonaro stab	0.8004	1505	47.97%
Jornal da Cidade Online	Politics and Money	0.8641	3238	47.25%
O Antagonista	Lula	0.2841	1844	46.31%
Eduardo Bolsonaro	Return of PT	0.4028	6219	45.57%
Jornal da Cidade Online	Politics	0.8224	1257	44.87%
Diário do Brasil	Politics and Money	0.1620	866	44.46%
Jornal da Cidade Online	Lula	0.1342	1042	44.34%
Caneta Desesquerdizadora	Politics and the Environment	0.8004	2720	44.12%
Carlos Bolsonaro	Lula	0.7899	1666	43.88%

Table 6.2: The ten posts with an above average number of comments with the highest proportion of toxic comments.

pages are more prevalent. This finding might indicate that even if left partisan pages can receive toxic content, right-leaning pages attract a larger audience and higher amounts of toxic messages. Accounting for content, we confirm the prevalence of politics as a topic. But, as the biggest story in the period considered was political, we could not generalize the results.

Between these top posts, only two were created by a public figure, written by Congressman *Eduardo Bolsonaro* and the alderman *Carlos Bolsonaro*, both sons of the current president. Eduardo Bolsonaro's page is also present in the subsequent analysis.

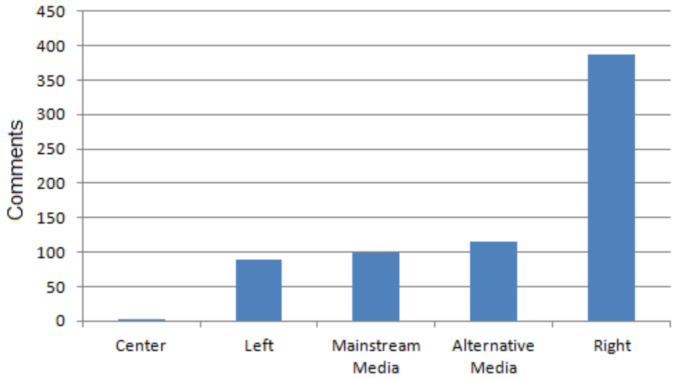
6.3 Toxicity in Comments

To delve deeper into what causes a user to publish toxic comments, we analyze all the messages whose toxicity score is 1 (the maximum score).

We first look into where those messages were collected among the five sub-groups (Figure 6.6). We see that right-wing pages have the largest proportion of the most toxic comments, and alternative media have a slightly more of the most toxic comments than mainstream media.

We find two relevant patterns by manual examination. First, as we said previously, Bolsonaro has a low proportion of toxic comments, but he still receives the extreme toxic

Figure 6.6: Distribution of comments with toxicity equal to 1 by page.



Source: The author.

Facebook Page	User Comment (Translated to English)	Toxicity
Magno Malta	Thug Son Of A B*tch! Fascist pastor! F*ck you!	0.99998
Jair Messias Bolsonaro	Bial asshole, always f*ck up d*uchebag, bial sh*tty communist dumb*ss	0.99998
Jornal da Cidade Online	Your Excellency, hyena son of a b*tch h*bo th*g garbage sewer rat	0.99996
Lula	Damn you old man damn people I want you all to go f*ck yourselves	0.99996
Ivan Valente	Die soon and go to hell, old bastard, disgusting communist !!!	0.99995
O Globo	you disgusting ridiculous plague go to hell !!!!!	0.99995
Jair Messias Bolsonaro	F*ck u you piece of sh*t cr*p president go do something for the country you piece of cr*p just know how to appear on TV	0.99994
Jornal da Cidade Online	Go sing in the bathroom, you idiot, f*ck you stop defending th*gs	0.99994
O Globo	This *sshole just talks sh*t.	0.99993
Eduardo Bolsonaro	This *sshole just talks sh*t.	0.99993

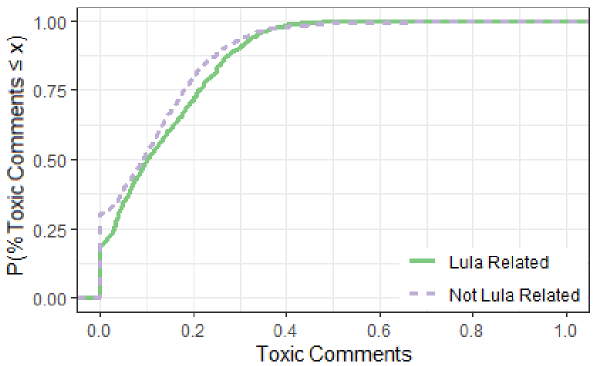
Table 6.3: Top 10 worst comments on the dataset with toxicity below 1.

messages (i.e., toxicity score is 1). It might imply that he does not actively remove the toxic comments, but due to his more extensive audience, the proportion of them stays at a low level. Or, those comments might not aim at Bolsonaro and thus are not flagged. Second, we also found that *Eduardo Bolsonaro* and *Jornal da Cidade Online*, an alternative media page, had more comments with maximum toxicity than the president. Eduardo's page has a quarter of the total messages of his father but three times more of these toxic comments. This result shows that the president's larger following agreeing with him might decrease his percentage.

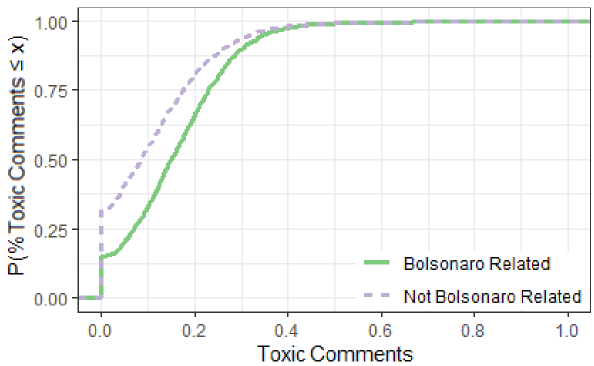
We also examine the content of the top 10 toxic comments with toxicity less than 1, as seen in Table 6.3. As we can see, the topic of politics is present in all of them, with Bolsonaro and Lula being in seven of the ten comments, with only one being about a supreme court judge instead. Nonetheless, hate towards these judges is driven by politics in Brazil. Given that Lula's page has very few toxic comments, it is surprising to see that he appeared in the messages with the highest toxicity. As our data collection covers the period close to the release of Lula, we searched him as a topic in the toxic posts by looking for his complete name and common alias in the post's text. We find that 8.24% of the posts of all pages are about him.

Figure 6.7a shows that the distribution of the proportion of toxic comments for posts about Lula is slightly higher, with a mean of 13.07% compared to an 11.35% mean for posts not related to him. This result indicates that people left toxic comments much more frequently in posts about him, and he becomes a controversial topic that attracts a lot of hate. Figure 6.7b shows that Bolsonaro has a similar effect. With 9.66% of all posts citing him, the average proportion of toxic comments in the posts mentioning him is 15.92%, which is higher than 11.02% from the publications without mentions about him.

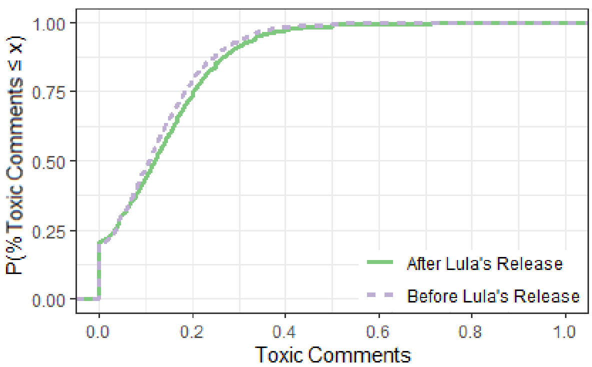
Figure 6.7: Cumulative Distribution Function (CDF) of toxic comments in relation to Lula and Jair Bolsonaro.



(a) Content related to Lula.



(b) Content related to Bolsonaro.



(c) Content before and after Lula's release.

Source: The author.

Thus, we find some characteristics of toxic comments, answering the **fourth research question**. It is possible to note that **politics is a common subject**, and despite the fact that pages of public figures receive fewer toxic messages, **comments about public figures tend to be more toxic**. And finally, such messages are not usually sent on the pages of these figures, but on **news pages talking about them**.

6.3.1 How Lula's Release affected the comments

Finally, we again analyze the **third question**, looking at how a political incident can change the proportion of toxic comments. As previously mentioned, this data collection is a valuable resource to see how online discussions were conducted around Lula's release.

Based on these data, the toxicity of the posts from a week before and a week after their departure was reported as eminent, on November 7, was compared. Figure 6.7c present the results. We found that the posts had, on average, 12.32 % of toxic comments before the event, and increased to 13.60 % after the decision. But when doing Pearson's chi-square test, as well as calculating the odds ratio of the data, we found that no significant correlation was found. This shows that a single political incident could alter the toxicity of the comments, but as our results are not significant, **we could not verify nor deny this possibility**.

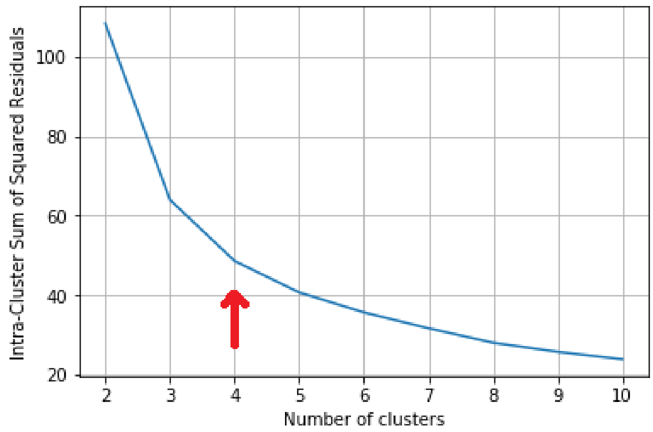
6.4 Posts Reactions

After looking at toxicity, in this section, we show the results of our examination of the posts reactions and how they correlate with the results of examining the toxicity. To accomplish that, we first create a profile of these reactions for each page of our more comprehensive dataset, the extended set that includes pages without interests.

We built this profile based on the proportion of each of these reactions from all received on each post. With that done, we group the pages of similar profiles using the K-means algorithm.

To define the number of clusters (K) in our data, we use the elbow method [61]. This method aims to apply an approximation of the variance explained by the clustering as a performance indicator. One possible approximation is the square of the distance between the points in each cluster and its centroid. This value can be seen as the sum of squared residuals (SSR). For each K-value, we calculate this metric. When K approaches the number of real clusters, the SSR shows a rapid decline, but when that number goes beyond that value, the same metric will continue to decline, but slower. Following this procedure, Figure 6.8 shows how the SSR changes for each value of K, where we can see the transition described when we have four clusters. To confirm this result we also evaluate the silhouette metric [51].

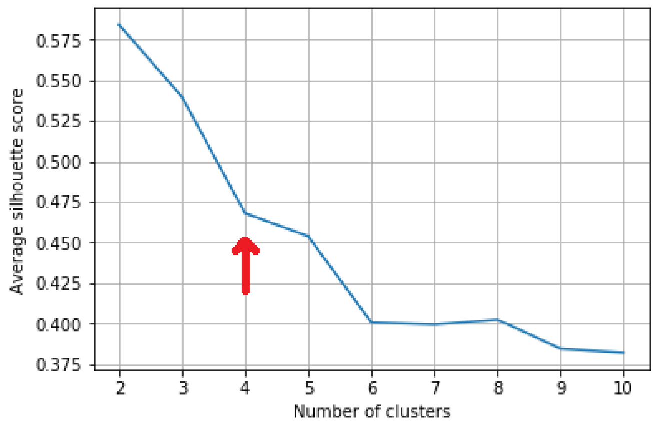
Figure 6.8: Intra-cluster sum of squared residuals for the best number of clusters for our pages.



Source: The author.

Figure 6.9 presents the variation of the average silhouette metric for each value of K. As in our analysis of the elbow method, there was a change in the curve's behavior when K equals four. As the value only diminishes when we increase K, we expect that our data does not have a high separation between all clusters, with some points being harder to place in one group or another. When we are dividing the data and reach K equal to four, all the cases of objects that are not divisible already affected the metric. From that point forward, we start breaking groups that have high tightness, only making separation worse.

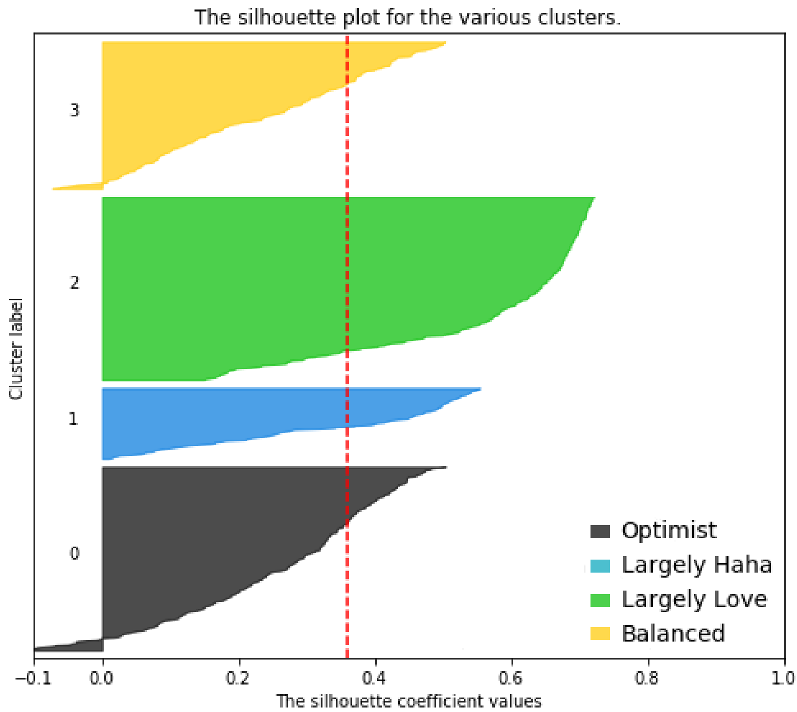
Figure 6.9: Silhouette metric for the best number of clusters for our pages.



Source: The author.

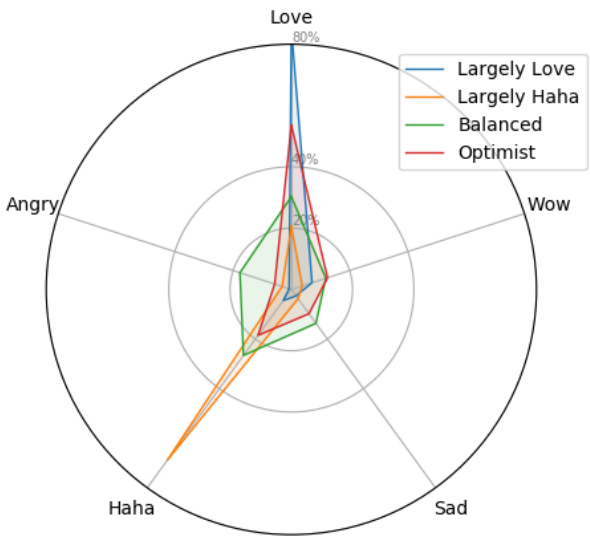
With the decision of using the number of K equal to four, Figure 6.10 shows the silhouette plot of these four clusters, and Figure 6.11 displays their reaction profiles. The silhouette plot shows one curve for each group, presenting the metric's value for each point in each of these groups and the average silhouette score of the whole clustering as a vertical line. What we can conclude is that the first and last clusters are the ones that contain the more difficult points to separate.

Figure 6.10: Final silhouette metric with the four clusters in total.



Source: The author.

Figure 6.11: Radar graph of all groups.



Source: The author.

Now analyzing the clusters, we can note that there is a group of pages that mainly receive the reaction of *Love*, one that mostly earn *Haha* as a response, and the last two types of pages receive a more balanced mix of all these reactions, with one still having a significant amount of Love reactions. To discover which pages are in each cluster, we analyze the page types, again adding to our comparison to previous sections.

6.4.1 Page Reaction Profiles relation with Page Types

Following our clustering, we now show the distribution of political bias and the type of the pages for the four clusters in Table 6.4.

In the first cluster, which we will call the **Largely Love** Cluster, we see a parallel to our results in toxicity. As mentioned in the previous analysis, pages from public figures receive very few toxic comments in the proportion of all messages. Following that trend, in this group, where the majority of reactions were *Love*, most pages are from public figures (44.81%). More specifically, this cluster had a large number of writers, artists, and authors, which were 71.52% of all these figures.

Reaction Profile	Political Bias	Page Type		
		Public Figures	Mainstream Media	Alternative Media
Largely Love	Left-wing	15,89%	20,97%	38,71%
	Centrist	17,22%	29,03%	18,55%
	Right-wing	66,89%	50,00%	42,74%
	Total Pages	151	62	124
Largely Haha	Left-wing	20,00%	7,69%	32,89%
	Centrist	4,00%	12,82%	10,53%
	Right-wing	76,00%	79,49%	56,58%
	Total Pages	25	39	76
Balanced	Left-wing	66,67%	15,15%	49,02%
	Centrist	9,52%	21,21%	9,80%
	Right-wing	23,81%	63,64%	41,18%
	Total Pages	21	33	51
Optimistic	Left-wing	32,76%	15,22%	35,80%
	Centrist	5,17%	13,04%	18,52%
	Right-wing	62,07%	71,74%	45,68%
	Total Pages	58	46	81

Table 6.4: Page reaction profile groups.

Nonetheless, the proportion of personalities directly linked to politics was also significant compared to other groups, with this cluster had the second largest proportion of politicians (30%). When we also look at the political bias, for all page types, the right-wing pages were the majority, possibly showing how popular pages with this politics is on Facebook. We discover fewer pages from mainstream media, circa 18.40%, significantly less compared to alternative media. This fact might be a reflection of the environment in these pages, that, as stated before, we expect to be cross-cutting across political lines more often than in the other pages.

In the second group, the **Largely Haha** cluster, we can see that the *Haha* reaction is prevalent. This group is mainly comprised of pages from media, with a considerable amount of them related to entertainment more than to news (64.35%). All comedians that comment on politics and pages that present headlines in a comedic format also exist in this group.

Although this result did not decidedly relate to a specific part of our analysis of toxicity, the high proportion of right-wing pages in this cluster again shows their popularity and might indicate a higher level of toxic posts considered toxic in these pages.

Now, examining the last group of pages, we found two groups with a balanced mix of responses. We call the first one the **Balanced** cluster. Because of our method to measure controversy by the entropy of the reactions, we expect this group to have the most controversial posts by definition. Interestingly enough, news media is mostly present here, with 71.43% of all the media pages exclusively related to news. This amounts to 37.31% of all pages focused only on news coverage in all clusters. This group also has the largest proportion of mainstream media at 31.43%. As disagreements of world view can generate controversy, this large proportion of media pages again reinforces our notion that the posts produced by these pages are present in an ecosystem where different political sides meet. This situation possibly creates more controversy and even toxic messages.

Finally, we examine one group very close to both the Balanced and Largely Love clusters, the **Optimist** group of pages, which is made of pages with a similar balanced proportion of reactions, with a more pronounced amount of *Love*.

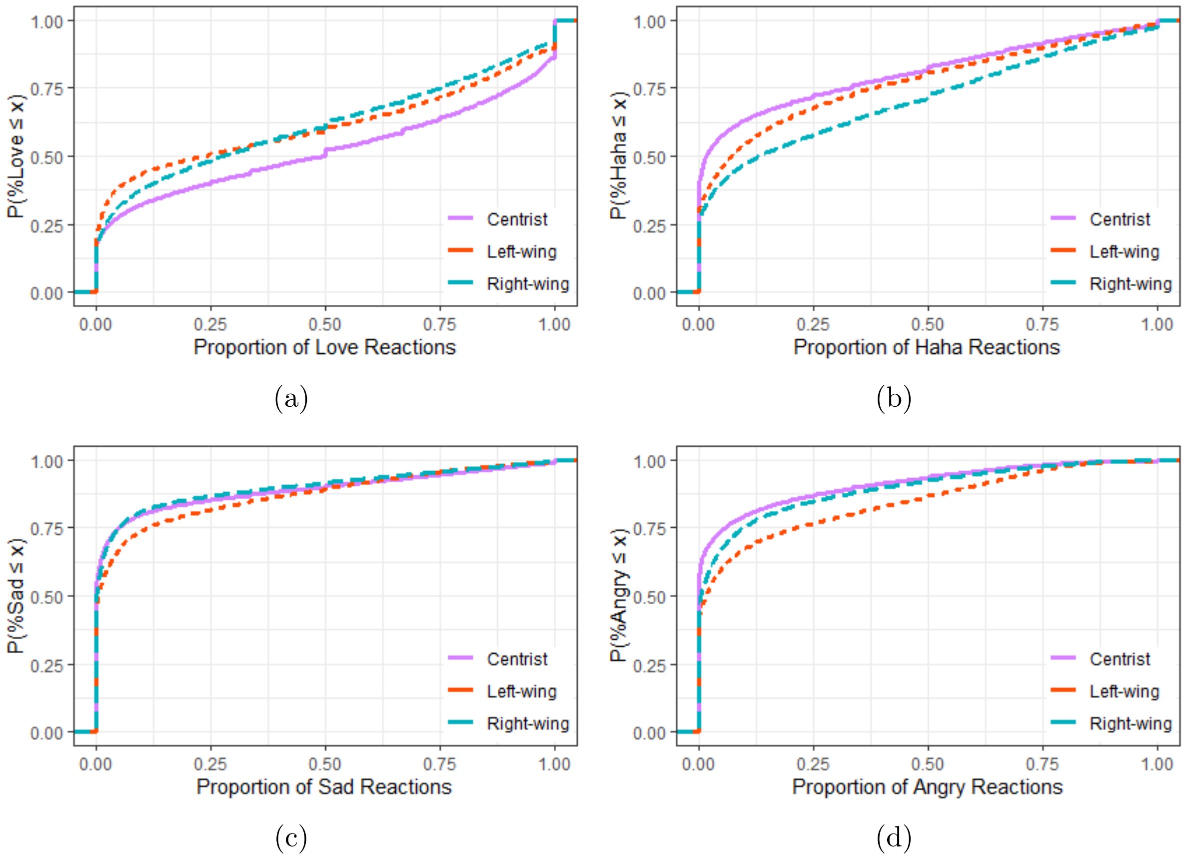
Different from the Balanced cluster, there is more public figures proportionally (31.35%), especially considering the presence of 41.67% of all politicians. It also has a higher proportion of entertainment media pages, with 62.70% of all media of this subset being more focused on entertainment than only news. What we see in this cluster is the same high frequency of Love towards public figures and less restrained alternative media. However, in this case, the pages receive more mixed responses in their reactions, showing a higher controversy, like pastor *Silas Malafaia* and the TV show host and Journalist *Ratinho*.

In resume, we have two clusters with more public figures which received more *Love* reactions, *Largely Love* and *Optimistic*, and ome cluster with more proportion of mainstream media, which received a more diverse set of reactions, showing a higher level of controversy, and finally, the *Largely Haha* set, which is mainly composed of right-wing pages.

6.4.2 Reactions relation to Political Biases and Page Types

Beyond the clusters of the reaction profiles, we also analyze the distribution of the posts by political bias, which Figures 6.12a to 6.12d show. For more negative emotions, like *Angry* and *Sad*, posts from left-wing pages receive significantly more of those reactions than the other political sides.

Figure 6.12: Cumulative Distribution function (CDF) of the reactions proportions grouped by political bias



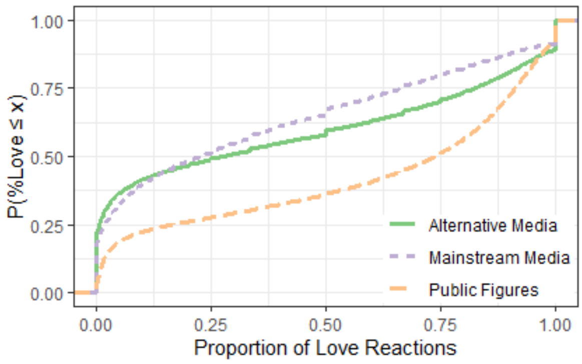
Source: The author.

When we consider only the *Angry* reaction, we see that the users give statistically less amounts of it to right-wing pages compared to centrist ones. The same right-leaning pages have, with statistical significance, the posts with the most of *Haha* reactions of all three ideologies. Finally, for the *Wow* reaction, we did not find any statistically significant differences.

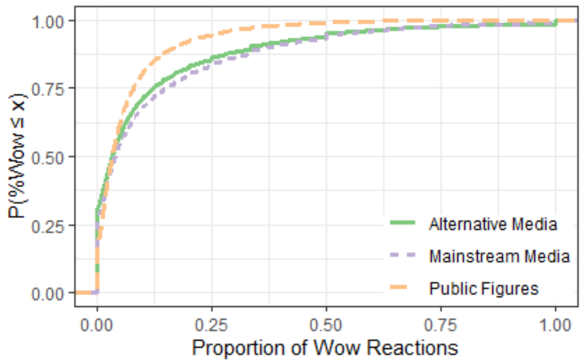
Meanwhile, when we look into the relationship between reactions and page types, the statistical significance is not found for the two more negative responses. For *Haha* and *Wow*, we have the following relation of the proportion of the reactions: Mainstream Media > Alternative Media > Public Figures, as shown in Figures 6.13a to 6.13c. The exact reversed relationship appears for the *Love* reaction. This discrepancy between *Haha* and *Love* might indicate that the *Haha* reaction might be used as a negative reaction sometimes.

Finally, analyzing our measure of controversy, the entropy of the reactions, we found that mainstream media posts have a higher entropy than both other types, and centrist pages have less entropy than pages of other biases, in both cases with statistical significance ($p < 0.001$). Figures 6.14a and 6.14b show the CDFs for these results.

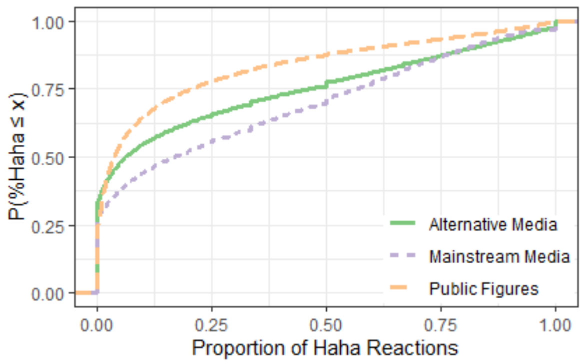
Figure 6.13: Cumulative density Function (CDF) of the reactions proportions grouped by page type



(a)



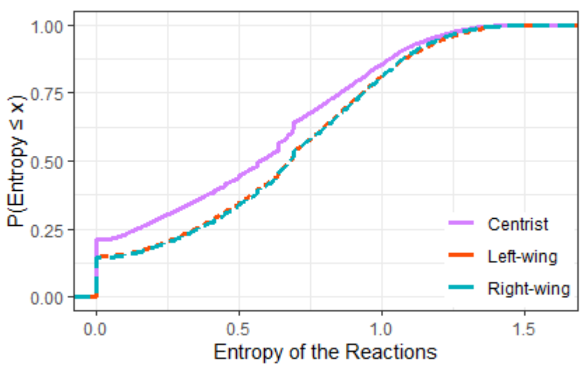
(b)



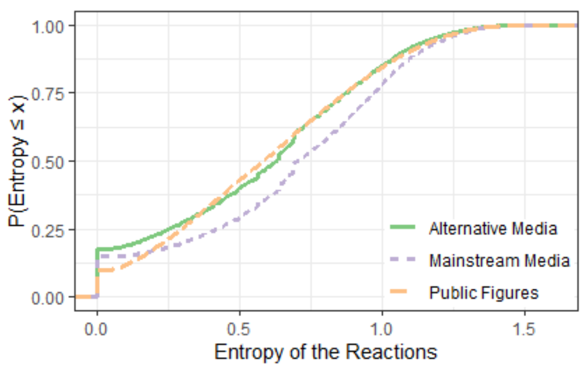
(c)

Source: The author.

Figure 6.14: Cumulative Distribution function (CDF) of the entropy grouped by political bias and page type



(a)



(b)

Source: The author.

6.4.3 Reactions and Toxicity

Lastly, we can unite the toxicity analysis with the reaction data, using the intercession between the preliminary dataset and the extended one. Table 6.5 shows this interposition between the data with toxicity measured and the page profiles we found using reactions.

Cluster	Pages	Avg. Toxic Comments	Avg. Toxic Posts	Avg. Entropy	Worse Toxic Comments	Worse Toxic Posts
Largely Love	11	10.12%	1.00%	0,61468	Silas Malafaia (20.73%)	Carlos Bolsonaro (6.00%)
Optimist	9	12.88%	1.10%	0,716406	Maria do Rosário (21.46%)	Ciro Gomes (2.67%)
Balanced	18	17.11%	0.73%	0,816313	Ivan Valente (23.33%)	Jornal da Cidade Online (5.00%)
Largely Haha	14	19.14%	1.91%	0,808381	Falando Verdades (25.28%)	Olavo de Carvalho (15.00%)

Table 6.5: Page reaction profiles of the pages with toxicity data

What we found is that the page profiles can be ranked by the average percentage of toxic comments, with *Largely Love* having the least proportion of these messages and the *Largely Haha* with the most.

As we saw that alternative media has the largest proportion of toxic comments of all three types, and indeed the order *Largely Love*, *Optimist*, *Balanced*, and *Largely Haha* follows how much of their pages are alternative media which is, respectively, 36.80%, 43.78%, 48.57%, and 54.29%. The inverse happens to public figures that are more prevalent in the *Largely Love* cluster and have the least toxic comments.

Along with the clusters' information, we also present the pages with the worst proportion of toxic comments and posts by group. Table 6.5 shows that the pages that receive the most toxicity were mainly left-wing pages, with most of these toxic messages being on alternative media pages. Meanwhile, the pages with more toxic posts were, for the most part, right-wing, with the one that produced the most toxic posts being from a right-wing public figure.

Another thing showed in the Table 6.5 is the average entropy. There is a clear correlation of the average entropy to the average percentage of toxic comments, which can be measured as $\rho = 0.957$. However, it is not significant due to the large confidence interval this small sample has. Nonetheless, as it indicates the possibility of a real correlation, we examine the same metric by page. In that case, the correlation drops to 0.377, but now with a confidence interval from 0.116 to 0.589 and a p-value of 0.00589. Finally, going to the level of the post itself, the correlation diminishes to 0.14. However, using Pearson's chi-square test, we were able to find that posts with entropy above the average also had toxic comments above the mean in our data, with significance (χ^2=374.01, $p < 0.005$).

At last, we also searched for a relationship with individual types of responses. From the five reactions we collected, *Angry*, *Love*, and *Haha* had the best correlation with the percentage of toxic messages, respectively, a correlation of 0.581, -0.572, and 0.461 when looking by page. Similarly to entropy, when applied to posts, all correlations drop below 0.3.

Nevertheless, we find by using the Pearson's chi-square test that posts with one of these reactions above the average also have toxic comments above the mean. In this case, the post's popularity might be what creates this dependency. Still, it is possible to use these reactions as indicators of which publications we need to prioritize when searching or inhibiting hateful content.

To summarize, when we analyze reactions, we find similar findings to our analysis of toxicity. There are significant differences between the three types of pages, with **public figures having a lower proportion of toxic messages and high levels of positive reactions**. Meanwhile, **mainstream media receives more toxic comments and a more diverse set of reactions by post**. Analyzing the correlation between toxicity and post reactions, **the alternative media impacts which page reaction profiles have the highest levels of toxicity**.

Chapter 7

Concluding Discussion and Future Work

Social media platforms have changed news consumption patterns. Alternative Media proliferates in this new environment, and together with public figures official pages, they affect public perception of affairs, sometimes having politically biased coverage. Especially, Brazil sees a surge in the usage of social networks as news-gathering tools, with a great focus on Facebook. Still, little is know about their political leanings and audiences. To bridge this gap, we present the following contributions:

- A novel method to estimate the ideological bias of news related Facebook pages by graph-based semi-supervised learning;
- An examination of the audience of these pages;
- A detailed characterization of the toxicity of 4,095,919 comments online in 15,825 posts from 63 Facebook pages, collected between October 27th, 2019 and November 16th, 2019;
- An analysis of the reactions given to posts of 767 pages from May 15th, 2019 to November 30th, 2019.

Our methodology has an advantage in its applicability, which applies to any other country where Facebook marketing API is available. We tested four different learning algorithms and compared them with multiple datasets of different methods for estimating ideological leaning [47, 3, 38, 6]. This test showed a high correlation of our method with most of them, particularly with other Facebook-based data [47, 3]. In the audience analysis, we found an older and more male audience for public figures, confirming similar trends from other works, now adding an examination of relationship status. Finally, we also found a mild correlation between political bias and the education level, which showed pages further to the left having an audience that self-described as more educated.

After dividing the pages into Public Figures, Mainstream Media, or Alternative Media, we analyzed their toxicity score in various levels of granularity. Our results disclosed a series of relevant trends.

We find that the toxic comments and posts are the minority, being 13.27% and 0.84% of the total, respectively, but there was a concentration. About 20% of pages were responsible for nearly 60% of toxic comments and 56% of all toxic posts.

In general, comments replying to other commentaries are more toxic than conventional responses, and there are more of these messages on pages with more toxic publications, even when the post and toxicity score of comments do not have a high correlation. In our division, news media pages receive more toxicity than pages from public figures, but when a post cites public figures, the proportion of toxic messages increases. The political affiliation of the public figures did not affect the percentage of toxic comments or posts, but mainstream media receives more toxic comments than the alternative media. Finally, we found that Lula's Release shows that a determined political event might increase toxicity, but the effect might not be significant. We expect that these insights could help improve and guide Content Policies for Facebook and comment sections in general because it shows where the toxicity appears, how concentrated it is, what factors affect it, and the typical characteristics of its content.

Examining the posts reactions, we confirm a series of trends we found in our toxicity results. Using the most comprehensive dataset we constructed, we see that correlations between the toxicity and the reactions exist. We present a group of page reaction profiles that showed relevant distinctions between the pages, with some of them also related to our division of three types of pages. In the end, we also used the concept of entropy from information theory to measure the controversy of a post. We calculate this metric with the reaction percentages. Similar to the post's toxicity, the entropy also showed to have a dependency on the proportion of toxic messages the post receives.

With these results, we expect that the main factors that influence the toxic comments associated with the news can help social media platforms in the design of content policies capable of minimizing this problem, especially in Brazil. In particular, the methodology shown can easily be replicated in other countries. This replication is possible as the division of the pages into the five subcategories presented and the use of *Audience Insights* do not depend on the analyzed country being Brazil. Also, although our results are not concerned with finding an easy path to detect or predict the toxicity or hate on a post, many of the analysis and metrics used here could turn into features that machine learning algorithms can use to fill this research gap.

For future research, we expect to improve some aspects of our political bias algorithm with the intent to increase the method's scalability so more pages can be analyzed, especially in other countries. Besides that, a more thorough study of the toxicity metric judgments and what are the alternatives could also add to our contributions.

Bibliography

[1] Jisun An, Haewoon Kwak, Oliver Posegga, and Andreas Jungherr. Political discussions in homogeneous and cross-cutting communication spaces. In *Proceedings of the 13th International AAAI Conference on Web and Social Media*, pages 68–79. AAAI Press, 2019.

[2] Matheus Araujo, Yelena Mejova, Ingmar Weber, and Fabricio Benevenuto. Using facebook ads audiences for global lifestyle disease surveillance: Promises and limitations. In *Proceedings of the 9th ACM on Web Science Conference*, pages 253–257. Association for Computing Machinery, 2017.

[3] Eytan Bakshy, Solomon Messing, and Lada A Adamic. Exposure to ideologically diverse news and opinion on facebook. *Science*, 348(6239):1130–1132, 2015.

[4] Samuel Barros and Rodrigo Carreiro. A discussão pública e as redes sociais online: o comentário de notícias no facebook. *Fronteiras-estudos midiáticos*, 17(2):174–185, 2015.

[5] Angelo Basile, Tommaso Caselli, and Malvina Nissim. Predicting controversial news using facebook reactions. In *Proceedings of the 4th Italian Conference on Computational Linguistics*, pages 12–17. Accademia University Press, 2017.

[6] Ceren Budak, Sharad Goel, and Justin M Rao. Fair and balanced? quantifying media bias through crowdsourced content analysis. *Public Opinion Quarterly*, 80(S1):250–271, 2016.

[7] Cambridge Dictionary. Profanity. `https://dictionary.cambridge.org/dictionary/english/profanity`, 2019. Accessed May 5th, 2019.

[8] Cambridge Dictionary. Hate speech. `https://dictionary.cambridge.org/dictionary/english/hate-speech`, 2020. Accessed Nov 25th, 2020.

[9] Jacob Cohen. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46, 1960.

[10] Michael D Conover, Bruno Gonçalves, Jacob Ratkiewicz, Alessandro Flammini, and Filippo Menczer. Predicting the political alignment of twitter users. In *Proceedings of the 2011 IEEE International Conference on Privacy, Security, Risk and Trust and*

2011 IEEE International Conference on Social Computing, pages 192–199. IEEE, 2011.

[11] Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27, 1967.

[12] Dave D'Alessio and Mike Allen. Media bias in presidential elections: A meta-analysis. *Journal of communication*, 50(4):133–156, 2000.

[13] Thomas Davidson, Dana Warmsley, Michael Macy, and Ingmar Weber. Automated hate speech detection and the problem of offensive language. In *Proceedings of the 11th International AAAI Conference on Web and Social Media*, pages 512–515. AAAI Press, 2017.

[14] Rogers Prates de Pelle and Viviane P Moreira. Offensive comments in the brazilian web: a dataset and baseline results. In *Proceedings of the 6th Brazilian Workshop on Social Network Analysis and Mining*, pages 510–519. SBC, 2017.

[15] James N Druckman and Michael Parkin. The impact of media bias: How editorial slant affects voters. *The Journal of Politics*, 67(4):1030–1049, 2005.

[16] Mai ElSherief, Shirin Nilizadeh, Dana Nguyen, Giovanni Vigna, and Elizabeth Belding. Peer to peer hate: Hate speech instigators and their targets. In *Proceedings of the 12th International AAAI Conference on Web and Social Media*, pages 52–61. AAAI Press, 2018.

[17] Robert M Entman. Framing bias: Media in the distribution of power. *Journal of communication*, 57(1):163–173, 2007.

[18] Facebook. Facebook reports fourth quarter and full year 2019 results. https://investor.fb.com/investor-news/press-release-details/2020/Facebook-Reports-Fourth-Quarter-and-Full-Year-2019-Results/default.aspx, 2020. Accessed March 10th, 2020.

[19] Robert W Floyd. Algorithm 97: shortest path. *Communications of the ACM*, 5(6):345, 1962.

[20] Paula Fortuna and Sérgio Nunes. A survey on automatic detection of hate speech in text. *ACM Computing Surveys (CSUR)*, 51(4):85, 2018.

[21] Antigoni Founta, Constantinos Djouvas, Despoina Chatzakou, Ilias Leontiadis, Jeremy Blackburn, Gianluca Stringhini, Athena Vakali, Michael Sirivianos, and Nicolas Kourtellis. Large scale crowdsourcing and characterization of twitter abusive behavior. In *Proceedings of the 12th International AAAI Conference on Web and Social Media*, pages 491–500. AAAI Press, 2018.

[22] Iginio Gagliardone, Danit Gal, Thiago Alves, and Gabriela Martinez. *Countering online Hate Speech*. Unesco Publishing, 2015.

[23] Sarah Anne Ganter and Fernando Oliveira Paulino. Between attack and resilience: The ongoing institutionalization of independent digital journalism in brazil. *Digital Journalism*, pages 1–20, 2020.

[24] Felipe Taliar Giuntini, Larissa Pires Ruiz, Luziane De Fatima Kirchner, Denise Aparecida Passarelli, Maria De Jesus Dutra Dos Reis, Andrew Thomas Campbell, and Jo Ueyama. How do i feel? identifying emotional expressions on facebook reactions using clustering mechanism. *IEEE Access*, 7:53909–53921, 2019.

[25] Jennifer Golbeck, Zahra Ashktorab, Rashad O Banjo, Alexandra Berlinger, Siddharth Bhagwan, Cody Buntain, Paul Cheakalos, Alicia A Geller, Quint Gergory, Rajesh Kumar Gnanasekaran, et al. A large labeled corpus for online harassment research. In *Proceedings of the 9th ACM Conference on Web Science*, pages 229–233. Association for Computing Machinery, 2017.

[26] Tony Harcup. "i'm doing this to change the world": journalism in alternative and mainstream media. *Journalism Studies*, 6(3):361–374, 2005.

[27] Brad Haynes. Facebook retira do ar rede ligada ao mbl antes das eleições. https://br.reuters.com/article/idBRKBN1KF1MI-OBRDN, 2018. Accessed September 30th, 2020.

[28] Sanne Hille and Piet Bakker. Engaging the social news user: Comments on news sites and facebook. *Journalism Practice*, 8(5):563–572, 2014.

[29] Kristoffer Holt. Alternative media and the notion of anti-systemness: Towards an analytical framework. *Media and Communication*, 6(4):49–57, 2018.

[30] Kristoffer Holt, Tine Ustad Figenschou, and Lena Frischlich. Key dimensions of alternative news media. *Digital Journalism*, 7(7):860–869, 2019.

[31] Jigsaw. Perspective api. https://www.perspectiveapi.com/, 2017. Accessed May 24th, 2019.

[32] Thorsten Joachims. Transductive learning via spectral graph partitioning. In *Proceedings of the 20th International Conference on Machine Learning*, pages 290–297. AAAI Press, 2003.

[33] Shreeharsh Kelkar. Post-truth and the search for objectivity: political polarization and the remaking of knowledge production. *Engaging Science, Technology, and Society*, 5:86–106, 2019.

[34] Sabih Ahmad Khan and Hsien-Tsung Chang. Comparative analysis on facebook post interaction using dnn, elm and lstm. *PloS ONE*, 14(11):1–26, 2019.

[35] Varada Kolhatkar and Maite Taboada. Constructive language in news comments. In *Proceedings of the First Workshop on Abusive Language Online*, pages 11–17. Association for Computational Linguistics, 2017.

[36] Ramona Kreis. The "tweet politics" of president trump. *Journal of Language and Politics*, 16(4):607–618, 2017.

[37] Aibek Makazhanov and Davood Rafiei. Predicting political preference of twitter users. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, page 298–305. Association for Computing Machinery, 2013.

[38] Amy Mitchell. *Political polarization & media habits.* Pew Research Center, 2014.

[39] Tobias Moers, Florian Krebs, and Gerasimos Spanakis. Semtec: social emotion mining techniques for analysis and prediction of facebook post reactions. In *Proceedings of the 10th International Conference on Agents and Artificial Intelligence*, pages 361–382. Springer International Publishing, 2018.

[40] Mainack Mondal, Leandro Araújo Silva, and Fabrício Benevenuto. A measurement study of hate speech in social media. In *Proceedings of the 28th ACM Conference on Hypertext and Social Media*, pages 85–94. Association for Computing Machinery, 2017.

[41] Marcio Moretto and Pablo Ortellado. Quanto mais velhos, mais polarizados: Perfil dos usuários que interagem com páginas de notícias no facebook. Technical Report 1, Monitor do Debate Político no Meio Digital, 2018.

[42] Sendhil Mullainathan and Andrei Shleifer. Media bias. Technical report, National Bureau of Economic Research, 2002.

[43] Nic Newman. Mainstream media and the distribution of news in the age of social media. RISJ Report 2011, Reuters Institute for the Study of Journalism, 2011.

[44] Nic Newman, Richard Fletcher, Antonis Kalogeropoulos, and Rasmus Nielsen. *Reuters Institute digital news report 2019.* Reuters Institute for the Study of Journalism, 2019.

[45] Julio Reis, Fabricio Benevenuto, Pedro Vaz de Melo, Raquel Prates, Haewoon Kwak, and Jisun An. Breaking the news: First impressions matter on online news. In *Proceedings of the 9th International AAAI Conference on Web-Blogs and Social Media*, pages 357–366. AAAI Press, 2015.

[46] Filipe N. Ribeiro, Fabrício Benevenuto, and Emilio Zagheni. How biased is the population of facebook users? comparing the demographics of facebook users and census data to generate correction factors. In *Proceedings of the 12th ACM Web Science Conference*, pages 325—-334. Association for Computing Machinery, 2020.

[47] Filipe N Ribeiro, Lucas Henrique, Fabricio Benevenuto, Abhijnan Chakraborty, Juhi Kulshrestha, Mahmoudreza Babaei, and Krishna P Gummadi. Media bias monitor: Quantifying biases of social media news outlets at large-scale. In *Proceedings of the 12th International AAAI Conference on Web and Social Media*, pages 290—-299. AAAI Press, 2018.

[48] Filipe N Ribeiro, Daniel Kansaon, and Fabrício Benevenuto. Leveraging the facebook ads platform for election polling. In *Proceedings of the 25th Brazillian Symposium on Multimedia and the Web*, pages 305–312. Association for Computing Machinery, 2019.

[49] Richard Romancini and Fernanda Castilho. Strange fruit: The rise of brazil's 'new right-wing'and the non-partisan school movement. *Journal of Alternative and Community Media*, 4(1):7–22, 2019.

[50] Björn Ross, Michael Rist, Guillermo Carbonell, Ben Cabrera, Nils Kurowsky, and Michael Wojatzki. Measuring the reliability of hate speech annotations: The case of the european refugee crisis. In *Bochumer linguistische Arbeitsberichte 17, 3rd Workshop on Natural Language Processing for Computer-Mediated Communication*, pages 6–9. Sprachwissenschaftliches Institut, 2016.

[51] Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.

[52] Adrienne Russell. The arab spring— extra-national information flows, social media and the 2011 egyptian uprising. *International Journal of Communication*, 5:10, 2011.

[53] Joni Salminen, Hind Almerekhi, Milica Milenković, Soon Gyo Jung, Jisun An, Haewoon Kwak, and Bernard James Jansen. Anatomy of online hate: Developing a taxonomy and machine learning models for identifying and classifying hate in online news media. In *Proceedings of the 12th International AAAI Conference on Web and Social Media*, pages 330–339. AAAI press, 2018.

[54] Elisa Shearer. Social media outpaces print newspapers in the u.s. as a news source. https://www.pewresearch.org/fact-tank/2018/12/10/social-media-outpaces-print-newspapers-in-the-u-s-as-a-news-source/, 2018. Accessed December 15th, 2019.

[55] Elisa Shearer and Katerina Eva Matsa. News use across social media platforms 2018. https://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/, 2018. Accessed December 15th, 2019.

[56] Yongren Shi, Kai Mast, Ingmar Weber, Agrippa Kellum, and Michael Macy. Cultural fault lines and political polarization. In *Proceedings of the 9th ACM on Web Science Conference*, pages 213–217. Association for Computing Machinery, 2017.

[57] Leandro Araújo Silva, Mainack Mondal, Denzil Correa, Fabrício Benevenuto, and Ingmar Weber. Analyzing the targets of hate in online social media. In *Proceedings of the 10th International AAAI Conference on Web and Social Media*, pages 687–690. AAAI Press, 2016.

[58] Kate Starbird. Examining the alternative media ecosystem through the production of alternative narratives of mass shooting events on twitter. In *Proceedings of the 11th International AAAI Conference on Web and Social Media*, pages 230–239. AAAI Press, 2017.

[59] Natalie Jomini Stroud, Emily Van Duyn, and Cynthia Peacock. Survey of commenters and comment readers. https://mediaengagement.org/research/survey-of-commenters-and-comment-readers/, 2016. Accessed February 21th, 2020.

[60] Leona Yi-Fan Su, Michael A Xenos, Kathleen M Rose, Christopher Wirz, Dietram A Scheufele, and Dominique Brossard. Uncivil and personal? comparing patterns of incivility in comments on the facebook pages of news outlets. *New Media & Society*, 20(10):3678–3699, 2018.

[61] Robert L Thorndike. Who belongs in the family? *Psychometrika*, 18(4):267–276, 1953.

[62] Ye Tian, Thiago Galery, Giulio Dulcinati, Emilia Molimpakis, and Chao Sun. Facebook sentiment: Reactions and emojis. In *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*, pages 11–16. Association for Computational Linguistics, 2017.

[63] Luis Fernando Toledo. Maior rede pró-bolsonaro do facebook é excluída após denúncia do 'estado'. https://politica.estadao.com.br/noticias/eleicoes,maior-rede-pro-bolsonaro-do-facebook-e-excluida-apos-denuncia-do-estado,70002558430, 2018. Accessed May 15th, 2020.

[64] Luis Fernando Toledo. Rede pró-bolsonaro engaja mais do que madonna e neymar. https://politica.estadao.com.br/noticias/eleicoes,

rede-pro-bolsonaro-engaja-mais-do-que-madonna-e-neymar,70002544629, 2018. Accessed May 15th, 2020.

[65] Carolina Vieira, Filipe Ribeiro, Pedro Olmo Vaz de Melo, Fabricio Benevenuto, and Emilio Zagheni. Using facebook data to measure cultural distance between countries: The case of brazilian cuisine. In *Proceedings of The Web Conference 2020*, pages 3091–3097. Association for Computing Machinery, 2020.

[66] Zeerak Waseem, Thomas Davidson, Dana Warmsley, and Ingmar Weber. Understanding abuse: A typology of abusive language detection subtasks. In *Proceedings of the First Workshop on Abusive Language Online*, pages 78–84. Association for Computational Linguistics, 2017.

[67] Brian E Weeks and R Lance Holbert. Predicting dissemination of news content in social media: A focus on reception, friending, and partisanship. *Journalism & Mass Communication Quarterly*, 90(2):212–232, 2013.

[68] Dengyong Zhou, Olivier Bousquet, Thomas N Lal, Jason Weston, and Bernhard Schölkopf. Learning with local and global consistency. In *Proceedings of the 16th International Conference on Neural Information Processing Systems*, pages 321–328. MIT Press, 2003.

[69] Xiaojin Zhu, Zoubin Ghahramani, and John D Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. In *Proceedings of the 20th International Conference on Machine Learning*, pages 912–919. AAAI Press, 2003.

Appendix A

List of Analyzed Pages

Page Name	Page Type	Page Category	Toxic Posts	Toxic Comments	Comemnts Avg. Toxicity
RENOVA Mídia	Alternative Media	News & Media Website	0,00190	0,19435	0,43628
Mídia Ninja	Alternative Media	Nonprofit Organization	0,00400	0,08641	0,28362
Brasil 247	Alternative Media	Broadcasting & Media Production Company	0,00800	0,21978	0,46385
Brasil de Fato	Alternative Media	Broadcasting & Media Production Company	0,00000	0,21638	0,41722
Pragmatismo Político	Alternative Media	Media/News Company	0,04000	0,22764	0,46444
Caneta Desesquerdizadora	Alternative Media	Media	0,08000	0,20828	0,42317
Jornalistas Livres	Alternative Media	News & Media Website	0,01600	0,12061	0,33529
O Antagonista	Alternative Media	News & Media Website	0,00640	0,16628	0,42248
Falando Verdades	Alternative Media	Media/News Company	0,04000	0,25282	0,46753
Spotniks	Alternative Media	Media/News Company	0,00000	0,11048	0,28157
Diário do Brasil	Alternative Media	Interest	0,02400	0,25164	0,47803
Diário Causa Operária	Alternative Media	Media/News Company	0,00000	0,22915	0,46048
Jornal da Cidade Online	Alternative Media	Media/News Company	0,05000	0,21863	0,44516
Crítica Nacional	Alternative Media	Media/News Company	0,00000	0,10468	0,36000
Diário do Poder	Alternative Media	News & Media Website	0,00000	0,16686	0,41020
Diário do Centro do Mundo	Alternative Media	Media/News Company	0,01500	0,20258	0,44038
Senso Incomum	Alternative Media	Media/News Company	0,05333	0,20218	0,43564
Terça Livre TV	Alternative Media	Magazine	0,02000	0,16500	0,37836

Table A.1: Preliminary Dataset with toxicity data

Page Name	Page Type	Page Category	Toxic Posts	Toxic Comments	Comemnts Avg. Toxicity
Esquerda Diário	Alternative Media	News & Media Website	0,00000	0,23108	0,46956
The Intercept Brasil	Alternative Media	Media/News Company	0,01333	0,19642	0,42143
Conexão Política	Alternative Media	Broadcasting & Media Company	0,03200	0,16593	0,38749
Geraldo Alckmin	Centrist Public Figure	Public Figure	0,00000	0,06354	0,24844
Marina Silva	Centrist Public Figure	Politician	0,00000	0,12789	0,38604
Ana Amélia Lemos	Centrist Public Figure	Politician	0,00000	0,03418	0,20548
Michel Temer	Centrist Public Figure	Politician	0,00000	0,03346	0,23583
Manuela D'Ávila	Left-wing Public Figure	Politician	0,01000	0,08423	0,29973
Marcelo Freixo	Left-wing Public Figure	Public Figure	0,02000	0,20982	0,44106
Jaques Wagner	Left-wing Public Figure	Politician	0,00000	0,06115	0,23351
Gleisi Hoffmann	Left-wing Public Figure	Politician	0,00571	0,06458	0,25373
Maria do Rosário	Left-wing Public Figure	Politician	0,00000	0,21461	0,42923
Ivan Valente	Left-wing Public Figure	Public Figure	0,02000	0,23332	0,45693
Lula	Left-wing Public Figure	Politician	0,00000	0,02404	0,18437
Alessandro Molon	Left-wing Public Figure	Politician	0,00000	0,16362	0,40991
Eduardo Suplicy	Left-wing Public Figure	Politician	0,00000	0,04043	0,22854
Guilherme Boulos	Left-wing Public Figure	Public Figure	0,02667	0,14836	0,36993
Fernando Haddad	Left-wing Public Figure	Politician	0,00000	0,10949	0,30482
Ciro Gomes	Left-wing Public Figure	Public Figure	0,02667	0,06670	0,26783
Marco Antonio Villa	Right-wing Public Figure	Public Figure	0,00000	0,15950	0,37006
João Doria	Right-wing Public Figure	Politician	0,00000	0,11450	0,33369
Flavio Bolsonaro	Right-wing Public Figure	Public Figure	0,00000	0,06439	0,24338
Marco Feliciano	Right-wing Public Figure	Politician	0,00000	0,14516	0,38445
Alvaro Dias	Right-wing Public Figure	Public Figure	0,00000	0,10006	0,33129
Jair Messias Bolsonaro	Right-wing Public Figure	Public Figure	0,00000	0,05775	0,22980
Eduardo Bolsonaro	Right-wing Public Figure	Public Figure	0,01000	0,15531	0,35415
Silas Malafaia	Right-wing Public Figure	Public Figure	0,00000	0,20729	0,40364
Olavo de Carvalho	Right-wing Public Figure	Author	0,15000	0,13788	0,36264
Carlos Bolsonaro	Right-wing Public Figure	Politician	0,06000	0,16125	0,38293
Magno Malta	Right-wing Public Figure	Politician	0,04000	0,14880	0,33584

Table A.2: Preliminary Dataset with toxicity data (Part II)

Page Name	Page Type	Page Category	Toxic Posts	Toxic Comments	Comemnts Avg. Toxicity
VEJA	Mainstream Media	Magazine	0,00000	0,13764	0,36594
Exame	Mainstream Media	News & Media Website	0,00190	0,16111	0,39318
Portal R7	Mainstream Media	News & Media Website	0,00545	0,15207	0,34628
Época	Mainstream Media	Magazine	0,00000	0,21494	0,45738
Jovem Pan News	Mainstream Media	Broadcasting & Media Production Company	0,00571	0,12085	0,34990
BBC News Brasil	Mainstream Media	Media/News Company	0,00000	0,14616	0,37715
DW Brasil	Mainstream Media	Media/News Company	0,00500	0,17569	0,43037
O Globo	Mainstream Media	Media/News Company	0,00000	0,20798	0,44604
Estadão	Mainstream Media	News & Media Website	0,00000	0,20640	0,45650
UOL Notícias	Mainstream Media	Media/News Company	0,00462	0,17603	0,41896
Revista ISTOÉ	Mainstream Media	Magazine	0,00190	0,19286	0,42314
G1 - O Portal de Notícias da Globo	Mainstream Media	Media/News Company	0,00190	0,12289	0,34414
Valor Econômico	Mainstream Media	Media/News Company	0,00000	0,17607	0,41039
CartaCapital	Mainstream Media	Media/News Company	0,00444	0,20567	0,44840
EL PAÍS Brasil	Mainstream Media	News & Media Website	0,00000	0,16851	0,39827

Table A.3: Preliminary Dataset with toxicity data (Part III)

Page Name	Page Type	Page Category	Political bias	Std. of bias	Political leaning
Catraca Livre São Paulo	Alternative Media	Media/News Company	-0,456764176	0,08259712	Left-wing
China Xinhua News	Alternative Media	Media/News Company	-0,143672013	0,315627112	Centrist
Migalhas	Alternative Media	Media/News Company	-0,179794994	0,126266629	Left-wing
Quebrando o Tabu	Alternative Media	Media/News Company	-0,289737453	0,06539438	Left-wing
Revista Fórum	Alternative Media	Media/News Company	-0,669837228	0,061867007	Left-wing
Pragmatismo Político	Alternative Media	Media/News Company	-0,562727514	0,066033907	Left-wing
Brasil 247	Alternative Media	Broadcasting & Media Production Company	-0,714513425	0,049352333	Left-wing
iG	Alternative Media	Broadcasting & Media Production Company	0,036672275	0,080392217	Centrist
Le Monde Diplomatique Brasil	Alternative Media	Newspaper	-0,906577899	0,024201836	Left-wing

Table A.4: Extended Dataset with political bias data

Page Name	Page Type	Page Category	Political bias	Std. of bias	Political leaning
The Intercept Brasil	Alternative Media	Local Business	-0,747561202	0,042030644	Left-wing
Partido Comunista do Brasil	Alternative Media	Local Business	-0,47825876	0,05746958	Left-wing
PT - Partido dos Trabalhadores	Alternative Media	Political Party	-0,290142499	0,066880036	Left-wing
PSDB	Alternative Media	Political Party	0,213095277	0,091143169	Right-wing
Anti-PT	Alternative Media	Political Party	0,530644495	0,110962735	Right-wing
CQC na Band	Alternative Media	TV Show	0,276476725	0,061865743	Right-wing
The Noite com Danilo Gentili	Alternative Media	TV Show	0,351131691	0,048930611	Right-wing
VICE Brasil	Alternative Media	Magazine	-0,209605774	0,090471451	Left-wing
Brasil de Fato	Alternative Media	Business Service	-0,73952862	0,049288401	Left-wing
SPIEGEL ONLINE	Alternative Media	Website	-0,570941851	0,075477116	Left-wing
Terra	Alternative Media	Website	0,053898831	0,087439426	Centrist
Sensacionalista	Alternative Media	News & Media Website	-0,193887054	0,07554334	Left-wing
Yahoo Brasil	Alternative Media	News & Media Website	0,12321334	0,087411172	Right-wing
Portal Administradores	Alternative Media	News & Media Website	0,078193003	0,115549531	Centrist
BuzzFeed Brasil	Alternative Media	News & Media Website	-0,218954213	0,071416115	Left-wing
Catraca Livre	Alternative Media	News & Media Website	-0,132858522	0,083318896	Left-wing
MSN Brasil	Alternative Media	News & Media Website	0,085718915	0,086485245	Centrist
Congresso em Foco	Alternative Media	News & Media Website	-0,554290197	0,063777129	Left-wing
Ivete Sangalo News	Alternative Media	News & Media Website	0,264127993	0,18213303	Right-wing
Feminismo Sem Demagogia - Original	Alternative Media	Society & Culture Website	-0,7222553	0,051222566	Left-wing
PSOL 50	Alternative Media	Political Party	-0,82647333	0,028657659	Left-wing
Partido Social Liberal - PSL	Alternative Media	Political Party	0,334433155	0,058876747	Right-wing
Canal Brasil	Mainstream Media	TV Channel	-0,378788216	0,067800431	Left-wing
GloboNews	Mainstream Media	TV Channel	0,080640214	0,082791099	Centrist
DW Brasil	Mainstream Media	TV Channel	-0,527740431	0,08502766	Left-wing
Rede Aparecida	Mainstream Media	TV Channel	0,538546453	0,157587017	Right-wing
Forbes Brasil	Mainstream Media	Media/News Company	0,069716693	0,149322461	Centrist
Correio Braziliense	Mainstream Media	Media/News Company	-0,074230541	0,073116224	Left-wing
National Geographic	Mainstream Media	Media/News Company	0,061017312	0,069871189	Centrist

Table A.5: Extended Dataset with political bias data (Part II)

Page Name	Page Type	Page Category	Political bias	Std. of bias	Political leaning
BBC News Brasil	Mainstream Media	Media/News Company	-0,109332345	0,078879371	Left-wing
CBN	Mainstream Media	Media/News Company	-0,107512133	0,109282775	Centrist
O Globo	Mainstream Media	Media/News Company	0,089073413	0,068973064	Right-wing
UOL Notícias	Mainstream Media	Media/News Company	0,094274904	0,067785448	Right-wing
Época NEGÓCIOS	Mainstream Media	Media/News Company	0,046924079	0,137452258	Centrist
EL PAÍS Brasil	Mainstream Media	Media/News Company	-0,486057199	0,073588148	Left-wing
UOL	Mainstream Media	Media/News Company	0,058833215	0,076978103	Centrist
Grupo RBS	Mainstream Media	Media/News Company	-0,03104192	0,088610569	Centrist
Diário Catarinense	Mainstream Media	Media/News Company	0,037058938	0,088894206	Centrist
G1 - O Portal de Notícias da Globo	Mainstream Media	Media/News Company	0,106080145	0,076754371	Right-wing
Metrópole Estadão	Mainstream Media	Media/News Company	-0,239265164	0,100583326	Left-wing
Folha Poder	Mainstream Media	Media/News Company	-0,12555298	0,108967194	Left-wing
Valor Econômico	Mainstream Media	Media/News Company	-0,09783285	0,117169941	Centrist
Folha de S.Paulo	Mainstream Media	Newspaper	-0,016081007	0,080178051	Centrist
Rede Record	Mainstream Media	Newspaper	0,444925974	0,093463881	Right-wing
Jornal da Record	Mainstream Media	Local Business	0,469249815	0,046748114	Right-wing
GaúchaZH	Mainstream Media	Local Business	-0,032660988	0,074423881	Centrist
Correio do Povo	Mainstream Media	Local Business	-0,265511069	0,119997207	Left-wing
TV Canção Nova	Mainstream Media	Local Business	0,33843548	0,149411509	Right-wing
Pequenas Empresas & Grandes Negócios	Mainstream Media	TV Show	0,088250745	0,1528855	Centrist
Jornal Hoje	Mainstream Media	TV Show	0,122955728	0,101433621	Right-wing
Fantástico - O Show da Vida	Mainstream Media	TV Show	0,179896891	0,080065653	Right-wing
Profissão Repórter	Mainstream Media	TV Show	0,189691069	0,078572019	Right-wing
Jornal da Globo	Mainstream Media	TV Show	0,229113949	0,092533517	Right-wing
Jornal Nacional	Mainstream Media	TV Show	0,131740494	0,102620147	Right-wing
Panico na tv	Mainstream Media	TV Show	0,475224323	0,051539013	Right-wing
Programa Silvio Santos	Mainstream Media	TV Show	0,731862066	0,061578413	Right-wing
Roda Viva	Mainstream Media	TV Show	-0,394655175	0,097465974	Left-wing
Programa Pânico	Mainstream Media	TV Show	0,479036472	0,052750594	Right-wing

Table A.6: Extended Dataset with political bias data (Part III)

Page Name	Page Type	Page Category	Political bias	Std. of bias	Political leaning
Globo Repórter	Mainstream Media	TV Show	0,119218264	0,09432438	Right-wing
Agora é Tarde	Mainstream Media	TV Show	0,577502955	0,047799354	Right-wing
Bom Dia Brasil	Mainstream Media	TV Show	0,190232116	0,076698077	Right-wing
Domingo Espetacular	Mainstream Media	TV Show	0,358583916	0,087769196	Right-wing
Programa Eliana	Mainstream Media	TV Show	0,361395605	0,149316493	Right-wing
TV Cultura	Mainstream Media	TV Network	-0,28221374	0,111632379	Left-wing
Rede Globo	Mainstream Media	TV Network	0,108386868	0,077476835	Right-wing
SBT	Mainstream Media	TV Network	0,321965368	0,084863369	Right-wing
CartaCapital	Mainstream Media	Magazine	-0,403696184	0,072615509	Left-wing
VEJA	Mainstream Media	Magazine	0,11015921	0,076805608	Right-wing
Globo Rural	Mainstream Media	Magazine	0,211479619	0,053151055	Right-wing
CARAS Brasil	Mainstream Media	Magazine	0,217157318	0,086288476	Right-wing
Revista ISTOÉ	Mainstream Media	Magazine	0,139132036	0,078330047	Right-wing
Época	Mainstream Media	Magazine	0,102814513	0,079819736	Right-wing
Exame	Mainstream Media	Magazine	0,051556014	0,086379412	Centrist
INFO	Mainstream Media	News & Media Website	-0,130104661	0,113487094	Left-wing
Economia Estadão	Mainstream Media	News & Media Website	-0,1263626	0,115027837	Left-wing
DER SPIEGEL	Mainstream Media	News & Media Website	-0,703551493	0,055588736	Left-wing
Portal R7	Mainstream Media	News & Media Website	0,223912606	0,089760122	Right-wing
Contigo!	Mainstream Media	News & Media Website	0,291414459	0,181519833	Right-wing
Veja São Paulo	Mainstream Media	News & Media Website	0,039575553	0,099576115	Centrist
Estadão	Mainstream Media	News & Media Website	0,068515204	0,077910919	Centrist
InfoMoney	Mainstream Media	News & Media Website	0,085848634	0,122124632	Centrist
Geração de Valor	Public Figure	Public Figure	0,220299189	0,087471595	Right-wing
Dilma Bolada	Public Figure	Arts & Entertainment	-0,585445174	0,047164809	Left-wing
Tatá Werneck	Public Figure	Artist	0,573255501	0,039237499	Right-wing
Ana Hickmann	Public Figure	Artist	0,375445671	0,142287408	Right-wing
Bispo Edir Macedo	Public Figure	Author	0,330898967	0,136544451	Right-wing
Tico Santa Cruz	Public Figure	Author	-0,360899612	0,0610671	Left-wing
Silas Malafaia	Public Figure	Author	0,567742245	0,047460628	Right-wing
Padre Paulo Ricardo	Public Figure	Author	0,77840336	0,145290923	Right-wing
Fábio Porchat	Public Figure	Writer	0,072102651	0,058204928	Right-wing
Guilherme Boulos	Public Figure	Public Figure	-0,713116959	0,043314359	Left-wing

Table A.7: Extended Dataset with political bias data (Part IV)

Page Name	Page Type	Page Category	Political bias	Std. of bias	Political leaning
Marco Antonio Villa	Public Figure	Public Figure	0,284593799	0,132239271	Right-wing
David Miranda	Public Figure	Public Figure	-0,806948529	0,030340861	Left-wing
Jaques Wagner	Public Figure	Public Figure	-0,872490041	0,020746975	Left-wing
Maria do Rosário	Public Figure	Public Figure	-0,868933727	0,02151415	Left-wing
Ratinho	Public Figure	Public Figure	0,715302243	0,063138966	Right-wing
Ana Amélia Lemos	Public Figure	Public Figure	0,466210641	0,099148649	Right-wing
Tarso Genro	Public Figure	Public Figure	-0,981770996	0,007736623	Left-wing
Ana Maria Braga	Public Figure	Public Figure	0,297834196	0,134776985	Right-wing
Marcelo Rubens Paiva	Public Figure	Public Figure	-0,901671828	0,026625009	Left-wing
Olavo de Carvalho	Public Figure	Public Figure	0,581483118	0,095564623	Right-wing
Jean Wyllys	Public Figure	Public Figure	-0,729732792	0,044254518	Left-wing
Ricardo Amorim	Public Figure	Public Figure	0,186777882	0,107802862	Right-wing
Alvaro Dias	Public Figure	Public Figure	0,349763767	0,105332105	Right-wing
Jair Messias Bolsonaro	Public Figure	Public Figure	0,342114027	0,059315419	Right-wing
Ivan Valente	Public Figure	Public Figure	-0,841066055	0,035515875	Left-wing
Marcelo Freixo	Public Figure	Public Figure	-0,673582661	0,04951238	Left-wing
Flavio Bolsonaro	Public Figure	Public Figure	0,555393659	0,056631792	Right-wing
Rachel Sheherazade	Public Figure	Public Figure	0,392138288	0,064553822	Right-wing
Eduardo Bolsonaro	Public Figure	Public Figure	0,332251905	0,058879426	Right-wing
Eliane Brum	Public Figure	Journalist	-0,929703421	0,016375415	Left-wing
Reinaldo Azevedo	Public Figure	Journalist	-0,087532671	0,135438879	Centrist
Gilberto Dimenstein	Public Figure	Journalist	-0,725735586	0,055385459	Left-wing
Ricardo Boechat	Public Figure	Journalist	-0,143603887	0,109346419	Left-wing
Rafinha Bastos	Public Figure	Journalist	0,219526073	0,057709638	Right-wing
Marcelo Tas	Public Figure	Journalist	0,073175655	0,080012972	Centrist
Luiz Felipe Pondé	Public Figure	Book	0,19366246	0,178843103	Right-wing
Eduardo Jorge	Public Figure	Politician	-0,000338265	0,105826772	Centrist
João Doria	Public Figure	Politician	0,259269304	0,098419246	Right-wing
Geraldo Alckmin	Public Figure	Politician	0,091344063	0,112351334	Centrist
Romário Faria	Public Figure	Politician	0,216137452	0,063148192	Right-wing
Marina Silva	Public Figure	Politician	-0,032856354	0,08596121	Centrist
Michel Temer	Public Figure	Politician	-0,018872991	0,101586877	Centrist
Aécio Neves	Public Figure	Politician	0,225388846	0,085178527	Right-wing
Marcelo Crivella	Public Figure	Politician	0,755463708	0,075125573	Right-wing
Eduardo Suplicy	Public Figure	Politician	-0,691179865	0,052990629	Left-wing
Dilma Rousseff	Public Figure	Politician	-0,334160008	0,070153571	Left-wing
Fernando Haddad	Public Figure	Politician	-0,323696256	0,069669202	Left-wing
Fernando Gabeira	Public Figure	Politician	-0,017627115	0,133983951	Centrist
Ciro Gomes	Public Figure	Politician	-0,465224172	0,054447698	Left-wing
Fernando Pimentel	Public Figure	Politician	-0,893299956	0,018008207	Left-wing
Fernando Henrique Cardoso	Public Figure	Politician	0,224822532	0,083243474	Right-wing
Lula	Public Figure	Politician	-0,311393446	0,059136145	Left-wing
Cristovam Buarque	Public Figure	Politician	-0,214736791	0,104179233	Left-wing
Eduardo Campos	Public Figure	Politician	0,125063492	0,185416909	Centrist
Gleisi Hoffmann	Public Figure	Politician	-0,47182918	0,054376449	Left-wing

Table A.8: Extended Dataset with political bias data (Part V)

Page Name	Page Type	Page Category	Political bias	Std. of bias	Political leaning
Chico Alencar	Public Figure	Politician	-0,733228694	0,037190278	Left-wing
Marco Feliciano	Public Figure	Politician	0,536083899	0,050510021	Right-wing
Manuela D'Ávila	Public Figure	Politician	-0,629687025	0,060507516	Left-wing
Kátia Abreu	Public Figure	Politician	-0,93789422	0,008647607	Left-wing
Carlos Bolsonaro	Public Figure	Politician	0,599861252	0,055800851	Right-wing
Marcos Do Val	Public Figure	Politician	0,338371055	0,074741126	Right-wing
Alessandro Molon	Public Figure	Politician	-0,866406174	0,026993272	Left-wing
Roberto Requião	Public Figure	Politician	-0,915672895	0,012391955	Left-wing
Randolfe Rodrigues	Public Figure	Politician	-0,845414039	0,030118331	Left-wing

Table A.9: Extended Dataset with political bias data (Part VI)

Interest Name	Political Bias	Political Bias (Ribeiro et al.)	Political Bias (Bakshy et al.)	Political Bias (Budak et al.)	Political Bias (Mitchell et al.)
Colorlines	-0,361031736	-1,739	-	-	-
BlackNews.com	-0,484472254	-1,69	-	-	-
Jacobin magazine	-0,855210891	-1,68	-	-	-
Afro-American Newspapers	-0,836543146	-1,65	-	-	-
Black Enterprise	-0,863652308	-1,638	-	-	-
Black in America	-0,37486864	-1,613	-	-	-
TheGrio	-0,718675474	-1,588	-0,8326	-	-
Hunto.com	-0,618122625	-1,538	-	-	-
NewsOne	-0,860055131	-1,529	-	-	-
The American Prospect	-0,412014203	-1,52	-	-	-
BUST Magazine	-0,59035639	-1,505	-	-	-
The Electronic Intifada	0,229192141	-1,498	-0,8074	-	-
Now With Alex Wagner	-0,468339238	-1,497	-	-	-
The Ed Show	-0,909964747	-1,46	-	-	-
Center for American Progress	-0,511071873	-1,452	-	-	-
The Root	-0,669779391	-1,45	-0,7792	-	-
In These Times	-0,645285024	-1,405	-	-	-
Free Speech TV	-0,442749236	-1,389	-	-	-
The New York Review of Books	-0,453238502	-1,385	-	-	-
Longreads	-0,361490748	-1,381	-	-	-
Chicago Reader	-0,062807813	-1,378	-	-	-
Democracy Now!	-0,746895862	-1,374	-0,934	-	-
The Rachel Maddow Show	-0,616731895	-1,352	-	-	-
Truthdig	-0,50382026	-1,336	-0,8565	-	-
Mother Jones	-1	-1,326	-0,8663	-	-
Seattle Weekly	-0,876142903	-1,321	-	-	-
Queerty	-0,708510422	-1,316	-	-	-
The New Civil Rights Movement	-0,579220348	-1,291	-	-	-
Fresh Air with Terry Gross	-0,658564307	-1,272	-	-	-
YES! Magazine	-0,484418116	-1,27	-	-	-

Table A.10: Dataset for validation of our method with political bias

Interest Name	Political Bias	Political Bias (Ribeiro et al.)	Political Bias (Bakshy et al.)	Political Bias (Budak et al.)	Political Bias (Mitchell et al.)
Truthout	-0,358569035	-1,255	-0,8937	-	-
Latino Rebels	-0,567965575	-1,254	-	-	-
Ms. Magazine	-0,50189002	-1,254	-	-	-
Andrea Mitchell Reports	-0,560984278	-1,248	-	-	-
ProPublica	-0,665920431	-1,22	-	-	-
The Stranger	-0,74531323	-1,219	-	-	-
ThinkProgress	-0,685680671	-1,218	-0,8615	-	-
Daily Kos	-1	-1,217	-0,8972	-0,25	-
Modern Bear	-0,922505498	-1,19	-	-	-
The Raw Story	-0,753663777	-1,19	-0,8534	-	-
Edge Media Network	-0,677699817	-1,184	-	-	-
PoliticsNation with Al Sharpton	-0,573927525	-1,18	-	-	-
Talking Points Memo	-0,659331526	-1,178	-0,8614	-	-
The Young Turks	-0,616084259	-1,177	-	-	-
AlterNet	-0,498237722	-1,175	-0,8804	-	-
Hardball with Chris Matthews	-0,925801012	-1,168	-	-	-
NPR All Things Considered	-0,56223073	-1,16	-	-	-
Gay Rights Media	-0,257632225	-1,16	-	-	-
PRI Public Radio International	-0,56670718	-1,134	-	-	-
Aeon	-0,781828637	-1,119	-	-	-
Gay Star News	-0,421872539	-1,093	-	-	-
Guardian US	-0,576618896	-1,065	-	-	-1,07
Washington Post Politics	-0,601159024	-1,057	-	-	-
HuffPost Politics	-0,610363731	-1,043	-	-	-
PolitiFact	-0,928745796	-1,039	-0,4802	-	-
NPR Morning Edition	-0,495845725	-1,029	-	-	-
LGBTQ Nation	-0,434516026	-1,029	-0,8501	-	-
Tablet Magazine	-0,373711329	-1,028	-	-	-
Last Week Tonight with John Oliver	-0,74879858	-1,012	-	-	-
PBS NewsHour	-0,869329196	-1,002	-	-	-
The Real News Network	-0,711534491	-0,988	-	-	-
FiveThirtyEight	-0,811685711	-0,979	-0,5225	-	-
PinkNews	-0,730551711	-0,973	-	-	-
The Late Show with Stephen Colbert	-1	-0,957	-	-	-
The New Yorker	-0,692787561	-0,957	-0,7584	-	-1,23
The Economist, Asia	-0,343104269	-0,93	-	-	-
Morning Joe	-0,641775649	-0,929	-	-	-
The Daily Show	-0,687662656	-0,921	-0,6704	-	-1,09
World Politics Review	-0,44197534	-0,902	-	-	-
Sojourners	-0,354044976	-0,896	-	-	-
The Atlantic	-0,548597534	-0,881	-0,5424	-	-
Anderson Cooper 360	-0,533133263	-0,875	-	-	-

Table A.11: Dataset for validation of our method with political bias (Part II)

Interest Name	Political Bias	Political Bias (Ribeiro et al.)	Political Bias (Bakshy et al.)	Political Bias (Budak et al.)	Political Bias (Mitchell et al.)
EcoWatch	-0,568969584	-0,872	-0,8554	-	-
FactCheck.org	-1	-0,871	-	-	-
Vox	-0,604079958	-0,87	-0,6591	-	-
Mundo Hispánico	-0,355216928	-0,855	-	-	-
FRONTLINE — PBS	-0,744451008	-0,853	-	-	-
Chronicle of Higher Education	-0,667761321	-0,842	-	-	-
Miami New Times	-0,725075747	-0,836	-	-	-
VICE News	-0,593425409	-0,831	-0,4284	-	-
The Daily Beast	-0,595377373	-0,827	-0,4561	-	-
Addicting Info	-0,615005752	-0,822	-0,8894	-	-
Haaretz.com	-0,495609976	-0,817	-0,504	-	-
THE WEEK	-0,516591264	-0,814	-	-	-
MPR News	-0,821231292	-0,799	-	-	-
Foreign Affairs	-0,289330279	-0,796	-	-	-
The Economist	0,416415702	-0,757	-0,3173	-	-0,71
Shepherd Express	-0,092525525	-0,75	-	-	-
POLITICO	-0,678436105	-0,744	-0,1334	-	-0,7
HuffPost Women	-0,601488548	-0,74	-	-	-
Denver Westword	-0,802447153	-0,726	-	-	-
ClickHole	-0,727427861	-0,724	-	-	-
All In with Chris Hayes	-0,922018528	-0,723	-	-	-
Daily Camera	-0,793329734	-0,723	-	-	-
The Open Mind	-0,663111851	-0,687	-	-	-
Latina Magazine	-0,711518777	-0,675	-	-	-
Late Night with Seth Meyers	-0,672525497	-0,664	-	-	-
True Activist	-0,5115466	-0,65	-0,6032	-	-
Financial Times	-0,578077835	-0,642	-	-	-
BBC News	-0,673392325	-0,636	-0,2609	-0,0225	-0,74
Miami Herald	-0,417122725	-0,634	-0,3418	-	-
The Advocate	-0,321381039	-0,633	-0,9276	-	-
Los Angeles Times	-0,00420358	-0,625	-0,3995	-0,045	-
The Onion	-0,498664221	-0,616	-0,5516	-	-
Meet the Press	-0,610344455	-0,612	-	-	-
The Seattle Times	-0,605725804	-0,609	-	-	-
The Jewish Week	-0,224202373	-0,603	-	-	-
Elephant Journal	-0,577185057	-0,577	-	-	-
The New York Times	-0,557829006	-0,569	-0,5469	-0,05	-0,9
The Washington Post	-0,00792508	-0,555	-0,2568	-0,0225	-0,69
AP	-0,315547743	-0,549	-	-	-
WGN Morning News	-0,40992553	-0,53	-	-	-
Austin Chronicle	-0,752740035	-0,526	-	-	-
One Green Planet	-1	-0,509	-	-	-
The Hill	-0,858633413	-0,504	0,1661	-	-
KTLA 5 Morning News	-0,063402678	-0,5	-0,2287	-	-
Bloomberg	-0,430050271	-0,482	-0,1565	-	-0,24

Table A.12: Dataset for validation of our method with political bias (Part III)

Interest Name	Political Bias	Political Bias (Ribeiro et al.)	Political Bias (Bakshy et al.)	Political Bias (Budak et al.)	Political Bias (Mitchell et al.)
ABC7	-0,281469802	-0,48	-0,31	-	-
The Huffington Post	-0,444334435	-0,476	-0,6176	-0,05	-0,7
Newsweek	0,234936821	-0,468	-	-	-
The Boston Globe	-0,59330704	-0,45	-0,4446	-	-
Bloomberg Businessweek	-0,221792683	-0,448	-0,2615	-	-
Reuters	-0,517359229	-0,446	-0,0945	-0,01	-
Texas Observer	-0,756488963	-0,445	-	-	-
Exposing The Truth	-0,647960696	-0,444	-	-	-
HuffPost Weird News	-0,573149827	-0,437	-	-	-
CBS Los Angeles	-0,128644346	-0,415	-	-	-
Esquire	-0,231955715	-0,4	-	-	-
Columbus Underground.com	-0,606160155	-0,397	-	-	-
RedEye	0,421019933	-0,388	-	-	-
WUSA 9	0,270858602	-0,386	-0,4447	-	-
NBC News World	-0,323145102	-0,382	-	-	-
Chicago Tribune	-0,403670261	-0,359	-0,3117	-0,01	-
Yahoo Finance	-0,110077498	-0,355	0,0777	-	-
Thrillist	-0,047872163	-0,355	-	-	-
NBC Nightly News with Lester Holt	-0,258390713	-0,35	-	-	-
The Moscow Times	-0,110053153	-0,347	-	-	-
CNN	-0,106039358	-0,342	-0,2705	-0,015	-0,42
Jimmy Kimmel Live	-0,504972321	-0,341	-	-	-
The Mind Unleashed	-0,185341904	-0,338	-0,4008	-	-
60 Minutes	-0,419597371	-0,314	-	-	-
The Diplomat Magazine	-0,489046198	-0,307	-	-	-
Nightline	-0,045089797	-0,307	-	-	-
Orlando Weekly	-0,499039318	-0,306	-	-	-
Elite Daily	-0,110035555	-0,306	-0,2558	-	-
Riverfront Times	-0,480789681	-0,305	-	-	-
snopes.com	-0,537166563	-0,3	-	-	-
CBS Sunday Morning	-0,24930848	-0,289	-	-	-
Boston Herald	-0,436406737	-0,288	-	-	-
The Palm Beach Post	-0,008119532	-0,28	-	-	-
C-SPAN	-0,276288561	-0,274	-0,0442	-	-
PETA (People for the Ethical Treatment of Animals)	-0,076942697	-0,272	-	-	-
Mother Nature Network	-0,480675287	-0,27	-	-	-
World News Now	-0,379740679	-0,254	-	-	-
Inside Edition	0,26199902	-0,253	-	-	-
Upworthy	-0,524147281	-0,247	-	-	-
CBS This Morning	-0,241770834	-0,237	-	-	-
Quartz	-0,422554812	-0,22	-	-	-
Texas Tribune	-0,842095805	-0,219	-	-	-
NBC Chicago	-0,34611807	-0,216	-0,3416	-	-

Table A.13: Dataset for validation of our method with political bias (Part IV)

Interest Name	Political Bias	Political Bias (Ribeiro et al.)	Political Bias (Bakshy et al.)	Political Bias (Budak et al.)	Political Bias (Mitchell et al.)
FOX 5 / Fox5NY.com	0,277477218	-0,212	-	-	-
Al-Monitor	-0,927420271	-0,211	-	-	-
March Against Monsanto	-0,542682912	-0,205	-	-	-
Pioneer Press	-0,677675076	-0,187	-	-	-
Foreign Policy	-0,64976446	-0,174	-0,1516	-	-
Wall Street Journal	-0,211694613	-0,164	0,2754	0,05	-0,13
AboveTopSecret.com	0,424075042	-0,162	-	-	-
St. Louis Magazine	-0,478754099	-0,152	-	-	-
NBC News	0,220823955	-0,148	-0,2735	-0,01	-0,34
New York Post	-0,114366947	-0,138	0,2497	-	-
MSNBC	-0,621635579	-0,129	-0,8102	-	-0,48
Houston Press	-0,155546894	-0,128	-	-	-
WGN Radio	-0,306248387	-0,126	-	-	-
The Denver Post	-0,832728615	-0,121	-	-	-
ABC News	0,380487958	-0,117	-0,1605	-	-0,24
WIRED	-0,334531664	-0,096	-0,4027	-	-
TheStreet	0,038066625	-0,08	-	-	-
MSN News	-0,313082502	-0,076	-0,0623	-	-
San Antonio Current	-0,840619921	-0,07	-	-	-
Forbes	-0,123267787	-0,062	0,0563	-	-
Slate.com	-0,679857614	-0,045	-0,6813	-	-1,21
Star Tribune	-0,461248801	-0,022	-	-	-
Milwaukee Journal Sentinel	-0,027951613	0,012	-	-	-
USA Today	0,313699362	0,014	-	0	-0,23
AJC	0,425007393	0,019	-	-	-
FOX31 KDVR.com	0,183079562	0,03	-	-	-
The Columbus Dispatch	-0,393021619	0,047	-	-	-
Yahoo! News	0,26850211	0,052	0,0493	-0,015	-0,16
Richmond Times-Dispatch	-0,556695014	0,054	-	-	-
Tampa Bay Times	-0,352948497	0,057	-0,4745	-	-
Digital Trends	-0,176219348	0,066	-	-	-
The News & Observer	-0,399167071	0,072	-	-	-
FOX 10 Phoenix	0,488392014	0,079	0,2386	-	-
The Charlotte Observer	-0,530090241	0,082	-	-	-
The Virginian-Pilot	0,129815309	0,085	-	-	-
Good Morning America	0,348064924	0,09	-	-	-
Houston Chronicle	0,165024731	0,11	-	-	-
San Antonio Express-News	-0,317619171	0,113	-	-	-
La Crosse Tribune	-0,477541922	0,116	-	-	-
TODAY	0,26870989	0,123	-	-	-
The Dallas Morning News	0,31220658	0,163	-	-	-
KARE 11	-0,590702144	0,164	-	-	-
Orlando Sentinel	-0,322976772	0,164	-	-	-

Table A.14: Dataset for validation of our method with political bias (Part V)

Interest Name	Political Bias	Political Bias (Ribeiro et al.)	Political Bias (Bakshy et al.)	Political Bias (Budak et al.)	Political Bias (Mitchell et al.)
Squawk Box	0,161046624	0,167	-	-	-
St. Louis Post-Dispatch	-0,491361185	0,185	-	-	-
The Epoch Times	0,333537061	0,204	-	-	-
Mad Money	0,151450667	0,212	-	-	-
KHOU 11 News	0,161677872	0,228	0,1182	-	-
WSB-TV	0,439552764	0,238	-0,0872	-	-
CBS News	0,212335327	0,244	-0,127	-	-0,3
KVUE	-0,218696157	0,253	-	-	-
Dayton Daily News	0,43653251	0,276	-	-	-
WVUE FOX 8 News	0,43988975	0,323	-	-	-
Today Show	0,368977588	0,364	-0,1668	-	-
Reason Magazine	0,436681915	0,391	0,3502	-	-
WAFB Channel 9	0,34269581	0,403	-	-	-
WFLA News Channel 8	0,465529988	0,429	0,041	-	-
The Inquisitr	0,143631023	0,44	0,0977	-	-
The Idea Room	-0,46014735	0,444	-	-	-
No Labels	0,231239822	0,456	-	-	-
The Post and Courier	0,479080046	0,457	-	-	-
Opposing Views	0,15141859	0,492	0,2685	-	-
11Alive	0,345389688	0,509	0,0686	-	-
Countdown with Keith Olbermann	-0,718503202	0,533	-	-	-
Colorado Springs Gazette	0,420637602	0,574	-	-	-
Ron Paul .com	0,299585039	0,653	-	-	-
Navy Times	0,415603683	0,658	0,3365	-	-
WDBJ7	0,135912176	0,685	-	-	-
Army Times	0,308478815	0,713	0,4171	-	-
The Times of Israel	0,493119317	0,73	0,3685	-	-
The Jerusalem Post / JPost.com	0,473047118	0,76	0,4149	-	-
The Jewish Press	0,364751307	0,761	0,5559	-	-
al.com	0,424005949	0,761	0,009	-	-
Military Times	0,52912809	0,761	0,4459	-	-
Ynetnews	0,396487961	0,781	0,3335	-	-
The Christian Post	0,479232218	0,807	0,6722	-	-
Fox News Channel	0,49341199	0,825	0,7754	0,11	0,43
Texas Monthly	0,172959879	0,831	-	-	-
Stars and Stripes	0,393895804	0,895	0,3658	-	-
30A	0,331192181	0,908	-	-	-
USMC Life	0,763943935	0,922	-	-	-
Marine Corps Times	0,375970332	0,925	-	-	-
Proud to be an American	0,084843481	0,928	-	-	-
Christian News Network	0,48906466	0,947	0,807	-	-
Soldier of Fortune Magazine	0,844476755	0,977	-	-	-
The Washington Times	0,469551443	1,003	0,6975	-	-
The Libertarian Republic	0,480542359	1,031	-	-	-

Table A.15: Dataset for validation of our method with political bias (Part VI)

Interest Name	Political Bias	Political Bias (Ribeiro et al.)	Political Bias (Bakshy et al.)	Political Bias (Budak et al.)	Political Bias (Mitchell et al.)
Casper Star-Tribune	-0,459389294	1,084	-	-	-
Saturday Down South	0,38180421	1,107	-	-	-
The New Republic	0,227094371	1,116	-0,6695	-	-
Fox & Friends	0,369838685	1,13	-	-	-
Human Events	0,456357848	1,14	0,9303	-	-
Fox Business	0,514923912	1,14	-	-	-
Anything About Guns	0,44889367	1,144	-	-	-
SOFREP	0,463497517	1,155	-	-	-
Guns & Ammo	0,187189711	1,164	-	-	-
Prophecy in the News	0,381039781	1,201	-	-	-
Washington Examiner	0,564089154	1,22	0,8285	-	-
Crosswalk.com	0,299846793	1,233	-	-	-
NumbersUSA	0,10511852	1,254	-	-	-
Texas Country Reporter	0,596812063	1,32	-	-	-
Fox News Politics	0,511519609	1,338	-	-	-
The Weekly Standard	0,747559591	1,347	0,9074	-	-
TheBlaze	0,547647865	1,364	0,8887	-	1,25
VirtualJerusalem.com	0,388389152	1,37	0,512	-	-
Fox News Video	0,853986846	1,422	-	-	-
National Review	0,431756016	1,455	0,9009	-	-
Fox News Sunday	0,730041645	1,464	-	-	-
The Daily Caller	0,330281151	1,484	0,8697	-	-
Newsmax	0,42659493	1,5	0,7697	-	-
FOX News Radio	0,668281797	1,519	0,6973	-	-
American Thinker	0,708399038	1,522	0,9112	-	-
Being Conservative	0,464720844	1,551	-	-	-
PJ Media	0,802542769	1,554	-	-	-
African-American Conservatives	0,560614529	1,566	-	-	-
Fox News Opinion	0,804649959	1,571	-	-	-
The Five	0,624118395	1,572	-	-	-
The Blacksphere	0,529599372	1,575	-	-	-
American Patriot	0,385319145	1,583	-	-	-
MRCTV	0,222477787	1,597	0,8636	-	-
The Patriot Post	0,504199115	1,599	-	-	-
Young Conservatives	0,435601771	1,602	0,9782	-	-
The Political Insider	0,231176809	1,606	0,8998	-	-
Conservative Lady	0,747658596	1,632	-	-	-
Media Research Center	0,359706738	1,633	-	-	-
Judicial Watch	0,265879957	1,644	0,8287	-	-
100 Percent FED Up	0,342906104	1,667	0,856	-	-
Hot Air	0,439079339	1,673	0,9247	-	-
LifeNews.com	0,422584879	1,681	0,9664	-	-
Conservative News Today	0,26454917	1,682	-	-	-
NewsBusters.org	0,294879359	1,691	0,9168	-	-
The Daily Signal	0,365564572	1,699	-	-	-

Table A.16: Dataset for validation of our method with political bias (Part VII)

Interest Name	Political Bias	Political Bias (Ribeiro et al.)	Political Bias (Bakshy et al.)	Political Bias (Budak et al.)	Political Bias (Mitchell et al.)
CNSNews.com	0,366180917	1,715	0,8979	-	-
National Pro-Life Alliance	0,346824561	1,72	0,9108	-	-
American Conservative	0,461907478	1,733	-	-	-
Townhall.com	0,55250333	1,748	0,9263	-	-
Washington Free Beacon	0,334481928	1,755	-	-	-
Gateway Pundit	0,335281862	1,767	0,9389	-	-
Conservative Angle	0,615575947	1,774	-	-	-
RedState	0,417830646	1,776	0,9624	-	-
Patriot Update by WJ	0,362071788	1,794	0,8892	-	-
Legal Insurrection	0,406834216	1,862	0,8557	-	-

Table A.17: Dataset for validation of our method with political bias (Part VIII)