INFERÊNCIA DE DADOS DEMOGRÁFICOS A PARTIR DE PLATAFORMAS DE PROPAGANDA BASEADAS EM REDES SOCIAIS

FILIPE NUNES RIBEIRO

INFERÊNCIA DE DADOS DEMOGRÁFICOS A PARTIR DE PLATAFORMAS DE PROPAGANDA BASEADAS EM REDES SOCIAIS

Tese apresentada ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Doutor em Ciência da Computação.

ORIENTADOR: FABRÍCIO BENEVENUTO DE SOUZA

Belo Horizonte

Março de 2019

FILIPE NUNES RIBEIRO

INFERENCE OF DEMOGRAPHIC DATA FROM DIGITAL ADVERTISING PLATFORMS BASED ON SOCIAL MEDIA

Thesis presented to the Graduate Program in Computer Science of the Universidade Federal de Minas Gerais – Departamento de Ciência da Computação. in partial fulfillment of the requirements for the degree of Doctor in Computer Science.

ADVISOR: FABRÍCIO BENEVENUTO DE SOUZA

Belo Horizonte

March 2019

Ribeiro, Filipe Nunes

R484i Inference of demographic data from digital advertising platforms based on social media / Filipe Nunes Ribeiro. — Belo Horizonte, 2019
xxv, 121 f. : il. ; 29cm

Tese (doutorado) — Universidade Federal de Minas Gerais – Departamento de Ciência da Computação.
Orientador: Fabrício Benevenuto de Souza

1. Computação – Teses. 2. Redes sociais on-line. 3. Marketing digital. 4. Mídia digital - Dados demográficos. I. Orientador. II. Título.

CDU 519.6*22(043)

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

FOLHA DE APROVAÇÃO

Inference of Demographic Data from Digital Advertising Platforms Based on Social Media

FILIPE NUNES RIBEIRO

Tese defendida e aprovada pela banca examinadora constituída pelos Senhores:

PROF. FABRÍCIO BENEVENUTO DE SOUZA - Orientador
Departamento de Ciência da Computação - UFMG

PROF. ANTONIO ALFREDO FERREIRA LOUREIRO
Departamento de Ciência da Computação - UFMG

PROF. CLODOVEU AUGUSTO DAVIS JÚNIOR
Departamento de Ciência da Computação - UFMG

PROF. BERNARDO LANZA QUEIROZ
Departamento de Demografia - UFMG

PROFA. JULIANA FREIRE DE LIMA E SILVA
Departamento de Ciência da Computação e Engenharia - New York University

PROF. THIAGO HENRIQUE SILVA
Departamento Acadêmico de Informática - UTFPR

Belo Horizonte, 29 de Março de 2019.

To my beloved mother, Julieta Nunes Nunes Ribeiro (in Memoriam).

Acknowledgments

First and foremost, I would like to thank GOD. To Him all honor and all glory, forever and ever! My sincere gratitude to my beloved wife, for her unconditional support, complicity and endless love. I am also thankful to my daughter Elisa, while still trying her first steps, makes my heart overflows with joy when I get home and she is waiting for me with opened arms and a smile in her face. I also thank my mother Julieta, who would certainly be one of the happiest people on earth with my victory, and did not forget to teach me the principles that govern my life, not a single day. In spite of being not here anymore, her legacy and teachings will endure forever. I would also like to give a special thank to my father Anderson, for being my example of man, God's servant, and human being. Thank you, my beloved parents, for standing by my side all the time, especially when my strength seemed to vanish. You have shown with your life that "when I am weak, then I am strong". I also thank my dear sisters, Débora and Maressa, who always stand by my side with unconditional support and always give to me a special affection. I am also thankful to my niece and nephews, grandparents, uncles and aunts, brothers-in-law, and cousins that despite the distance always encouraged me.

I would like to express my deep gratitude to Professor Dr. Fabrício Benenenuto, my research supervisor, for his guidance, enthusiastic encouragement, and patience. His advice was crucial to the completion of this work. I could not have imagined having a better advisor and mentor for my Ph.D. study. I would also like to thank Dr. Krishna Gummadi for his valuable contributions to this work. I also thank my friends from LOCUS Lab for the stimulating discussions and for all the fun we have had in the last four years. I am also thankful to all my coauthors for the friendship and for the hard working together before deadlines in particular, Matheus Araújo, Lucas Henrique, Johnnatan Messias, Koustuv Saha, and Abhijnan Chakraborty. I am particularly grateful to my friends Júlio Reis and Rodrigo Silva for their support and encouraging words.

Besides my advisor, I would like to thank the other members of my thesis committee: Prof. Dr. Antônio Loureiro, Prof. Dr. Bernardo Lanza, Prof. Dr. Clodoveu Augusto, and Prof. Dr. Juliana Freire, and Prof Dr. Thiago Henrique Silva for their insightful comments

and encouragement, and also for the questions which helped me improve my research.

My thanks also go to the Federal University of Ouro Preto (UFOP), that provided me an opportunity to leave my academic attributions for four years to pursue my Ph.D. I also thank the Coordination of Superior Level Staff Improvement (CAPES) for supporting me in my stay in Germany. I would also like to thank the professors of the Computer Science department of UFMG for the singular contribution to my overall formation. Finally, I thank all of those who contributed directly or indirectly to the execution of this work.

Abstract

The growth of Online Social Networks (OSN) in the last years is impressive. Only Facebook, the most popular OSN, attracted more than 500 million new users in the last two years, reaching the massive amount of 2.32 billion monthly active users. The revenue of OSN is concentrated on their marketing platforms, which evolved substantially in comparison with the traditional advertising model. By using OSN ad platforms, an advertiser is able to explore micro-targeting advertising, which means that the advertiser may select users with very particular characteristics, including thousands of different attributes such as race, gender, interests, and behaviors. In this work, we propose and develop a framework to infer demographics based on the attributes available on OSN Advertising Platforms. Social networks provide the ideal environment to infer demographics of online populations since users share a large number of personal information as well as behavioral signals such as likes and shares of content they enjoy. In our framework, we leverage the aggregate information about users provided by Facebook advertising platform to advertisers to build new applications. Particularly, we conducted four case studies in which we apply our framework. In the first, we studied the extent to which demographics data extracted from OSN advertising platforms are reliable and similar to the offline data. Our analysis will considers seven demographic categories including gender, race, and political leaning. Next, we applied our methodology to the US news ecosystem and we show that the ideological (liberal or conservative) leaning of a news source can be accurately estimated by the extent to which liberals or conservatives are over-/under-represented among its audience. Then, we build and deploy a system, called "Media Bias Monitor", which exposes the biases in audience demographics for over $20,000$ news outlets on Facebook to any Internet user. In the third study case, we examine a specific case of malicious advertising, exploring the extent to which political ads from the Russian Intelligence Research Agency (IRA), run prior to 2016 U.S. elections, exploited Facebook's targeted advertising infrastructure to efficiently target ads on divisive or polarizing topics (e.g., immigration, race-based policing) at vulnerable sub-populations. In the last case study, we leveraged our framework in the election context to check if the popularity fluctuation verified in the election polls before the elections are captured in the online perspective. For

this analysis, we considered the Brazilian presidential election context.

Keywords: Demographics, Social Networks, Systems, Social Media Advertising Platforms, Media Bias.

List of Figures

List of Tables

List of Abbreviations and Acronyms

ACS *American Community Survey*

API *Application Programming Interface*

CPS *Current Population Survey*

CTR *Click-Through Rate*

FMA *Facebook Marketing API*

IOS *IPhone OS device*

IRA *Internet Research Agency*

OS *Operating System*

OSN *Online Social Networks*

OTF *Original Targeting Formula*

PII *Personally Identifiable Information*

WWW *World Wide Web*

Contents

Chapter 1

Introduction

The widespread adoption of Online Social Networks (OSN) in the last years is impressive. Only Facebook, the most popular OSN, attracted more than 500 million new users in the last two years, reaching the massive amount of 2.32 billion monthly active users as of December 31, 2018[1]. Facebook, as well as Snapchat, Linkedin, Twitter, and other social networks, represents an important part of people's life nowadays. A recent survey indicates that two-thirds of US adults use social networks, out of which 42% say it would be hard to give up social media [Smith and Anderson, 2018]. The same report shows that 51% of US Facebook users access their accounts several times per day. One deep change, for example, may be verified in the manner people access news [Kwak et al., 2010]. Keeping informed by just watching TV news or subscribing to some specific newspaper or magazine is being continuously replaced by getting informed about the latest news by only scrolling their social network feeds. In the US, for instance, 62% of adults consume news primarily from social media sites [Mitchell, 2016].

As clearly stated by Facebook's founder in a response to a question made by a senator, they run ads to sustain their "business in which users don't pay for the service" [2]. Indeed, advertising underpins much of the Internet's economy and has a key role in the business model of OSN. Consolidated multinational companies or even local small businesses around the world can take advantage of the advertising infrastructure provided by OSN. With an enormous and global customer list, the revenue of companies like Facebook and Twitter increased substantially and their market cap reached values of 541[3] and 24[4] billion of dollars as of June 2018, respectively. Not surprisingly, Online Social Networks revolutionize how ads are created and how to attract the user's attention and engagement. Viral marketing

[1] https://newsroom.fb.com/company-info/
[2] https://www.nbcnews.com/card/we-run-ads-n864606
[3] https://www.forbes.com/companies/facebook/
[4] https://www.forbes.com/companies/twitter/

techniques, close contact with customers, low costs and the possibility of targeting very specific niches of the population attracted advertisers from many different areas and sizes.

One of the keys in the success of OSN advertising platforms is the vast possibilities to reach users such as by providing a list of personally identifiable information (name, phone number, email, etc) or by configuring targeting options from a huge list of fine-grained attributes such as race, income level, interests and behaviors. As an example of attribute targeting, one may select users who live in a Brazilian state (Minas Gerais, Amazonas, etc) and that are interested in one particular candidate in the Brazilian presidential election. In a more detailed targeting, the owner of a beverage shop that sells high-quality wine may exhibit a bunch of ads to OSN users interested in wine, aged between 25 and 45, that live less than 5 miles from the store and with high income. Observe that microtargeting can be useful for a wide range of businesses, since thousands of other users' interests are available, including physical exercise, games, movies, pets, and even religious ones. In a political example of usage, a liberal US organization might try to spread their ideas by showing ads to young African-American Facebook users who are conservative and live in a particular city.

In this scenario, we propose a method that explores the OSN advertising platforms to infer high-quality demographic data and develop novel applications that explore such data. Our method consists of adding a new layer in the targeting formula, i.e., the set of attributes, used in the OSN advertising platforms to define the audience to which an advertiser wants to deliver an ad. We also developed a framework that implements our method, automates the requests and deals with scalability requirements.

Demographic information about different populations is very helpful and, in some cases, it may be of utmost importance for the overall well-being of a modern society. Census data, for instance, is crucial to define priority investments for education, infrastructure and other public policies of a country. Another example is the demographic information about the audience of media sources, which is critical to understand who are the readers of the media outlets and how users might be influenced by biased news, especially during election periods. Notice that surveying demographics in the offline world to carry on census or election polls, or to conduct other sorts of analysis is, in general, time and money consuming and most of the times executed with large intervals. A traditional approach to detect the audience bias of a media source, for example, would require selecting a sample of the readers and asking them to answer surveys. With the help of our method, we can derive the political leaning of the audience of many media outlets and even create a system that makes it transparent to readers. Similarly, our approach allows the reproduction of the Census and election polls quickly and with almost no cost. In the Census case particularly, alternatives

to the traditional collection of data have also been explored by the United Nations.[5].

1.1 Motivation

Recent studies have shown that demographic data extracted from online environment may be able to revolutionize much of the studies conducted in the demography research field in the next years [Weber et al., 2018; Cesare et al., 2018].

Despite this recent claim, the inference of demographic data from the online world has received much attention from researchers since the beginning of the World Wide Web (WWW). As an example, back in 1997, researchers created the Lifestyle Finder [Krulwich, 1997], a fortune teller web application that asked questions about demographics, interests, leisure activities, and other topics to recommend web pages the user would likely enjoy accessing. Since its beginning, the WWW is going through a huge growth in terms of number of users as well as in terms of the variety of available services. In the same direction, the collection of a large variety of data about users increased exponentially and along with it the possibilities to extract demographic information from online data.

The services and useful insights that can be leveraged using demographic data are wider than only recommending web pages that fit the user's interests. Among efforts touching this topic there are studies about worldwide gender inequality [Garcia et al., 2018], movement of migrants [Zagheni et al., 2017] and detection of gender to help forensic investigations [Vel et al., 2002], not to mention the importance of this personal data to commerce, by allowing the creation of ads that target specific groups or by recommending products that best fit user interests. On the other hand, several works have studied privacy concerns and exposed the huge amount of data that has been gathered [Ackerman et al., 1999; Graeff and Harmon, 2002; Andreou et al., 2018]. In spite of uninterrupted concerns about privacy, the use of demographic data has taken the advertising platforms of OSN to a new level, allowing targeting very specific niches of users. OSN have access to plenty of additional information about users, such as the workplace, visited venues, published posts, and 'likes', which are explored to infer user demographics at a fine-grained level. Our motivation in this thesis lies in exploring this new source of data and provide useful mechanisms to delivery valuable demographic data.

[5]https://unstats.un.org/UNSD/demographic/sources/census/alternativeCensusDesigns.htm

1.2 Goals

In this thesis, we aim at exploring the richness of information provided by Online Social Network advertising platforms to create a methodology and a framework that extracts advanced demographic data including race, income, gender, and location. We hypothesize that the data extracted from such platforms may be valuable to infer demographics from online and even real-world populations. We leverage the presented framework to build a set of novel applications. Particularly, we validate our methodology by employing it in four case studies that explore demographics by making use of our framework.

- Inferring Census from online data.
- Inferring the political alignment of news media outlets.
- Understanding the demographics of people targeted by the potentially divisive ads before and during the American Election on the year of 2016.
- Inferring the demographics of candidates on the period preceding elections, similar to election polls.

1.3 Organization

This thesis is organized as follows:

- Chapter 2 presents the background and the previous initiatives that inspired and served as the foundation for our work.
- Chapter 3 details the layers methodology we employed to collect advanced demographics. We also present the overall execution and the architecture of the framework in order to make our approach scalable. The case studies are presented in the following chapters.
- Chapter 4 - Case study 1: Inferring Demographics of data similar to Census. The main reason for this case study is providing a detailed evaluation of the demographic data obtained with our framework in comparison with reliable data. The basic idea consists of gathering similar data from the US Census and from our framework and check the extent to which both data are similar.
- Chapter 5 - Case study 2: Inferring Demographics of Media Outlets Audience. Previous knowledge about the political leaning of publishers may help readers interpret

news stories, especially political ones. One of the techniques used to derive this information is based on detecting if the audience of the media source is biased to any political leaning. In this case study, we plan to infer the political alignment of media outlets by detecting the political leaning distribution of the audience with our framework [Ribeiro et al., 2018].

- Chapter 6 - Case study 3: Demographic study of the Russian Ads in the US Elections. In this case study, we intend to check in detail the demographics of users targeted by a set of potentially divisive ads before and during the American Election on the year of 2016. By using our framework, we identify if the ads published by the Russian Internet Research Agency (IRA) were delivered to vulnerable populations. In addition to this, we access the content of the ads and inspect how divisive they are [Ribeiro et al., 2019].

- Chapter 7 - Case study 4: Inferring Demographics of Election Polls. In this case study, we check if the data obtained with our framework reflects the demographic variation in short periods of time for specific scenarios. Particularly, we intend to check if abrupt demographic fluctuations in voting intentions, obtained with election polls, are captured by data collected with our framework. Is common that specific subpopulations, represented by people from a particular region or from a specific gender, have their interest changed about a particular candidate during the electoral process.

- Finally, in Chapter 8 we present the concluding remarks and future work.

Chapter 2

Background and related work

Demographic information is, in many cases, hard to be obtained and the inference of demographic data from the online environment frequently deals with limited and missing data. Despite being challenging, the scientific community developed plenty of techniques that tried to discover specific demographic attributes of online users given the broad application of this information in many online services. The input data used to derive demographics is originated from a diverse set of sources ranging from browsing behaviors to mobile phone calls. In this chapter, we present the evolution of online demographic inference since the popularization of the World Wide Web. We also analyze the usage of demographic data in particular contexts related to our case studies. Firstly, in Section 2.1, we describe traditional approaches that infer demographics from online data. Then, Section 2.2 sheds light on the approaches that derive demographic information by exploring OSN advertising platforms data whereas Section 2.3 describes the abuses on ads platforms, a related theme to our case study on Russian Ads (Chapter 6). Finally, we present the efforts that tried to infer the political leaning of media sources in Section 2.4, which is related to our case study presented in (Chapter 5).

2.1 Inferring demographics from online data

In the early days of the Web, demographic information extracted from users relied mostly upon surveys [Pitkow and Recker, 1994] or direct questions [Krulwich, 1997]. With the global popularization of the Web and the appearance of many services, several methods to gather information about age, gender, income and many other demographic attributes from users have emerged. Next, we present those different methods and show how they were used.

Some previous works gather demographic information about users by looking at their behavior in the Web. Murray and Durrell [2000] used neural modeling to find the age, gender,

income, and other three demographic aspects by inspecting Web pages accessed and search terms typed by users into their modeling. Hu et al. [2007] derived age and gender of Web users by employing a supervised regression model to define the probability distribution of these demographic attributes for a given webpage using an initial set of users with previously known demographics. Then, the authors used a Bayesian framework to predict other users' gender and age based on the demographic tendency distribution of the webpages accessed by the users. The authors lastly deal with webpages and users with sparse information. They used latent semantic indexing to identify similar webpages according to their content and similar users according to the accessed content.

Whereas some works relayed upon Web page click-through data, other related studies have used content evaluation to link the demographics of users. Kabbur et al. [2010] also employed machine learning techniques to explore the demographics, but the authors purposely avoided using any information about users (Web-browsing and Web-search histories, etc) by arguing that the increasing concern about privacy may lead users to deny the usage of their personal information and undermine approaches that rely on these types of information. Then, the features were based on the categorization of Websites' content and structure. Finally, Ulges et al. [2012] correlated the demographics of video viewers with the semantic concepts that appear in the clips. The concepts were extracted with the help of a supervised visual learner that extracts the concepts from the frames presented in the video.

The mobile environment has also been explored as a source of demographic attributes about users. Modern smartphones are shipped with advanced sensors and high processing power hardware, which enables several services and applications to be executed and explored. Recent work has claimed that user activity logs may be closely related to their interests and demographics. Zhong et al. [2013] created a supervised learning framework that used many features extracted from a huge log provided by Nokia Mobile Data Challenge (MDC) 2012. The authors cleaned the data to extract information such as the most used applications, the most visited places and the duration of calls to create its model. They also included an additional step called prediction adjustment, to detect correlation among the demographic categories they predicted (gender, job type, marital status, age, and the number of family members) and include the relationship in their model. In another approach, the authors collected more than 1 billion calls and text messages and constructed two undirected and weighted graphs where the users are the nodes and the communication events are the edges (one graph for calls and another for messages). Then, they extracted features mostly based on the topological organization such as degree, clustering degree, and the number of connections to learn a model and predict gender and age [Dong et al., 2014].

The writing peculiarities were also used to group users into demographic categories. One of the studies [Vel et al., 2002], extracted features from textual characteristics and style

aspects of a large set of emails to create an SVN classifier that was able to predict the gender and language background of email writers. Schler et al. [2006] used content published by bloggers to detect their gender and age interval. The authors leveraged the text published (after removing stop-words), part-of-speech tags and 'blog words' (*lol*, *haha*, etc) as features to train a classifier with Multi-Class Real Winnow (MCRW). Similarly, Otterbacher [2010] explored features such as text style, content, and metadata to infer the gender of movie reviewers from an IMDB dataset. Finally, Vikatos *et al.* [Vikatos et al., 2017] characterized language usage across demographic groups on Twitter, unveiling clear differences in their writing styles and topics of interest.

In the retail market scenario, one particular study derived demographic information about users employing different techniques. Wang et al. [2016] extracted gender, age, marital status, income, and education level from users by implementing the stochastic gradient descent (SGD) method to create a model that learns the demographics of users based on a 'bag-of-item' representation of the purchases.

Many other works leveraged demographics from OSN users. Mislove et al. [2011] provided one of the first studies in this environment, by looking at the gender and racial demographics of Twitter users, and analyzing how the demographics vary across different US states leveraging self-identified descriptions in users' profiles. The authors inferred the location of users by mapping the self-reported location field into geographic coordinates. Gender and race were inferred by leveraging the name of the users, the first name was used to define the gender and the last name used to define the race. Kosinski et al. [2013] have shown how Facebook 'likes', one of the most popular features to express positive association with some content on Facebook, can reveal much about the users. They basically proved that the association of multiple likes with a single user can expose personal attributes such as political views, ethnic origin, and even sexual orientation.

Pennacchiotti and Popescu [2011] used linguistic attributes to predict political affiliation and ethnic origin from the content of published tweets whilst another approach inferred basic demographics and even blood type and zodiac sign based on the location check-ins of users on Sina Weibo, a Chinese social media Website [Zhong et al., 2015]. Culotta et al. [2015] collected the demographic distribution for Websites (from quantcast.com) and assumed that the followers of the respective Twitter page have the same distribution. Then, they create a regression model to predict six demographic variables of Twitter users. Some works derived the age, gender, and race through analyzing the physical aspects of the users' profile pictures [Chakraborty et al., 2017; Reis et al., 2017; Messias et al., 2017; An and Weber, 2016]. They explored image processing tools that derive the physical characteristics of the user's face available in the profile picture (when a face is detected). In this direction, a group of researchers studied the accuracy of most popular tools to infer demographic aspects

based on face images [Jung et al., 2018] and they concluded that the tool Face ++ [1] reached the best accuracy for the detection of gender and race (above 90%). They also verified that inferring age from face images is still a complicated task for all analyzed tools, with accuracy below to 50%.

In spite of existing several efforts that infer demographics from online data, there is still space for advancement. It is clear that the techniques evolved together with technology improvements. The OSN advertising platforms, for instance, emerged as a new paradigm change since aggregate data is freely available to advertisers, taking the demographics inference to a new level, even though, the usage of this kind of data to infer demographics has been explored only in a few specific contexts. The first research gap explored in this thesis is the absence of an efficient way to convert the aggregate data provided by OSN advertising platforms into valuable demographic datasets. We then propose and develop an innovative framework that explores an OSN advertising platform to derive demographic data on a large scale. The framework is detailed in Chapter 3.

2.2 Inferring demographics from advertising platforms

A few concurring efforts have explored similar strategy to obtain demographic. In fact, social networks deeply changed the scenario of demographics inference in two aspects mainly. Firstly, OSN have access to a bunch of personal data about users rather than browser behavior only, which allows the inference of a much more detailed set of demographic information. The second aspect refers to making aggregate data about the targeted audiences freely available to advertisers. In this way, researchers may take advantage of this feature to build demographic cuts with no/low cost. Next, we present the studies that used the OSN advertising platforms as source of data.

Araujo et al. [2017] and Mejova et al. [2018b] used the Facebook advertising platform to estimate global lifestyle disease surveillance. The basic idea is checking if the amount of online people interested in lifestyle disease related themes (diabetes: diabetic diet, insulin, etc; obesity: bariatrics, obesity awareness, etc; tobacco: cigarette, smoking, etc) somehow correlates with the real prevalence of the diseases in some selected countries. In a first analysis, they found a reasonable correlation of users interested in diabetes with the prevalence of diabetes in the countries according to data from the World Health Organization (WHO). For other diseases, the correlation was low. However, they verified a within-country correlation, for example, more online men are interested in tobacco topics than women, similar to the

[1] http://www.faceplusplus.com/

prevalence of men who smoke in comparison with women. Also in a health-related work, the authors inspected the awareness of different demographic groups around schizophrenia-related on Facebook [Saha et al., 2017] and they found, for instance, that only 1.03% of Facebook users in the United States have interest on schizophrenia-related themes and that women, people with lower education levels and Hispanics are more aware of this disease.

Zagheni et al. [2017] analyzed the movement of migrants in the US by counting the number of immigrants from 52 countries in the U.S. according to Facebook and comparing with real data from the American Community Survey (ACS - provided by Census Bureau). The correlation found was very high even considering different age intervals and gender. Zagheni et al. [2018] extended the migration analysis work to predict migration of Mexicans to the US by combining historical data from ACS with online data using a Bayesian hierarchical model. The American Community Survey was also used in combination to Gallup surveys as baselines to investigate if the sexual preferences obtained with Facebook advertising platform (especially those from the LGBT community) are similar to the real incidence in the US [Gilroy and Kashyap, 2018]. They found a high correlation in the results, especially for sexual identity across the states. In addition to this, they also identified a higher willingness to disclose sexuality among younger cohorts than among older and also a large number of younger bisexual women, which is consistent with sociological work on gender and sexuality.

Another study investigated the relationship in the gender gap verified on Facebook and four indices of gender equality measured by the World Economic Forum (WEF) that includes economic opportunity, political participation, and education [Garcia et al., 2018]. They verified that the higher the proportion of men compared to women on Facebook, inside a country, the lower the metrics curated by the WEF for that country. The correlation of gender inequality on Facebook was also studied in India [Mejova et al., 2018a], where states with higher GDP (Gross Domestic Product) per capita, literacy and internet penetration are associated with lower gender gaps in Facebook. Mueller et al. [2017] studied the gender preferences for venues in the physical world through analyzing the check-ins in Location Based Social Networks (LBSN) such as Foursquare and Yelp. Another group of researchers proposed a methodology to identify cultural aspects across populations based on LBSN check-ins and compared with data from the World Values Surveys (WVS) [Silva et al., 2017].

Finally, some researchers used data from LinkedIn advertising platform to explore the professional gender gaps across US cities [Haranko et al., 2018], proving that online data is fairly similar across locations but varies strongly across industries and, to a lesser extent, across skills.

These particular studies reinforce the importance of the framework developed in this thesis and suggest that our methodology tends to be widely employed in the future.

2.3 Abuses on advertising platforms

Despite being the base of our framework, the advertising platforms have been employed in several forms of abuses. In Chapter 6, we study one particular misuse of this kind of platform, and in this section we discuss other abuses recently studied.

Prior studies highlighted abuses that ranged from inappropriately exposing private information on users to advertisers and for allowing discriminatory advertising (e.g., to exclude users belonging to a certain race or gender from receiving their ads). Andreou et al. [2018] pointed out that the function "Why I am seeing this" created by Facebook to let users know which attributes were used to target him/her is not transparent. They proved that Facebook's ad explanation may omit from their users many of the real characteristics that linked their accounts with the advertisers and sometimes is also misleading. Venkatadri et al. [2018] unveiled how an attacker can use one of the ad types available on Facebook, namely PII (Personally Identifiable Information), to infer full phone number by using an email address and also de-anonymizing all the visitors to a Website and infer their phone numbers.

A recent work elucidated the potential for discrimination for all the types of targeting in the Facebook advertising platform [Speicher et al., 2018]. In one of the malicious targeting options explored, authors exposed that an attacker may retrieve a list of the criminal records (available online for 18 US states with no downloads restriction) to track users on Facebook and eventually show harassing content. The criminal database contains, along with the specific criminal record, the name, the race, date of birth among other personal information that may be used to seek that user on Facebook through PII advertising. They also have shown that by exploring look-alike targeting available on Facebook, it is possible to expand a small discriminatory source audience in a considerably larger discriminatory audience.

Facebook advertising platform has also dealt with media criticism [Angwin and Parris Jr., 2016] and a civil rights lawsuit for allowing advertisers to exclude users by race, gender and other sensitive groups from receiving ads related to sale or rental of a dwelling and employment recruitment, which is forbidden by Federal law. In spite of assuming the possibility of discriminatory ads [Facebook, 2017] and agreeing to not allow ads related to housing and employment to eliminate sensitive groups from their audience, Facebook was found to be accepting discriminatory ads half a year later [Angwin et al., 2017a], having eliminated that possibility only in a later occasion. In another disruptive episode of its platform [Angwin et al., 2017b], Facebook allowed advertisers to target users by anti-semitic terms such as "How to burn jews", "Jew hater", "Nazi Party" and many others. Facebook remove the attributes and explained that those target options were created automatically and that campaigns using it were not common or widespread.

Finally, a group of researchers conducted a deep study on the ads viewed by Facebook

users on their timeline [Andreou et al., 2019], captured by a browser plugin installed by nearly six hundred users. They disclosed that a considerable number of advertisers belongs to potentially sensitive categories such as politics and religion. They also emphasized the need for better mechanisms to audit this kind of advertising platforms.

A rich body of prior work has focused on understanding emergent phenomena in social media. Some researchers studied that OSN users tend to select information that is similar to their beliefs, characterizing the echo chambers [Del Vicario et al., 2016; Garimella et al., 2018; Lima et al., 2018]. The filter bubbles, a situation in which algorithms, especially those exploring machine learning, amplifies the segregation when recommending content an individual is likely to agree with, were also covered [Pariser, 2011; Flaxman et al., 2016]. The polarization was also addressed, in particular when related to the ideological discourse [Castillo et al., 2014; Guerra et al., 2013; Sharma et al., 2017]. We provide a complementary perspective on the topic by examining how echo chambers and polarization can be engineered on social media through targeted advertising. A recent work conducted a detailed study about Facebook advertising environment by analyzing thousands of ads collected through a browser plugin [Andreou et al., 2019]. More closely related to our work, Kim et al. gathered Facebook ads from individuals and analyzed who are behind divisive ad campaigns, reporting suspicious foreign entities [Kim et al., 2018]. Differently, we focus on understanding the disruptive ability of microtargeting for providing divisive political ad campaigns.

Our effort may also be considered complementary to prior work that attempts to understand the abuse of social media by misinformation campaigns, especially along political elections [Lazer et al., 2018; Vosoughi et al., 2018]. Our work provides a better comprehension about a key dissemination mechanism of fake news stories, highlighting how advertising platforms allow injection of misinformation in social systems and choose vulnerable people as the target.

More specifically, we analyze the demographics of people reached by a particular set of potentially divisive ads. We intend to understand more deeply the OSN advertising environment and identify if malicious publishers may reach vulnerable populations with the micro-targeting options available. We aim at exploring our framework to identify the specific populations that received ads published by a Russian agency called Internet Research Agency (IRA) before and during the American Election of 2016 [2],[3]. We also expect to provide an in-depth analysis of the publicly released dataset by accessing if the published content seems to be divisive. This research gap is addressed in Chapter 6.

[2]https://www.wsj.com/articles/you-cant-buy-the-presidency-for-100-000-1508104629
[3]https://www.nytimes.com/2017/11/01/us/politics/russia-2016-election-facebook.html

2.4 Exploring OSN data to infer politics-related information

The data from Online Social Networks was also explored to extract information about topics in politics. In fact, OSN represents a valuable source of data about politics since in addition to the high presence of the population in these platforms, it is also a place where people directly influence each other. In this section, in particular, we present studies that try to infer the political leaning of OSN users, predict elections and detect the ideological bias of media outlets.

2.4.1 Political polls and elections prediction

The OSN allowed users to actively participate in public debates and expose their opinions publicly. Several works explored content produced by the crowd and applied in many different scenarios such as the detection of earthquakes [Sakaki et al., 2010], identifying areas with more incidence of dengue, a tropical infection with severe flu-like illness [Gomide et al., 2011], or detecting flu epidemics [Lee et al., 2013]. In common, all the above-mentioned studies leverage posts on Twitter about particular topics to achieve their results. Particularly, the political bias of OSN users has received much attention from researchers in the last years. Many researchers investigated data produced by OSN users to detect their political bias based on the content published by users [Fang et al., 2015; Sylwester and Purver, 2015], based on the bias of known people users follow on Twitter or Facebook [Golbeck and Hansen, 2011; Bond and Messing, 2015], or by filtering the topics of interest of users and matching this interests with republicans or democrats [Kulshrestha et al., 2017]. Some of the previous studies have tried to infer the political leaning of users on Twitter through analyzing the social links a specific user has with other users with known biases [Conover et al., 2011a; Golbeck and Hansen, 2011]. Another approach focused on the language used by users with different partisan inclinations [Makazhanov and Rafiei, 2013; Sylwester and Purver, 2015]. Conover et al. [2011b] compared distinct machine learning approaches to detect the political leaning of Twitter users. The supervised method based on the community structure of diffusion networks outperformed those based on the hashtags as features as well as the version that used full text of the tweet.

The content produced by OSN users has also been examined to predict elections. In one of the most controversial works, researchers argued that Twitter could predict the election results [Tumasjan et al., 2010]. The authors collected the tweets that mentioned the parties and/or candidates in the month prior to the German national election to infer the ranking of candidates/parties based simply on the tweet's volume. They also derived the sentiments

of tweets and confirmed some previous findings that states positive emotions consistently outweigh negative emotions.

Jungherr et al. [2012] directly refuted the results that indicate the correlation between the total amount of tweets and the election's outcomes. The authors pointed out many methodological issues such including the lack of information about the collection and other method details that prevent other researchers to reproduce the experiment. Additionally, they conducted a similar collection and concluded that the real winner of the election would be the Pirate Party (Piraten), whose tweets were excluded from the original analysis without explicitly explaining the reason. Other researchers tried to reproduce the method that counted tweets and predicted elections, and it partially worked as in the case of the Singapore election [Skoric et al., 2012]. But, in general, the method proved to be flawed such as the 2010 US Senate special election in Massachusetts [Chung and Mustafaraj, 2011] and the Dutch Senate elections of 2011 [Sang and Bos, 2012].

Williams and jeff Gulati [2009], analyzed the US presidential election of 2008 and found that the number of Facebook supporters, in general, indicates electoral success. Conversely, Giglietto [2012] tried to predict the results from elections in the Italian cities using the number of likes of candidates on Facebook and obtained poor results.

Wang et al. [2014] conducted a series of daily voter intention polls through Xbox gaming platform, a highly biased towards male and young people. They then adjusted the estimates by using linear regression and partitioning the population into cells that considered the regions with more prevalence of each one of candidates and regions with more players.

Gayo-Avello raised many concerns in predicting elections and elicited many recommendations to drive this kind of research [Gayo-Avello, 2012; Gayo-Avello et al., 2011]. Among concerns, that can also be applied to studies that consider other Social Networks as the source of data, we highlight the self-selection bias issue, meaning that those studies do not capture the political insights of all users since only the politically active ones produce data. The author also emphasizes that the demographic bias in social media, an issue also inspected by Mislove et al. [2011], is often ignored. The online population does not represent the real world population and this must be considered. The main recommendations include acknowledging the demographic bias and defining clearly what constitutes a "vote" in the OSN.

On the other hand, many studies showed that some reliable data about politics can be extracted from Social Networks. DiGrazia et al. [2013] pointed out that the vote share for candidates had a high correlation with the popularity of the candidates on Twitter in many district elections. Another study presented some interesting findings for the 2010 Swedish election such as the over-represented participation in the online debate for a few users, being those more active users identified mostly as journalists, consolidated bloggers or politicians,

although the authors also detected some influential anonymous users [Larsson and Moe, 2012]. Bermingham and Smeaton [2011] employed a linear regression that considered the volume of tweets together with sentiment analysis scores to predict the 2011 Irish General Election. Finally, some researchers created a methodology to define the fluctuation of candidate preference by users based on their history of visited websites [Comarela et al., 2018].

In this thesis, we check if the popularity fluctuation verified in the election polls before the elections are captured in the online perspective. We conduct a comparison of the demographics provided by election polls with data collected by our framework. Particularly, we verify if the increase/decrease of voting intentions for one candidate in general or among specific groups have some correlation with the variation of interest on Facebook. For this analysis, we use the Brazilian presidential election as a case study. The Brazilian political scenario is very turbulent in last years due to many controversial and uncommon situations such as a president impeached, a former president arrested, and even a candidate that was stabbed. In addition to this, the number of unemployed citizens is increasing in a devastated economy. This situation has transformed the presidential election in a very turbulent period, which is clearly reflected in the social networks. This case study is presented in Chapter 7.

2.4.2 Political leaning of news media sources

As explained in previous sections, the advertising platforms data took demographic inference to a new level by giving advertisers access to aggregate data of targeted audiences. Another interesting point in using this new source of data is that it allows measuring the distribution of subpopulations with specific interests, such as people who are interested in a specific newspaper such as The New York Times. In one of the case studies, we derive the political bias of media sources by checking the political leaning distribution of the audience for each specific media outlet. In this section, we discuss studies of political leaning in online media sources.

Traditionally, news media organizations played an important role in societal evolution by acting as *gatekeepers of information*, and by deciding and regulating what news is consumed by the common people [Shoemaker et al., 2009]. With this powerful role played by them, media researchers have long worried that an ideologically partisan and deregulated media can have a high impact on the political outcomes, and ultimately on our society [Groseclose and Milyo, 2005; Chiang and Knight, 2011]. Therefore, a large number of research studies as well as media watchdog groups like FAIR (`fair.org`) and AIM (`aim.org`) have investigated *news media bias*, and evaluated the content produced by different news organizations for fairness, balance, and accuracy in news reporting.

Most of the efforts have focused on studying political bias in traditional news me-

dia [Budak et al., 2016; Gentzkow and Shapiro, 2010; Groseclose and Milyo, 2005; Munson et al., 2017]. Particularly, Groseclose *et al.* [Groseclose and Milyo, 2005] linked media sources to the members of the US Congress utilizing the co-citation of political think-tanks, and assigned them political bias scores to media sources based on the ADA scores of Congress members given by the political watchdog group 'Americans for Democratic Action'[4]. Gentzkow *et al.* [Gentzkow and Shapiro, 2010] inferred 'media slant' based on whether the language used by a media source is more similar to congressional Republicans or Democrats. Budak *et al.* [Budak et al., 2016] used a combination of crowdsourcing and machine-learning methods to study the selection and framing of political issues by different news organizations.

As online news sources are continuously gaining popularity, Munson et al. [2017] assigned political bias scores to popular news websites; whereas Babaei et al. [2018] proposed a system called "purple feed" to show users news which are likely to have high consensus between both republican and democrat leaning readers. In a recent work, Le et al. [2017] presented a method to measure ideological slant of individual news articles by monitoring their consumption on Twitter. They analyzed the connectivity of the users tweeting an article to label them as republican or democrat leaning.

While political bias of news media has received a lot of attention, other forms of media biases have also been analyzed (*e.g.,* demographic bias [Chakraborty et al., 2017] such as gender [Shor et al., 2015] and racial biases [Ramasubramanian, 2007]) to address concerns about these biases in news coverage, which can reinforce or even create certain forms of racial, gender, and ethnic stereotypes [Gilliam Jr et al., 1996]. Similarly, efforts have been made to understand the topical coverage biases in news dissemination [Chakraborty et al., 2016] or recommendations [Bakshy et al., 2015; Chakraborty et al., 2015], and whether they can lead to 'filter bubbles' [Pariser, 2011]. Being aware of such biases of different news media outlets is crucial for the society, since awareness can play a critical role in shaping readers' assimilation of news published by these outlets [Dooling and Lachman, 1971].

Inferring the political bias for the increasing amount of media outlets that emerge every day is a huge challenge. Traditional approaches to detect this information rely on content analysis, demanding human verification of the published news stories. An alternate approach has shown that the bias in the audience reflects the bias of the specific media outlet. Both approaches have proved to be hard to execute, time-consuming and hard to scale, delivering the bias of at most 500 media outlets. We infer the bias of thousands of media outlets by detecting the political leaning distribution of the audience with our framework and compare it with current approaches. Additionally, we aim at expanding our analysis and show the

[4]www.adaction.org

biases of the news sources along with other demographic aspects such as race, gender, and age. This research gap is addressed in Chapter 5.

Chapter 3

Framework to infer demographics

Broadly speaking, social networks operation relies upon the content production and sharing by users through their public profiles. Users are able to publish videos and pictures, share news and post their ideas to her friends/followers. In this context, it is common for businesses, public figures and institutions to join the social networks by creating public profiles and start spreading content aiming at attracting engagement from users, which end up aggregating value to their brands. The ordinary posts on social networks are shown to the users who follow the owner of the published content. However, social networks allow publishers to pay for boosting their post/content and reach much more users than only those who follow them. This sponsored content model is exactly how the advertisement ecosystem of social networks works. In order to provide an attractive environment to advertisers, the social networks provide a wide range of demographics and behaviors to select the users to be targeted by those aiming at creating ads. As depicted in figure 3.1, OSN track users actions such as posts and accessed content to infer their main characteristics, which will be later available to advertisers when selecting the audience they want to target. In the example picture, we highlighted some users exploring content related to Donald Trump and dogs, entities that are next derived as attributes, together with the type of device used and the location where access came from. By exploring these attributes, the owner of a pet shop located in Los Angeles would be able to pay the OSN to show a post describing their services to users living in Los Angeles and interested in dogs, for instance. Obviously, the OSN explores as most as information they can about users which includes posts, browsing, 'likes', public profile data, and so on. Therefore, the attributes embrace plenty of categories such as the device or the Operating System used by users (IOS, for instance), the places users check-in, and many others

In this thesis, we explore this process of creating an ad to extract demographics, as detailed in the remaining of this chapter.

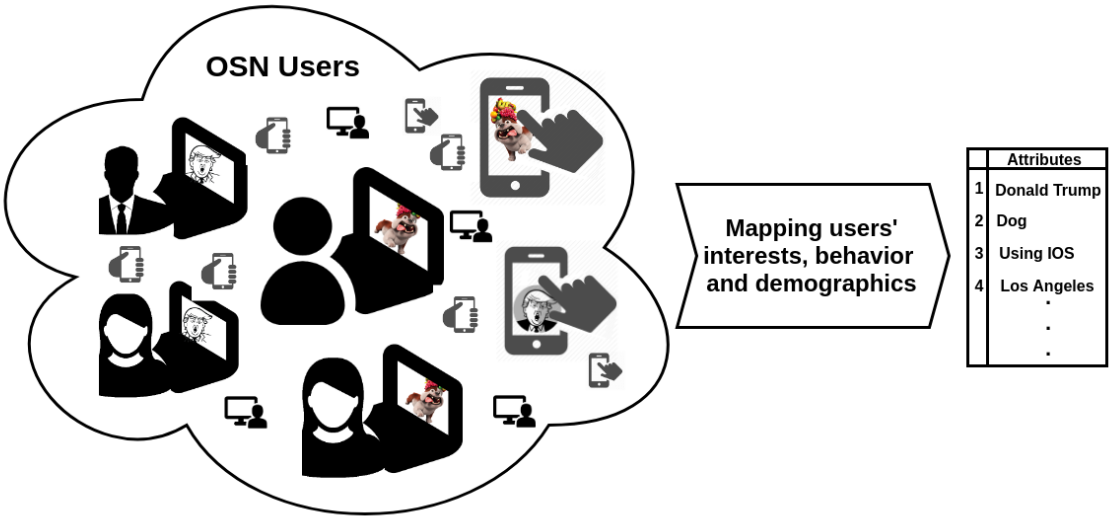


Figure 3.1: **OSN derives attributes from users**

The chapter is organized as follows. Firstly, we describe in detail the OSN advertising platforms and their options to target users. Then, we describe our novel methodology to infer demographics employed in this thesis as well as the architecture built to scale our approach. Next, we detail how we deployed our methodology in the Facebook advertising platform. Finally, we present the limitations of our approach.

3.1 OSN advertising platforms

OSN advertising platforms have evolved significantly in the last years. With access to personal information and activities of millions of people around the world, the Online Social Networks allow advertisers to target very specific niches of users considering personal information such as name, email address, demographic aspects, behaviors, and many others.

The new type of advertising platform provides basically three forms to define the audience an ad should target:

- **Personally Identifiable Information (PII) targeting**: in this case, advertisers provide a list containing information that can link the customer with his/her public account such as email or phone number, out of with OSN tries to match the respective online contacts. The list of allowed PIIs includes different types of data such as email, phone number, the name of the user, date of birth, and ZIP code).

  This mechanism is quite useful for businesses that keep a list of customers and want to exhibit ads for them with offers of products related to their last purchase, for instance.

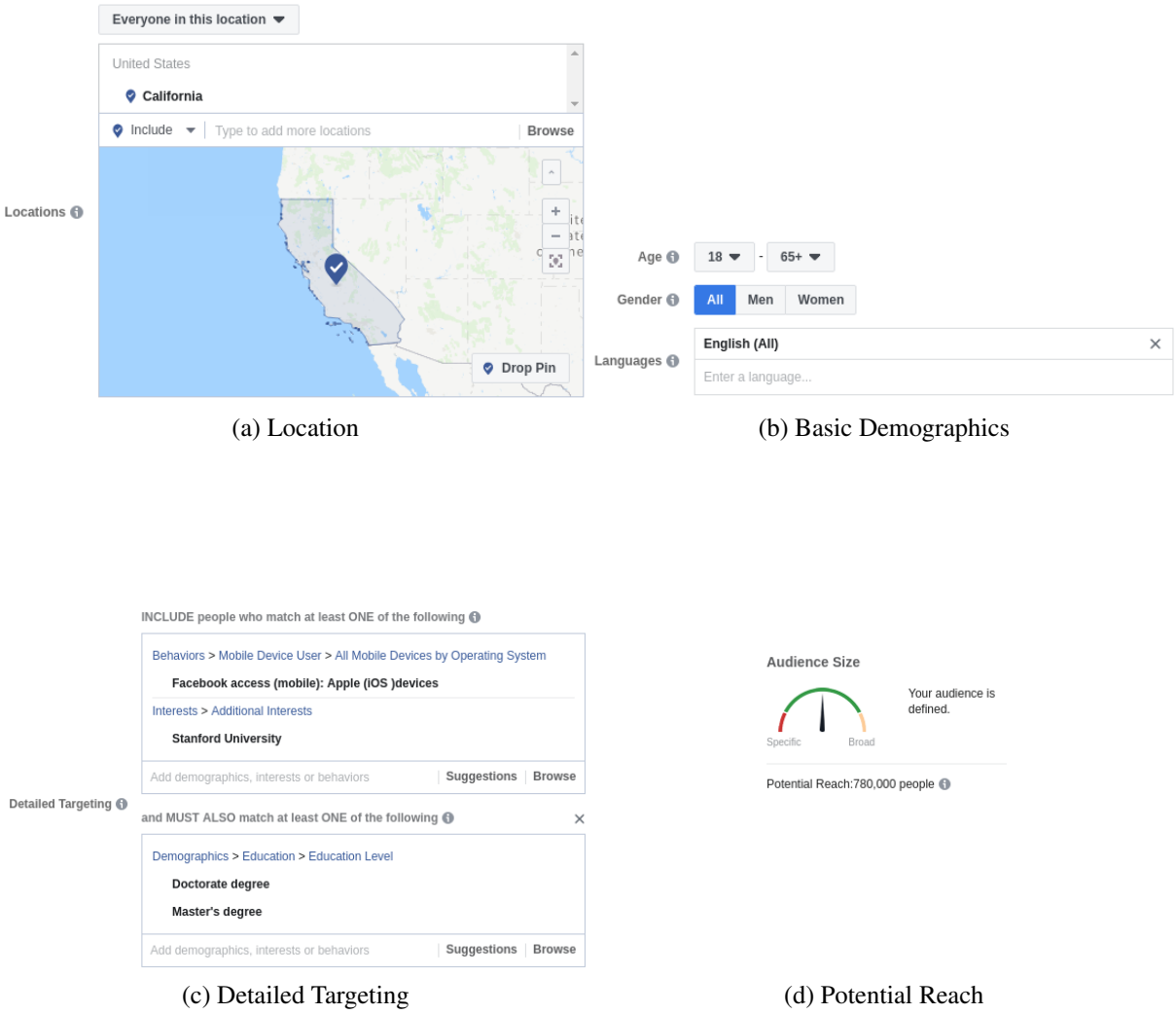
An experiment has shown that Facebook, for example, was able to target above 60% of the provided list in most of the test cases [Speicher et al., 2018].

- **Look-alike audience targeting**: this target option is based on finding a similar audience to an initial set of customers, namely the *source audience*. As an example of usage, a page owner can attract more followers by showing ads for users with similar characteristics to the fans of the page, inviting them to follow the page. The conversion rate, i.e. the percentage of users who end up following the page after seeing the ad, is likely to be higher if the ads target people with similar demographic characteristics of the page fans.

- **Attribute-based targeting**: the advertiser can define the targeted audience based on a range of attributes that varies from user interests to the device he/she uses when accessing the OSN. More specifically, these attributes can be organized into five main categories: 1 - Basic demographics - that includes the most basic info like gender, age, location, and language; 2 - Interests - entities in which user demonstrate interest on Facebook and can range from color preferences (black, blue), to religious orientations (Protestantism, Catholicism), artists and politicians, and many other entities. This kind of information through tracking user activities online; 3 - Behaviors - behavioral characteristics like 'Business Travelers' or 'Charity Donors', type of device and platforms used to access the OSN (mobile platforms, browser, etc); 4 - Advanced demographics - includes a wide range of demographic attributes like political leaning, income level, formation school or university, job title, new parents, parents with children in different ages (preschoolers, teenagers, etc), and so on; 5 - Connections - a filter that allows the advertiser to include or exclude from the audience people who follow its page. The inclusion of people narrows the audience, meaning that the resulting targeted audience includes people that are connected to the page and also match the other attributes. In the case of exclusion, the audience is also narrowed, but in this case, people who follow the page are eliminated from the final audience. The main reason for the exclusion filtering is to avoid reaching people that are already engaged with the page.

In this work, we explore the attribute-based targeting to gather demographic information, so next, we show some examples of this particular mechanism of targeting by exploring different platforms. It allows the targeting of very large and generic populations such as users who live in Mercosul countries, as well as very specific populations.

Figure 3.2 depicts an example of a target audience that may be constructed using some of the attributes provides by the Facebook advertising platform. At first, we defined the location as California state (see figure 3.2 (a)). Notice that on the top of the location selection,



(a) Location (b) Basic Demographics

(c) Detailed Targeting (d) Potential Reach

Figure 3.2: **Attribute targeting example for Facebook advertising platform.**

we choose to target 'everyone on this location' but there are also other options like 'people who live in this location' and 'people traveling in this location'. Next, as shown in figure 3.2 (b), we selected the default age interval (that includes people above eighteen years old - 18-65+), the default gender (both, men and women) and the language field as English.

After defining the basic demographic attributes we select more specific attributes to refine our audience. First of all, in the first box of figure 3.2 (c) we include in our audience people interested in Stanford University and who access the OSN with IOS. In the second box, we narrow the audience by selecting only users with a Master's or Doctorate degree.

Summing up the final selected audience is composed by the Facebook users aged above 18 accessing the OSN from IOS devices, with high levels of education (Master's or Ph.D.), that are at California state, interested in Stanford University, and have English as their language. The potential target for this particular combination of attributes is shown by Facebook before running the ad, in this case, $780K$ users (3.2 (d)). This number represents the max-

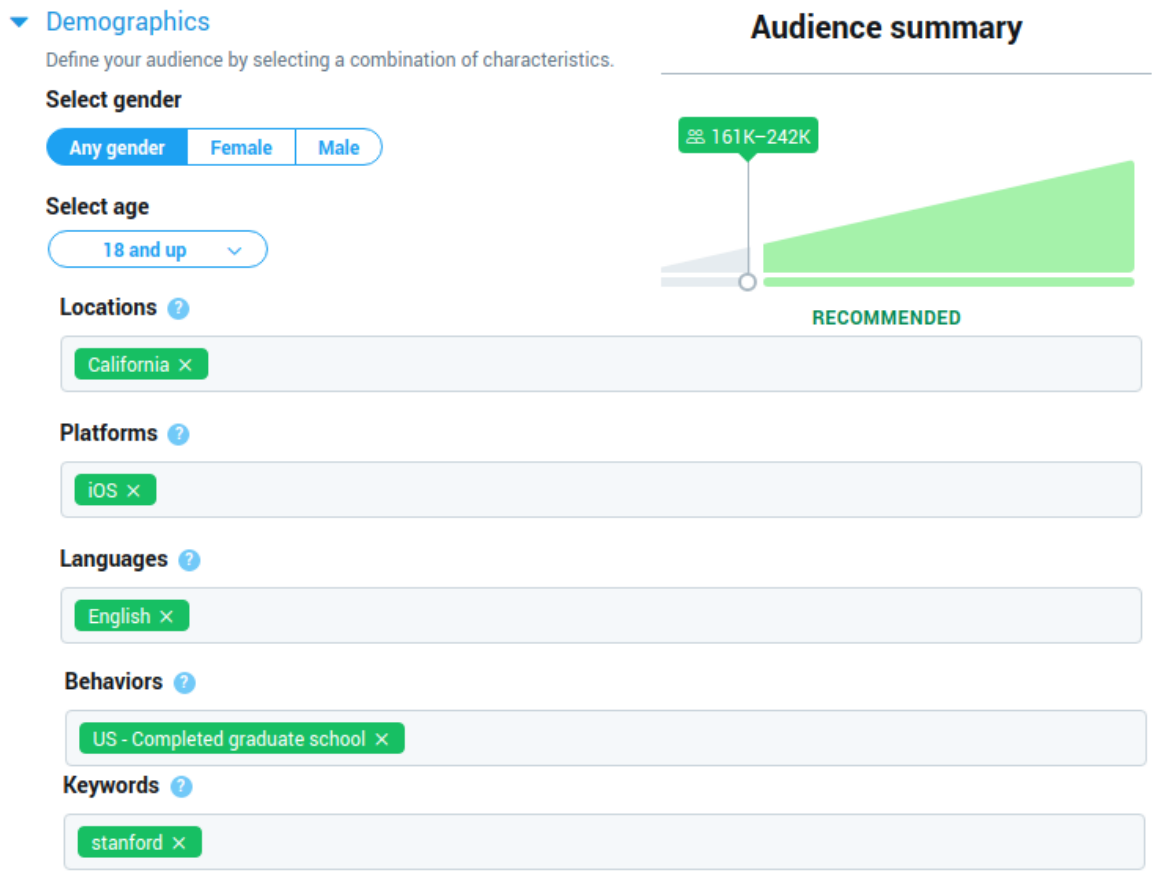


Figure 3.3: **Attribute targeting example for Twitter advertising platform.**

imum number of users an ad with this target options might reach and it is a key for our methodology to extract demographics as we will detail in the next subsection. The actual number of users the ad will reach after its publication is not necessarily that potential target, but rather depends on how much money the advertiser decides to pay.

The same audience reproduced with Twitter advertising platform, depicted in figure 3.3, results in $242K$ Twitter users. Finally, we calculated the audience size for a similar audience with the LinkedIn advertising platform, resulting in 29.000 users. When selecting the audiences in different platforms, we verified some divergence in the name or in the presence/absence of attributes. On the Twitter platform, for instance, the highest education level includes users with Master's and Doctorate degree together, and the attribute is considered as a behavior instead of a demographic field. The LinkedIn platform has no option to select people interested in Stanford University and the device used when accessing the OSN. Then, to complete the audience attributes in this last platform, we selected those users who studied at Stanford University. This variation in the attributes is expected since every company may follow their own rules to select the most important ones to their systems, but the fact is that

all systems provide a high range of attributes with detailed targeting possibilities.

Notice that for the audience selection presented before we used basic demographic attributes (location, age, gender), advanced demographics (education level), and behavioral information (the device used to access the social network). In addition to these attributes, an advertiser can also include interests such as Donald Trump, beer, The New York Times and thousands of others in its audience creation. This abundance of attributes certainly provides targets for a very diverse range of advertisements: a newly installed factory of craft beer could create an ad to publicize its beverage to adult users with interest in beer and with high-income levels who live in neighboring cities; with a very distinct goal, an oppositionist of Donald Trump may reach white Texan males users who are between 20 and 50 years old and are interested in the president Donald Trump, to convince them that the opponent is a better option.

As expressed in the previous examples, the attributes can be combined in very different manners, similarly to a Boolean formula, in which an attribute or the negation of an attribute can be combined with other attributes using operator OR as well as operator AND, *i.e.,* a predicate. We will refer to this combination of attributes as *Targeting Formula*.

3.2 Methodology to infer demographics

We leverage the attribute-based targeting to gather demographics by including a new layer of attributes selection that allows us to stratify the data in a comprehensive manner. In order to make clear the characteristics of data we intend to infer in this thesis, we considered the definition of 'demographic' as follows. A particular sector of a population or a group of people who are similar in one or more characteristics.

As an example, let's take a simple targeting formula, used in a particular OSN advertising platform to select all users that live in the US. This formula includes people from both genders aged above 13 who live in the US. It is named in our framework as Original Targeting Formula (OTF) and represents the entity from which we intend to infer the demographics, in this case, the population of the US.

Remember that, for each targeting formula, the advertising platform calculates the number of users who match that attributes' selection (potential target). Let's say that X is the potential target for the example OTF defined above, *i.e.,* the number of users of a particular OSN that live in the US.

We can derive five new targeting formulas in which we include a new attribute that limits the audience according to their political alignment. For example, a new targeting formula would select people from both genders aged above 13 who live in the US and additionally

with the conservative political leaning. It would represent a subpopulation from the original targeting formula with an audience size (potential target) of Xc (conservative). We can also calculate the audience size of the population in the US with liberal political leaning (Xl) after replacing the attribute conservative by the liberal political leaning. Finally, we can derive the audience size of people that live in the US with other three different political leaning categories: very liberal (Xvl), moderate (Xm), and very conservative (Xvc).

Based on the amount of five subpopulations with different political alignments, we can derive the political leaning distribution for the particular OSN users who live in the US by dividing the amount of a particular subpopulation by the sum of all subpopulations. As a general formula to calculate the percentage of the population for a particular demographic dimension, pd, we need to divide the number of users for that dimension by the sum of all dimensions in the same demographic attribute (n represents the number of demographic dimensions for a particular attribute).

$$pd = \frac{Xd}{\sum_{i=1}^{n} Xi}$$

By applying the general formula to the political leaning example, we can calculate the percentage of conservative users in the US as shown below.

$$pc = \frac{Xc}{Xvl + Xl + Xm + Xc + Xvc}$$

The calculus is similar for other political leaning dimensions. Notice that the divisor is the sum of all demographic dimensions instead of only X (the original population size), because the demographic attribute may not be inferred for all OSN users. We may also extend our initial targeting formula to narrow the audience and include users from both genders, aged above 13, who live in the US and that are also interested in the New York Times, for instance. And again, by computing the potential targeting for each one of the five subpopulations we generate the political leaning distribution for The New York Times audience among the OSN users in the US.

The underlying principle of the methodology is, given the original targeting formula (OTF), we add another layer of derived targeting formulas that splits the total targeted population into smaller sets filtered by a particular demographic attribute. For the sake of clarification, we assume here that demographic attribute or demographic category is the type of the demographic element, i.e. Race, Political Leaning, etc; and demographic dimension represents the values available in that attribute, for instance, Hispanic, Conservative, Male, etc.

Figure 3.4 shows this new layer that allows the derivation of demographic attributes by

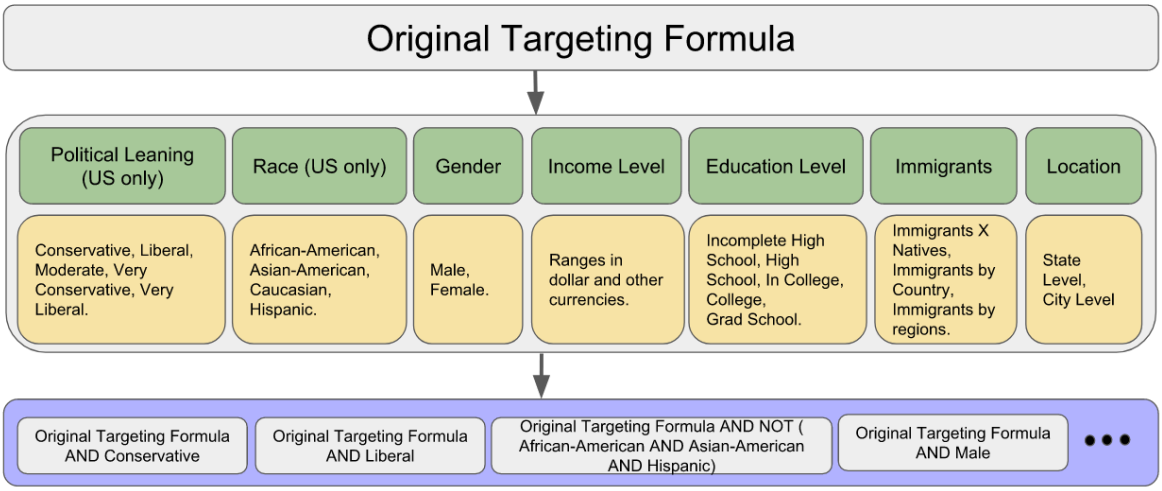


Figure 3.4: **New layer to process demographic data**

employing our approach. The green boxes show the demographic attributes whereas the yellow boxes detail the demographic dimensions. In this research, we applied this methodology to gather the demographic distributions for gender, education level, age intervals, immigrants and location in addition to the aforementioned categories.

In order to extract the demographics, we combine the original targeting formula with every demographic dimension and create a resulting formula that limits the audience to include only the desired dimension. In general, this combination is performed by concatenating the OTF with the specific attribute using the operator AND that narrows the audience. This combination results in a new targeting formula, whose result is the count of records (*i.e.,* the number of users) for which the predicate is True. The bottom box in the figure pinpoints some combinations of the OTF with demographic dimensions.

In addition to the combination of the OTF with one layer of demographics, we may also include a second layer or a third layer and treat demographic categories in parallel to retrieve more detailed data. For instance, we can use two layers and infer the percentage of each race filtered by location (state or city), or three demographic layers to obtain the age distribution of immigrants grouped by gender.

We should also emphasize that, although the examples presented before considered the OTF as the users who live in the US, the targeting formula may be considered the entity we want to unveil the demographic distribution of the audience: a country or other geographic location (state, city, region), a person (celebrity, politician or other public figure), a media outlet (newspaper, magazine, blog, TV Channel). Furthermore, the methodology can also be employed to disclose the actual audience of a particular ad. Section 6, for instance, present a case study in which the real audience of the ads were explored to shed light on how socially

divisive ads might impact vulnerable people.

Although the demographics can be calculated by selecting the targeting formula in the online systems provide by OSN advertising platforms as depicted in figures 3.2 and 3.3, we need to use programmatic access through Application Programming Interfaces (APIs) provided by the OSNs to scale our approach. This particular APIs, known as Marketing APIs, can be used for many purposes including running ads, creating personalized audiences and automatically checking the impact of the ads (number of people who visualized or interact with the ad, for instance). Next, we describe the architecture we employed in this work to scale even more our crawler capabilities.

3.3 A scalable demographics crawler

The methodology to infer demographics is based on making one request to the Marketing API for every demographic dimension to be analyzed. Therefore, depending on the size of the entities to be accessed the number of requests may grow very fast, reaching millions of API calls. For instance, we need to make a huge amount of requests if we intend to derive the demographics of the audience of thousands of media outlets (as one of our case studies). On the other hand, the call limits of Marketing APIs are strict and collecting a large amount of data may take months.

In order to expand our crawler capabilities, we created a distributed approach that allows the automatic parallelization of the collections task as depicted in Figure 3.5. The Crawler Coordinator is supplied with the entities to be collected, checks the Crawler Agents that are not busy with other collections and send them the entities each one is intended to collect and the developer tokens that should be used to complete the requests. The Crawler Agents then, make the requests to the OSN Marketing API and store the responses in a remote database. The crawler task is subject to many failures and the Crawler Agent must address every eventual error and ensure that the responses for all requests are properly delivered.

Finally, the Crawler Coordinator checks when all Crawler Agents finished their collections and then processes the raw data containing the potential target number for each request to calculate the demographic distributions. After implementing this parallel approach that takes advantage of multiple user tokens, we were able to make around 2.5 million requests per day. The tokens are provided by the Marketing platforms and the number of tokens available to developers depends on platforms policies and developers settings.

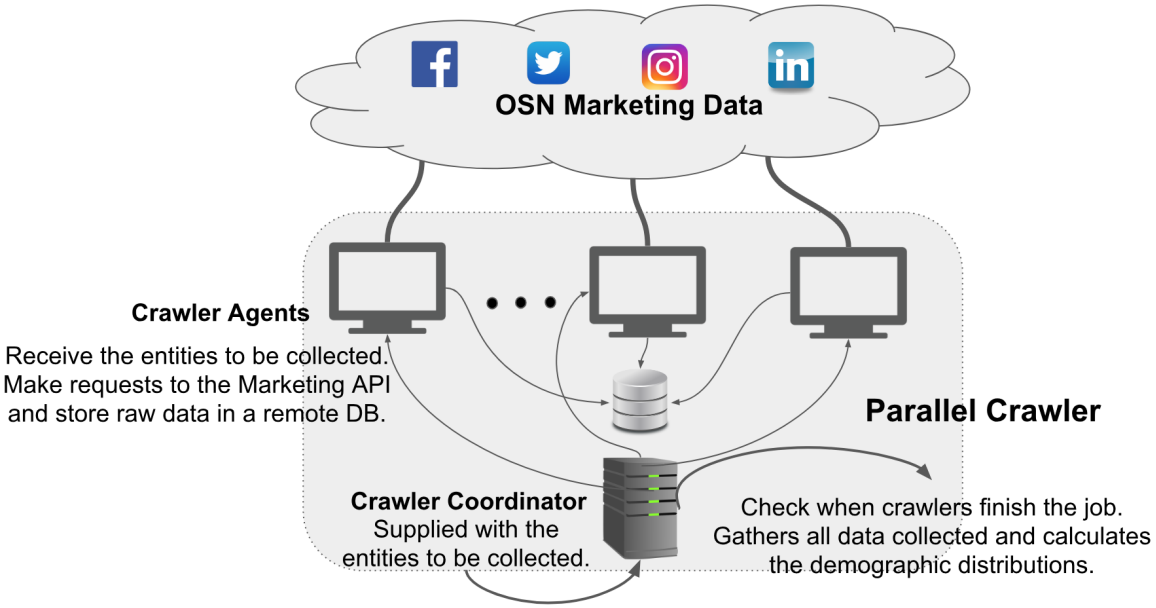


Figure 3.5: **Parallel crawler architecture**

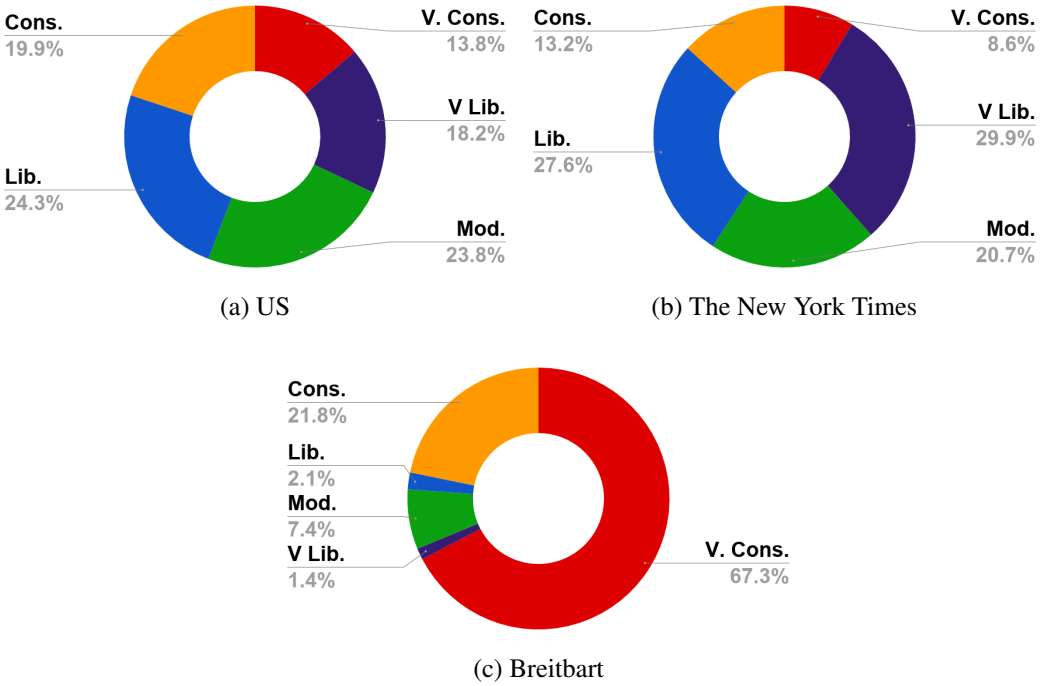
3.4 Facebook advertising platform

In spite of being able to employ our approach across different OSN platforms, we decided to use Facebook due to the following reasons.1) The impressive number of active users across the globe, with more the 2.32 billion as of December 31, 2018[1]. 2) The platform also maintains the Facebook Marketing API[2], a well-documented API that allows developers to make plenty of requests in a very simple way, including the requests for the potential target, given a specific formula. 3) The huge amount of attributes available, nearly 250 thousand [Speicher et al., 2018], which allows the study of a broad spectre of scenarios.

We should also notice that Instagram is also owned by Facebook and the Facebook Marketing platform provides data for both social networks. In addition to this, the number of active users is still growing on both platforms.

Today, anyone with a Facebook account can access the system designed by Facebook[3], select one of the listed options (Brand awareness, Reach, Traffic, Engagement, and so on) and boost posts that will reach a larger audience than the limited circle of friends. The ads must be linked to a Facebook public page and cannot be hosted and published by users with personal accounts only. However, any Facebook user can access the system and simulate the creation of an advertisement and selection of targeting options before effectively running the ad.

[1]https://newsroom.fb.com/company-info/ - accessed on 31 January, 2019.
[2]https://developers.facebook.com/docs/marketing-apis
[3]https://www.facebook.com/adsmanager/creation



(a) US

(b) The New York Times

(c) Breitbart

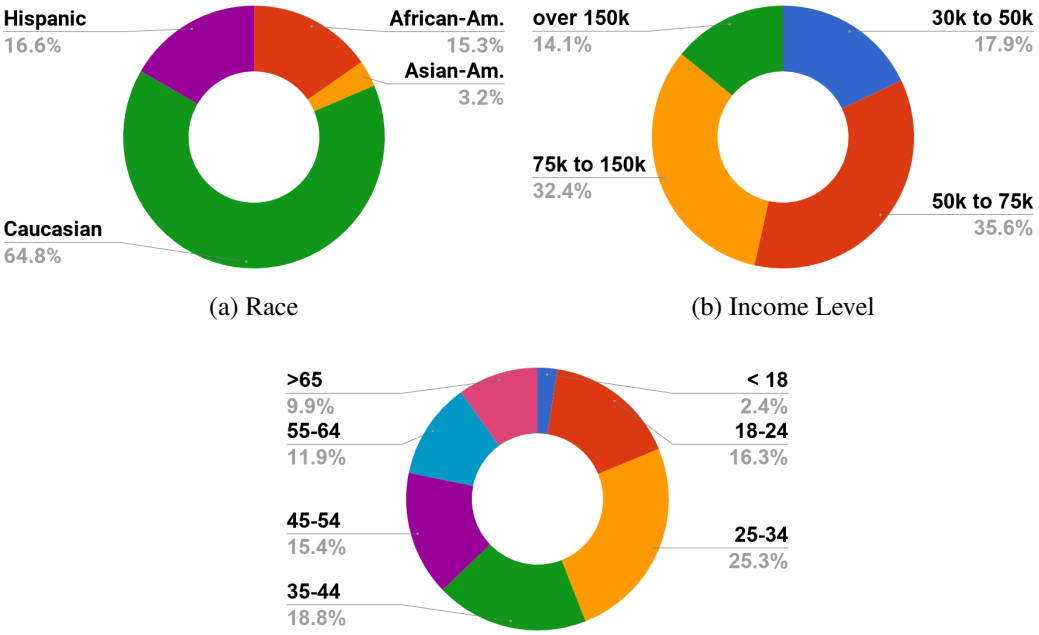
Figure 3.6: **Political leaning distribution for different targeting formulas.**

We accessed the system as a regular user to employ our methodology and derive some demographic attributes from the US population on Facebook. Figure 3.6 depicts the political leaning distribution in the US, as well as for The New York Times and Breitbart, influential media outlets in the US. The last one is a known conservative media outlet with a relevant role in the US Presidential election of 2016 [4,5].

Figure 3.7 shows the distribution of income level, race, and age in the US Facebook users. For the race distribution, in particular, we used three 'behaviors' available on Facebook Marketing API, named as multicultural affinities that identify people who live in the US whose activity on Facebook aligns them with three different races: African-American, Asian-American or Hispanic. Despite the absence of a 'behavior' that identifies the predominant race in the US, we compute this percentage by including in the targeting formula the negation for the other 3 'behaviors'. As a final example, we used our methodology with three layers to derive the age distribution of immigrants grouped by gender as depicted in figure 3.8 (a) (also considering Facebook users who live in the US). For comparison, figure 3.8 b) shows the age distribution for US Natives. The immigrants have a high proportion of adults, especially on the interval that ranges from 30 to 49, which is consonance with a recent study

[4]https://www.nytimes.com/2016/08/27/business/media/breitbart-news-presidential-race.html

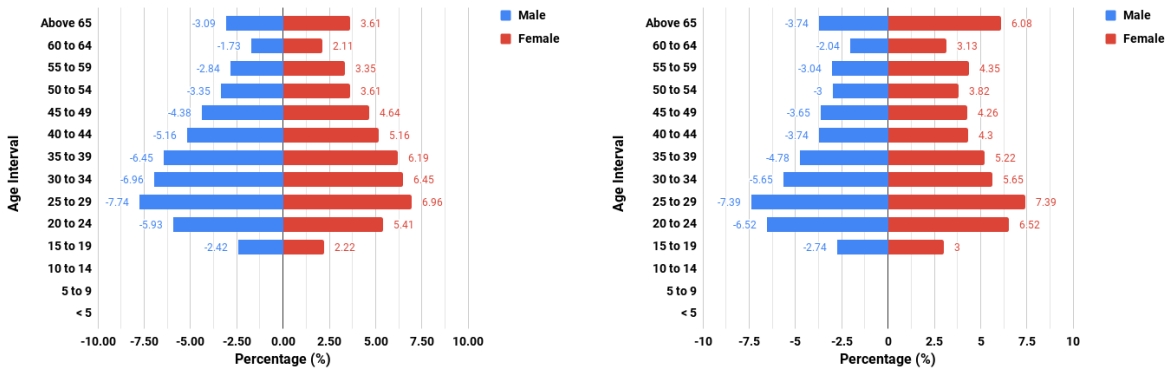
[5]https://www.npr.org/2017/03/14/520087884/researchers-examine-breitbart-s-influence-on-misleading-information



(a) Race (b) Income Level

(c) Age

Figure 3.7: **Demographic distributions from the US Facebook users.**



(a) Immigrants (b) US Natives

Figure 3.8: **Age intervals by gender among the Facebook users in the US.**

published by Pew Research[6].

[6]http://www.pewhispanic.org/2017/05/03/facts-on-u-s-immigrants-current-data/

3.5 Mapping interests, behaviors, and demographic attributes

Different from the Facebook advertising web interface, in which we can manually select the attributes in combo-boxes and by typing the attribute in searching tools, the usage of programmatic API requires the attributes to be converted into their specific IDs. So, we need to map interests, behaviors and demographic attributes used in the OTF into the ID recognized by the API.

The attributes that make up the OTF may be obtained through two different approaches. In some particular applications of our methodology, the attributes come from external sources. For example, the case study presented in Section 5, calculated the demographics of media outlets in the United States. The source of the entities in that study was not Facebook, but rather a set of efforts that tried to enumerate the media outlets in the US, the entity under study. In other applications, the entity represented by the OTF may include a set of attributes that are known to be part of the Facebook platform. The demographic study of the audience reached by specific ads for instance (as presented in section 6), in which each OTF contains the exact ad attributes, certainly includes only elements provided by Facebook.

Independently of the source, the attributes that compose the OTF from which we intend to infer the demographics are commonly provided in a textual format, and that is not accepted by the Facebook Marketing API. The most direct way to convert the input attributes into their respective IDs is comparing the attribute provided with the name of the attribute as registered in the API. We then created a dictionary that maps the attribute name to its respective ID for all Facebook attributes we were able to collect. The collection of Facebook Marketing API attributes is described next.

All the behaviors, demographics as well as a basic set of interests (such as Pizza, Physical exercise and Yoga) are available to be selected in the user interface (figure 3.2 b)) and can be collected by a specific function provided by the API. Nevertheless, thousands of other interests, not shown in the selection boxes as options, may be selected by using an autocomplete approach, in which the user types some characters and the system shows a list of interests that include those characters.

In order to collect as many interests as possible, we design an extensive (though not exhaustive) strategy that explored two useful calls available in the API: i) given a piece of text, it provides possible interests that match the given text; and ii) given a user's interest, it provides a set of other related suggestions. For instance, the interests related to 'The New York Times' includes 'The Washington Post', 'The Wall Street Journal', and 'The Economist'. We execute a snowball sampling on users' interests following these interest suggestions.

As seed we use a set of news outlets names extracted from three different sources: Google News [Leskovec et al., 2009], List of Newspapers[7] and the top 1000 newspapers from Alexa.[8] Then, we identified around 3000 interests that exactly match the names of the external news outlet and we execute a snowball sampling on these users' interests, using Facebook's suggestions recursively starting from them. This process resulted in nearly 255K interests.

To estimate the fraction of all existing interests to which this sample corresponds, we measure the percentage of the entire graph that a snowball crawler is able to gather. We select a random list with 5,000 entities from DBpedia [9], a crowd-sourced community effort to extract structured information from Wikipedia, and crawl all interests related to those entities using the first useful API function mentioned above, that returns the interests on Facebook related to a specific text. We then verify that our snowball database missed only 71 out of the 13,996 interests collected, suggesting that our sample corresponds to nearly 99.5% of the total amount the existing user's interest. Thus, although using a biased seed we were able to collect the entire connected component.

Our resulting mapping dictionary includes $255,712$ interests, 1060 behavioral attributes and 77 demographic attributes. The mapping of an OTF attribute into an ID basically check if the source attribute perfectly matches their respective attribute name in the mapping dictionary. If the attribute is not found we try a different approach that compares the similarity[10] of the given attribute with all strings in the mapping dictionary. In case the OTF attribute is still not found in the dictionary we try to find it by seeking the Marketing API for its name. If none of the results match, the attribute is considered to be absent.

The reason for a missing attribute may be explained by two main factors. First of all, similar to all social networks, Facebook is highly dynamic and the topics of interest may eventually change. Apart from the dynamism of topics, the commercial factor may also impact the decision to remove unused or seldom used attributes.

[7] http://www.listofnewspapers.com/

[8] https://www.alexa.com/topsites/category/Top/News/Newspapers

[9] http://wiki.dbpedia.org/about

[10] This similarity is calculated by recursively finding contiguous matching subsequences that contain no junk elements. We considered in our methodology a very high similarity score (above 90%). This relaxation compared to the full matching of the string is important to avoid missing attributes due to minimum differences between the strings. 'The Washington Post' interest for instance that refers to the homonyms newspaper is found without the article 'The' in some cases. By using the relaxation we can match The Washington Post and some other particular cases.

3.6 Data limitations

First of all, it is important to point out that, despite the several privacy issues that surround the OSN advertising platform as described previously in Chapter 2, our methodology does not represent a threat to users' privacy. One may argue that our methodology may lead to the leakage of personal information about users, however, our work uses only aggregate information. It is only a number that represents the quantity of users that match each particular set of attributes. In addition to this, there is no need to run any ad in our approach, meaning that all gathered data is returned before any cost is incurred. Therefore, we do not collect any personally identifiable information. Among the millions of requests we performed to the Marketing API, none of them were able to gather and link any personal information to any particular user.

Though the Facebook advertising platform can be explored to infer demographics from the offline world, the mechanisms behind the tool are not publicly known, which is a limitation of our method. As a black box, make it difficult for the researchers to check if and to what extent the data is reliable. Furthermore, the population of Facebook is known to be biased towards gender, age, and other aspects. Therefore, it is important that conducted studies relying on this methodology validate their data. On the other hand, these issues open new research avenues on the statistics and demographics that might apply their artifacts to deal with noise and imperfect data in order to improve the confidence of the data.

The lack of control in the attributes set is another limitation of this research. As stated before, missing attributes may occur because Facebook has no interest in creating or keeping it. This situation may produce some inconsistencies in the final dataset, since a very popular entity, such as a largely known newspaper may have no interest related to it, whereas a low popular media outlet has a related attribute. Furthermore, some attributes are summarily discontinued without explanation or previous warning, which means that studies aiming to explore the evolution of demographic audiences for certain entities or studies that evaluate the audience of ads ran in the past may eventually face lack of data.

Chapter 4

Case Study: Inferring Census from online data

Censuses have been used for many centuries to estimate demographics about the actual population of countries. They are necessary and of utmost importance for the orderly functioning of modern societies. Censuses are crucial to defining priority investments for education, infrastructure and other public policies of a country. Despite its importance, the cost and time consumed to obtain these data are quite high. A recent report published by the US Census Bureau agency estimates the cost for the 2020 decennial Census in 15 billion dollars [1].

As also pointed by a United Nations report [2] alternatives to the traditional Census have been tested by different countries. In Norway, for instance, authorities conducted the Census with a register-based approach, which uses information from an existing administrative source and gather information about households, dwellings and individuals to complement data about the population demographics. This technique depends on a unique identification number across different administrative sources and may not be used in many countries due to legal restrictions of using these data with statistics purposes. An alternative, tested by Spain mix the register-based with the traditional approach. France has tested an approach that relies upon collecting data in a cumulative survey that covers the country for years instead of a short period. In addition to this, researchers have proposed alternative/complementary approaches to infer demographic aspects from different sources.

In this case study, we aim at reproducing some basic Census information with our framework. In particular, we conduct an in-depth analysis of seven demographic categories collected through advert platform and compare them to estimates from Census data to eval-

[1] https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/2020-cost-estimate1.pdf

[2] https://unstats.un.org/UNSD/demographic/sources/census/alternativeCensusDesigns.htm

uate to what extent the data gathered online approximate Census figures. The study provides an overall evaluation of different demographic attributes and the relationship between online and offline data. More specifically, our analysis takes into account the following demographic categories: gender, race, age, income, education, political leaning, and country of previous residence.

4.1 Methodology

In order to compare the Facebook Census with real Census, we turn to the US official authority in this topic. The United States Census Bureau provides two annual reports in addition to the decennial Census. The "American Community Survey" (ACS) and the "Current Population Survey" (CPS) are official surveys, curated by the official US agency and have some significant differences in their methodologies[3]. ACS deals with a small number of indicators such as major income sources, however, the ACS data collection use a self-response mail questionnaire with an internet response option and with mandatory response, similar to the decennial census form. Conversely, the CPS provides much more detailed data including more comprehensive coverage of all potential income sources, but the data collection is conducted by interviewers via Computer Assisted Telephone Interviewing and the participation is not mandatory. In order to provide more reliable data we used the 2013-2017 ACS 5-year Estimates (ACS 2017)[4], released on December 8, 2018[5].

More specifically, we used the following ACS tables to obtain the official Census demographic data: S0101(age and gender), DP05(race), S2001(income), and S1501(education attainment), B05006 (immigrants). For the political leaning attribute we used a Gallup study based on party affiliation by state[6] as the baseline, since Census do not include this attribute in their reports.

All data obtained from the original Census were collected in three granularity levels: country, state, and city level. We collected the demographic distribution for the 50 most populated cities in the US to provide a comparison in a more fine-grained level. In order to compare Facebook Census with real Census, we calculated the Pearson correlation to check the linear correlation between each one of the demographic dimensions.

A critical challenge in this analysis relies on differences in the fields nomenclature. For instance, relationship status includes many more options in Facebook Social Network when compared with the Census, such as 'engaged' and 'in a domestic partnership'. For education

[3]https://www.census.gov/topics/income-poverty/poverty/guidance/data-sources/acs-vs-cps.html
[4]https://factfinder.census.gov
[5]https://www.census.gov/programs-surveys/acs/news/data-releases/2017/release.html
[6]https://news.gallup.com/poll/226643/2017-party-affiliation-state.aspx

Category	Census	Facebook
Incomplete High School	Less than high school graduate(18-24), Less than 9th grade(above 25), 9th to 12th grade, no diploma(above 25)	In high school,Some high school
High School	High school graduate (includes equivalency) (18-24),High school graduate (includes equivalency)(above 25)	High school grad
Some College	Some college, no degree(above 25)	In college, Some college
College	Associate's degree(above 25), Bachelor's degree(above 25), Some college or associate's degree(18-24), Bachelor's degree or higher(18-24)	College grad
Grad Degree	Graduate or professional degree(above 25)	Some grad school, Master degree, Doctorate degree, Professional degree, Studying grad school

Table 4.1: **Educational attainment mapping.**

attainment, in particular, we need to group different categories from Census data, since they provide separate categories for people between 18 and 24 years old and above 25[7]. Finally, age is limited on Facebook since the platform only allows users above 13 years old. Table 4.1 details the education attainment fields of Census and Facebook used to compose the total audience in each category.

Another issue in the Census reproduction concerns small sized targeted populations. For subpopulations smaller than one thousand users, the Facebook advertising platform returns the value 1000. This is a mechanism to prevent advertisers to succeed in unveiling the identity of a certain user by creating a target formula that leads to a single specific user. As we focused on the most populous cities, this limitation represented no problem in our study. However, this privacy protection mechanism could represent a limitation for obtaining demographic data from Facebook in small cities.

4.2 Analysis

In this section, we aim at comparing the demographic distributions collected with our framework with consolidated offline results. For most of the validation in this study, we used recent baselines provided by Census Bureau estimation studies.

[7] https://factfinder.census.gov/bkmk/table/1.0/en/ACS/17_5YR/S1501/0100000US

The analysis is presented in three steps. Firstly, we characterize the distribution of selected demographic attributes in the US as a whole. In a second analysis, we dig into states and cities to check the demographic distribution of the Facebook population with a fine-grained perspective. Finally, we present a report about immigrants in the US.

4.2.1 Country-level analysis

The population of the US according to the ACS 2017 survey is roughly 321 million people, whereas Facebook registers 230 million active users who live in the US (July 2018). It indicates that nearly 71% of the American population have a public profile on Facebook and/or Instagram. This percentage is in consonance with a recent Pew Research report that pointed out two-thirds of US adults use social networks [8]. Figure 4.1 details the population size by different age intervals. Not surprisingly, the Facebook population for people under 19 and above 65 are significantly lower than the real US population calculated by Census analysis. This may be explained because the younger group does not include people under 13 since Facebook does not allow children to register. In spite of increasing their participation in Social Networks in the last years, people above 65 years old are, in general, less inclined to use OSNs than young people as also highlighted by the same Pew Research report.

In opposition to these underrepresented groups, Facebook overestimates the population with ages between 20 to 39 years old in comparison with the Census. This large population of adults raised some criticism about the way Facebook calculates its audience size, and some suggested that Facebook might be inflating the numbers in order to increase their revenue [9] [10]. Facebook alleged in a statement that "Reach estimations are based on a number of factors, including Facebook user behaviors, user demographics, location data from devices, and other factors. They are designed to estimate how many people in a given area are eligible to see an ad a business might run. They are not designed to match population or census estimates. We are always working to improve our estimates".

Figure 4.2 depicts the distribution of age by gender in a pyramid bar chart. We grouped all the intervals with persons older than 65 in the above 65 bar since Facebook does not allow stratifying users above 65 years old more accurately. Additionally, the Facebook bar chart does not contain information for the population under 15, since Facebook does not allow users younger than 13. As shown in the previous age distribution figure, the more represented interval on Facebook ranges from 20 to 39 years old. Curiously, the number of male and female users are exactly the same for the most populous three ranges: 20 to

[8] www.pewinternet.org/2018/03/01/social-media-use-in-2018/

[9] www.dataiq.co.uk/article/news-analysis-facebook-v-census-out-count

[10] www.businessinsider.com/facebook-tells-advertisers-reaches-25-million-more-people-than-exist-us-census-data-2017-9

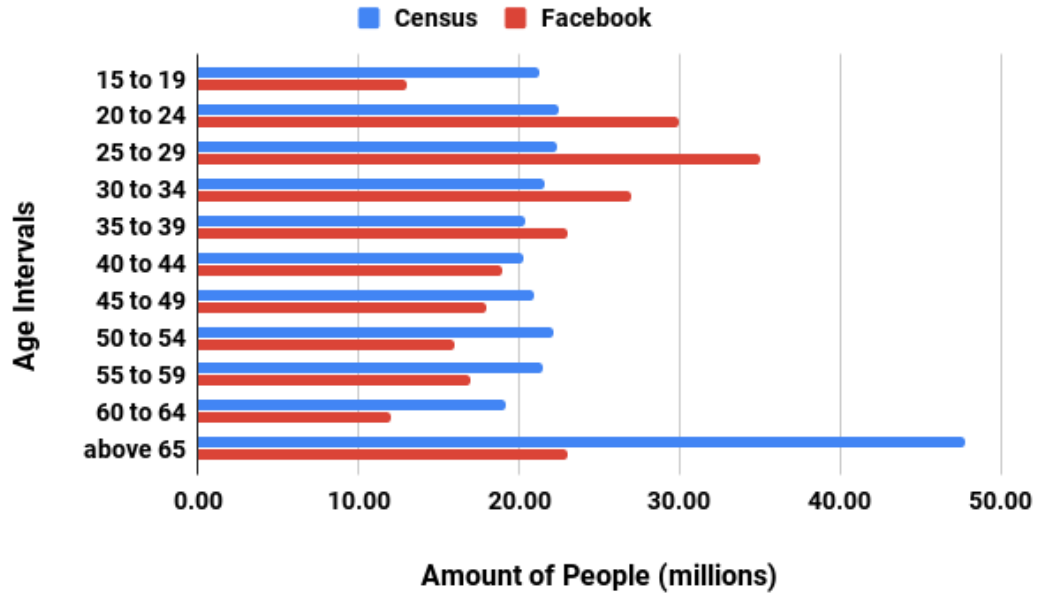
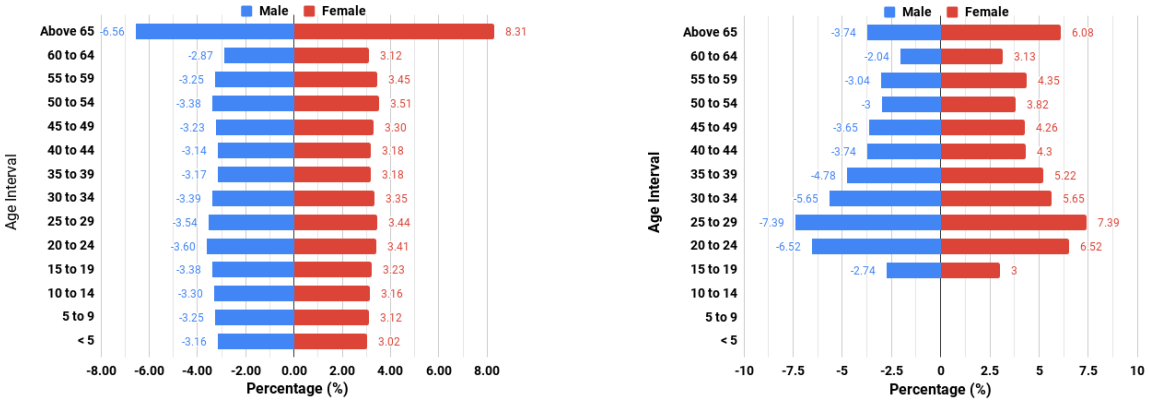


Figure 4.1: **Population grouped by age**



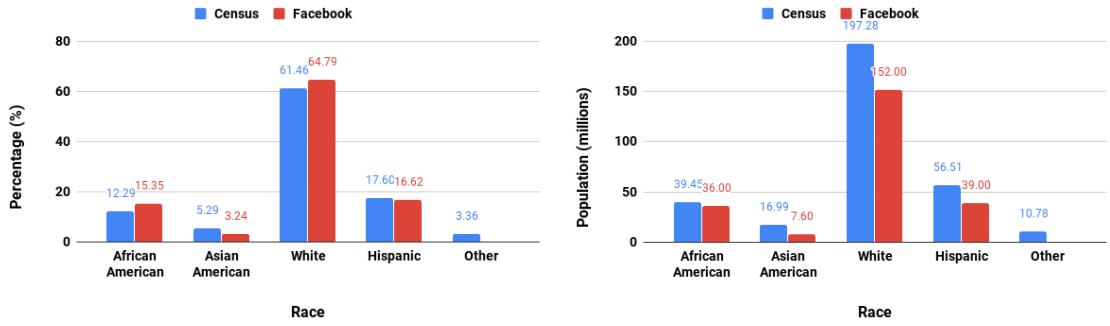
(a) Census (b) Facebook

Figure 4.2: **Age distribution by gender.**

24 years old (6.52%), 25 to 29 (7.39%) and 30 to 34 years old (5.65%). The figure also shows that women are overrepresented in the older intervals, especially, above 65 years old, in which the number of women is 62% superior to the male presence. This difference is only 24% in the census distribution.

The overall gender distribution of Facebook is slightly biased towards women. While men comprise 49.2% of the United States population and women account for 50.8% in the ACS survey, the women population on Facebook is 52.8%.

Figure 4.3 compares the distribution of the US population in terms of race. Facebook Marketing API include called an attribute called racial affinities for which they identify the



(a) Percent values

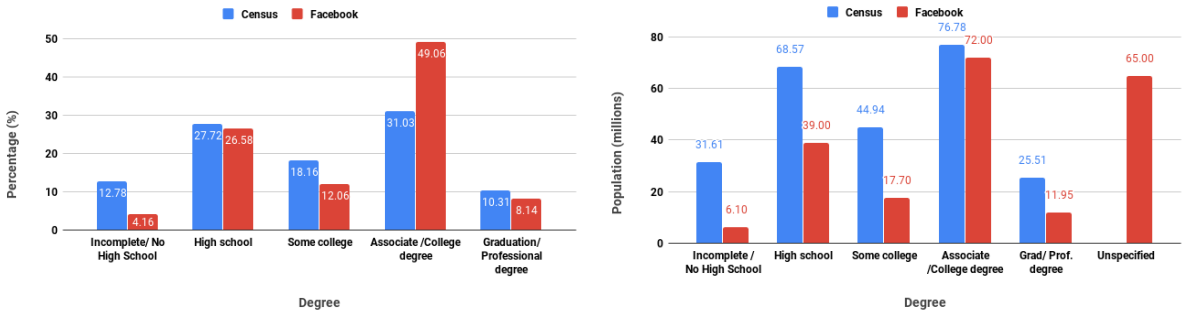
(b) Raw values

Figure 4.3: **Race distribution in the US**

affinity of their users with specific races: Hispanic, African-American and Asian-American. We then counted the audience size for each one of these categories and we considered the remainder as white. We can notice that the distribution of races across Facebook is quite similar to the Census distribution, being slightly overrepresented by African-American and white, and underrepresented by Asian-American and Hispanic (see Figure 4.3 (a)). When analyzing raw values, depicted in Figure 4.3 (b) we can check that the overpopulation found in the age intervals category is not verified in the race distribution, at least not directly. The African-American population on Facebook is only 3 million less than the African-American population in the Census. Considering that Facebook population includes only users above 13 years old, the 36 million population of African-American may lack at least some millions of African-American under 13 years old, which would also characterize an overpopulation of this particular ethnicity.

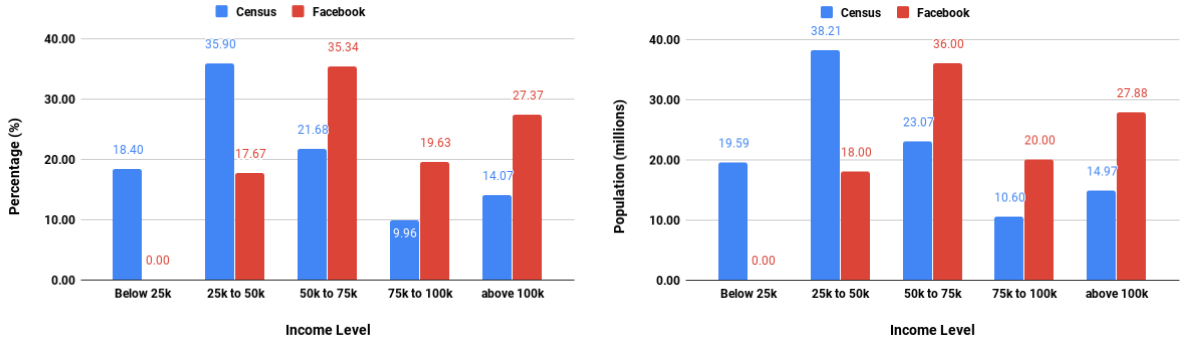
An important challenge when considering data produced by OSN users is that there is no guarantee the information is correct. In many cases, users insert information in their profiles to mock some situation or subject and sometimes they include some information to avoid leaving the field blank. Creation of fictitious job titles or colleges may be found with relative frequency. Another situation occurs when the users do not fill out their public profile due to privacy concerns or simply do not wish to spend their time doing this. The education level field, for instance, is not filled out by 65 million users as can be seen in figure 4.4 (b). This figure depicts the educational attainment in the US. Note that the number of people with the associate or college degree on Facebook also overpasses the amount informed by the census authority. The percentages are depicted in figure 4.4 (a).

In terms of income level, data obtained from Facebook partially differs from Census. Firstly, Facebook only infers the income with values above 30 thousand dollars a year. Another observation is that the Facebook population is much richer than the real population with an overestimation of the number of people who earn more than 50 thousand dollars.



(a) Percent values (b) Raw values

Figure 4.4: **Education level distribution in the US**



(a) Percent values (b) Raw values

Figure 4.5: **Income level distribution in the US.**

The income level is provided by one of the Facebook partners that help the OSN to provide more detailed targeting options to advertisers, especially regarding the purchase and offline behavior. However, data provided by some of these partners is no longer available since October 2018[11]. It is not clear how Facebook and partners classify the users by the amount of money they earn, but the bias toward the richer, again, may raise some criticism on Facebook, since it would inflate the audience most targeted by advertisers. As the baseline for income level, we considered full-time, year-round workers with earnings in the Census table named ACS_17_5YR_S2001[12].

4.2.2 Finer granularity - states and cities

Next analysis aims at checking if the demographic data obtained with Facebook Marketing API captures the variation across different locations. Firstly, we compared the total population in each one of the 50 US states and the federal capital according to FMA and Census,

[11] https://newsroom.fb.com/news/h/shutting-down-partner-categories/

[12] https://factfinder.census.gov/bkmk/table/1.0/en/ACS/17$_5$YR/S2001/0100000US

Figure 4.6: **Population by state**

and we find a very high Pearson correlation (0.9988) depicted in figure 4.6. The District of Columbia has the highest proportion of the population on Facebook compared to the Census population with a higher population on Facebook than in the real world, one million on Facebook compared to less than 700 thousand official number. This may be due to border characteristics of the US capital that lead to a misleading inference of location from Facebook. Apart from the US capital, the states with the higher proportion of population on Facebook are New York with 76%, and Alaska and Texas with 75%, whereas the states less represented online are New Mexico with 64% and South Dakota with 65%.

In a second analysis, we compared the fluctuation of race, income level, political leaning, and educational attainment across all the 50 US states and capital again. Table 4.2 (state level) synthesizes how correlated the data collected with our framework is when compared with data from ACS for the states by calculating Pearson correlation with 95% confidence interval (CI). Notice that, in terms of race, the correlation is very high for African-Americans, Asian-Americans, and Hispanics, which means that Facebook accurately infers the origins of a user. Recall that Facebook does not classify white people and for the calculation of this attribute, we excluded the other three races, which may explain, in part, the lower accuracy in this case.

Figure 4.7 plots the correlation for the white population across states. By analyzing the most dissonant points we figured out that they include Hawaii and Alaska, both states with particular ethnicities calculated by Census and left unconsidered on Facebook, Native Hawaiian, and Alaska Natives. For Hawaii, by using our framework we found almost 70% of white people whereas there are less than 23% according to Census data. The difference for Alaska calculus is about 21% (83% with our framework rather than 62% with Census).

Category	Dimension	State Level		City Level	
		Pearson C.	CI (95%)	Pearson C.	CI (95%)
Race	African-American	0.97	[0.95,0.98]	0.94	[0.90,0.97]
	Asian-American	0.97	[0.95,0.98]	0.94	[0.89,0.96]
	Hispanic	0.97	[0.95,0.98]	0.96	[0.94,0.98]
	White	0.82	[0.71,0.90]	0.86	[0.77,0.92]
Income Level	25k to 50k	0.76	[0.62,0.86]	0.69	[0.51,0.81]
	50k to 75k	-0.33	[-0.55,0.06]	-0.13	[-0.39,0.16]
	50k to 75k (*)	0.83	[0.72,0.90]	0.73	[0.56,0.84]
	75k to 100k	0.67	[0.49,0.80]	0.55	[0.31,0.72]
	above 100k	0.93	[0.88,0.96]	0.83	[0.72,0.90]
Educational Attainment	Incomplete High School	0.34	[0.08,0.57]	0.36	[0.09,0.58]
	High School	0.87	[0.77,0.92]	0.71	[0.54,0.83]
	Some College	0.55	[0.32,0.71]	0.51	[0.27,0.69]
	College	0.62	[0.41,0.76]	0.57	[0.35,0.73]
	Grad School	0.98	[0.97,0.99]	0.86	[0.77,0.92]
Political Leaning	Left leaning	0.87	[0.79,0.93]	-	-
	Moderate	0.02	[-0.26,0.29]	-	-
	Right leaning	0.91	[0.85,0.95]	-	-

Table 4.2: **Correlations for demographic categories across US States and Cities.**

Alabama is the third state with the highest difference, less than 50% on Facebook against 66% in Census data.

In terms of income level, the best correlation is for people with high earnings (above 100k dollars per year). For other levels, we find a poor correlation, except for an interesting finding in this particular attribute - the percentage of Facebook users with annual income level between 50k and 75k dollars are highly correlated with 25k to 50k distribution across states according to census(see 50k to 75k (*) in table 4.2 and figure 4.8), whilst the correlation with the same income level is negative.

By checking the education attainment rows in table 4.2 we find a similar result, *i.e.,* only two attributes presented a high correlation, including people with high school degree and with a graduation degree, especially the most educated people. Figure 4.9 depicts the high correlation for the grad school education level. This suggests that when a Facebook user fills out its education level with some grad school it is correct, in opposition to college grad, a case where apparently users tend to fill out with wrong information.

The last attribute checked across all states was political leaning, for which we found a high correlation for left leaning and right leaning and a poor correlation for moderate. The moderate lower correlation may be explained by the baseline we used, that is based on annual state averages of party affiliation from Gallup Daily tracking. This dataset is not ideal to detect the proportion of moderates in each state.

In order to provide a comparison at a more fine-grained level, we conducted similar comparisons for the 50 most populous cities in the US with results presented in table 4.2 (city level). It also presents the Pearson correlation as well as the values with 95% confidence interval for each category, except for political leaning. The last one was not included due

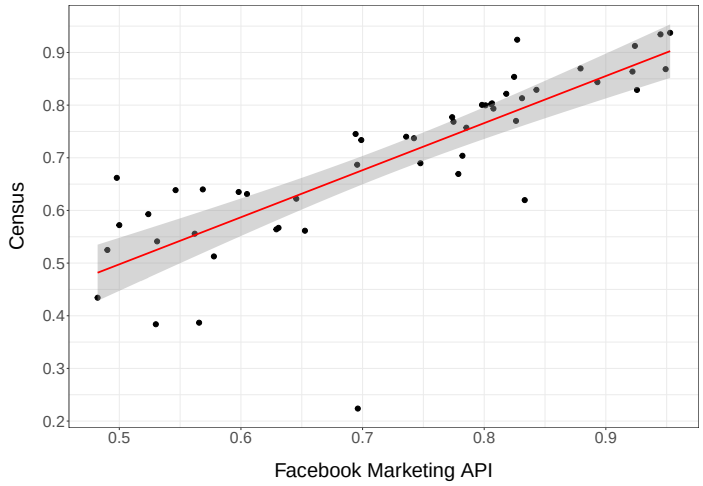


Figure 4.7: **Facebook Marketing API x Census across US states - Race - White.**

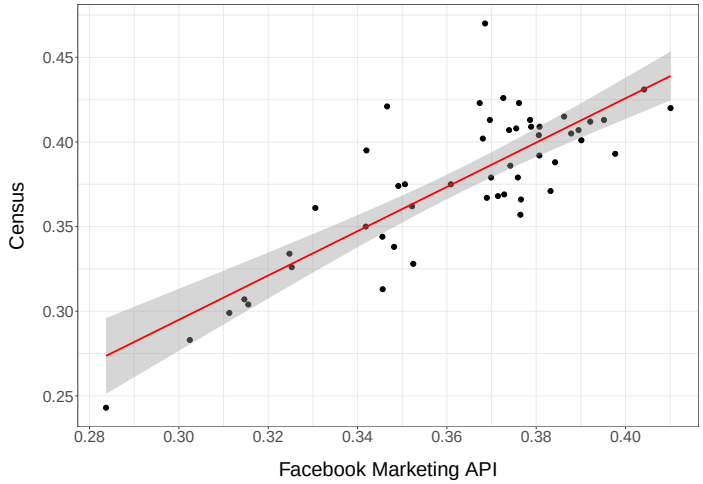


Figure 4.8: **Facebook Marketing API x Census across US states - Income - 50k to 75k(*).**

to the lack of available information on that fine-granularity level. We figured out that the correlation is often a little lower than at the state level. However, the inference of race is still high when considering the finer granularity.

One of the factors that may explain the lower correlation in comparison with the state level analysis is due to the Facebook collection. When selecting the city on Facebook advertising platform, we must define the name of the city and the radius of the collection that limits the population included in the target audience. The default radius is 30 miles and the lowest radius available is 10 miles. In our collection, we used the 10-mile radius option, which does not match the official borders of the city, meaning that the calculated demographics may include users from neighbors regions or exclude users that were supposed to be included in the audience. The census population of Arlington, in Texas, next to Fort Worth (a large city with

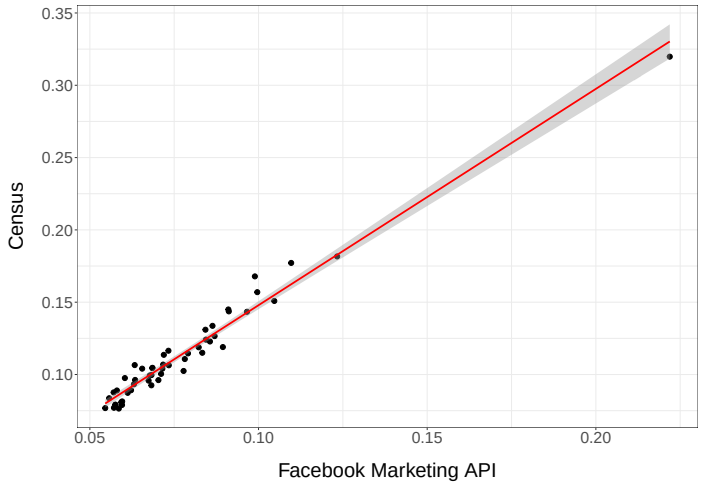


Figure 4.9: **Facebook Marketing API x Census across US states - Education Level - Grad School.**

874 thousand inhabitants) is roughly 390 thousand people (ACS 2017) whereas the population on Facebook is 1 million. The same issue is verified for Minneapolis (neighbor to the large city of St. Paul, state capital) with a 411 thousand population according to census and 1.1 million as counted by Facebook. In both cases, the final audience includes people from outside the city borders. Conversely, for the New York City, the population size is similar in both measurements, 8.5 million people.

4.2.3 US immigrants analysis

In this analysis, we compared the population size of immigrants in the US. We used the table B05006 from the ACS 5 Year Estimates as the baseline. For the Facebook data, we collected the number of immigrants for all available countries on the platform.

Figure 4.10 depicts the number of immigrants living in the US with origins in different regions around the World. Notice that the population size on Facebook is inferior to Census data in all regions except in Central America, in which Facebook population is nearly 550 thousand larger than in the Census. There are different gaps between both measurements in the other regions. For immigrants from South and East Asia, for instance, the census population size is roughly 4.8 million larger than Facebook population. This may be explained by the banishment of Facebook from China, meaning that the largest OSN is not a good site to get in touch with compatriots that still live in their origin country. On the other hand, the gap in South America is small, with 2.6 million immigrants according to Facebook and 2.9 million on Census.

Figure 4.11 allows us to check the difference in the country level for the top 25 coun-

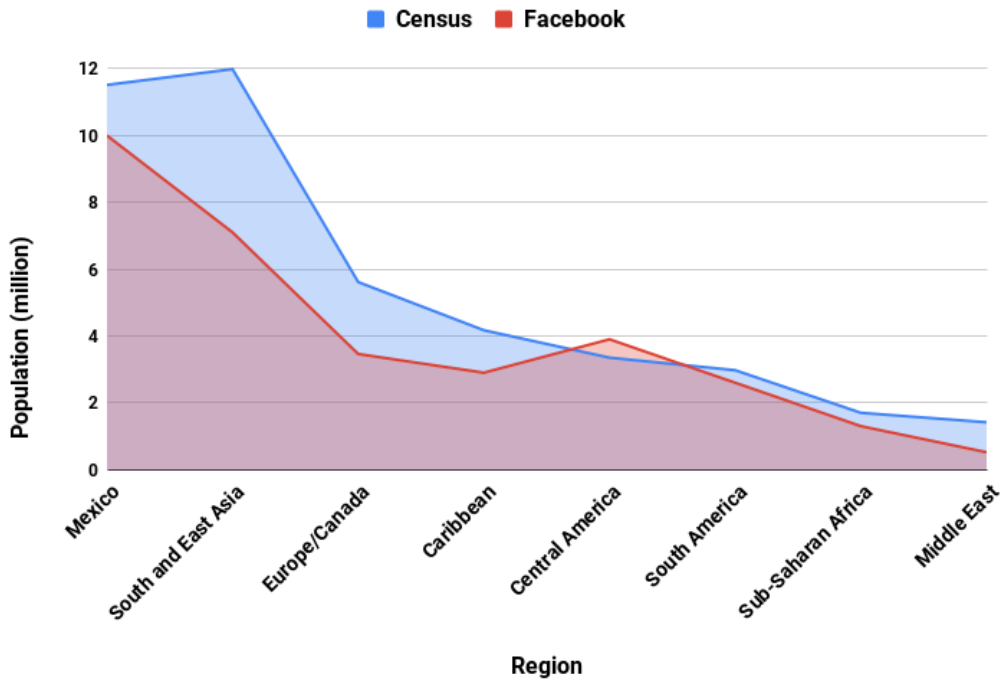


Figure 4.10: **Population of immigrants by region.**

tries with more immigrants in the US. We can notice, for instance, that the gap is huge for Chinese Facebook users, 2.64 million according to census and only 0.66 million on Facebook. On the other hand, Central and South America countries such as Guatemala, Honduras, Brazil, and Venezuela have more immigrants on Facebook than those calculated by the census. In other examples for the same origin region, the numbers are very similar in both measurements, for example, El Salvador, Dominican Republic, and Peru. This finding might indicate that Census is underestimating the population of immigrants with origins in countries with a high incidence of unauthorized people. For the sake of simplicity, Mexico was included in the regions' figure and not in the top 25 countries due to the high number of immigrants from this country.

We should mention that we were not able to count the immigrants from some particular countries on Facebook and they were excluded from our analysis. Those countries excluded from the top 25 list are Iran, Pakistan, Ukraine, and Ecuador. For the regions figure, we were not able to include a considerable number of countries due to the absence of information about these countries on Facebook. The percentage of missing countries per region are the following: South and East Asia (48%), Europe (31%), Caribbean (83%), Central America (43%), South America (50%), Middle East (60%) and Sub-Saharan Africa (69%).

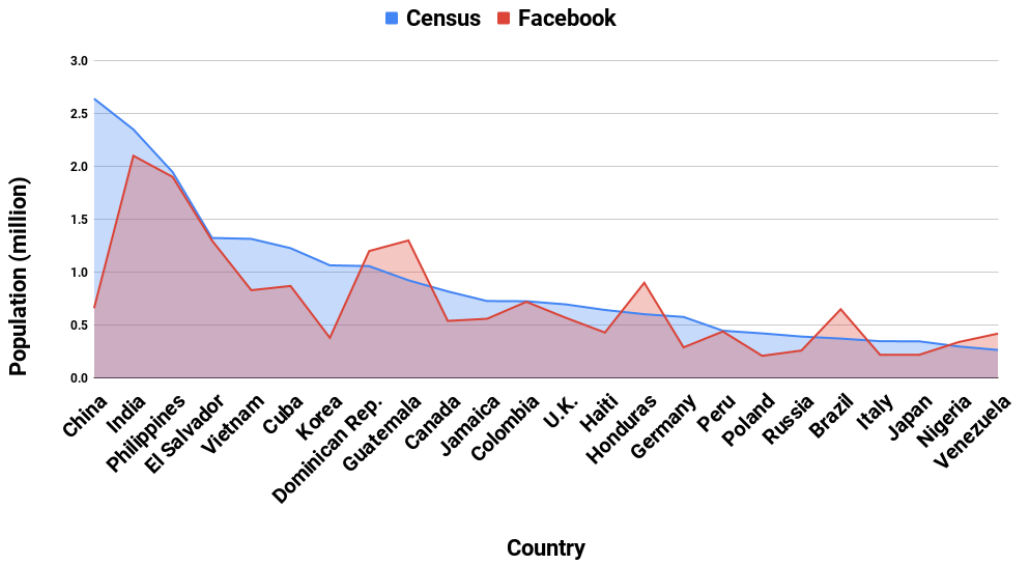


Figure 4.11: **Population of immigrants by country.**

4.3 Correction factors

As another contribution of this case study, we plan to release a dataset containing all demographic distributions collected from Facebook. In addition to the Census and Facebook distribution for each demographic attribute, we also computed a correction factor that allows one to multiply it by the Facebook distribution to obtain the Census distribution as a result.

The correction factors, computed for each demographic dimension and for all levels (country, state and city level) can be very useful for demographic research. One particular use is deriving the real population for some distribution of interest calculated previously through the Facebook advertising platform. Suppose someone wants to identify how many people are interested in an activity, brand or any other entity in a particular geographic region, stratified by gender. She can collect the distribution in the Facebook advertising platform (by manually selecting the audiences on the ad creator graphic interface) and derive the population interested in that entity after multiplying the numbers by the correction factor. Recall that the audience estimation does not require the publication an ad and do not incur at any expense. Facebook provides more than 250 thousand attributes [Speicher et al., 2018] that can be used to select a huge range of audiences that can be further extrapolated to the real world.

In addition to the statistic, there is also the sociological value of the corrections factors, that allows accounting for groups that are over and underrepresented in the online world. It is widely known that certain groups are more or less represented on Facebook, and by using

US State	% Facebook	% Census	CF
West Virginia	14.061	3.507	0.24939
Montana	1.256	0.396	0.31546
Hawaii	4.216	1.687	0.40007
District of Columbia	46.829	46.871	1.0009
Massachusetts	6.598	6.682	1.01279
South Dakota	1.495	1.671	1.11739

Table 4.3: **Correction factor for African-American dimension (most biased states).**

the correction factors, we can quantify this bias. Table 4.3 shows the percentage of African-American measured by Facebook and Census as well as the correction factor (CF) in each case for six US states. Notice that, the Census value can be obtained by multiplying the CF by the Facebook value, indicating that the lower is the CF the most overrepresented is the Facebook percentage in comparison with Census percentage. The top three rows show the three most overrepresented states on Facebook in this demographic dimension whereas the bottom rows present the states underrepresented by African-American in comparison with Census. For this demographic dimension, in particular, the Facebook percentage is overrepresented in 48 out of 51 states.

4.4 Summary

In this case study, we leveraged our Framework to derive the Facebook Census, containing the demographic distribution for seven different attributes: gender, race, age, income, education level, political leaning, and country of previous residence. We calculate the demographic distributions in different granularity levels in the US: country, state, and city level.

We analyzed the Facebook Census by comparing it with official data provided by the US Census Bureau and Gallup consulting company. Among the main findings in our study, we confirmed the bias in the online population towards young people and women. We also verified that the distributions of race, in particular, are very similar to the real distribution in all granularity levels. The education level obtained online seems to be oversized for the college degree level, that is likely to be caused by wrong information provided by users. However, for high school and grad school degree, it provides similar information to the offline data at the state level. The same occurs for income level, in which the higher income level (above 100k per year) provides accurate data. We also checked that the Facebook Census data for educational attainment and income level in city level are not as good as data for the state level, probably due to the issues on identifying the city borders. In terms of immigration, the online data seems to follow the same tendency of official data, except

for immigrants from South America and Central America that outnumbered the numbers calculated by Census, which may indicate the underestimation of official sources, especially due to illegal immigration. Finally, Facebook Census also provides accurate distributions in state level for conservative and liberal people.

Our framework showed to be valuable and complementary to the original Census. The key advantage of our approach is that it could be computed periodically and at a very low cost. As a final contribution of this study, we also plan to release our Facebook census data along with the computed correction factors. We expect that our dataset can open many avenues of research and help researchers and practitioners to explore OSN Advertising platforms data with novel goals.

Chapter 5

Case Study: Inferring demographics of the audience of News Media Outlets

In this chapter, we apply our framework to bring more transparency to the media ecosystem, a key problem in our modern society. We derive the political bias of media outlets by calculating the political leaning of the audience for each one, similar to some related works that assume that the bias in the audience reflects the bias in the source. We then compare our results with different approaches and deploy a system to Internet users.

Recent years have witnessed a radical change in the way news are being produced and consumed. Online social media sites like Facebook and Twitter have emerged as popular media for users to receive, share, and discuss news about the world around them. A recent survey by Pew Research Center estimates that 62% of the adults in the U.S. consume news primarily from social media sites [Mitchell, 2016], and this number is still growing. Similar to the traditional news media, the news stories disseminated over social media can also have a considerable impact on shaping people's opinions and influencing their choices, having the potential to sway the outcomes of political elections [Allcott and Gentzkow, 2017]. Another recent study has shown that being exposed to news media makes Americans increase their participation in conversations about national politics and helped them standing publicly on specific issues [King et al., 2017].

A key characteristic of news on social media is that anyone can register as a news publisher without any upfront cost (*e.g.,* anyone can create a Facebook page claiming to be a newspaper or news media organization). Consequently, not only traditional news corporations are increasingly migrating to social media, but many social-media-only news outlets are emerging [Lella, 2016]. With this recent transition, not surprisingly, there are growing

concerns about 'fake' news publishers posting 'fake' news stories, and often disseminating them widely using 'fake' followers [Allcott and Gentzkow, 2017; Vosoughi et al., 2018; Lazer et al., 2018; Reis et al., 2019].

Even when the accounts used to publish or promote news stories are not 'fake bot' accounts (*i.e.,* they actually correspond to real persons or organizations), readers of news on social media are often not aware of the biases of these accounts. This situation is in sharp contrast to the news consumption over traditional news media channels, since the constant monitoring by media studies scholars and watchdog groups, ensures that at least well-informed consumers are aware of the biases of different news publishers.

For traditional media, two broad strategies have been used to quantify the biases of a given news outlet:

(i) The first strategy is to analyze the readership of the news outlets, which assumes that the content and attitudes of a news outlet end up driving the biases of its audience. Although this approach has been used by both researchers [Bakshy et al., 2015; Gentzkow and Shapiro, 2010; Zhou et al., 2011] and thinktanks like Pew Research [Mitchell et al., 2014], they often rely on readership surveys, and thus can not cover more than a few dozen mainstream news outlets.

(ii) The second class of approaches quantifies media bias directly, by inspecting the published content [Covert and Wasburn, 2007; Budak et al., 2016], specifically focusing on the coverage of important events by the media organizations. As there are significantly more news publishers on social media (with a constantly expanding list) than in the traditional media scenario, such strategies for measuring media bias do not scale for the current news ecosystem.

Thus, there is no mechanism available today for users to know the biases of the growing amount of publishers on social media.

In this chapter we used our framework to assess the biases of thousands of social media news outlets in the United States. We demonstrate the scalability of our approach by building and publicly deploying a system called **Media Bias Monitor**[1], which quantifies the ideological biases of **20,448** news outlets in Facebook. Media Bias Monitor also provides demographic information along five other dimensions: gender, income level, racial affinity, national identity, and age. We hope that our system can bring more transparency to the biases of news publishers on social media, not only to the most popular ones, but also to small, niche news outlets.

The rest of this chapter is organized as follows. The next section describes the strategy for gathering the media outlets to be used in the OTF, *i.e.,* the entity from which we want to

[1] `twitter-app.mpi-sws.org/media-bias-monitor`

Dimension	Attributes
Gender	Male, Female
Racial Affinity	African American, Asian American, Caucasian, Hispanic
Age	13-17, 18-24, 25-34, 35-44, 45-54, 55-64, above 65
National Identity	Australia, Africa, Canada, East Asia, Europe, Latin America, Mexico, Middle East, Russia, South Asia
Income Level	30-40K, 40-50K, 50-75K, 75-100K, 100-125K, 125-150K, 150-250K, 250-350K, 350-500K, >500K
Political Leaning	Very Conservative, Conservative, Moderate, Liberal, Very Liberal

Table 5.1: **Different demographic dimensions and attributes gathered from with our framework.**

derive demographics. We then compare our approach for inferring ideological bias with four state-of-the-art methods, with the aim of checking if our approach presents similar results. Next, we investigate other demographic aspects of news outlets with different ideological biases. We end with a brief discussion of our system design including examples of its application, before finally concluding the chapter with the discussion about potential future research directions.

5.1 Methodology

As explained previously in our Framework description, the OTF includes the attributes that represent an entity to be evaluated. In this case study particularly, we used the interest attributes available on the Facebook advertising platform to represent the the media outlets. Among the categories covered by the interests we can find newspapers, websites, TV shows, and many other topics related to media outlets. Those categories embrace popular media outlets such as 'The New York Times' and 'The Economist' as well as regional media sources *e.g.,* 'Pittsburgh City Paper' and 'East Bay Express'.

Before mapping the media outlets to their respective interests we executed an intermediary step that aimed at finding the Facebook page related to the media outlet, as described next.

5.1.1 Finding news outlets on Facebook

We start with a list of newspapers whose ideological biases we want to infer. To populate this list, we consider the news outlets used in the following prior efforts:

- 36 news outlets considered in the Pew Research survey on media habits [Mitchell et al., 2014].
- 15 news outlets from [Budak et al., 2016].
- 500 outlets used in [Bakshy et al., 2015].
- 112 news outlets analyzed by the media bias monitoring website `AllSides.com`.[2]

To identify the Facebook pages of these news outlets, we took the following approach:
(i) First, we crawled the news media websites and searched for references to their corresponding Facebook pages. If we found such a reference, we fetched the name and URL of the referred page and compared with the name and URL of the newspaper to validate the mapping between the Facebook page and the media outlet.
(ii) If we did not get a match in the first step, we searched for their domains (`nytimes.com`, `cnn.com`) using Facebook Graph API[3], and compared the name and URL in the returned Facebook page with the name and URL of the news media outlet.
(iii) If we did not succeed in establishing the mapping with the above steps, we searched for the news outlet's name using Facebook Graph API, and only included the pages where the names and URLs matched exactly.

In total, we were able to identify 32 news outlets (out of 36) from the Pew Research study [Mitchell et al., 2014], all 15 outlets from [Budak et al., 2016], 360 (out of 500) outlets from [Bakshy et al., 2015], and 81 (out of 112) from `AllSides.com`.

After identifying the Facebook pages for these media outlets, we used the page names to search for their corresponding 'interests' in our 250k base of interests. As the above process of matching newspaper names using different methods may result in errors, we conducted a manual validation for the mapping of news media sites to Facebook Pages, using a sample of 150 randomly selected outlets. We found the precision of the mapping to be 94.3% with 90% recall. Thus, we can conclude that using the steps described earlier, we could identify the Facebook pages belonging to different media outlets with high accuracy. Once the interests are identified we used our Framework to collect the political leaning distribution for the audience of the selected media outlet.

5.2 Comparison with previous work

In this section, we first describe how we quantified the ideological bias of different news outlets. Then, to verify whether our inference strategy properly captures the news media

[2] `allsides.com/media-bias/media-bias-ratings`
[3] `developers.facebook.com/docs/graph-api`

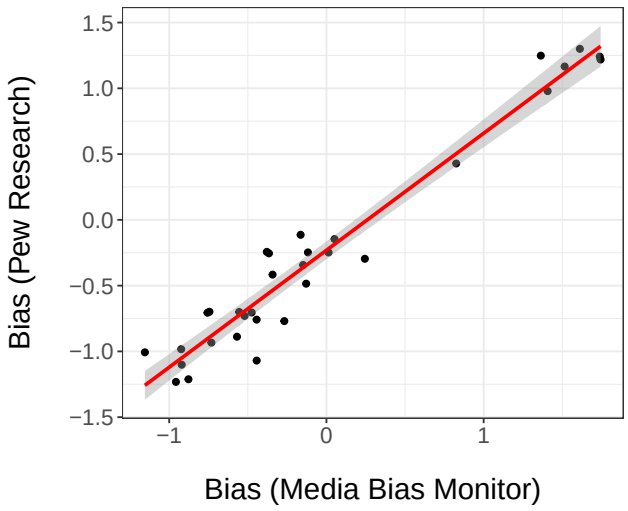


Figure 5.1: **Ideological leaning inferred by Media Bias Monitor in comparison with the bias inferred by the study from Pew Research [Mitchell et al., 2014].**

bias, we compare our results with four very different approaches to measure media bias.

5.2.1 Measuring bias using Facebook audience demographics

We explore the percentage of each one of the five categories of political leaning distribution to measure the ideological bias score of an outlet. Specifically, we multiply the fraction for each category with the following values: very liberal (-2), liberal (-1), moderate (0), conservative (1), and very conservative (2). The resultant sum gives the bias score which can vary from -2 to 2, where a negative number indicates a liberal audience while a positive number indicates a conservative audience for the media outlet. Accordingly, the media outlet is labeled as liberal leaning or conservative leaning. We utilized the above approach to quantify the bias of different media outlets and built the system 'Media Bias Monitor' to make these biases more transparent to social media users (details are presented in later sections). Next, we compare our approach with different approaches used to infer media bias.

5.2.2 Comparison with survey based approach

We begin by comparing our approach with a study conducted by the Pew Research Center [Mitchell et al., 2014]. Pew Research classified the audience of popular news media outlets based on a ten-question survey covering a range of issues like homosexuality, immigration, economic policy, and the role of government. In that study, the authors inferred the political leaning of the audience in a 5-point scale that are conceptually similar to those returned by Facebook Audience API – consistently liberal, mostly liberal, mixed, mostly

News Outlet	Source	V. Lib	Lib	Mod	Con	V. Con
NPR	Pew Research	0.41	0.26	0.21	0.09	0.03
	Facebook	0.34	0.29	0.19	0.10	0.07
BBC	Pew Research	0.32	0.28	0.26	0.08	0.05
	Facebook	0.24	0.33	0.22	0.13	0.08
NYTimes	Pew Research	0.4	0.25	0.23	0.09	0.03
	Facebook	0.30	0.28	0.21	0.13	0.09
CNN	Pew Research	0.19	0.25	0.4	0.12	0.04
	Facebook	0.23	0.27	0.22	0.15	0.12
Breitbart	Pew Research	0.03	0.04	0.14	0.31	0.48
	Facebook	0.01	0.02	0.07	0.22	0.67
Fox News	Pew Research	0.04	0.14	0.37	0.27	0.19
	Facebook	0.07	0.10	0.17	0.27	0.40

Table 5.2: **Pew Research results in comparison with our Facebook audience-based approach for measuring political leaning of different news media.**

conservative, and consistently conservative. In total, they evaluated 36 mainstream news media outlets, and we were able to gather the composition of their audience in Facebook for 32 of them.

To compare the bias inferred by Pew Research with ours, we compute the bias score from their data similar to how we compute the score for our method. For each category, we multiplied its fraction by its respective value in the scale ranging from -2 (consistently liberal) to 2 (consistently conservative). Figure 5.1 shows the scores obtained by our method for each news outlets along with the scores for the pew research study. Computing Pearson's correlation coefficient [Lee Rodgers and Nicewander, 1988] between the scores obtained by both methods, we found the correlation coefficient to be **0.97** (which is very high), with a 95% confidence interval of $[0.952, 0.986]$. This high correlation indicates that the results from both methods match almost perfectly.

Table 5.2 highlights the inferences from the two approaches for some popular news outlets. We can observe that both methods lead to the same conclusions about the political leaning of all these news outlets. Overall, the mean difference between the results of the two studies is 0.052 ± 0.016 for very liberal, 0.034 ± 0.012 for liberal, 0.070 ± 0.023 for moderate, 0.061 ± 0.022 for conservative, and 0.099 ± 0.034 for very conservative. We observe the highest divergence occurs for very conservative users, which can be explained by the possibility that number of conservative may have grown in the US since 2014 (when the Pew Research study was conducted). We also note that in, 26 out of the 32 media outlets, the number of moderate-leaning users decreased, which is also expected given the high polarization of the news discourse around the 2016 presidential election.

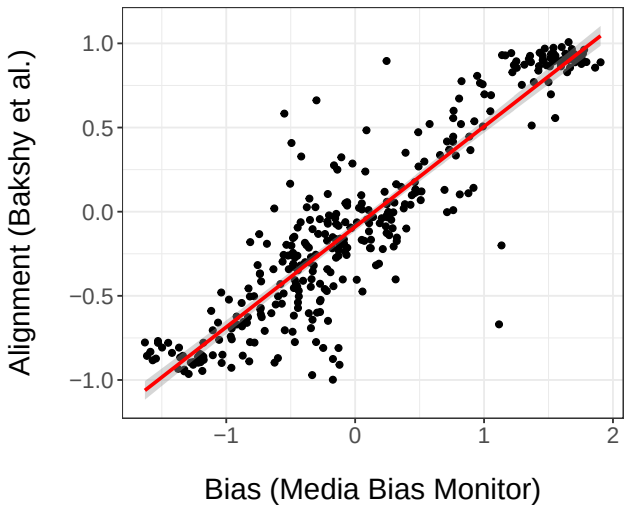


Figure 5.2: **Ideological leaning inferred by Media Bias Monitor in comparison with the bias inferred by [Bakshy et al., 2015].**

5.2.3 Comparison with the news sharing approach

In [Bakshy et al., 2015], the authors derived the alignment score of 500 media outlets by first identifying the political leaning of over 10 million Facebook users based on self-declarations, and then considering how users with different political leanings shared the stories published by these outlets. Similar to us, the authors measured the ideological leaning of the outlets on a scale ranging from -2 (Very Liberal) to $+2$ (Very Conservative). They identified the leaning of 500 news outlets, out of which we were able to find the Facebook pages (and thus identify the biases) for 342 outlets.

There are two reasons for not finding the remaining outlets in Facebook: (i) we found that the domains of some outlets considered in their study (e.g., `dcbeacon.com`, `scgnews.com` etc.) are no longer active and hence could not be reached; (ii) we could not find the Facebook pages or Interest IDs for the remaining outlets, without which we can not gather the composition of Facebook users interested in those outlets.

Figure 5.2 shows the scatter plot of the scores obtained by two methods, where each dot in the figure is a news outlet and the scores of each method can be seen on the axes. Overall, the Pearson correlation coefficient for the scores obtained by our method and the method proposed by [Bakshy et al., 2015] is **0.91**, with a 95% confidence interval of $[0.891, 0.927]$. Thus, we can note that despite the large number of news outlets considered, inferred ideological biases from both approaches are very close.

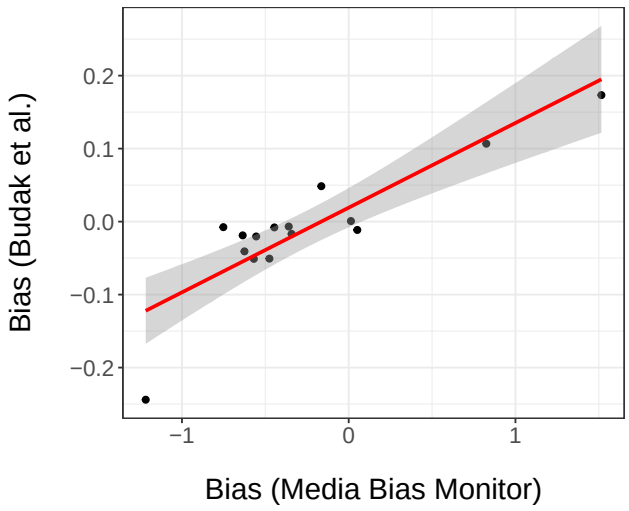


Figure 5.3: **Ideological leaning inferred by Media Bias Monitor in comparison with the bias inferred by [Budak et al., 2016].**

5.2.4 Comparison with the content based approach

Budak *et al.* [Budak et al., 2016] used a content-based approach to identify the slant of the top 13 U.S. news outlets and two popular political blogs. They sampled two political stories per day for each outlet, and used Amazon Mechanical Turk platform[4] to ask human judges if the article was positive, negative or neutral towards Democrats or Republicans. The answer was encoded in separate 5-point scale with the values $\{-1, -0.5, 0, 0.5, 1\}$ for Democrats, and $\{1, 0.5, 0, -.0.5, -1\}$ for Republicans. Therefore, a negative average score implies the article is positive toward Democrats, while a positive average score indicates Republican leaning. Finally, the slant for each news outlet is calculated as an average of individual news' leaning scores.

Figure 5.3 shows the scatter plot between the bias scores obtained by us and by [Budak et al., 2016]. Overall, the Pearson Correlation Coefficient between the scores obtained by these two methods is **0.87**, with a 95% confidence interval of $[0.650, 0.956]$. This implies that our approach inferred results similar to their content-based approach.

5.2.5 Comparison with the crowdsourcing approach

Finally, we compare our approach with a crowdsourcing-based method to infer media bias deployed at the website `AllSides.com`. It encourages its users to rate different news outlets in one of the five categories: left, lean left, center, lean right, and right[5], using any of three different strategies: (i) blind surveys, in which users rate the bias of stories without

[4] `mturk.amazon.com`
[5] `allsides.com/media-bias/media-bias-ratings`

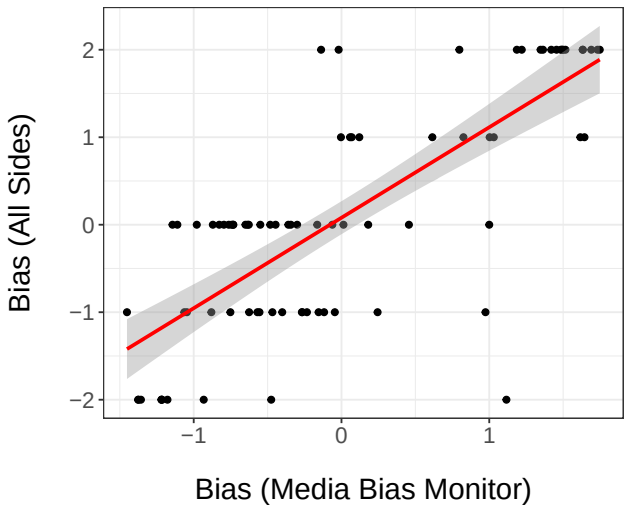


Figure 5.4: **Ideological leaning inferred by Media Bias Monitor in comparison with the bias inferred by `Allsides.com`.**

knowing the news source; (ii) showing them the bias of the source as inferred by previous research efforts (e.g., the work by [Groseclose and Milyo, 2005]), and (iii) showing them the past feedback from the other users. In (iii), a user can agree or disagree with the past ratings of the news outlets and can suggest new ones.

In total, `AllSides.com` presents bias of 112 media outlets, out of which we were able to identify the Facebook audiences for 81 outlets. Similar to the previous approaches, we defined a fixed bias score for each one of the five categories assigned by AllSides: left (-2), lean left(-1), center(0), lean right (1) and right (2). Figure 5.4 compares bias scores inferred by our approach vis-a-vis the scores from AllSides. The Pearson correlation coefficient obtained for the results of the two methods is **0.77** with a 95% confidence interval of $[0.658, 0.843]$.

Diving deeper into the mismatches, we observed that the main difference occurs when AllSides classifies a news outlet as center biased, whereas our approach assigns it a liberal bias score. Among those media outlets, we found BBC, CNN and Reuters, with the following liberal scores according to our approach: -0.521, -0.342, and -0.446 respectively. In order to verify the correctness of our score for these specific cases, we contrasted them with the results from [Bakshy et al., 2015] and from [Budak et al., 2016]. Both methods assigned liberal scores to these news outlets. Additionally, we verified that media outlets like Al Jazeera, FiveThirtyEight, and NPR have strong liberal bias according to the method by [Bakshy et al., 2015] as well as to our method, whereas AllSides marked them as moderate.

In summary, Table 5.3 presents the Pearson Correlation Coefficients (PCC) for the comparison between our approach with four existing state-of-the-art methods that use very

Method	Total	Identified	PCC	CI (95%)
Pew Research	32	36	0.97	[0.946,0.987]
Bakshy *et al.*	500	342	0.91	[0.891,0.927]
Budak *et al.*	15	15	0.87	[0.650, 0.956]
AllSides.com	112	81	0.77	[0.658, 0.843]

Table 5.3: **Summary of the comparison between our approach to infer ideological bias and four previous efforts.**

different inference methods. Our method closely matches most of these studies, thus, validating our methodology to obtain the ideological bias. In the next section, we discuss two key benefits of using our approach over the existing ones.

5.3 Media Bias Monitor

In the previous section, we showed that our approach to quantify media bias can produce inferences similar to four very different state-of-the-art methods. However, the key advantage of our method over the existing approaches is that our approach is highly scalable, and can infer the ideological bias of several thousands of news media outlets that exist today. As a show case, we built a system, named *Media Bias Monitor* [6], which makes the biases in audience demographics for **20,448** news outlets in Facebook transparent to users. The number of news outlets we cover is at least two orders of magnitude above any other existing efforts.

5.3.1 Scaling bias inference

In order to expand our list of news outlets and take the assessment of political bias of media sources in the US to a new level, we select, among $255k$ interests collected with the snowball approach, those whose categories are related to News and Media (*e.g.,* 'Newspaper', 'Media/News Company', 'News & Media Website', 'Journalist', 'Magazine', 'Broadcasting & Media Production Company', 'Website', 'Publisher' *etc.*). After filtering out other categories, our final dataset of news outlets contains **20, 448** Facebook pages and their corresponding Interest IDs. Then, we gathered the demographics of the audiences of all these outlets by following the procedure detailed in the Methodology section.

Table 5.4 shows the number of news outlets in each category, as well as a few examples that help us to understand what kind of news outlets are grouped into each of these categories. The most popular category is TV show, which contains TV news programs such as NBC's Today Show. We note that there are also TV shows from outside the US, but they have a large following among US users. The second most popular category corresponds to external

[6] `twitter-app.mpi-sws.org/media-bias-monitor`

Category	(#)	Example News outlets
Magazine	2,565	In Touch Weekly, Country Living, UNILAD
Newspaper	1,099	The Washington Post, The Daily Caller
Journalist	750	Bill O'Reilly, Lester Holt, Megyn Kelly
News Company	1,346	BuzzFeed Food, Conservative daily
Website	3,687	Topix, GroupMe, Delish
Other	900	BuzzFeed, Yahoo! News
Radio Station	992	2Day FM, Radio One Lebanon, Radio Disney
TV Show	4,447	NBC today show, The voice
Sports Team	2,615	Dallas Cowboys, Pittsburgh Steelers
TV Channel	2,047	ABC, CBS Sports

Table 5.4: **Number of news outlets in different categories covered by Media Bias Monitor.**

websites, followed by sports teams and magazines. We also observe smaller yet considerable fractions of radio stations, newspapers, and individual journalists. Interestingly, although the number of individual journalists is small in comparison to other categories, some journalists have quite large audiences. For example, Bill O'Reilly, with an audience of 2.2 million users is at the top of the list, followed by Lester Holt (1.8M) and Megyn Kelly (1.3M).

5.3.1.1 Importance of Measuring Bias at Scale

Figure 5.5 shows the audience size of all news outlets gathered using the above steps. We can observe from Figure 5.5 that although a small number of the most popular news outlets on Facebook reach a large number of news readers, while a large number of news outlets cater to small niche audiences, which taken together, account for a non-negligible fraction of the overall news audience.

Interestingly, we find that the news outlets with smaller audiences are also those that are most ideologically biased. For example, among the 10-percent most biased (i.e., either most conservative or most liberal) news outlets, 58% outlets have audience size less than $10,000$ users, whereas, among the 10-percent least biased outlets, only 31% outlets have less than $10,000$ audience. This suggest that the most biased news outlets are usually those that reach niche and smaller audiences, thereby highlighting the importance of monitoring the news published by these outlets and not only those published by the mainstream news publishers.

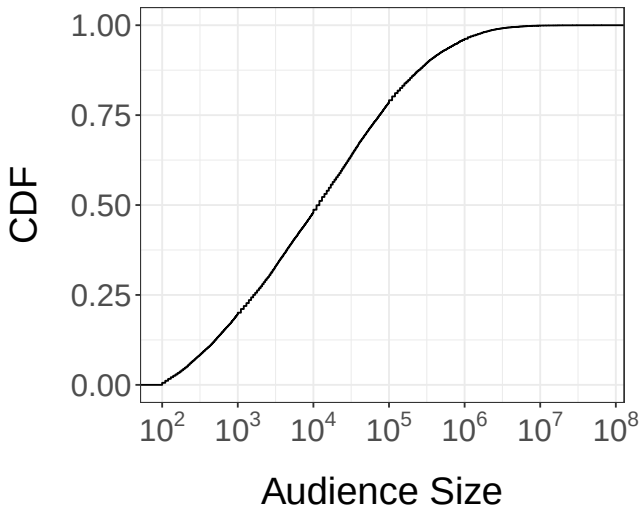


Figure 5.5: **Distribution of the audience size of news outlets.**

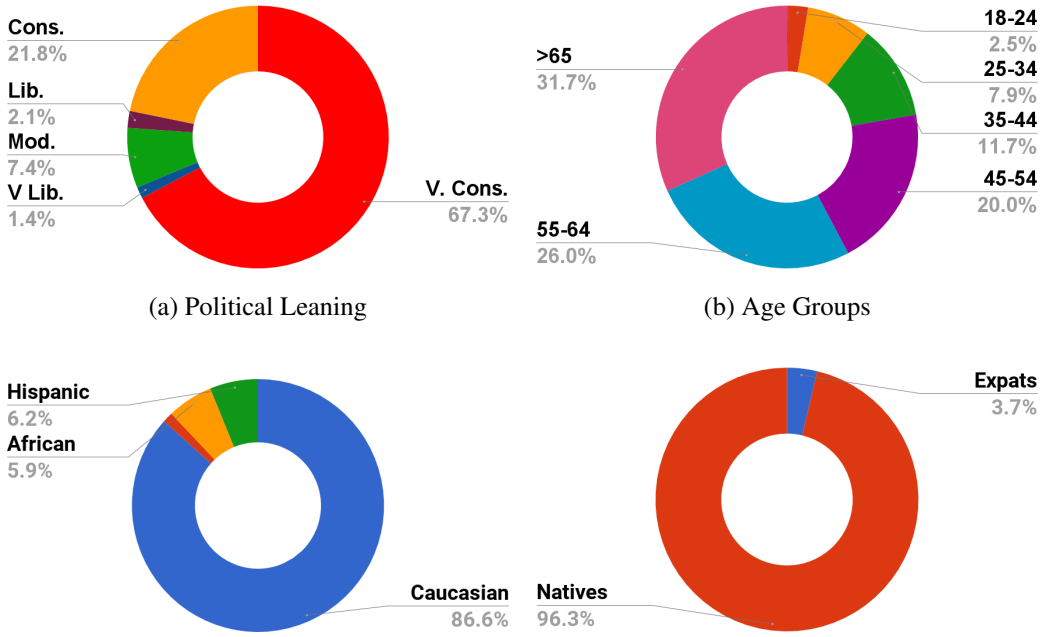
5.3.2 Quantifying biases in a finer granularity

Most of the prior works on news media bias restricted themselves to political bias, leaving out other dimensions (e.g., racial bias, gender bias or age bias) that can be very helpful in providing a more fine-grained perspective of the complex news ecosystem. One might wonder, for instance, whether a highly conservative media outlet mostly has a young Black audience, or does their audience have a high prevalence of old Caucasian people. Next, we briefly discuss, through a series of examples, the benefits of incorporating the measurement of other demographic attributes as part of our system.

5.3.2.1 Breakdown of Demographic Dimensions

A key feature we have incorporated in Media Bias Monitor is a breakdown of the audience across different demographic attributes. For example, Figure 5.6 shows the breakdown of four demographic dimensions for Breitbart, a well-known conservative media outlet. As a reference for comparison, Table 5.5 shows the distribution for these demographic attributes for all Facebook users in the US.

As can be noted in Figure 5.6(a), number of conservative and very conservative users constitute more than 89% of the Breitbart audience. Figures 5.6(b), (c), and (d) present the breakdown of Breitbart audience along age, racial affinity and national identity. These figures show that the audience of Breitbart consists of 96.3% US natives, 86.6% Caucasians, and 57.7% of users older than 55 years. These values are much higher compared to the Facebook population in the US. Additionally, the proportion of men among Breitbart audience is 55%



(a) Political Leaning (b) Age Groups

(c) Racial Affinities (d) National Identity

Figure 5.6: **Breitbart and its bias across four demographic dimensions.**

compared to 46% in the Facebook population in the U.S. However, in terms of Income Level, the distribution is quite similar to the overall Facebook population.

Observing the demographic dimensions for The Economist (see Figure 5.7), a liberal biased outlet (more than 65% interested users are Liberal and Very Liberal), we find that it has a higher fraction (21%) of high income audience, earning more than $150K, against 14.1% in the US-based Facebook population. Men and expats (immigrants) are also higher compared to the overall population, while in terms of age and racial affinities, we don't see much difference.

Taking a closer look at other media outlets, we find other examples in which conservative news outlets have audiences that are over-represented by Men - Drudge Report (63%) and Rush Limbaugh Show (60%); by older people (aged above 55) - Rush Limbaugh Show (51%) and Sean Hannity Show (67%); and by Native Americans - The Blaze (97%). On the liberal side, we see an over-representations of women - ABC News (70%) and BuzzFeed (71%); African American - Daily Show (23%) and Al Jazeera America (35%); older people (aged above 55) - Politico (38%) and PBS (33%); younger people (aged between 25-34) - Daily Show (34%).

Apart from these well known media outlets, we can expand our analysis to a wider range of outlets. For example, we found a set of conservative media outlets that are more biased towards men. Such outlets include publishers of news related to Guns (e.g., Guns.com

Gender	**Male**	46%
	Female	54%
Racial Affinities	**African American**	16.1%
	Asian American	3.5%
	Caucasian	64.3%
	Hispanic	16.1%
Income	**30k to 50k**	17.9%
	50k to 75k	35.6%
	75k to 150k	32.4%
	over 150k	14.1%
Age Groups	**under 18**	2.4%
	18-24	16.3%
	25-34	25.3%
	35-44	18.8%
	45-54	15.4%
	55-64	11.9%
	above 65	9.9%
National Identity	**Natives**	84.3%
	Immigrants	15.7%

Table 5.5: **The composition of the US-based Facebook users along different demographic dimensions and their corresponding attributes.**

(92%), FourGuysGuns (94%)), or containing military news (e.g., The Fire Critic (83%), publishing stories from Fire Service, and SOFREP.com (92%), with news written and curated by former CIA and Veterans). Conservative women, in turn, have interest in media outlets publishing religious articles (e.g., PrayAmerica (82%), Breaking Christian News (71%)), or health related stories (e.g., Lifenews.com (76%), a website that post stories with topics against abortion and euthanasia). On the other hand, Liberal outlets with major predominance of men include gay magazines (like Gay Times (84%) and Instinct (88%)) and a left-wing magazine (Jacobin - (66%)). Liberal women are over-represented in the audience of feminist and liberal magazines: The Man Repeller (96%), Feministing.com (91%) and Everyday Feminism (90%).

These examples show that our bias inference approach and the deployed system allows one to get a deeper understanding of bias in different media outlets. It not only presents the political bias of a large number of news outlets, but it also provides an interesting way of understanding other intrinsic biases (i.e., gender bias, age bias, etc.) of the audience interested in a certain news outlet.

5.3.2.2 Search and Ranking Functions

As a final contribution, the system 'Media Bias Monitor' includes a search function, which allows users to search for news outlets by name, as well as a ranking function that allows the users to find news outlets in which a particular demographic attribute is over-represented. For instance, a user belonging to a particular demographic can look up what news sources other fellow group members are subscribing to. First two rows in Table 5.6 show news outlets highly gender-biased towards either men or women. We notice that many sports-specific outlets are highly biased towards men, whereas outlets related to fashion, makeup, and pregnancy tend to be most biased towards women.

Similarly, the most racially biased outlets (third and fourth rows in Table 5.6) in terms of African-American and Asian-American users are clearly focused on these specific racial demographic groups. In terms of high age bias (as presented in Table 5.6), for people under 18 years, we can note TV channels like Disney that target adolescents, as well as outlets related to games. For the 18 to 24 years age group, we find many news outlets associated with dating, music, TV series, TV shows, and games. Interestingly, the outlets most biased towards the 25 to 34 years old users are associated with business, professions, and job seeking. For age groups higher than that, the most biased news outlets are related to parenting and family. Finally, the last two rows in Table 5.6 show very conservative and very liberal news outlets, which are all focused on politics, with their political leaning being expressed via their own names.

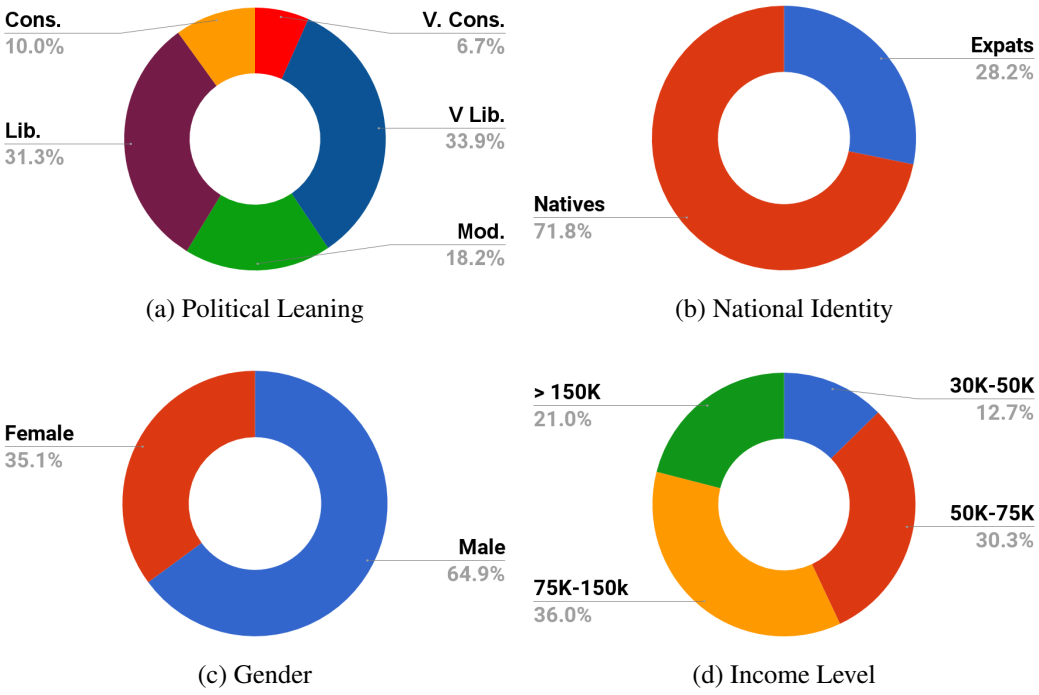
5.4 Summary

In this case study, we proposed a novel methodology to quantify the political biases of thousands of news outlets on social media. To do so, we utilized the leaning of their audience which can be obtained from our proposed framework. Specifically, for this work, we collected $20,448$ pages categorized as news by Facebook, and then leveraged the Facebook audience API to obtain demographic information for their audiences. Such audience demographics allowed us to cover a large number of media outlets, which are at least two orders of magnitude more than what existing efforts have covered. Additionally, we also identified news outlets biased along five other axes: age, gender, income level, racial affinity, and national identity. Finally, we built and publicly deployed a system, called *Media Bias Monitor* [7], which makes the biases in audience demographics for these $20,448$ news outlets transparent to any Internet user.

[7] `twitter-app.mpi-sws.org/media-bias-monitor`

Demographic Dimension	Demographic Attributes	Sample of the Most Over-represented in News Outlets
Gender	**Male**	Velocity RC Magazine(99%), myGayTrip.com(99%), Best Motoring(99%), The Gentleman's Journal(99%)
	Female	Styletoday(100%), Makeuptalk.com(99%), Pregnancy and newborn(99%), Proud single Moms(98%)
Racial Affinities	**African American**	BlackamericaWeb(92%), BlackNews.com(89%), Black Men Magaz.(80%)
	Asian American	Hoa hoc Tro Magazine(97%), Kenh14.vn(97%), Sportsoho(100%)
Age Groups	**Under 18**	Fox Action Movies(25%), Disney Channel(23%), MuchGames.com (24%), BeingGirl(26%)
	18-24	Disaster Date(68%), Fairy Tail Fans(65%), Insert Gamer(76%), Speed and Sound Magazine(66%)
	25-34	JobTopGun(70%), Canadian Business(72%), Marketing na Cozinha(65%), WeddingSutra(76%)
	35-44	Fans of Being a Mom(50%), Scholastic Parents(47%), Growing Without Schooling(44%)
	45-54	Rush is a Band (59%), Ultimate Classic Rock(38%), Yahoo! Sports Radio(58%)
	55-64	SmartMoney(61%), The new avengers(42%), The Monkees(38%), I Love Being a Grandma(37%)
Political Leaning	**Very Conservative**	Legal Insurrection(90%), RedState(84%), Patriot Update(84%), Conservative Angle(82%), Fox Nation(75%)
	Very Liberal	Sister 2 Sister(76%), The Alaska Quarterly Review(66%),Democracy Now(59%)

Table 5.6: **Examples of highly biased news outlets in Facebook along different demographic dimensions. The percentage of audience belonging to the respective demographic groups are shown in parenthesis.**



(a) Political Leaning (b) National Identity

(c) Gender (d) Income Level

Figure 5.7: **The Economist and its bias across four demographic dimensions.**

Chapter 6

Case Study: Inferring demographics of Russian Ads

In this chapter we leverage our framework to determine the targeted audience of the controversial ads that were run in the 2016 US Elections period. We also intend to provide a deeper analysis of the content of the ads in order to understand how divisive they were by taking into account the perception of people with different ideologies about the ads.

More specifically, in this case study, we analyze the potential for a new form of abuse on targeted ad platforms, namely, *socially divisive advertising*, where malicious advertisers incite social conflict by publishing ads on divisive societal issues of the day (e.g., immigration and racial-bias in policing in the lead up to 2016 US presidential elections). We focus on how ad targeting on social media sites such as Facebook can be leveraged by selective target groups on different sides of a divisive issue with (potentially false) messages that are deliberately crafted to stoke their grievances and, thereby, worsen social discord. We also investigate whether targeted ad platforms allow such malicious campaigns to be carried out in stealth, by excluding people who are likely to report (i.e., alert site administrators or media watchdog groups about) such ads.

Our study is based on an in-depth analysis of a publicly released dataset of Facebook ads run by a Russian agency called Internet Research Agency (IRA) before and during the American Election on the year of 2016 [1] [2]. Our analysis is centered around three high-level research questions:

RQ 1: How divisive is the content of the IRA ads? We quantify the divisiveness of an ad by analyzing the differences in reactions of people with different ideological persuasions to

[1] *www.wsj.com/articles/you-cant-buy-the-presidency-for-100-000-1508104629*

[2] *www.nytimes.com/2017/11/01/us/politics/russia-2016-election-facebook.html*

the ad. Specifically, using US census-representative surveys, we look at how conservative- and liberal-minded people differ in (a) how likely they are to report the ad, (b) how strongly they approve or disapprove the ad's content, and (c) how they perceive truthfulness (or falsehood) in ad's claims. Our analysis shows that IRA ads elicit starkly different and polarizing responses from people with different ideological pursuasions.

RQ 2: How effective was the targeting of the socially divisive ads? We find that the "Click Through Rate" (CTR), a traditional measure of effectiveness of targeting, of the IRA ads are an order of magnitude (10 times) higher than that of typical Facebook ads. The high CTR suggests that the ads have been targeted very efficiently. A deeper analysis of the demographic biases in the targeted audience reveals that the ads have been targeted at people who are more likely to approve the content and perceive fewer false claims, and are less likely to report.

RQ 3: What features of Facebook's ad API were leveraged in targeting the ads? We also analyze the construction or specification of "targeting formulae" for the ads, i.e., the combination of Facebook user attributes that are used when selecting the audience for the ads. We find widespread use of interest attributes such as "Black Consciousness movement" and "Chicano movement" that are mostly shared by people from specific demographic groups such as African-Americans and Mexican-Americans. We show how Facebook ad API's suggestion feature may be exploited by the advertisers to find interest attributes that correlate very strongly to specific social demographic groups.

6.1 Characterization of Russia-linked facebook ads dataset

On May 10th, 2018 the Democrats Permanent Select Committee on Intelligence released a dataset containing 3,517 Facebook advertisements[3] from 2015, 2016, and 2017 that are linked to a Russian propaganda group: Internet Research Agency (IRA).

Each ad is composed of an image and text, both of which were shown to Facebook users (Figure 6.1 shows an example). Additionally, each ad contains a landing page, which is a link to the host of the ad, as well as an ad ID; an ad targeting formula, which is a combination of demographic, behavioral and user interest aspects used to target Facebook users; the cost for running the ad in Russia Rubles[4]; the number of impressions, which is the number of users who spent some time observing the ad; the number of clicks received by the

[3] *democrats-intelligence.house.gov/facebook-ads/social-media-advertisements.htm*

[4] We converted currency of the costs to USD as of May 15th, 1 USD = 61.33 RUB.

Figure 6.1: **Example of an Ad from the Dataset.**

ad; and, finally, the ad creation and end dates. This section provides an overview of these ads.

6.1.1 Time distribution

The ads in the dataset were run between June 2015 and August 2017. From the 3,517 advertisements, we found that 617 (17.5%) were created in 2015, 1,867 (53.1%) in 2016, and 1,033 (29.4%) in 2017. Figure 6.2 shows the distribution of these ads over time in terms of the number of ads created per month, cost to run the ads, and impressions and clicks received. Note that the y-axis is in log scale. We observed that the number of impressions, and clicks, increases almost an order of magnitude around the election period (shaded region). There is also another peak in February, just after the newly elected U.S. President Donald Trump assumed office.

6.1.2 Landing pages

We first explore the ad landing pages: the *urls* to which users who clicked on the ads were redirected. There are 462 unique landing pages corresponding to all the ads. Figure 6.3 shows the top 10 landing pages per number of ads posted. The most popular landing page (`fb.com//Black-Matters-1579673598947501/`) posted 259 advertisements. Interestingly, one of the top landing pages, the *musicfb.info*[5], invites users to install a browser extension, which was reported to send spam to the Facebook friends of those who installed it[6]. This landing page received 24,623 impressions, 85 clicks, and spent around US$112.38.

[5] `https://web.archive.org/web/20161019155736/https://musicfb.info/`
[6] `https://www.wired.com/story/russia-facebook-ads-sketchy-chrome-extension/`

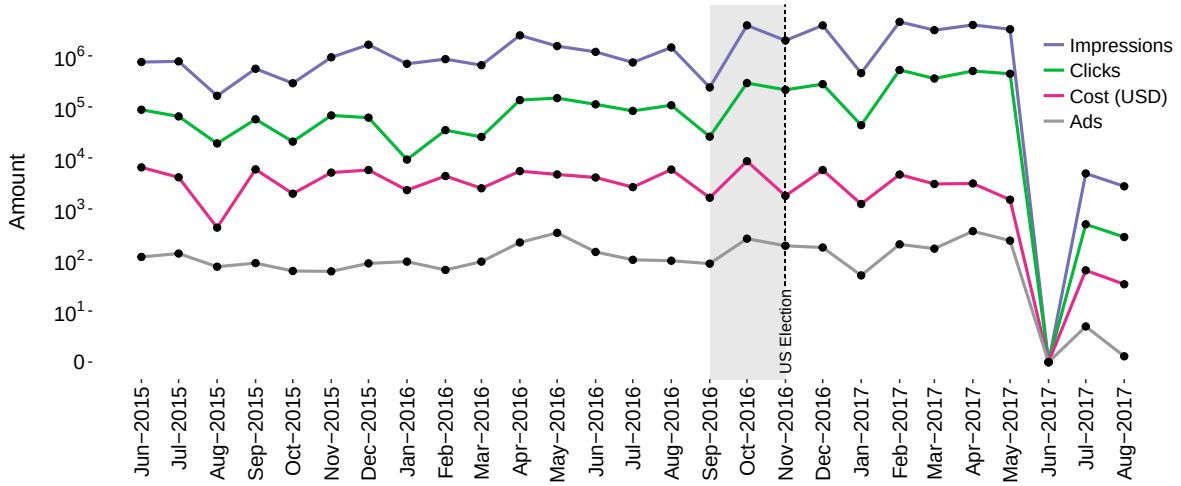


Figure 6.2: **Number of ads created, their impressions, cost, and received clicks over time. Shaded region shows 2-month period just before the 2016 U.S. Election.**

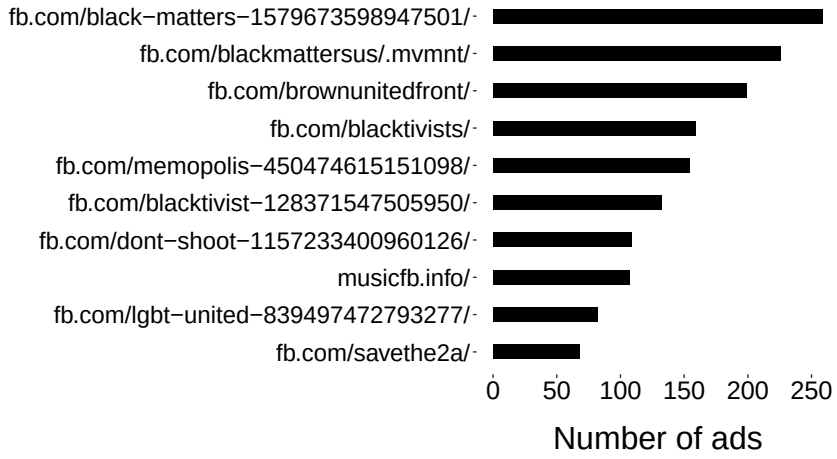


Figure 6.3: **Top 10 Landing Pages based on the number of ads.**

The domain *musicfb.info* was also promoted by other pages, accounting for 3% of all ads. We also find that the most popular landing pages are Facebook pages, accounting for 84% of all ads, followed by *blackmattersus.com* (7%), and Instagram (3.4%). For 28 ads, we were not able to identify their landing pages because these pages were obfuscated, possibly to protect some personally-identifiable information.

6.1.3 Cost, impressions, and clicks

Figure 6.4 (left) shows the Cumulative Distribution Functions (CDFs) for the number of impressions, clicks, and amount spent to advertise all of the ads in the dataset. The most expensive ad cost $5,307$ USD, whereas the highest number of impressions generated was

Impressions		Clicks	
fb.com/brownunitedfront/	14.3%	fb.com/brownunitedfront/	18.8%
fb.com/blacktivists/	10.8%	fb.com/Blacktivist-128371547505950/	13.8%
fb.com/Blacktivist-128371547505950/	10.5%	fb.com/blacktivists/	11.9%
fb.com/blackmattersus.mvmnt/	4.7%	fb.com/blackmattersus.mvmnt/	7.0%
fb.com/Woke-Blacks-29423460095/	3.3%	fb.com/Dont-Shoot-1157233400960/	3.6%
fb.com/copsareheroes/	3.3%	fb.com/blackmattersus/	2.5%
fb.com/blackmattersus/	3.1%	fb.com/patriototus/	2.5%
fb.com/South-United-177703736255/	2.7%	fb.com/Memopolis-4504746151510/	2.4%
fb.com/Dont-Shoot-1157233400960/	2.2%	fb.com/Woke-Blacks-29423460095/	2.3%
fb.com/patriototus/	1.7%	fb.com/South-United-177703736255/	2.0%

Table 6.1: **Most popular landing pages per impressions and clicks.**

$1,335,000$ and the maximum number of clicks was $73,060$.

Nearly 25% of the landing pages spent more than 100 dollars, 26.8% of the pages received more than 1,000 clicks, and around 36.1% had more than $10,000$ impressions. On the other hand, more than 25% of the ads had no impressions, clicks, and cost, suggesting these ads were not launched or ran for a very short period of time.

An average ad cost 34.5 USD, was seen by 11,536 users, and received 1,062 clicks. The average value is increased to 38 USD for cost, 16,482 for impressions, and 1,521 for the number of clicks if we exclude those ads that appeared not to have been run. The Pearson's correlation coefficient among cost, impressions, and clicks is very high, particularly between impressions and clicks (0.89). We also noted that this dataset is quite skewed, as 10% of the ads accumulate 85.18% of the total cost, 71.93% of the total number of impressions, 69.47% of the total number of clicks.

However, there were notable exceptions to this correlation: higher investment (cost) did not always lead to higher return (e.g., impressions, clicks). Table 6.1 shows the most popular landing pages per impressions, clicks, and cost of the ads. For example, `fb.com/brownunitedfront/`, received the largest number of impressions (5,817,734), corresponding alone to 14.3% of impressions obtained by all ads, but cost only 6.5% of the total cost of all ads in the dataset.

6.1.4 Click-through rate

Finally, we compute the click-through rate (CTR) of these ads, which is a typical metric to measure the effectiveness of an ad. It is computed as a ratio between the number of clicks and the number of impressions received by an ad. Figure 6.4 (right) shows the cumulative distribution function of the CTR of the ads, excluding those with 0 values for clicks, impressions, and cost. The median CTR is 10.8% and 75% of the ads have a CTR higher than

Cost (USD)	
fb.com/patriototus/	6.5%
fb.com/blacktivists/	5.4%
fb.com/blackmattersus/	5.3%
fb.com/timetosecede/	4.7%
fb.com/Igbtun/	4.3%
fb.com/BlackJourney2Justice/	4.1%
fb.com/MuslimAmerica/	3.28
fb.com/South-United-177703736255/	3.2%
fb.com/blackmattersus.mvmnt/	2.7%
fb.com/savethe2a/	2.5%

Table 6.2: **Most popular landing pages per cost.**

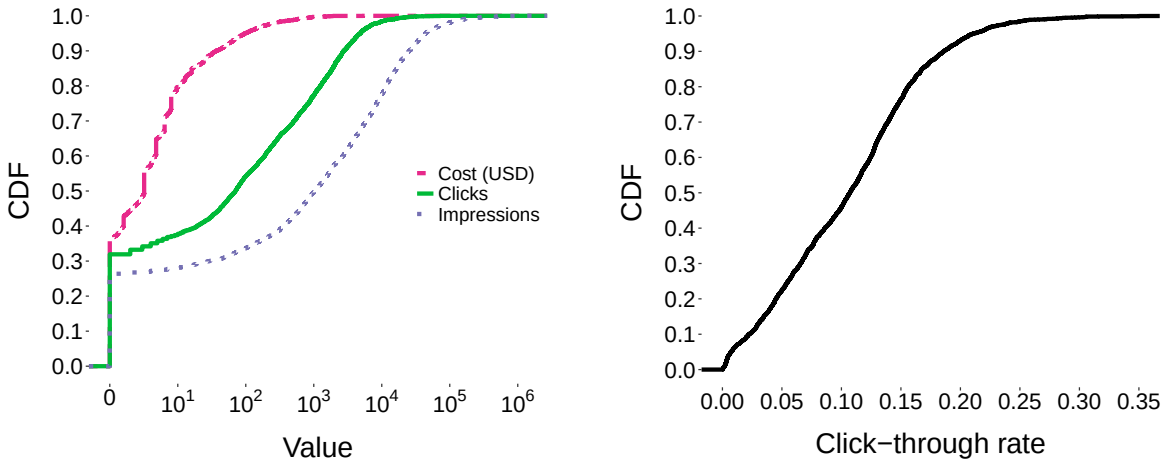


Figure 6.4: **Cumulative Distribution Function (CDF) of clicks, impressions, and costs (left), and click-through-rates of the ads (right)**

5.6. The average CTR is 10.8%. These are very high values for CTR. As a comparison, WordStream released a report as of April 2018[7] which shows the average CTR for Facebook ads across all industries is 0.9%. As an example, Retail is 1.6%, Fitness is 1%, Health care 0.8%, and Finance is 0.56%. This means that these political ads have a CTR that is about an order of magnitude higher than a typical Facebook ad.

6.1.5 Ad campaigns

Next, we identify and quantify ad campaigns: sets of ads with similar content (text). To group ads, we use a simple text-based similarity measurement.[8] We compute the Jaccard's

[7] *wordstream.com/blog/ws/2017/02/28/facebook-advertising-benchmarks*
[8] We could not include 33 (9.38%) ads in this analysis as they did not contain text.

Campaign	#Ads	#Impres.	#Clicks	Cost (USD)
Join the event, bring your friends, It's free! Organized on a donation basis. Free Self-Defense class in Los Angeles, CA	60	97,905	4,204	982.99
The community of 2nd Amendment supporters, guns lovers & patriots	58	564,949	40,576	2,021.39
Free online player! Jump in the world of free music! Click and download for ur browser Unlimited, free and rapid app for you - listen music online on ur Facebook! musicfb.info FaceMusic, Stop A.I.	49	3,190	12	10.57

Table 6.3: **Top 3 campaigns, using with a text similarity of 60%, based on the number of ads in each campaign. The impressions, clicks, and cost (USD) are based on the aggregated sum for each ads in a Campaign. We use the smallest text for each campaign due to space limitation.**

Index for the text[9] of all the possible pairs of ads. We consider two ads to be part of a campaign if the text of those ads was more than 60% similar. By manually inspecting the obtained campaigns we validate the results of this approach: it grouped ads with the same objectives, but slightly different text.

In total, we found 376 campaigns. Table 6.3 shows the 3 most popular campaigns (those containing the most ads). The campaign containing the highest number of ads consisted of 60 ads that cost a total of US$$982.99$ to run, and generated $97,905$ impressions and $4,204$ clicks. The campaign with the highest number of impressions consisted of 7 ads, that generated a total of $1,351,594$ impressions, $66,949$ clicks, and which cost a total of US$$1,798.42$. The campaign that generated the most clicks consisted of 49 ads, generating $1,016,420$ impressions and $157,066$ clicks, and cost US$$160.19$ to run. Finally, the most expensive campaign consisted of 2 ads that cost US$$5,885.37$ to run. It generated $629,151$ impressions and $83,727$ clicks.

6.1.6 High impact ads

Our analysis reveals that only a few ads are responsible for most of the cost, impressions, and clicks. Considering this, we defined a set of high impact ads as the union of the top 10% ads in terms of cost, impressions, clicks, and CTR. We obtained 905 high impact ads, corresponding to 27.7% of the entire dataset. These ads account together for 83.9% of the total number of impressions, 81.8% of clicks, 88.5% of the cost, and 46.9% of the CTR. For the purposes of our study, where we require manual inspection of the ads (to identify their targets and to run surveys), our ensuing analyses concern those high impact ads run before the 2016 U.S. elections: 485 ads.

[9]We first performed basic text preprocessing, including the remotion of stop words and punctuation, tokenization, and stemming.

6.2 Analyzing the divisiveness of the ads

To investigate whether these ads were designed to be ideologically divisive – that is, designed to elicit different reactions from people with different political viewpoints – we conducted three online surveys on a U.S. census-representative sample (n=2,886). We used each survey to measure one of three axes along which ads could potentially be divisive: 1) *reporting*: whether respondents would report the ads, and why, 2) *approval and disapproval*: whether they approve or disapprove the content of the ad, and 3) *false claims*: if they are able to identify any false claims in the content of the ad.

Our surveys considered only those 485 *high impact* ads which were run before the elections. Each survey showed ten ads followed by demographic questions. More detail on the specific questions used to assess each axis is provided in the corresponding axis subsections that follow. The survey questions were pre-tested using cognitive interviews and all survey questions included a "I don't know" or "Prefer not to respond" answer choice to ensure internal measurement validity [Beatty and Willis, 2007].

To obtain a demographically representative sample, and ensure that we captured a wide variety of American perceptions, we deployed the surveys using the Survey Sampling International survey panel[10], a non-probabilistic census-representative survey panel. For each survey, we sampled at least 730 respondents (15 responses per ad) whose demographics were representative of the U.S. within 5% and who had a range of political views (40% liberal, 40% conservative, and 20% moderate or neutral); across the three surveys we obtained a total sample of 2,886 respondents.

We measured overall ideological divisiveness on the three axes (reporting, approval, and false claims) using two metrics:

Within-group divisiveness. Within-group divisiveness measures the extent to which respondents' answers about a particular ad are consistent with their political ideology. That is, do all liberals answer similarly about a particular ad. For each ad, we first calculate the standard deviation of *all* the responses, and then we calculate the standard deviation of the responses within a particular ideological group. Next, we compute within-group divisiveness as the fraction of within-group standard deviation to the overall standard deviation. Therefore we interpret values lower than 1 as lower divisiveness (and greater agreeableness) within a group than overall, and values greater than 1 as greater within-group divisiveness than overall.

Between-group divisiveness. Between-group divisiveness measures the extent to which answers from respondents of one political ideology differ from answers of respondents who align with another political ideology. That is, do liberals answer differently about a particular ad than conservatives? For an ad, we calculate the difference between the mean responses

[10]https://www.surveysampling.com/audiences/consumer-online/

Measure (Group)	Reporting		Approval		False Claims	
	Mean	Stdev.	Mean	Stdev.	Mean	Stdev.
Within-group divisiveness						
Liberals	0.87	0.47	0.92	0.36	0.66	0.69
Conservatives	0.90	0.43	0.98	0.31	0.86	0.63
Between-group divisiveness						
Political	0.24	0.18	0.34	0.24	0.17	0.14

Table 6.4: **Divisiveness measures of the high impact ads.**

per ideological group, and then compute the fraction of this difference over the maximum possible difference given the range of values to obtain the between-group divisiveness measure. This limits the range of between-group divisiveness measure between 0 and 1, where higher values indicate greater divisiveness between ideological groups.

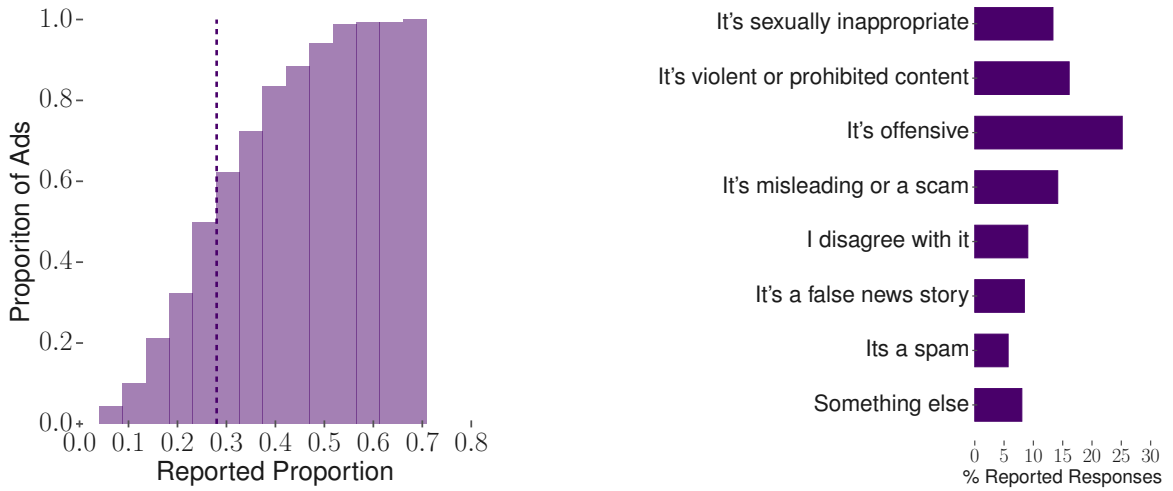
Table 6.4 summarizes the divisiveness of the high impact ads. We find that the within-group divisiveness measure is lower than 1 for all our surveys. This indicates high agreeableness within the ideological groups. In addition, about 20% of the ads show between-group divisiveness higher than 0.5, indicating severe divisiveness between ideological groups for those ads.

6.2.1 Likelihood of reporting the ads

The first axis of divisiveness that we explored was "reporting". We surveyed respondents regarding: 1) Whether they would report the ad shown?[11], and 2) If they would, why do they find the ad inappropriate? Answer choices given, drawn directly from Facebook's reporting interface [Facebook Help Center, 2018], were: *sexually inappropriate*, *violent*, *offensive*, *misleading*, *disagree*, *false news*, *spam*, and *something else*.

Figure 6.5 shows the reporting responses for the high impact IRA ads. For over 73% of these ads, at least 20% of the respondents responded that they would have reported the ads. We observe that the majority of the ads were reported on the grounds of being offensive (25%), violent (15%), and misleading (15%). Additionally, a substantial proportion (9%) of the reported responses belonged to the *something else* category. In such cases, the respondents entered free-text to explain their reason for inappropriateness. Out of the 61 responses that we received in the free-text box, the pre-dominant reasons were that the ad incites racism (20%), and that the ad creates division (5%) in the society.

[11] Specifically, we asked "Some social media platforms allow you to report content by clicking "report". Would you report this ad (e.g., Mark it as inappropriate or offensive)" With answer choices "Yes", "No", "I don't know".



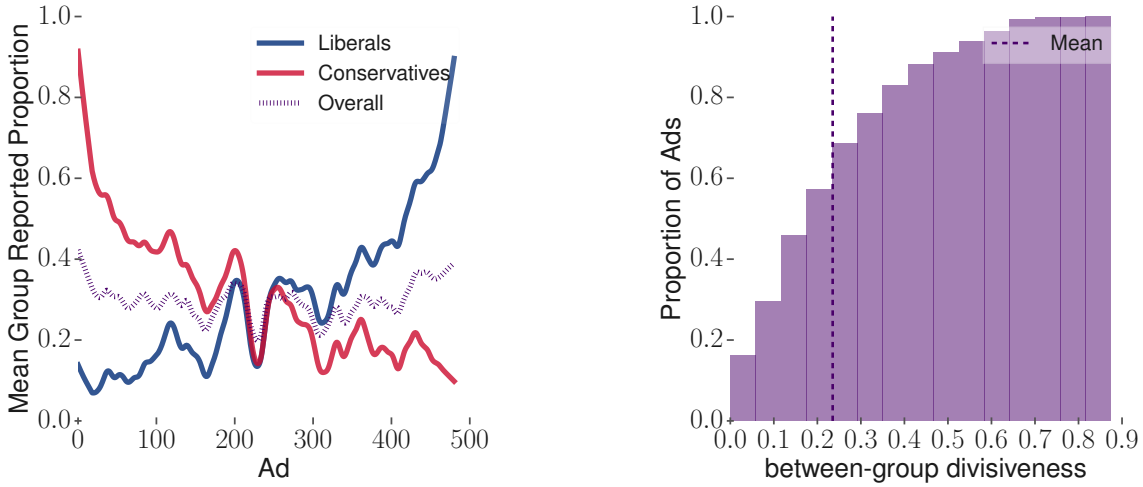
(a) Reported Ads

(b) Reasons of Inappropriateness

Figure 6.5: **Distribution of the high impact ads on the (a) proportion of reported ads in our dataset, (b) reasons of inappropriateness.**

Next, to examine ideological divisiveness, we find that the mean within-group divisiveness is 0.87 (stdev = 0.47) for liberals and 0.90 (stdev = 0.43) for conservatives. Both of these within-group divisiveness measures being less than 1, suggests that the likelihood with which individuals within the same ideological group agree about reporting an ad is higher than that when compared against individuals across ideological groups.

Figure 6.6 (a, b) shows the distribution of the reporting across ideological groups. We find significant differences in terms of the reporting behavior across political ideologies. Defining a median threshold for divisiveness, we find that in over 50 percent of the ads, liberals and conservatives completely disagreed with each other (eg. conservatives showed *more* than their median reported proportion and liberals showed *less* than their median reported proportion, and the vice versa). Table 6.5 shows a few examples of the ads which showed the greatest differences in the reporting behavior by the respondents of two political ideologies. These ads typically mention politically-charged topics. For example, immigration — "TAG YOUR PHOTOS WITH #TXagainst Send us the reason why don't you want illegals in Texas. Comments, photos, and videos are welcomed!" — in this case, presenting a viewpoint associated with the Republican Party, Or police brutality — "Police are beyond out of control, help us make this viral! Follow our account in order to spread the truth!" — in this case, presenting a viewpoint associated with the democratic party. Finally, figure 6.7 depicts two controversial ads. The first one was mostly reported by liberals and not reported by conservatives (a), whereas the second one had the opposite analysis depending on the ideological view, *i.e.,* reported by conservatives and not reported by liberals (b).



(a) Ideological differences

(b) Between-group divisiveness

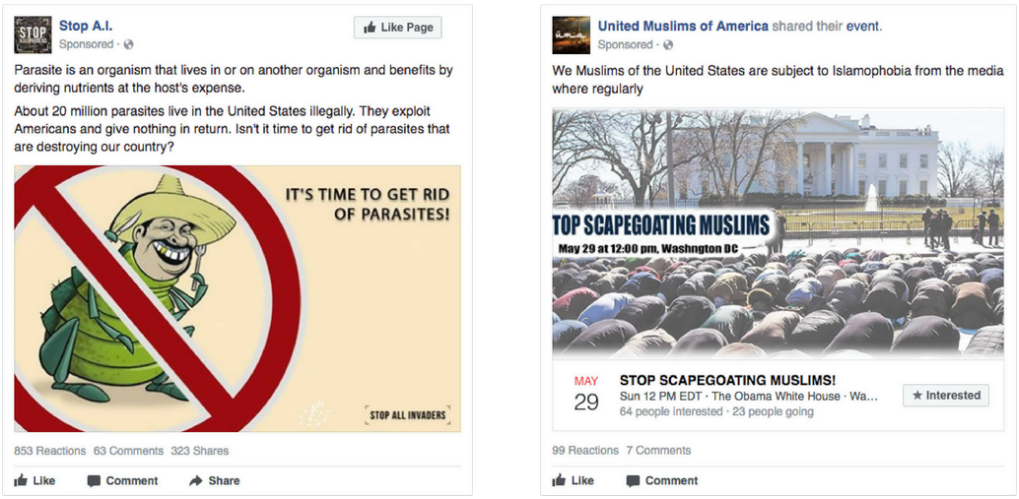
Figure 6.6: **Distribution of reporting across ideological groups. (a) shows the distribution o proportion of the ads being reported by either of political ideology, with x-axis containing each of the high impact ads, (b) plots the between-group divisiveness for the high impact ads**

Reported by both the liberals and the conservatives
TAG YOUR PHOTOS WITH #TXagainst Send us the reason why don't you want illegals in Texas. Comments, photos, and videos are welcomed!
Counter-protest against 'White Power' Confederate rally at Stone Mountain Not My Heritage
Wakiesha Wilson, 36, was found dead on Easter Weekend at the LAPD's detention center in jail cell. According to ABC7 report, the black woman, Wakiesha Wilson, had a disagreement with officers before she was found died. Wilson spoke to her lovely family that v Black Woman Found Dead In Jail Cell After Arguing With Detention Officers I Black Matters Black Matters
Police are beyond out of control, help us make this viral! Follow our account in order to spread the truth!
Everything you wanted to know about Clinton's dark side. Clinton FRAUDation
Reported predominantly by the liberals.
Join us to learn more! Why aren't white hoods and white supremacist propaganda illegal here in America? Why are Germans ashamed of their bigotry, while America is proud of it? Black America(@black Blacklivessss
We simply can't allow Muslims to wear burga, otherwise everybody who wants to commit a crime or terror attack would wear this ugly rug and hide his or hers identity behing it. The risk is too highl Burga and other face covering cloth should be banned from wearing in public!
Five police officers were killed in an organized attack during the protest in Dallas this Blue Lives Matter
Black intelligence is one of the most highly feared things in this country.
Parasite is an organism that lives in or on another organism and benefits by deriving nutrients at the host's expense. About 20 million parasites live in the United States illegally. They exploit Americans and give nothing in return. Isn't it time to get rid of parasites that are destroying our country?
Reported predominantly by the conservatives.
Come and march with us on 16 April. Stand with Baltimore. Let's make change! Freddie Gray Anniversary March
Click Watch More to join us! Let's fight against police brutality together! donotshoot.us Donotshoot.us Don't Shoot
The USA is exactly the place where cops can't care less about people's civil rights. They are cynical toward the rule of law and disrespectful of the rights of fellow citizens. Details: http://donotshoot.us/
We Muslims of the United States are subject to Islamophobia from the media where regularly STOP SCAPEGOATING MUSLIMS!
People, our race is in danger! Together we are an invincible power. Just say your word! Join us! Black Pride

Table 6.5: **Example ads on the basis of reporting behavior by the respondents from two political ideologies.**

6.2.2 Approving content of the ads

As another characterization of people's reactions to the ads, we asked respondents in a second survey whether they approve or disapprove a particular ad, and how strongly they approve



(a) Reported by Liberal (b) Reported by Conservative

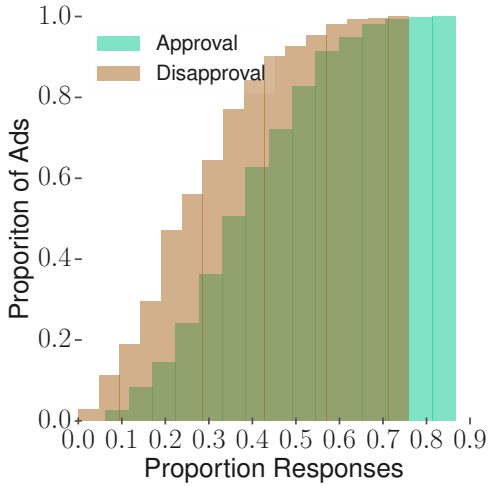
Figure 6.7: **Ads with controversial content**

or disapprove.[12] These questions in the survey were constructed based on questions about political preference that have been extensively pre-tested by Pew Research for previous surveys about political polarization Beatty and Willis [2007]. We find that 87% of the adds were approved and 63% of the ads were disapproved by at least 20% respondents (see Figure 6.8 (a)). To quantify the received responses, we assigned an approval score on a 5 point scale with values of -2 (strong disapproval), -1 (weak disapproval), 0 (neither approve or disapprove), +1 (weak approval), and +2 (strong approval). While computing the mean approval score for a group, we dropped the 0 responses to ensure that a mean approval score close to 0 corresponds to similar weights from approval and disapproval. Table 6.6 lists some example ads along with their approval tendencies by the two ideological groups within our dataset.

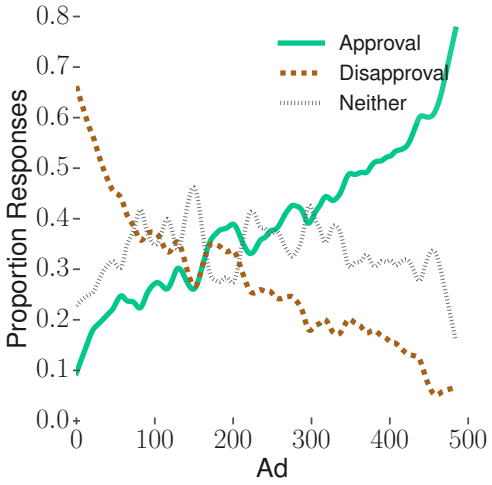
Figures 6.8 (c and d) show the relationship between respondents' ideology and approval of ad content. We observe that the mean within-group divisiveness for liberals is 0.92 (stdev = 0.36) and 0.98 (stdev = 0.31) for conservatives (Table 6.4). Both the within-group divisiveness values being lower than 1, suggests that the likelihood with which individuals within the same ideological group would agree about approving an ad is higher than that when compared against individuals across ideological groups.

The divisiveness in approval responses is further confirmed by the between-group divisiveness measure which ranges between 0 and 1 (mean = 0.34) across the high impact ads.

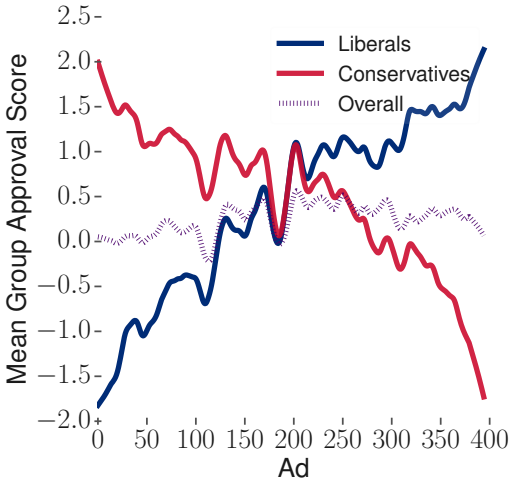
[12]Specifically, we asked "Do you approve or disapprove of what the ad says or implies?" Answer choices: Approve; Disapprove; Neither; There is nothing in this ad to approve or disapprove of; I don't know. Followed by a measure of strength "Do you [approve/disapprove] very strongly, or not so strongly?" if the prior question was answered with approve or disapprove.



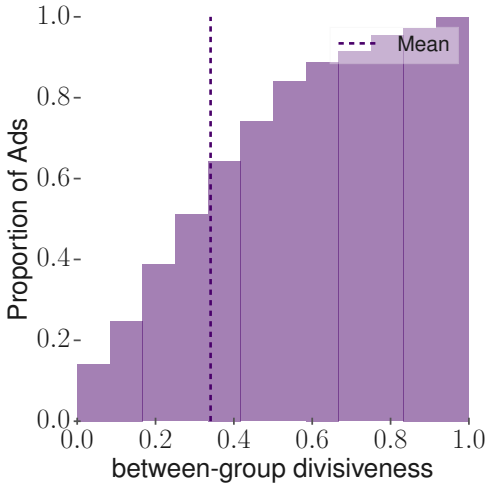
(a) CDF: Proportion Responses



(b) Approval per ad



(c) Ideological Differences



(d) Between group divisiveness

Figure 6.8: **Distribution of the ads on approval and disapproval: (a&b) overall, (c&d) across ideological groups. (a&d) plot the cumulative distribution functions (cdfs), (b&c) plot the differences in approval in each ad, where x-axis consists of all the ads**

6.2.3 Perceptions of false claims in the ads

To examine whether the high impact IRA ads contained any false claims, in another survey we asked the respondents if they could identify any false claims present in the ads.[13] We find that 89% (433 out of 485) of the high impact ads were identified to have at least one false claim, and about 45% of the ads contained false claims according to 10% of the respondents. Figure 6.9 (a) shows the cumulative distribution of the ads with the number of respondents

[13]Specifically, we asked respondents to "Please copy and paste any phrases or sentences in the advertisement that you think contain a "factual claim". That is, something that someone could verify as True or False. If you cannot identify any claims, please type "No Claims" in the first box." We then asked them to label the phrases they had identified as "True", "False" or "not sure whether they are True or False".

Approved by both the liberals and the conservatives
Show up, fight racism and take a stand for equality. Monday, May 2 at 4 PM at Erie County Holding Center Justice For India: Not 1 More!
Alton Sterling, an innocent 37-year-old Black male, was outrageously executed by two Baton Justice For Alton Sterling
Did you see this? Damn... We lost count of how many mentally ill citizens were murdered during encounters with violent cops. Here is another woman suffering from mental illness. Full story: http://bit.ly/10rglhk
Join Us! Support The Police!
Darkness cannot drive out darkness; only light can do that. Hate cannot drive out hate; only love can do that. Martin Luther King, Jr.
Disapproved by both the liberals and the conservatives
No wonder white boys don't get shot when they're arrested':
Anti-immigration is the only salvation!
O000ps seems like someone screwed up! Salute our brave and smart cops who mistake man's member for a deadly weapon. Follow US and stay WOKE!
It's ok they're women so they'll only find the kitchen
This man beat up police officer who tased his wife. Do you agree with the man who defended his woman?
Approved by the liberals and disapproved by the conservatives
Two years have passed since August 11, 2014, the date. when 25-year-old Ezell Ford was mur Justice For Ezell Ford And Donnell Thompson
We don't want to honor racism, slavery and hatred. This is what Confederate Heritage is. Not My Heritage Rally
Say it loud: I'm black and I'm proud!
We Muslims of the United States are subject to Islamophobia from the media where regularly STOP SCAPEGOATING MUSLIMS!
Click to Learn More! Everybody knows that Islam is against terrorism but not everyone believes this fact! Islam does not support terrorism under any circumstances. Terrorism goes against every principle in Islam. In fact if a Muslim engages in terrorism, he is not following Islam and so he is not a Muslim! America, stop insulting peaceful citizens, stop taking all of us as criminals, we don't deserve such attitude! #muslimvoice #muslim @muslim Voice Musliminst
Approved by the conservatives and disapproved by the liberals
Heritage not hate y'all! Our flag has nothing to do with racism! The Federal Government shouldn't be able to dictate what we can and cannot do. Go follow Confederate page #1 on Instagram south united if you are proud of our southern heritage. God bless Dixie! Confederate page #1 on Instagram! No racism, no hate! The south will rise again!
If we ever forget that we are One Nation Under God, then we will be a nation gone under. Ronald Reagan
Our country was drawing a blank for the last eight years. We need a strong leader who will March for Trump
America is at risk. To protect our country we need to secure the border.
Stop refugees! The're taking our jobs!

Table 6.6: **Example ads on the basis of the approval behavior by the respondents from two political ideologies.**

who identified at least one false claim in them.

Next, as in the other two content analyses, we examined whether respondents' ideology related to their perception of the presence of false claims (Figure 6.9). Both the within-group divisiveness values being lower than 1, suggests that the likelihood with which individuals within the same ideological group would agree about finding false claim in an ad is higher than that when compared against individuals across ideological groups. Table 6.7 shows a sample of ads and false claims identified by respondents from each ideological group (liberal, conservative).

6.3 Analyzing the targeting formula

Next, we focus on understanding how the target formula is created by advertisers and the role that Facebook interface plays on that.

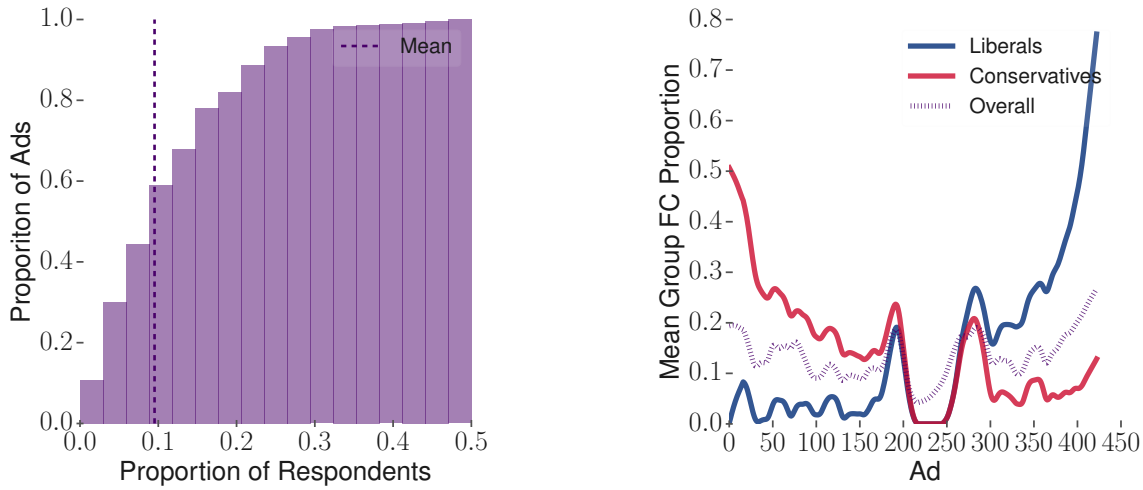
False Claims identified by both the liberals and the conservatives
Bernie Sanders has proven himself to be the best candidate in every way. He is fair and strong and he is the only one fighting for the black community. For more than 40 years he is advocating against any injustice and stays true to his moral values. Vote for Bernie!
Illegal immigrants are not only flooding our country with drugs and sinking our economy, they are a major threat to Americas integrity. They don't care about American laws, history and heritage. They just want our money and social benefits. Considering the fact that they multiply like rabbits, soon we will all wear sombreros. We need to stop this invasion.
When you live in Texas you know that you are the chosen one!
It is time to wake up and see the truth. Cops are not our friends and government doesn't care about you. Share this if you're awake!.
It's ok they're women so they'll only find the kitchen
False claims identified by the liberals.
Clinton said the United States needs to confront the 'systematic racism' in its law enforcement efforts. We're sick of politicians organizing and leading the systematical propaganda against our police. It is unfair and vilely to accuse our heroes of every sin and crime. In fact, the efficiency of our cops resulted in a decrease of the average amount of crimes, especially in large cities. Law-abiding citizens should never fear cops, but criminals do. And that's why Hillary is on the criminals' side. Join our rally on July, 23th in New York City, it's time to show Clinton that we will never let her become our next President!
It might sound like a cliche but "get a job" is a really good advice for young liberals protesting against everything in the world. Old man Ronald knew what he was talking about! Our college students should have an experience of paying taxes before standing for illegal immigrants' rights. They should rise their own children before standing for gay parenthood. It's no secret most active liberal's supporters are people about 20-25 years- old while most conservatives are older. Well; as they say; wisdom comes with ages.
His failed medical reform and unbelievable national debt is enough to put Obama behind bars. but that's not all. His greatest "accomplishment" is flooding America with countless criminals and giving them all an absolute omnipotence. Thanks to Barack Hussein Obama we have at least one big terror attack each year; not to mention illegals raging out and poisoning our country with drugs. For what he did to America Obama should rot in prison for the rest of his life.
Border Patrol agents in South Texas arrested an illegal alien from Honduras that had previously been deported and convicted of Rape Second Degree. Thanks to Obama's and Hillary's policy, illegals come here because they wait for amnesty promised. The wrong course had been chosen by the American government; but all those politicians are too far from the border to see who actually sneaks through it illegally. Rapists, drug dealers, human traffickers; and others. The percent of innocent poor families searching for a better life is too small to become an argument for amnesty and Texas warm welcome.
Anti-immigration is the only salvation!
False claims identified by the conservatives.
Don't Shoot is a community site where you can find recent videos about outrageous police misconduct, really valuable ones but underrepresented by mass media. We provide you with first-hand stories and diverse videos. Join us! Click Learn more!
We don't want to honor racism, slavery and hatred. This is what Confederate Heritage is. Not My Heritage Rally
The USA is exactly the place where cops can't care less about people's civil rights. They are cynical toward the rule of law and disrespectful of the rights of fellow citizens. Details: http://donotshoot.us/
Police are beyond out of control, help us make this viral! Follow our account in order to spread the truth!
Join us to study your blackness and get the power from your roots. Stay woke and natural! Nefertiti's Community

Table 6.7: **Example ads on the basis of false claims identified by the respondents from two political ideologies. Identified false claims are highlighted in pink.**

6.3.1 Russian ads targeting attributes and strategy

The vast majority of the *high impact* ads, 895 out of the 905, used the Attribute-based targeting provided by Facebook to choose the audience to be reached. The remaining 10 ads used Look-alike targeting. Particularly, we noted that 91.2% of these ads contain interests and behaviors, which are attributes that Facebook suggests as part of its interface. We found that 78% of the ads used 2 or more interests and behaviors in their formula, creating very complex formulas with up to 39 distinct attributes.

The leftmost table in table 6.8 shows the top 20 attributes used based on the number of times they appeared in different ads. There were 497 distinct attributes and the most present attributes interest were African-American history and African-American Civil Rights Movement (1954-68), being used in 295 (32%) advertisements. This list indicates a prevalence of attributes related to African-American (the top seven, eleven in total) and Hispanic Population, with interests like Mexico, 'Hispanidad' and 'Latin hip hop', accounting for seven



(a) Proportion of FCs Identified

(b) Political Ideology

Figure 6.9: **Distribution of the ads on false claims (FCs): (a) overall (as a cumulative density function), (b) across ideological groups (where each ad is plotted on the x-axis.**

Simple Count	
African-American history	295
African-American Civil Rights Movement (1954-68)	295
Malcolm X	196
Martin Luther King, Jr.	181
African-American culture	152
Black (Color)	135
Pan-Africanism	127
La Raza	103
Chicano Movement	103
Chicano rap	100
Mexico	97
Lowrider	97
Latin hip hop	97
Hispanidad	97
Black Consciousness Movement	79
Stop Police Brutality	73
Black nationalism	72
Martin Luther King III	70
Police misconduct	68
Behaviors:African American (US)	54

Impressions	
African-American Civil Rights Movement (1954-68)	32.61 %
African-American history	32.56 %
Black (Color)	21.71 %
Pan-Africanism	21.04 %
African-American culture	16.58 %
Malcolm X	16.45 %
La Raza	14.84 %
Chicano Movement	14.84 %
Chicano rap	14.55 %
Mexico	14.32 %
Lowrider	14.32 %
Latin hip hop	14.32 %
Hispanidad	14.32 %
Martin Luther King, Jr.	13.69 %
Black nationalism	13.30 %
Black Consciousness Movement	13.24 %
Stop Police Brutality	12.82 %
Martin Luther King III	12.74 %
Police misconduct	12.50 %
Behaviors:African American (US)	8.97 %

Table 6.8: **Top 20 Interests count and per impressions. The percentage represents the ratio between the aggregated sum for each interest and the total sum for each metric.**

interests. The remaining 2 interests are Stop Police Brutality and Police misconduct.

As the ranking presented considers the simple count of appearance in the Ad and may lack some useful information, we created other tables also as part of table 6.8 that shows the most popular interests per impressions, clicks, and cost (USD) of ads. For instance, the total number of impressions in the dataset is $34,054,044$, and the interest which achieved the highest aggregated number of impressions is the African-American Civil Rights Movement (1954-68), with $11,104,578$ impressions. Therefore, 32.61% of the total number of impres-

sions occurred for ads which targeted *African-American Civil Rights Movement (1954-68)*. Impressions and Clicks rankings do not present significant differences in the top elements whereas the ranking by cost has some interesting variations.

At first, the top two elements, Patriotism and Independence do not figure in the top 20 count ranking as they appeared in only 29 and 26 ads, respectively. However, they appeared in higher cost ads, as the one with more than five thousand dollars ran by the page Being Patriotic with the following text: "United We Stand! Welcome every patriot we can reach. Flag and news!" [14]. Similarly, some LGBT related interests such as Homosexuality and Gay Pride appeared only in the ranking by cost due to ads with high cost, in particular, ads published by LGBT United page [15] [16]. This difference in the rankings might suggest that the advertisers used distinct approaches to reach different categories. In order to reach African-American they ran many ads while for reaching Nationalists and the LGBT, there were a less amount of ads with a higher budget. As previously shown in figure 6.4, nearly 90% of the ads had a budget below 100 USD which supports the idea that a small number of ads used high budgets.

This characteristic might also indicate a strategy to first recruit as many followers as possible by publishing ads with generic posts and then influencing the followers by publishing posts (boosted or not) containing ideological positions. For instance, the page Being Patriotic has also launched ads related to illegal immigration [17] asking users to share if they support the idea and ads with criticism to the high issue of green cards [18]. Both examples had a very low budget (0.18 USD and 2.07 USD, respectively), and the cheaper was shared more than two hundred times (as identified in the ad's image), meaning that it might have reached all the friends of people who shared it. The same strategy was verified for the landing page Woke Black that published ads inviting people to like the page [19,20], spending 1300 USD for both publications, and also ran two ads, spending less than one dollar, with criticism to the polarization between two parties in the US [21,22]. It is important to note that this analysis included only posts that were boosted through Facebook Ads Platform. Many other posts may have reached thousands of people with no cost, being shared by the followers of the page.

[14] http://www.socially-divisive-ads.dcc.ufmg.br/app.php?query=450
[15] http://www.socially-divisive-ads.dcc.ufmg.br/app.php?query=588
[16] http://www.socially-divisive-ads.dcc.ufmg.br/app.php?query=590
[17] http://www.socially-divisive-ads.dcc.ufmg.br/app.php?query=506
[18] http://www.socially-divisive-ads.dcc.ufmg.br/app.php?query=437
[19] http://www.socially-divisive-ads.dcc.ufmg.br/app.php?query=3145
[20] http://www.socially-divisive-ads.dcc.ufmg.br/app.php?query=2815
[21] http://www.socially-divisive-ads.dcc.ufmg.br/app.php?query=3167
[22] http://www.socially-divisive-ads.dcc.ufmg.br/app.php?query=3117

Clicks	
African-American history	39.21 %
African-American Civil Rights Movement (1954-68)	39.13 %
Black (Color)	26.86 %
Pan-Africanism	25.65 %
African-American culture	20.58 %
La Raza	19.78 %
Chicano Movement	19.78 %
Chicano rap	19.41 %
Mexico	19.03 %
Lowrider	19.03 %
Latin hip hop	19.03 %
Hispanidad	19.03 %
Malcolm X	16.47 %
Black Consciousness Movement	16.39 %
Martin Luther King III	15.67 %
Black nationalism	15.63 %
Stop Police Brutality	15.61 %
Police misconduct	15.23 %
Martin Luther King, Jr.	14.76 %
Mexican american culture	8.43 %

Cost (USD)	
Behaviors:African American (US)	12.83 %
Patriotism	12.65 %
Independence	12.44 %
Malcolm X	11.97 %
African-American history	9.36 %
African-American Civil Rights Movement (1954-68)	9.29 %
Martin Luther King, Jr.	7.73 %
Cop Block	7.27 %
LGBT community	5.54 %
Human rights	5.53 %
BlackNews.com	5.08 %
Homosexuality	4.90 %
HuffPost Black Voices	4.81 %
Same-sex marriage	4.27 %
LGBT culture	4.08 %
Gay pride	4.08 %
Black (Color)	4.01 %
Gun Owners of America	3.95 %
Jesus	3.73 %
Pan-Africanism	3.41 %

Table 6.9: **Top 20 Interests per clicks, and cost (USD). The percentage represents the ratio between the aggregated sum for each interest and the total sum for each metric.**

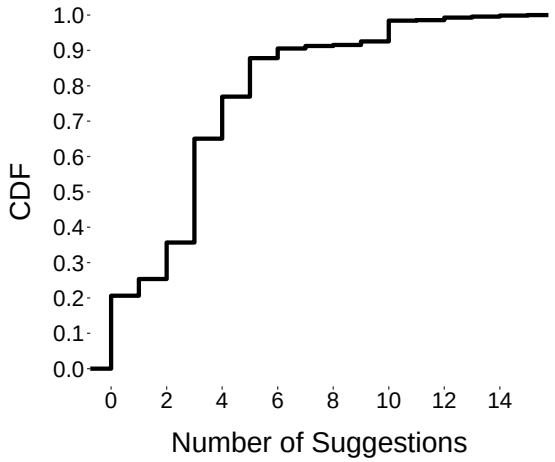


Figure 6.10: **Cumulative Distribution Function (CDF) for the number of suggestions.**

6.3.2 The role of attribute suggestions

Facebook provides a tool for advertisers that, given a target attribute, it presents a list of other attributes that target people with similar demographic aspects [Speicher et al., 2018]. For example, in the list of suggested targeting interests for 'Townhall.com', a page with an audience in which 79.5% of the users are very conservative users according to Facebook, there are other pages with similar bias towards very conservative users, i.e. 'The Daily Caller' (67.1%), 'RedState' (84.3%), and 'TheBlaze' (59.6%) [Ribeiro et al., 2018].

In order to investigate if the IRA ads have used suggestions to elaborate complex targeting formulas, we crawled the attribute suggestions for each attribute that appear in the

dataset of highly impact ads. Figure 6.10 shows the cumulative distribution function for the number of suggested attributes that appear in the same formula. We can see that around 64% of the ads that potentially used this feature because they have at least three target attributes suggested by Facebook as part of the same formula. There are 1.2% of ads with more than 10 suggested attributes in the same formula. As an example, all the 13 interests, including Islam, Ramadan, Islamism, used in the target formula of the ad ID 1915[23] appear as suggestions for at least one of the others in the formula. For ad ID 1840[24], we were able to find 9 out of 10 of the interests using the interest suggestion feature. This provides evidence that this feature may have been a key element used by the IRA campaign to choose the target audience.

6.4 Analyzing the targeted audience

In this section we intend to identify the demographics of users that where targeted by the ads. In order to do that we considered the attributes used in each ad as the Original Targeting Formula (OTF) in our framework. For our analysis, we considered seven demographic categories: political leaning, race, gender, education level, income, location (in terms of states), and age. As a baseline for comparison, we also gathered the demographic distribution of the United States Facebook population.

Only 11% of the attributes that appear in the IRA ads targeting formulas are not available for targeting anymore due to changes in the Facebook Marketing API. In most of these cases, we reproduced the ad target formula without the missing attribute, especially when the attribute looks redundant with the others in the formula. We were only unable to reproduce 6 targeting formulas.

6.4.1 Measuring audience bias

To assess the audience bias of each of the demographic aspects that we considered, we computed the differences between the fraction of the population with a demographic aspect and the same fraction of the population in the baseline distribution (i.e., the U.S. Facebook population), namely the *bias score*. For instance, if the percentage of African-Americans in the audience of a particular ad is 40%, the *bias score* for this dimension in the ad is 0.25 as the percentage of African-American in the U.S. Facebook population is nearly 15.5% $(0.4 - 0.155)$.

[23] http://www.socially-divisive-ads.dcc.ufmg.br/app.php?query=1915
[24] http://www.socially-divisive-ads.dcc.ufmg.br/app.php?query=1840

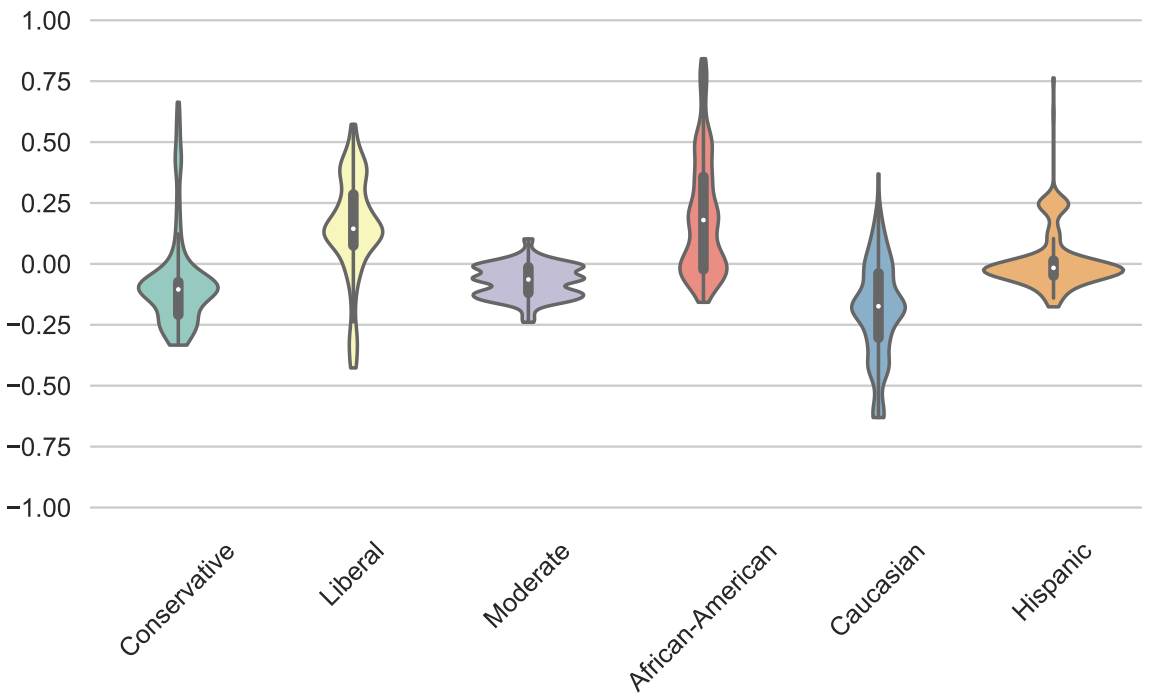


Figure 6.11: **Bias in demographic dimensions. Each violin represents the bias score for all high impact ads in a particular demographic dimension. The median is represented by a white dot in the center line of the violin graph. 50% of the data is present between the two thick lines around the center.**

Figure 6.11 depicts the distribution of the measured bias on political leaning and ethnic affinity. In comparison with all the demographic categories, these two showed to be the ones with the highest biases. We note that most of the ads target audiences that are more biased towards the African-Americans population and the Liberals. More specifically, about 70% of the IRA ads target an audience with a higher proportion of African-Americans than in the US Facebook distribution. This difference is even more accentuated for Liberals, with 82% more biased in comparison with the reference distribution. The percentage of ads with bias score superior to 0.15 is 52% for African-American and 41% for Liberals. Our dataset suggests the presence of those ads that target extremely biased populations of conservatives, Liberals, Hispanic, and especially African-Americans. The target audiences for the IRA ads are slightly biased towards women and young adults (18-34 years), which are omitted from Figure 6.11 due to space constraints.

6.4.2 Targeting audience and divisiveness

Next, we investigate if the advertisers target the ads towards audiences that are less likely to identify their inappropriateness due to their ideological perception bias. Additionally, we

Group	Report	Approval	False Claims
Liberals	-0.17***	0.41***	-
Conservatives	-0.15***	0.32***	-

Table 6.10: **Pearson's r correlation between targeting and the ideological divisiveness for the high impact ads (*** $p < 0.001$, correlation revealed no statistical significance in the case of false claims).**

examine if the ads directed to biased audiences could leverage the already existing societal divisiveness to further amplify it among the masses.

To understand these nuances of targeted advertising, in this section we focus on the relationship between the targeted population and the ideological divisiveness in reporting, approval, and false claim identifying behaviors for the ads. Table 6.10 reports the correlation values between the targeted population and the tendency of the population to report, approve, and identify false claims.

Reporting. We observe a negative correlation in the case of reporting for both Liberals and Conservatives (also see Figure 6.12 (a)). This suggests that the targeted population has a lower tendency to report than the non-targeted one. This is also evident in Figure 6.12 (b), where we find that the reporting by the targeted population carries way lower likelihood than the reporting by the overall (or non-targeted) population.

Approval. We observe a positive correlation in the case of approval for both Liberals and Conservatives (also see Figure 6.12 (c)). This suggests that the targeted population has a greater tendency to approve the ads as compared to the non-targeted population. This is also evident in Figure 6.12 (d), where we find that the approval score by the targeted population carries a greater score for most of the ads compared to the overall (or non-targeted) population.

False claims. For false claims, we do not find any significant correlation between the targeted population and divisiveness. However, in Figure 6.12 (e&f) we do find that the targeted population has a lower tendency to identify false claims.

Taken together, we can assume that the ads were "well-targeted" in a way towards that population which was more likely to believe and to approve then, and subsequently less likely to report or identify false claims in them.

6.5 Summary

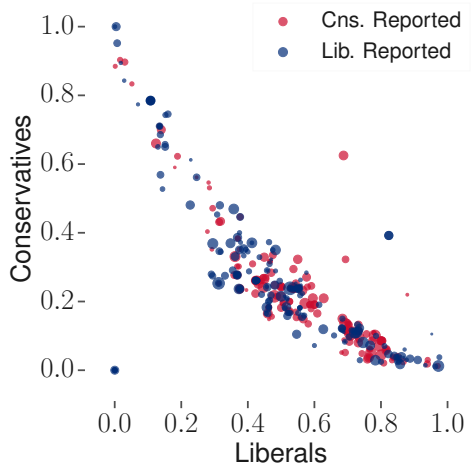
In this case study, we presented an in-depth study of the ads published before and during the 2016 US election. Firstly, we characterized the ads in the IRA dataset. Our analysis

highlights the landing pages that paid for the ads and identifies the most successful ads in terms of impressions and clicks. We find that the ad campaigns were intensified near to the U.S. election period. Among our main findings, we show that the typical CTR for these ads is an order of magnitude higher than typical values for Facebook, meaning that these ads were very effective.

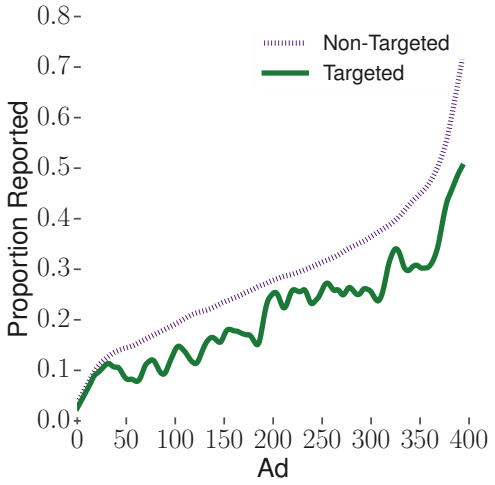
After this initial analysis, we focused on peoples' perceptions of the content of the 485 IRA ads we identified as high impact. To assess these perceptions along three axes – likelihood of being reported, approval and disapproval, and the presence of false claims – we conducted three U.S. census-representative surveys. Our analysis of the perceptions queried in these surveys shows that ideological opinions of individuals influence their perceptions of these ads. We find that many of these ads were severely divisive, and generated strongly varied opinions across the two ideological groups of liberals and conservatives (see Figure 6.6, 6.8, 6.9).

We then showed that the vast majority of the IRA ads use attribute-based targeting, containing complex target formula that includes interest and behavioral attributes that were likely suggested by Facebook. The combination of attributes in IRA ads reached audiences that are very biased towards African-Americans and liberals. More important, we concluded that ads were overall targeted towards a population that is more likely to believe, and approve and subsequently less likely to report or identify false claims in them.

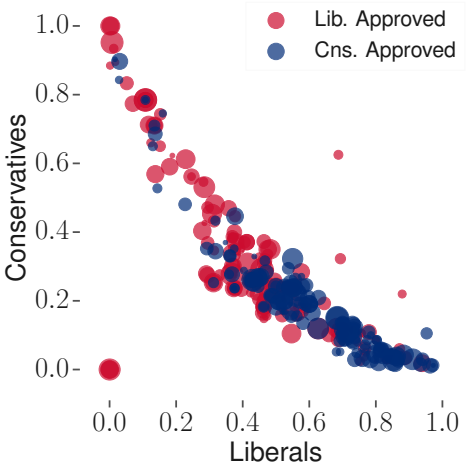
As a final contribution, we have deployed a system (available at *http://www.socially-divisive-ads.dcc.ufmg.br/*) that displays the ads and their computed information such as the demographics of their targeting audiences.



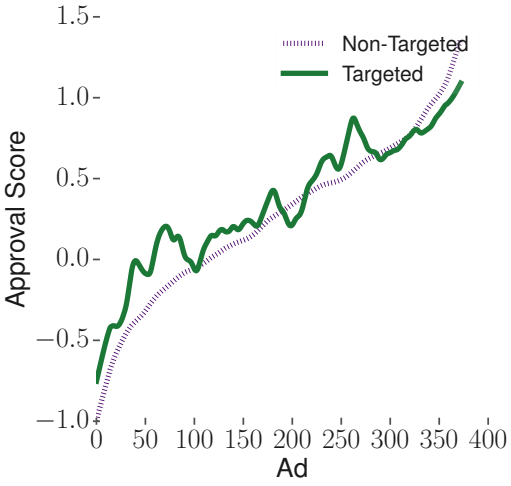
(a) Reporting: Targeted



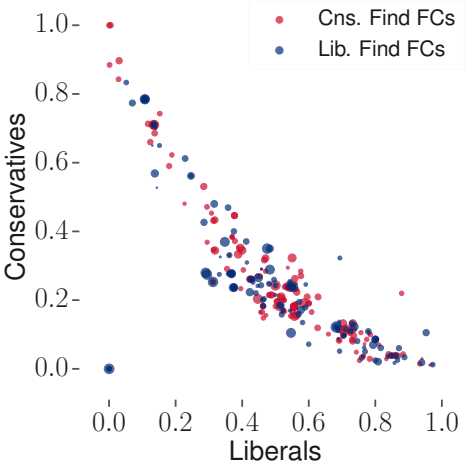
(b) Reporting: Bias



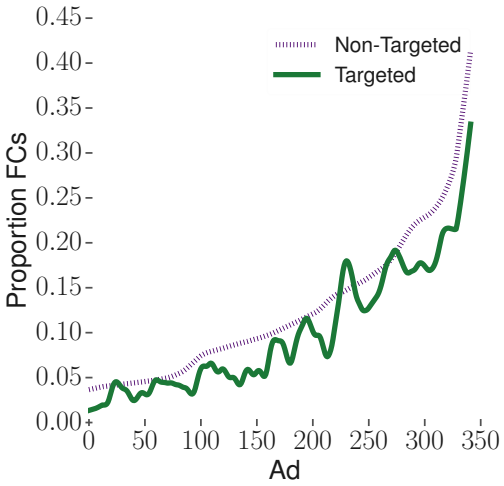
(c) Approval: Targeted



(d) Approval: Bias



(e) False Claims: Targeted



(f) False Claims: Bias

Figure 6.12: **Relationship between targeting and the responses by ideological groups. (a,c,e) show the proportion of population targeted and their tendency of response. Each circle represents an ad, and their size is proportionate with the between group disputability for that ad. (b,d,f) compares the mean responses of the targeted ads with their hypothetical non-targeted counterpart (i.e., overall responses), where each ad is represented on the x-axis**

Chapter 7

Case Study: Inferring Demographics of Election Polls

In our last case study, we employ our framework in the elections scenario. Particularly, we aim at inferring the audience demographics of politicians in the electoral race, in a similar way to election polls.

The polls surveyed during elections provide valuable insights about voting intention on candidates filtered by demographic characteristics. This information can be used to understand the dynamics of underlying electoral preferences and how it changes in the days or months during the campaign [Kenett et al., 2018]. Despite their importance, election polls are time and money consuming on many occasions, especially those with face-to-face surveys, that interview a representative share of the voter population across the entire country.

On the other hand, data extracted from OSN has been explored by researchers to shed light on the electoral context since many users actively participate in public debates and expose their opinions publicly, including on political topics. It is widely known that the population on OSN is biased towards young people and the sample is not well represented towards all demographic categories [Morstatter et al., 2013]. However, monitoring data from this niche of people may be highly elucidating since people influence and are influenced by discussions and debates on OSN, which may eventually impact the elections results.

In this chapter, we leverage our framework to calculate the demographic characteristics of the audience of candidates on Facebook. Then, we compared the online extracted data with election polls provided by major poll institutes in Brazil and taken in the same period. Our findings suggest that the candidate's popularity on Facebook, captured in terms of the number of likes, people talking about him/her, and the number of people interested in the candidate, is a good indicator of his/her variation on vote intention polls. Additionally, we figured out that the fluctuation in the demographic aspects of supporters detected by election

polls are captured by our methodology, obtaining more precision with popular candidates. In particular, we show that sharp variations caused by high impact events during the campaign, such as protests, are well captured by Facebook measurements. Finally, we present a system that exposes the audience demographics of Brazilian politicians on Facebook (available at *http://www.audiencia-dos-politicos.dcc.ufmg.br/*) and contributes to the understanding of the national political scene in Brazil.

7.1 Methodology

In our intent to compare the fluctuation of the candidates' audience in the OSN with the vote intention of candidates in the weeks before the election as estimated by election polls, we collect the distribution of two distinct demographic attributes: geographic region, gender. As a basis for comparison with our data, we gather election polls from two popular opinion public survey institutes, namely Datafolha[1] and IBOPE[2], both with results largely used in the mainstream media as part of the news coverage in Brazil. The election polls provided by these institutes stratify the vote intentions for each candidate into demographic groups, providing us similar data to that one we collect with our framework.

An important characteristic of Brazilian elections is the definition of the winner. In Brazil, a candidate to the executive branch gets elected after obtaining the majority of the votes *i.e.,* 50% of total valid votes plus one (blank and null votes are discarded). As a consequence, if none of the candidates reach this amount in the first round of elections, the running election is defined in a second round, that includes only the top 2 candidates.

The 2018 Brazilian presidential elections took place on October 7 (first round) and October 28 (second round), with 13 candidates, out of which we select the five better-positioned candidates in the initial voting intention polls: Ciro Gomes, Fernando Haddad, Geraldo Alckmin, Marina Silva and Jair Bolsonaro. We should notice that the candidate Ciro Gomes, who kept the third position in the polls during almost all the period before the first round had no attribute on Facebook advertising platform related to him in the first demographic collection on November 2017.

In order to provide an overview of the evolution in the audience demographics of candidates on Facebook, we performed a first collection in November 2017, considering the non-official list for the position of President. In July and August 2018 we ran the crawler on the second Tuesday of the month. From September, we started weekly collections every Monday until the day after the second turn. We also collected the demographic details of

[1] http://datafolha.folha.uol.com.br/
[2] http://www.ibope.com.br

candidates on the Fridays before both rounds (elections always take place on Sundays in Brazil).

Our baseline was built by tracking all election polls published by both institutes in the four weeks that preceded the election polls. We also included some polls published near the dates of our collections on months before elections. We should emphasize that we had no access to the schedule of election polls by institutes. Therefore, Facebook collections are not exactly on the same day as the polls. To tackle this issue, we present the results using the dates of our Facebook collections as a base and group the election polls with the nearest Facebook collection date. Except for the two first dates, the baseline values were published at most four days distant from online extracted data. In the oldest comparisons the distance was close to twenty days, which does not represent a great impact in the analysis since, due to the election remoteness, the variation in that time was not significant.

7.2 Brazilian elections in Facebook numbers

Before presenting the polls measured with our approach we need to (i) contextualize the political scenario in Brazil, (ii) understand the main noteworthy events that occurred along presidential campaign that might have had an impact in voters decision, and (iii) discuss the involvement and penetration of candidates on the social media, especially on Facebook. This section is dedicated to covering these three topics.

The 2018 presidential election in Brazil was surrounded by controversial and heated discussions suggesting manipulation from mainstream media, persecution to candidates and even an attempt on the life of a candidate. The dates of these events were included in the validation figures to make its impact clear. First of all, the former left-wing president Lula, one of the most popular presidents in the short democracy history of Brazil, was arrested after being found guilty of corruption by the court of appeal. Following the Brazilian law, he was judged and prevented from running in the presidential elections, in spite of being the best-positioned candidate in the election polls at that moment. The chosen candidate to replace Lula in the elections was Fernando Haddad, a former mayor of the largest city in Brazil, but mostly unknown by the Brazilian population in regions as north and northeast. The expectation of the Worker's party (Lula's party) was that the political capital of the arrested president would be transmitted to Fernando Haddad.

On the other hand, the right-wing candidate Jair Bolsonaro presented himself as the alternative to the corrupt political system in Brazil. Controversial and with inflamed speeches he expanded his supporting network after intensifying his presence on Online Social Networks and with organized groups to spread supportive messages [Resende et al., 2019]. Ac-

cused to be a misogynist, he was the main target of a series of protests on September 29, that emerged along the country. These protests named as 'EleNao' (not him), were proposed and coordinated by women unsatisfied with his increasing popularity. The next days were followed by generalized protests as a response to 'EleNao', organized by Bolsonaro supporters. Another wave of protests took place in many cities around the country in the weekend before the election. On October 20, thousands of people against Bolsonaro took the streets to protest. Again, his supporters also went to the streets to support him on the next day. In the end, Jair Bolsonaro won the 2018 presidential election.

The other three candidates that complete the top 5 list in the 2018 presidential elections are the following. Marina Silva has been running in the last three presidential elections (since 2010), and in all cases figured in the election polls as a prominent candidate, many times appearing in the first position according to voting intentions. However, when elections approach her voters seems to migrate to other candidates.

Ciro Gomes seemed to inherit the voters from former candidate Lula in the first polls, however he ended up in the third position and was not qualified to the second round. Finally, Geraldo Alckmin, former governor of the most populous state in Brazil, was expected to have expressive participation in the 2018 elections since his party made alliances with many parties and obtained the largest portion of the election advertising time on TV and Radio. This advertising time presented on TV and Radio programming is free of charge and guaranteed by law in Brazil in the weeks that precede elections and represented over the last elections an important resource to the winning candidates.

Next, we describe how some Facebook-derived metrics are correlated with the popularity fluctuation of candidates in the elections. Figure 7.1 shows the number of likes by each one of the top-five candidates and Lula. Notice that the election winner, Bolsonaro, had a large increase in the number of likes in its public profile, raising from 4.8 million to 8.2 million. Apparently, his popularity on Facebook started a sharp increase after he was stabbed on September 6.

Similarly, the public profiles of Ciro Gomes and Fernando Haddad presented a similar behavior in terms of likes, becoming much more popular as the campaign evolved. Ciro Gomes had about 250 thousand likes in its public page on November 2017 and ended the election with almost 800 thousand likes, whereas the growth of Haddad followers went from 330 thousand to 1.7 million. Notice that after the end of the first round, in which Haddad and Bolsonaro were qualified to the second round, Haddad left Ciro behind. This may be explained by the union of Ciro followers to Haddad's campaign, both with similar ideological alignments, aiming at defeating the right-wing candidate.

Conversely, Geraldo Alckmin and Marina Silva were, by far, more popular than Gomes and Haddad before the election campaign, began with 880 thousand and 2.3 million likes,

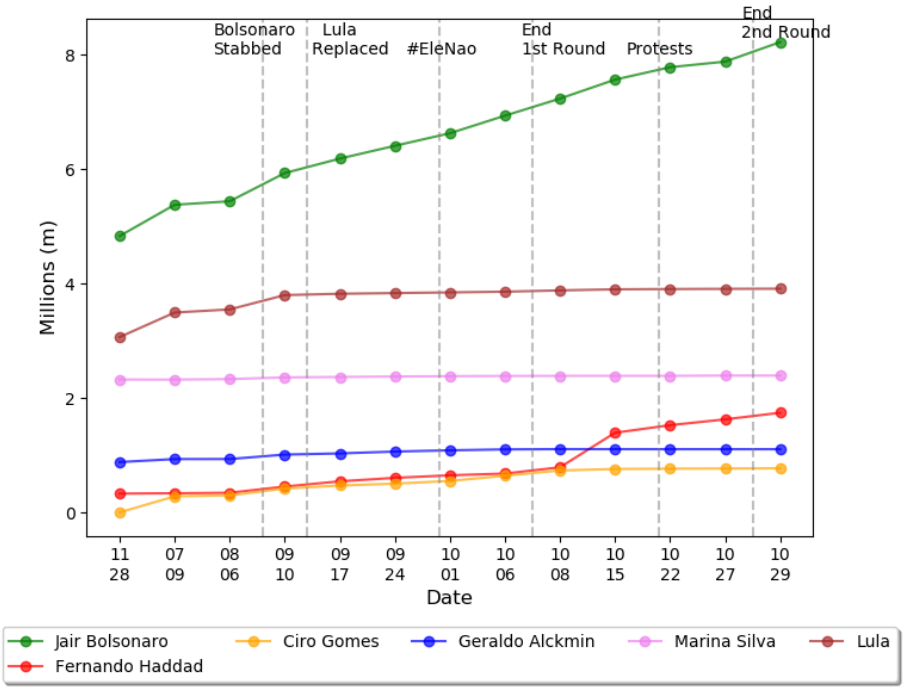


Figure 7.1: **Number of likes per candidate**

respectively. Whereas Marina's public profile attracted only 70 thousand new likes, the number of likes of Geraldo Alckmin increased only 25%, reaching 1.1 million likes. Finally, we can see that Lula's likes raised right after his trial, stabilizing as the campaign proceeded.

These findings indicate a correlation between the increase in the number of likes on Facebook and the rise in voting intentions. Figure 7.2 presents the variation for the top-five candidates in terms of vote intention according to IBOPE election polls. Notice that Bolsonaro, Haddad, and Gomes have their voting intentions variation correlated with the increase in likes. After calculating the Spearman correlation we found high values for them, 0.97, 0.98, 0.80, respectively. For Silva and Alckmin the correlation was negative (-0.48 and -0.18).

Other metrics provided by Facebook may clarify the relationship between the popularity on Facebook and the vote intentions. Figure 7.3 (a) depicts the variation of the number of people talking about candidates, that express the number of people interacting with the Facebook page (by commenting, sharing the published content, mentioning the page, etc). Figure 7.3 (b) depicts the total number of people interested in each candidate, as calculated by the Facebook Marketing API (described in the methodology). Notice that, the number of people interested in a candidate is sometimes much larger than the number of likes. The number of people interested in Bolsonaro was 15 million by the end of the election. This may be explained because Facebook calculates all people interested in Jair Bolsonaro, which may include people who post something about him as well as other Facebook pages related

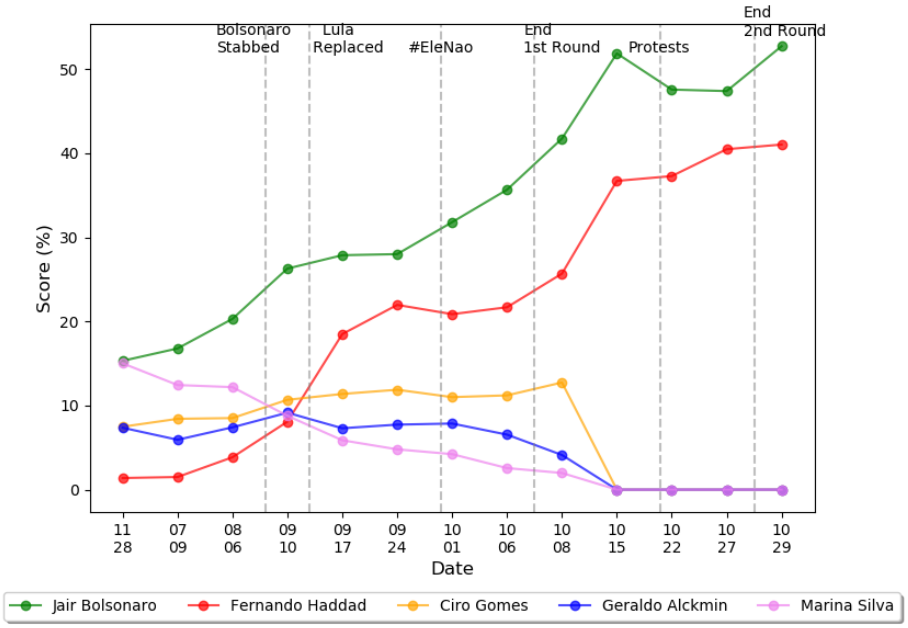
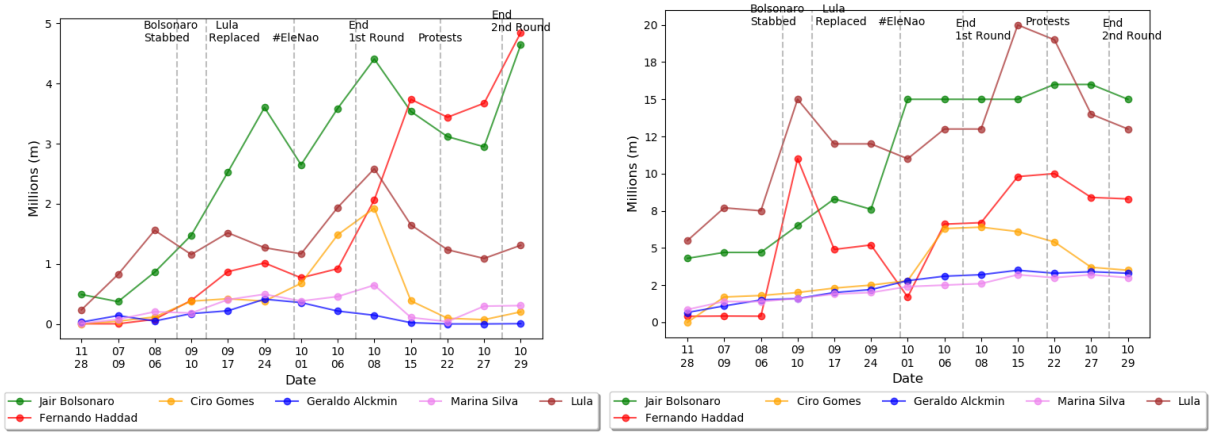


Figure 7.2: **Voting intentions per candidate (IBOPE)**



(a) Talking About

(b) People interested

Figure 7.3: **Number of people 'talking about' and total interest per candidate.**

to him. Especially in the election period, several supporting pages got popular, such as *Bolsonaro Opressor* (meaning Oppressive Bolsonaro, in English).

By analyzing these two metrics we can clearly see a distinction in the popularity of Bolsonaro, Haddad, and Gomes in comparison with Alckmin and Silva. The most voted candidates attracted much more attention on Facebook than less voted candidates. In particular, we can notice that the number of people talking about Bolsonaro, Haddad, and Gomes rose consistently after the 'EleNao' protests and the measurement just after the first round. After that, Gomes, unqualified to the second round, lost audience whereas Bolsonaro and Haddad kept the high number of people talking about them. The picture of the number of

people interested shows very similar behavior for Haddad and Lula audiences, whose lines follow the same pattern all over the period. In fact, most of the posts from widely known former candidate Lula made reference to Fernando Haddad. This clearly indicates that the replaced candidate inherited part of the political capital from Lula.

It is important to mention that the total number of likes alone does not represent a good source of information in these cases. In spite of the second position in the number of likes among candidates, with slightly more than 2 million likes, Marina Silva reached less than 1.1 million votes in the election. In a high-level analysis, the act of liking a page indicates that, at some point, the user liked the content or found that page aligned with his interests. On the other hand, the sharp increase of likes in a short period represents a rise in popularity and may be a stronger indication of a rise in voting intentions.

7.3 Comparing demographics from Facebook with election polls

After showing that the increase in candidate popularity on Facebook indicates a tendency of rising in the voting intentions polls, we turn now to the demographics aspects. In particular, we intend to check the extent to which the variation in the demographics of the audiences politicians on Facebook reflects the demographics of the election polls. In particular, we focus on gender and geographic (state level) aspects.

7.3.1 Gender

We begin our analysis by exploring the variation in the gender distribution of candidates audience. Figure 7.4 depicts the gender variation by the top three candidates. In a high-level analysis, the variation of the gender distribution appears to follow a similar trend between the Facebook data and election polls results. Notice that the proportion of women in Bolsonaro's audience was much smaller than the percentage of men in the first measurements, 72% against 28% according to Facebook. Nonetheless, the presence of women increased as the campaign evolved and remarkable events occurred. The stabbing of the candidate was clearly linked to an initial rise in the overall popularity of the candidate, and as a consequence, the percentage of women increased considerably in election polls as well as on Facebook.

More prominent peaks are detected soon after 'EleNao' protests and end of campaign protests, when the female presence in the winning candidate's audience on Facebook reaches 52% and 50%, respectively. The variation of the percentage on these specific dates when

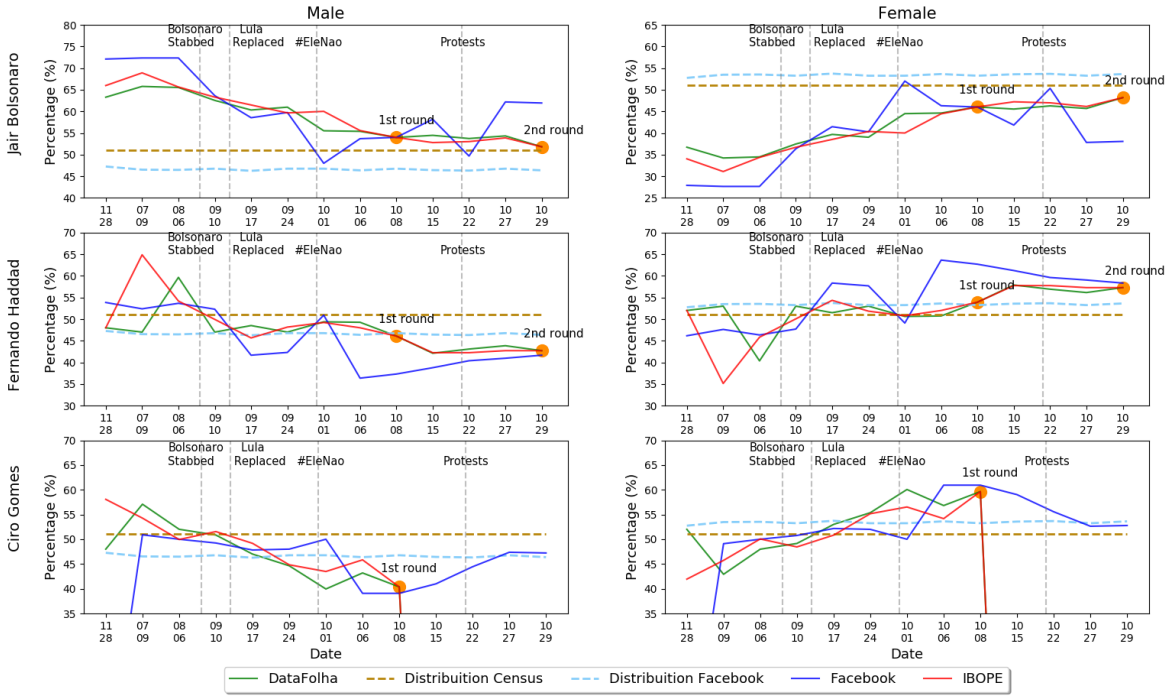


Figure 7.4: **Interest in candidates by gender (Top 3).**

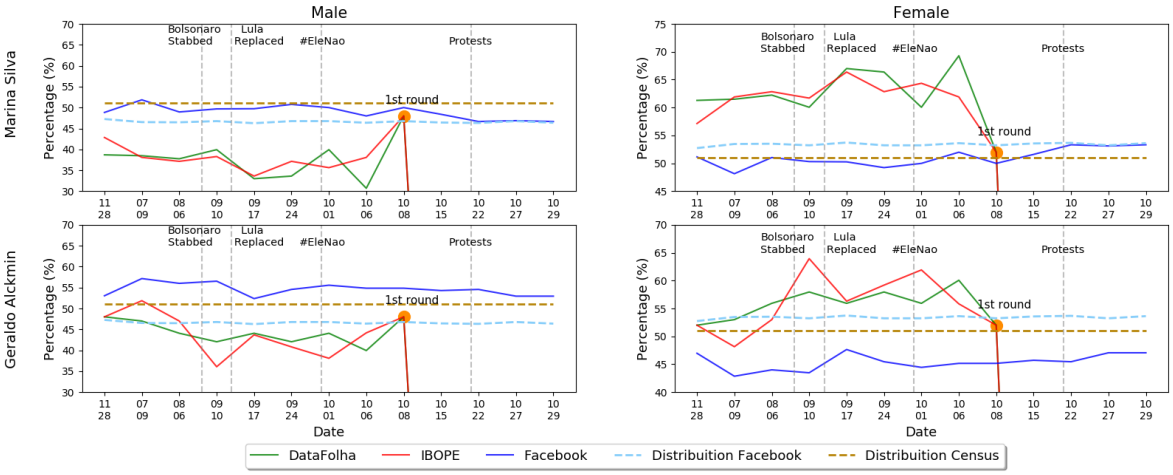


Figure 7.5: **Interest in Marina Silva and Geraldo Alckmin by gender.**

compared with the measure taken one week later was 12% and 8%, respectively. The protests also appeared to influence the percentage of women in the audience of Fernando Haddad and Ciro Gomes. In both cases, the women participation increased indicating that protests, in fact, impacted the overall scenario. This trend is detected in election polls and Facebook data. The gender variation of Geraldo Alckmin and Marina Silva also presented a similar trend, however, they are characterized by soft changes in contrast with other candidates. As a final observation, the official replacement of candidate Lula by Haddad was followed by a peak of women in Haddad's supporters.

Notice that Facebook data presents some abrupt changes in the proportion of gender when compared with the polls. This sharp increase detected by our methodology may be explained by the huge rise in the number of people interested in Bolsonaro as depicted in Figure 7.3 b), from 7.6 million to 15 million in the period of protests. As detailed before, the people interested in a particular entity (the candidate in this case) considers not only people who liked the entity's page, but also people who interact with the page, that mentions the candidate and other types of interactions. Bolsonaro was the center of discussions in the days before the 'EleNao' protests which probably attracted more interactions from women and in the end caused the abrupt change in its audience. The elections polls published after the protests confirmed that the popularity of Bolsonaro increased among women, however, Facebook seems to oversize this changes, which may be explained due to the highly polarized environment Social Networks become during this election in Brazil. The same occurs after the replacement of candidate Lula by Haddad where the total interest and the number of people talking about Lula's pupil sharply increased. On the other hand, the lower variations detected in the audience of Marina Silva and Geraldo Alckmin may be explained by the lower engagement as expressed by talking about and interest metrics (see figure 7.3).

Figure 7.4 also depicts the percentage of each demographic dimension (male, female) in the first and second turns highlighted with orange dots. These values were extracted from the election exit poll, carried out in the election day. By analyzing the orange dots we can see, for instance, that the exit poll percentage of men and women in the audience of Ciro Gomes were closer to the Facebook measurement than to the election poll value, in other words, the error of Facebook percentages was lower than election polls values for Ciro Gomes. In order to expand this error analysis, we used the exit poll values to calculate the gender error rate (e) by a candidate. We employed the following formula, where i is each one of the demographic dimensions, (pE) is the percentage as calculated by the exit poll (our baseline), and (pV) is the percentage obtained with Facebook or with the election poll that preceded the election.

$$e = \sum_{i=1}^{n} \frac{pVi - pEi}{pEi}$$

In a high level, the error represents the distance from the last measurement taken before the election day (first and second turns) to the exit poll value, including male and female percentages. Table 7.1 shows the error rates for each candidate and measurement in both election rounds. We can notice that the error from the Facebook measure is smaller than the other polls for Bolsonaro, Gomes, and Silva in the first round. In a general analysis, Facebook obtained lower error rates in three out of seven cases whereas IBOPE gets lower errors for the remaining four measurements.

Table 7.1: **Error rate (%) by gender.**

First Round	Facebook	IBOPE	DataFolha
Jair Bolsonaro	0.06%	3.21%	2.85%
Fernando Haddad	17.65%	3.86%	6.41%
Ciro Gomes	2.71%	11.35%	5.85%
Marina Silva	4.01%	19,84%	34.62%
Geraldo Alckmin	13.7%	7.67%	16.14%
Second Round	Facebook	IBOPE	DataFolha
Jair Bolsonaro	20.75%	4.12%	5.03%
Fernando Haddad	3.58%	0%	2.33%

7.3.2 Region

The analysis by region provides similar findings to the previous analysis. The variations seem to follow a similar tendency. Figure 7.6 depicts the variation of the audience by geographic regions for Jair Bolsonaro whilst figure 7.7 shows the results for Haddad. We grouped the values from mid-west with the north region as Datafolha, one of the survey agencies, employs this approach.

One of the clearest variations of supporters by regions suggests the transference of political capital from Lula to Haddad. As shown in figure 7.7, the percentage of Haddad supporters in southeast regions drops drastically from September 9 and sharply increase in northeast and north/mid-west. It matches perfectly the judgment of Lula that prevented him from running for the presidential election (August 30). The candidate Fernando Haddad officially became Lula's substitute on September 10. Notice that the same trend of decreasing in the southeast and increasing in other regions was detected by all measurements.

We also calculate the error in the election results for regions category by employing the same formula used on gender analysis, but in this case including 4 different demographic dimensions, north/center-west, northeast, south, and southeast. However, in this case, we used the election final result as the baseline. Different from the gender characteristics of voters that are not exposed in the elections due to anonymous voting, the number of votes per state is released, then there was no reason to use exit polls since the real values are public. The values calculated with our methodology have the lowest error rate for Marina Silva and a very close result to the lower for Ciro Gomes, both in the first round. IBOPE error was lower in three cases whereas Datafolha won in two cases.

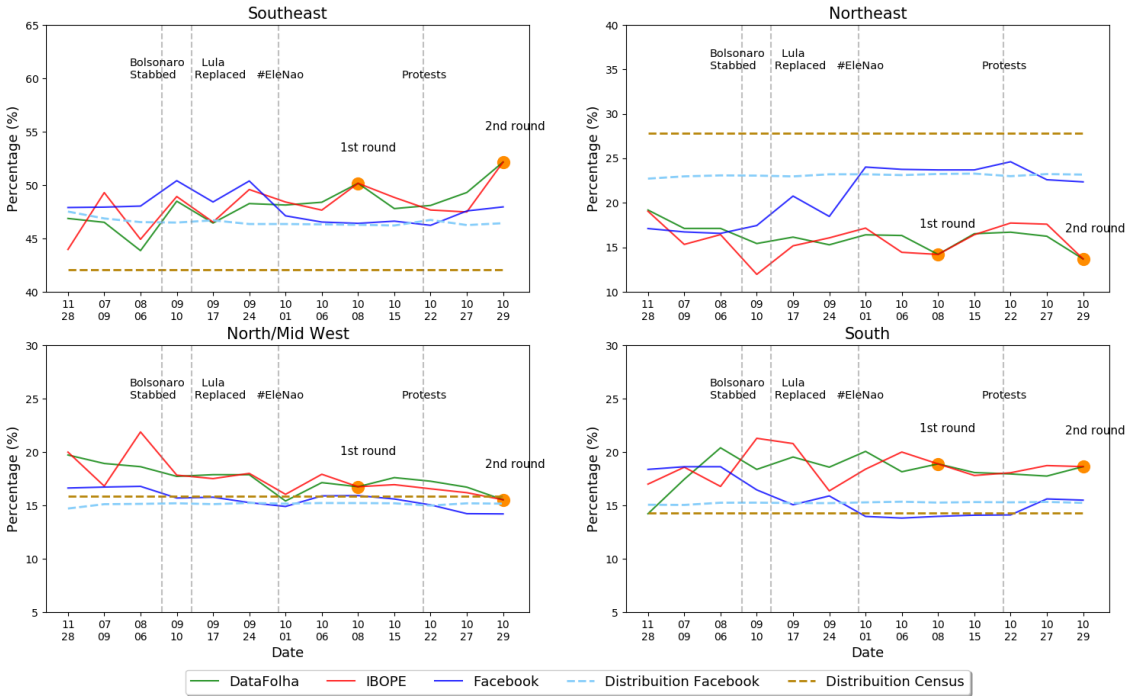


Figure 7.6: **Distribution by region - Jair Bolsonaro.**

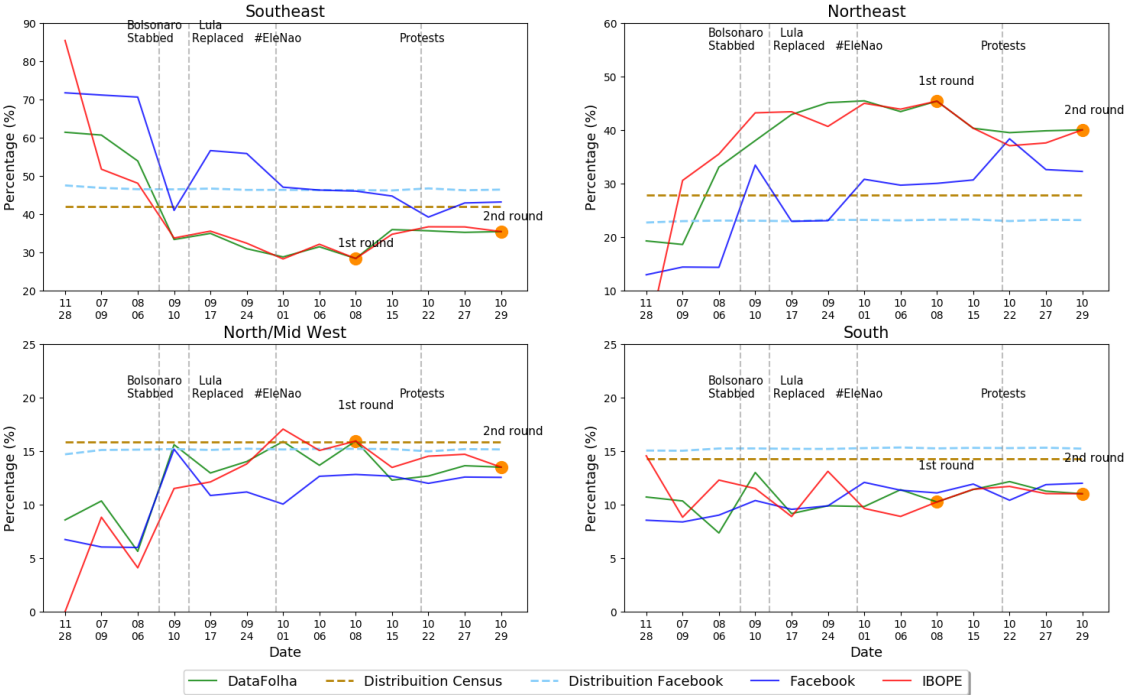


Figure 7.7: **Distribution by region - Fernando Haddad.**

7.4 Interactive tool

In order to provide a useful service as an outcome of this case study, we deployed a real system that computes the demographics of the politicians' audience according to our method-

Table 7.2: **Error rate(%) by region.**

First Round	Facebook	IBOPE	DataFolha
Jair Bolsonaro	26.33%	4.88%	6.18%
Fernando Haddad	30.69%	8.76%	10.14%
Ciro Gomes	8.5%	14.03%	8.47%
Marina Silva	8.26%	39.68%	14.6%
Geraldo Alckmin	19.0%	9.93%	11.22%
Second Round	Facebook	IBOPE	DataFolha
Jair Bolsonaro	24.6%	10.58%	9.11%
Fernando Haddad	13.57%	4.64%	1.06%

ology (*available at* `http://www.audiencia-dos-politicos.dcc.ufmg.br/`). This system was deployed months before the Brazilian election and contributed to a better understanding of the national political scene in Brazil as it attempted to elucidate demographic aspects of the supporters of each candidate. In addition to presenting the demographic distribution of candidates audience on Facebook, it also allows one to search for the most biased audiences through sorting the candidates by specific demographic dimension (gender, region, etc). We can realize, for instance, that 61% of the audience the candidate Jair Bolsonaro is male[3] and 37% of Alvaro Dias' audience are married against only 23% of Haddad's audience[4]. These features have attracted a lot of press coverage to our work [5], which we believe was one of the main reasons why Facebook added more interests related to the Brazilian politicians in its advertising platform. For example, Ciro Gomes was missing in the Facebook advertising platform, meaning that one could not target those interested in this candidate with political ads. So, after a few news reports have questioned publicly Facebook if that would be fair, Ciro Gomes was included as an interest to be used by targeting advertising[6].

7.5 Summary

In this case study, we leverage our framework to infer the demographics candidates' audience in the Brazilian Presidential election according to Facebook. Among the main findings, we demonstrate that the popularity of the candidates on Facebook, captured with three different

[3]https://bit.ly/2De2Xa1

[4]https://bit.ly/2UIRTIV

[5]https://www.bbc.com/portuguese/salasocial-43824463

[6]https://www.otempo.com.br/capa/pol%C3%ADtica/portal-elei%C3%A7%C3%B5es-sem-fake-desenvolve-ferramentas-para-fiscalizar-campanhas-1.1652546

metrics (number of likes, people talking about, and number of people interested in the candidate), may be a good indicator to detect his/her variation on vote intention polls. However, the number of likes alone can be misleading as in the case of a candidate with more likes on Facebook than real votes.

We also verified that the demographic fluctuation for gender and geographic regions in the audience of candidates on Facebook follows similar trends detected by the election polls, with the variations presenting analogous shape. In particular, we show that abrupt shifts occasioned by relevant events during the campaign such as protests and candidates prevented to run in the elections are well captured by Facebook measurements. However, peaks seems to be oversized in the Facebook perspective, which may be explained by the highly polarized and tense environment OSN became during the campaign.

Finally, we deployed a system that exposes the audience demographics of the Brazilian politicians on Facebook, available at *http://www.audiencia-dos-politicos.dcc.ufmg.br/*. We believe that systems such as ours are not only useful for the social media users, but also for journalists, as well as for social media researchers wanting to understand the election ecosystem on Online Social Networks. More important, our system provided us with a way to validate the utility of our system, which was intensively used by the Brazilian journalists to discuss the online audience of each politician. As future work, we aim at expanding our system to other countries, particularly those with upcoming elections.

Chapter 8

Conclusion

In this thesis, we presented a methodology that leverages OSN advertising platforms to extract demographic data. Our methodology explores the targeting formula, *i.e.,* the set of attributes used to define the audience to which an advertiser wants to deliver the advertisement. We include a new layer in the original targeting formula (OTF) that allowed us to calculate the distribution for eight demographic attributes by using the total audience calculated for each particular set of attributes before an ad is run and any cost is incurred. This innovative methodology is intended to unveil the demographics related to a large set of entities such as a country or other geographic location (state, city, region), a person (celebrity, politician or other public figure), a media outlet (newspaper, magazine, blog, TV Channel) or even the demographic distribution of the audience a particular ad targeted.

In order to automate the requests and demographics gathering, we developed a framework that explores the Facebook advertising platform through its Marketing API, a set of functions to automatically manage ads on Facebook, that is freely available, but with several rate limiting constraints. To deal with this, we employed a distributed approach in which a central node splits the requests into up to 50 remote machines and, after checking the end of the collections, calculates the demographics for the selected entities. We then used our framework to conduct four case studies that took advantage of demographics extracted from the online environment to provide transparency for internet users in different scenarios.

In the first case study, we used the framework to collect data that approximates the U.S Census aiming at validating our new methodology. We verified that the distributions of race and political leaning, in particular, are very similar to the distribution as calculated by official surveys in all granularity levels. Conversely, the education level obtained online seems to be oversized for the college degree level, that can be caused by wrong information provided by users. However, for high school and grad school degrees, it provides similar information to the offline data at the state level. The same occurs for income level, in which

the higher income level (above 100k per year) provides accurate data. We unveiled that immigrants from South America and Central America outnumber the numbers calculated by Census, which may indicate the underestimation of official sources, especially due to illegal immigration. As a final contribution of this case study, we computed correction factors that allows one to multiply it by the percentage of each demographic dimension as calculated by our framework and obtain the Census distribution as a result.

Next, in the second case study, we collected $20,448$ pages categorized as news by Facebook and then leveraged our framework to obtain the political leaning for their audiences. We validated the results by comparing it with four recently proposed approaches to measure ideological bias. Though the compared studies employed very different techniques (ranging from content analysis to surveys about topics such as immigration and homosexuality filled out by readers) the correlation with our results was very high. Our methodology to get audience demographics allowed us to cover a large number of media outlets, which are at least two orders of magnitude larger than existing efforts. Additionally, we also identified news outlets biased along five other axes: age, gender, income level, racial affinity, and national identity.

In a third case study, we provided an in-depth quantitative and qualitative characterization of the Russia-linked ad campaigns on Facebook. Our findings suggest that the Facebook advertising platform can be abused by a new form of attack, that is the use of targeted advertising to create social discord. These ads, that showed to be divisive, were 10 times more effective than a typical Facebook ad, were biased especially in terms of race and political leaning, and tended to target users who are less likely to identify their inappropriateness. We also provide strong evidence that these advertisers have explored the Facebook suggestions tool to engineer the targeted populations. It means that a malicious advertiser intending to reach vulnerable people may select one biased attribute and Facebook will probably suggest other attributes that confirm or even exacerbate the bias. Finally, we presented some suggestions to guide the ads that should be manually inspected to prevent discriminatory content to be spread. For instance, those ads that target extremely biased populations, on the basis of race, political leaning, and other sensitive topics have a greater likelihood of being divisive. Additionally, the ads that experience high click-through rates could also be flagged to be quickly inspected.

The last case study compared the demographics of candidates in the 2018 Brazilian presidential election during the campaign with the election polls surveyed by two popular opinion public survey institutes. We demonstrate that the popularity of the candidates on Facebook may be a good indicator to detect his/her variation on vote intention polls when captured with three different metrics, number of likes, people talking about, and number of people interested in the candidate. However, the number of likes alone can be mislead-

ing, showing the history of accumulated popularity in the past. We also verified that the demographic fluctuation of gender and geographic regions in the audience of candidates on Facebook presents analogous shape to the variations detected by the election polls. In particular, we show that abrupt shifts occasioned by relevant events during the campaign such as protests and candidates prevented to run in the elections are well captured by Facebook measurements.

As additional contributions of our work, we deployed four systems that give free access to our analysis to any internet user. Next lines summarize the developed systems:

1. ***Media Bias Monitor*** - Exposes the biases in audience demographics for $20,448$ news outlets in the United States to any Internet user (available at `twitter-app.mpi-sws.org/media-bias-monitor`).

2. ***Audiência das Notícias*** - Similar to the Media Bias Monitor system, applied to the Brazilian Media ecosystem (available at `http://www.audiencia-das-noticias.dcc.ufmg.br/`).

3. ***Audiência dos Políticos*** - Presents the demographics of the audience of Brazilian politicians. The Brazilian related systems have contributed to the understanding of the national political scene, which has attracted press coverage to our work [1] (available at `http://www.audiencia-dos-politicos.dcc.ufmg.br/`.

4. ***System to explore the Russian Ads*** - The last system displays the Russian ads details such as text, image, and attributes used as well as the demographics of their targeted audiences. (available at `http://www.socially-divisive-ads.dcc.ufmg.br/`).

The systems developed in the case studies of this thesis are innovative and very useful for the general population as well as for social media researchers. The Media Bias Monitor, for instance, exposes the bias of more than 20 thousand media outlets, which is unfeasible with the current approaches. *Audiência das Notícias* and *Audiência dos Políticos* are the first systems that addressed the issue of bias in the audience of politicians and media outlets in Brazil. Finally, the Russian ads monitor provided many details about the socially divisive ads published previous to the 2016 US presidential election. In spite of the huge attention and coverage of academics and mainstream media occasioned by this remarkable event, none of them published a fine-grained analysis as ours.

[1] https://www.bbc.com/portuguese/salasocial-43824463

We should emphasize that, in spite of using the Facebook advertising platform as the source of data, our methodology is not limited to the Facebook social media platforms (including Instagram) since it may be applied to other OSN platforms and even to other digital advertising systems. The application of our methodology in other platforms depends on adapting the framework to the APIs provided by the additional platforms, which includes mapping interests, behaviors and demographics do the IDs used by each different platform. It means that the prohibition of access to Facebook API or changes on their policies do not make our methodology unpractical since we can move to other platforms. In fact, the support to other platforms in our framework is the most important future step to be taken.

Our study forms the foundation for many research directions that can be pursued in the future to explore demographics with precision and low cost about different populations. As future work, we intend to expand our framework to collect data from other OSN advertising platforms such as Twitter and LinkedIn and minimize the dependency on Facebook. Avoiding being restricted to one single social media, will also allow us to compare demographic data from different sources and identify peculiarities across OSN with different purposes. This new source of data may also help us understand the different public of distinct social networks and study how the differences in the audiences impacts in their advertising systems.

We also intend to employ our methodology to extract demographic data similar to Census from other countries, especially in Brazil and other developing countries with a high number of people on Social Networks. In an optimistic scenario, we expect that our data may be employed as an auxiliary resource in Census surveys. This may save time and money, which is crucial for development and poor countries, where the economic resources are critical.

Finally, we aim at employing our framework to conduct other important demographic studies such as creating the Social Alexa, a similar system to Alexa[2], in which the rankings are created by demographic data from OSN instead of browser behavior. We also expect to create automatic scripts that feed an online repository with an historical data of the collected studies at least twice a year.

[2]https://www.alexa.com/siteinfo

Bibliography

Ackerman, M. S., Cranor, L. F., and Reagle, J. (1999). Privacy in e-commerce: Examining user scenarios and privacy preferences. In *Proceedings of the ACM Conference on Electronic Commerce*, EC '99, pages 1--8.

Allcott, H. and Gentzkow, M. (2017). Social media and fake news in the 2016 election. Technical report 2, National Bureau of Economic Research.

An, J. and Weber, I. (2016). greysanatomy vs. yankees: Demographics and hashtag use on twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, ICWSM'16, pages 523--526.

Andreou, A., Silva, M., Benevenuto, F., Goga, O., Loiseau, P., and Mislove, A. (2019). Measuring the facebook advertising ecosystem. In *Proceedings of the Annual Network and Distributed System Security Symposium*, NDSS'19.

Andreou, A., Venkatadri, G., Goga, O., Gummadi, K. P., Loiseau, P., and Mislove, A. (2018). Investigating ad transparency mechanisms in social media: A case study of facebook's explanations. In *Proceedings of the Annual Network and Distributed System Security Symposium*, NDSS'18.

Angwin, J. and Parris Jr., T. (2016). Facebook lets advertisers exclude users by race. https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race.

Angwin, J., Tobin, A., and Varner, M. (2017a). Facebook (still) letting housing advertisers exclude users by race. *https://www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin*.

Angwin, J., Varner, M., and Tobin, A. (2017b). Facebook enabled advertisers to reach 'jew haters'. *https://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters*.

Araujo, M., Mejova, Y., Weber, I., and Benevenuto, F. (2017). Using facebook ads audiences for global lifestyle disease surveillance: Promises and limitations. In *Proceedings of the ACM Conference on Web Science*, WebSci'17, pages 253--257.

Babaei, M., Kulshrestha, J., Chakraborty, A., Benevenuto, F., Gummadi, K. P., and Weller, A. (2018). Purple Feed: Identifying High Consensus News Posts on Social Media. In *Proceedings of the AAAI/ACM Conference on Artifical Intelligence, Ethics & Society*, AIES'18, pages 10--16.

Bakshy, E., Messing, S., and Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on facebook. *Science*, 348(6239).

Beatty, P. C. and Willis, G. B. (2007). Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly*, 71(2):287--311.

Bermingham, A. and Smeaton, A. F. (2011). On using twitter to monitor political sentiment and predict election results. In *Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology*, SAAIP'11, pages 2--10.

Bond, R. and Messing, S. (2015). Quantifying social media's political space: Estimating ideology from publicly revealed preferences on facebook. *American Political Science Review*, 109(1):62--78.

Budak, C., Goel, S., and Rao, J. M. (2016). Fair and Balanced? Quantifying Media Bias Through Crowdsourced Content Analysis. *Public Opinion Quarterly*, 80(S1):250–271.

Castillo, C., El-Haddad, M., Pfeffer, J., and Stempeck, M. (2014). Characterizing the life cycle of online news stories using social media reactions. In *Proceedings of the ACM conference on Computer supported cooperative work & social computing*, pages 211--223.

Cesare, N., Lee, H., McCormick, T., Spiro, E., and Zagheni, E. (2018). Promises and pitfalls of using digital traces for demographic research. *Demography*, 55(5):1979--1999.

Chakraborty, A., Ghosh, S., Ganguly, N., and Gummadi, K. P. (2015). Can trending news stories create coverage bias? on the impact of high content churn in online news media. In *Proceedings of the Computation and Journalism Symposium*, pages 559--562.

Chakraborty, A., Ghosh, S., Ganguly, N., and Gummadi, K. P. (2016). Dissemination biases of social media channels: On the topical coverage of socially shared news. In *Proceedings of the International AAAI Conference on Web and Social Media*, ICWSM'16, pages 559--562.

Chakraborty, A., Messias, J., Benevenuto, F., Ghosh, S., Ganguly, N., and Gummadi, K. P. (2017). Who makes trends? understanding demographic biases in crowdsourced recommendations. In *Proceedings of the International AAAI Conference on Web and Social Media*, ICWSM'17, pages 22--31.

Chiang, C.-F. and Knight, B. (2011). Media bias and influence: Evidence from newspaper endorsements. *The Review of Economic Studies*, 78(3):795--820.

Chung, J. and Mustafaraj, E. (2011). Can collective sentiment expressed on twitter predict political elections? In *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI'11, pages 1770--1771.

Comarela, G., Durairajan, R., Barford, P., Christenson, D., and Crovella, M. (2018). Assessing candidate preference through web browsing history. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD'18, pages 158--167.

Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Flammini, A., and Menczer, F. (2011a). Political polarization on twitter. In *Procedings of the International AAAI Conference on Weblogs and Social Media*, ICWSM'11, pages 89--96.

Conover, M. D., Goncalves, B., Ratkiewicz, J., Flammini, A., and Menczer, F. (2011b). Predicting the political alignment of twitter users. In *Proceedings of the IEEE Third International Conference on Social Computing*, SocialCom'11, pages 192–199.

Covert, T. J. A. and Wasburn, P. C. (2007). Measuring media bias: A content analysis of time and newsweek coverage of domestic social issues, 1975–2000. *Social science quarterly*, 88(3).

Culotta, A., Kumar, N. R., and Cutler, J. (2015). Predicting the demographics of twitter users from website traffic data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI'15, pages 72--78.

Del Vicario, M., Vivaldo, G., Bessi, A., Zollo, F., Scala, A., Caldarelli, G., and Quattrociocchi, W. (2016). Echo chambers: Emotional contagion and group polarization on facebook. *Scientific reports*.

DiGrazia, J., McKelvey, K., Bollen, J., and Rojas, F. (2013). More tweets, more votes: Social media as a quantitative indicator of political behavior. *PLOS ONE*, 8(11):1–5.

Dong, Y., Yang, Y., Tang, J., Yang, Y., and Chawla, N. V. (2014). Inferring user demographics and social strategies in mobile social networks. In *Proceedings of the ACM SIGKDD*

International Conference on Knowledge Discovery and Data Mining, KDD'14, pages 15--24.

Dooling, D. J. and Lachman, R. (1971). Effects of comprehension on retention of prose. *Journal of experimental psychology*, 88(2):216.

Facebook (2017). *https://newsroom.fb.com/news/2017/02/improving-enforcement-and-promoting-diversity-updates-to-ads-policies-and-tools*.

Facebook Help Center (2018). How to report things.

Fang, A., Ounis, I., Habel, P., Macdonald, C., and Limsopatham, N. (2015). Topic-centric classification of twitter user's political orientation. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '15, pages 791--794.

Flaxman, S., Goel, S., and Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public opinion quarterly*, 80(S1):298--320.

Garcia, D., Mitike Kassa, Y., Cuevas, A., Cebrian, M., Moro, E., Rahwan, I., and Cuevas, R. (2018). Analyzing gender inequality through large-scale facebook advertising data. *Proceedings of the National Academy of Sciences*, 115(27):6958--6963.

Garimella, K., De Francisci Morales, G., Gionis, A., and Mathioudakis, M. (2018). Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship. In *Proceedings of The Web Conference*, WWW'18, pages 913--922.

Gayo-Avello, D. (2012). "i wanted to predict elections with twitter and all i got was this lousy paper" – a balanced survey on election prediction using twitter data. *CoRR*, abs/1204.6441.

Gayo-Avello, D., Metaxas, P. T., and Mustafaraj, E. (2011). Limits of electoral predictions using twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, ICWSM'11, pages 490--493.

Gentzkow, M. and Shapiro, J. M. (2010). What Drives Media Slant? Evidence From U.S. Daily Newspapers. *Econometrica*, 78:35--71.

Giglietto, F. (2012). If likes were votes: An empirical study on the 2011 italian administrative elections. In *Proceedings of the International AAAI Conference on Web and Social Media*, ICWSM'12, pages 471--474.

Gilliam Jr, F. D., Iyengar, S., Simon, A., and Wright, O. (1996). Crime in black and white: The violent, scary world of local news. *Harvard International Journal of Press/Politics*, 1(3):6--23.

Gilroy, C. and Kashyap, R. (2018). Extending the Demography of Sexuality with Digital Trace Data. In *Proceedings of the Population Association of America Annual Meeting*, PAA'18, pages 1--25.

Golbeck, J. and Hansen, D. (2011). Computing political preference among twitter followers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI'11, pages 1105--1108.

Gomide, J., Veloso, A., Jr., W. M., Almeida, V., Benevenuto, F., Ferraz, F., and Teixeira, M. (2011). Dengue surveillance based on a computational model of spatio-temporal locality of twitter. In *Proceedings of ACM Conference on Web Science*, WebSci'11, pages 3:1--3:8.

Graeff, T. R. and Harmon, S. (2002). Collecting and using personal data: consumers' awareness and concerns. *Journal of Consumer Marketing*, 19(4):302–318.

Groseclose, T. and Milyo, J. (2005). A measure of media bias. *The Quarterly Journal of Economics*, 120:1191--1237.

Guerra, P. H. C., Meira Jr, W., Cardie, C., and Kleinberg, R. (2013). A measure of polarization on social media networks based on community boundaries. In *Proceedings of the International AAAI Conference on Web and Social Media*, ICWSM'13, pages 215--224.

Haranko, K., Zagheni, E., Garimella, K., and Weber, I. (2018). Professional gender gaps across US cities. In *Proceedings of the International AAAI Conference on Web and Social Media*, ICWSM'18, pages 604--607.

Hu, J., Zeng, H.-J., Li, H., Niu, C., and Chen, Z. (2007). Demographic prediction based on user's browsing behavior. In *Proceedings of the International Conference on World Wide Web*, WWW'07, pages 151--160.

Jung, S.-g., An, J., Kwak, H., Salminen, J., and Jansen, B. (2018). Assessing the Accuracy of Four Popular Face Recognition Tools for Inferring Gender, Age, and Race. In *Proceedings of the International AAAI Conference on Web and Social Media*, ICWSM'18, pages 624--627.

Jungherr, A., Jürgens, P., and Schoen, H. (2012). Why the pirate party won the german election of 2009 or the trouble with predictions: A response to tumasjan, a., sprenger, t. o.,

sander, p. g., & welpe, i. m. “predicting elections with twitter: What 140 characters reveal about political sentiment”. *Social Science Computer Review*, 30(2):229–234.

Kabbur, S., Han, E., and Karypis, G. (2010). Content-based methods for predicting web-site demographic attributes. In *Proceedings of the IEEE International Conference on Data Mining*, ICDM'10, pages 863--868.

Kenett, R. S., Pfeffermann, D., and Steinberg, D. M. (2018). Election pollsâa survey, a critique, and proposals. *Annual Review of Statistics and Its Application*, 5(1):1–24.

Kim, Y. M., Hsu, J., Neiman, D., Kou, C., Bankston, L., Kim, S. Y., Heinrich, R., Baragwanath, R., and Raskutti, G. (2018). The stealth media? groups and targets behind divisive issue campaigns on facebook. *Political Communication*, 35(4):515–541.

King, G., Schneer, B., and White, A. (2017). How the news media activate public expression and influence national agendas. *Science*, 358(6364):776--780.

Kosinski, M., Stillwell, D., and Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15):5802--5805.

Krulwich, B. (1997). LIFESTYLE FINDER: Intelligent User Profiling Using Large-Scale Demographic Data. *AI Magazine*, 18(2):37.

Kulshrestha, J., Eslami, M., Messias, J., Zafar, M. B., Ghosh, S., Gummadi, K. P., and Karahalios, K. (2017). Quantifying search bias: Investigating sources of bias for political searches in social media. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW'17, pages 417--432.

Kwak, H., Lee, C., Park, H., and Moon, S. (2010). What is twitter, a social network or a news media? In *Proceedings of the International Conference on World Wide Web*, WWW'10, pages 591--600.

Larsson, A. O. and Moe, H. (2012). Studying political microblogging: Twitter users in the 2010 swedish election campaign. *New Media & Society*, 14(5):729–747.

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., et al. (2018). The science of fake news. *Science*, 359(6380):1094--1096.

Le, H. T., Shafiq, Z., and Srinivasan, P. (2017). Scalable news slant measurement using twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, ICWSM'17, pages 584--587.

Lee, K., Agrawal, A., and Choudhary, A. (2013). Real-time disease surveillance using twitter data: Demonstration on flu and cancer. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD'13, pages 1474--1477.

Lee Rodgers, J. and Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42(1):59--66.

Lella, A. (2016). Traditional news publishers take non-traditional path to digital growth. *ComScore*.

Leskovec, J., Backstrom, L., and Kleinberg, J. (2009). Meme-tracking and the dynamics of the news cycle. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD'09, pages 497--506.

Lima, L., Reis, J. C. S., Melo, P., Murai, F., Araujo, L., Vikatos, P., and Benevenuto, F. (2018). Inside the right-leaning echo chambers: Characterizing gab, an unmoderated social system. In *Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ASONAM'18, pages 515--522.

Makazhanov, A. and Rafiei, D. (2013). Predicting political preference of twitter users. In *Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ASONAM'13, pages 298--305.

Mejova, Y., Gandhi, H. R., Rafaliya, T. J., Sitapara, M. R., Kashyap, R., and Weber, I. (2018a). Measuring subnational digital gender inequality in india through gender gaps in facebook use. In *Proceedings of the ACM SIGCAS Conference on Computing and Sustainable Societies*, COMPASS '18, pages 43:1--43:5.

Mejova, Y., Weber, I., and Fernandez-Luque, L. (2018b). Online Health Monitoring using Facebook Advertisement Audience Estimates in the United States: Evaluation Study. *JMIR Public Health Surveill*, 4:e30.

Messias, J., Vikatos, P., and Benevenuto, F. (2017). White, man, and highly followed: Gender and race inequalities in twitter. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, WI'17, pages 266--274.

Mislove, A., Lehmann, S., Ahn, Y.-Y., Onnela, J.-P., and Rosenquist, J. N. (2011). Understanding the demographics of twitter users. In Adamic, L. A., Baeza-Yates, R. A., and Counts, S., editors, *Proceedings of the International AAAI Conference on Web and Social Media*, ICWSM'11, pages 554--557.

Mitchell, A. (2016). Key findings on the traits and habits of the modern news consumer. *http://www.pewresearch.org/fact-tank/2016/07/07/modern-news-consumer/*. [Online; accessed 30-January-2019].

Mitchell, A., Gottfried, J., Kiley, J., and Matsa, K. (2014). Political polarization and media habits. *http://www.journalism.org/interactives/media-polarization/*. [Online; accessed 30-January-2019].

Morstatter, F., Pfeffer, J., Liu, H., and Carley, K. M. (2013). Is the sample good enough? comparing data from twitter's streaming API with twitter's firehose. In *Proceedings of the International AAAI Conference on Web and Social Media*, ICWSM'13, pages 400–408.

Mueller, W., Silva, T. H., Almeida, J. M., and Loureiro, A. A. (2017). Gender matters! analyzing global cultural gender preferences for venues using social sensing. *EPJ Data Science*, 6(1):5.

Munson, S., Chhabra, S., and Resnick, P. (2017). BALANCE - Tools for improving your news reading experience. *http://balancestudy.org/*.

Murray, D. and Durrell, K. (2000). Inferring demographic attributes of anonymous internet users. *Web Usage Analysis and User Profiling*, 1836:7--20.

Otterbacher, J. (2010). Inferring gender of movie reviewers: Exploiting writing style, content and metadata. In *Proceedings of the ACM International Conference on Information and Knowledge Management*, CIKM'10, pages 369--378.

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin UK.

Pennacchiotti, M. and Popescu, A.-M. (2011). Democrats, republicans and starbucks afficionados: User classification in twitter. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD'11, pages 430--438.

Pitkow, J. and Recker, M. (1994). Results from the First World-Wide Web user survey. *Computer Networks and ISDN Systems*, 27(2):243--254.

Ramasubramanian, S. (2007). Media-based strategies to reduce racial stereotypes activated by news stories. *Journalism and Mass Communication Quarterly*, 84(2):249–264.

Reis, J. C. S., Correia, A., Murai, F., Veloso, A., and Benevenuto, F. (2019). Supervised learning for fake news detection. *IEEE Intelligent Systems*, 34(2):1--8.

Reis, J. C. S., Kwak, H., An, J., Messias, J., and Benevenuto, F. (2017). Demographics of news sharing in the u.s. twittersphere. In *Proceedings of the ACM Conference on Hypertext and Social Media*, HT'17, pages 195--204.

Resende, G., Melo, P., Sousa, H., Messias, J., Vasconcelos, M., Almeida, J., and Benevenuto, F. (2019). (mis)information dissemination in whatsapp: Gathering, analyzing and countermeasures. In *Proceedings of The Web Conference*, WWW'19.

Ribeiro, F. N., Henrique, L., Benevenuto, F., Chakraborty, A., Kulshrestha, J., Babaei, M., and Gummadi, K. P. (2018). Media bias monitor: Quantifying biases of social media news outlets at large-scale. In *Proceedings of the International AAAI Conference on Web and Social Media*, ICWSM'18, pages 290--299.

Ribeiro, F. N., Saha, K., Babaei, M., Henrique, L., Messias, J., Benevenuto, F., Goga, O., Gummadi, K. P., and Redmiles, E. M. (2019). On microtargeting socially divisive ads: A case study of russia-linked ad campaigns on facebook. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*, FAT*'19, pages 140--149.

Saha, K., Weber, I., Birnbaum, M. L., and De Choudhury, M. (2017). Characterizing awareness of schizophrenia among facebook users by leveraging facebook advertisement estimates. *Journal of Medical Internet Research*, 19(5):e156.

Sakaki, T., Okazaki, M., and Matsuo, Y. (2010). Earthquake shakes twitter users: Real-time event detection by social sensors. In *Proceedings of the International Conference on World Wide Web*, WWW'10, pages 851--860.

Sang, E. T. K. and Bos, J. (2012). Predicting the 2011 dutch senate election results with twitter. In *Proceedings of the Workshop on Semantic Analysis in Social Media*, EACL'12, pages 53--60, Stroudsburg, PA, USA. Association for Computational Linguistics.

Schler, J., Koppel, M., Argamon, S., and Pennebaker, J. (2006). Effects of Age and Gender on Blogging. *Artificial Intelligence*, 86:199--205.

Sharma, E., Saha, K., Ernala, S. K., Ghoshal, S., and De Choudhury, M. (2017). Analyzing ideological discourse on social media: A case study of the abortion debate. In *Proceedings of the International Conference of The Computational Social Science Society of the Americas*, CSS'17, pages 3:1--3:8.

Shoemaker, P. J., Vos, T. P., and Reese, S. D. (2009). Journalists as gatekeepers. *The handbook of journalism studies*.

Shor, E., van de Rijt, A., Miltsov, A., Kulkarni, V., and Skiena, S. (2015). A paper ceiling: Explaining the persistent underrepresentation of women in printed news. *American Sociological Review*, 80(5):960--984.

Silva, T. H., de Melo, P. O. V., Almeida, J. M., Musolesi, M., and Loureiro, A. A. (2017). A large-scale study of cultural differences using urban data about eating and drinking preferences. *Information Systems*, 72(Supplement C):95 – 116.

Skoric, M., Poor, N., Achananuparp, P., Lim, E., and Jiang, J. (2012). Tweets and votes: A study of the 2011 singapore general election. In *Proceedings of the Hawaii International Conference on System Sciences*, pages 2583--2591.

Smith, A. and Anderson, M. (2018). Social media use in 2018. *http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/.* [Online; accessed 30-January-2019].

Speicher, T., Ali, M., Venkatadri, G., Nunes Ribeiro, F., Arvanitakis, G., Benevenuto, F., Gummadi, K. P., Patrick Loiseau, M.-s., and Mislove, A. (2018). Potential for Discrimination in Online Targeted Advertising. *Proceedings of Machine Learning Research*, 81(3):1--15.

Sylwester, K. and Purver, M. (2015). Twitter language use reflects psychological differences between democrats and republicans. *PLOS ONE*, 10(9):1–18.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., and Welpe, I. M. (2010). Predicting elections with Twitter: what 140 characters reveal about political sentiment. In *Proceedings of AAAI International Conference on Weblogs and Social Media*, ICWSM'10, pages 178--185.

Ulges, A., Koch, M., and Borth, D. (2012). Linking visual concept detection with viewer demographics. In *Proceedings of the ACM International Conference on Multimedia Retrieval*, ICMR'12, pages 24:1--24:8.

Vel, O. d., Corney, M., and Anderson, A. (2002). Language and gender author cohort analysis of e-mail for computer forensics. In *Proceedings of the Digital Forensics Research Workshop*, pages 1--16.

Venkatadri, G., Andreou, A., Liu, Y., Mislove, A., Gummadi, K. P., Loiseau, P., and Goga, O. (2018). Privacy risks with facebook's pii-based targeting: Auditing a data broker's advertising interface. In *Proceedings of the IEEE Symposium on Security and Privacy*, IEEE S&P'19, pages 221--239.

Vikatos, P., Messias, J., Miranda, M., and Benevenuto, F. (2017). Linguistic diversities of demographic groups in twitter. In *Proceedings of the ACM Conference on Hypertext and Social Media*, HT'17, pages 275--284.

Vosoughi, S., Roy, D., and Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380):1146--1151.

Wang, P., Guo, J., Lan, Y., Xu, J., and Cheng, X. (2016). Your Cart Tells You: Inferring Demographic Attributes from Purchase Data. In *Proceedings of the ACM International Conference on Web Search and Data Mining*, WSDM'16, pages 173--182.

Wang, W., Rothschild, D., Goel, S., and Gelman, A. (2014). Forecasting elections with non-representative polls. *International Journal of Forecasting*, 31(3):980--981.

Weber, I., Kashyap, R., and Zagheni, E. (2018). Using Advertising Audience Estimates to Improve Global Development Statistics. *ITU Journal: ICT Discoveries, Special Issue*, (2).

Williams, C. B. and jeff Gulati, G. (2009). What is a social network worth? facebook and vote share in the 2008 presidential primaries. In *Proceedings of the Annual Meeting of the American Political Science Association*, pages 1--17.

Zagheni, E., Polimis, K., Alexander, M., Weber, I., and Billari, F. C. (2018). Combining Social Media Data and Traditional Surveys to Nowcast Migration Stocks. *Proceedings of the Population Association of America Annual Meeting*, pages 1--17.

Zagheni, E., Weber, I., and Gummadi, K. (2017). Leveraging facebook's advertising platform to monitor stocks of migrants. *Population and Development Review*, 43(4):721–734.

Zhong, E., Tan, B., Mo, K., and Yang, Q. (2013). User demographics prediction based on mobile data. *Pervasive and Mobile Computing*, 9(6):823--837.

Zhong, Y., Yuan, N. J., Zhong, W., Zhang, F., and Xie, X. (2015). You Are Where You Go: Inferring demographic attributes from location check-ins. In *Proceedings of the ACM International Conference on Web Search and Data Mining*, WSDM'15, pages 295--304.

Zhou, D. X., Resnick, P., and Mei, Q. (2011). Classifying the political leaning of news articles and users from user votes. In *Proceedings of the International AAAI Conference on Web and Social Media*, ICWSM'11, pages 417--424.