

Representativeness of Abortion Legislation Debate on Twitter: A Case Study in Argentina and Chile

Eduardo Graells-Garrido
Barcelona Supercomputing Center
(BSC)
Barcelona, Spain
&
Universidad del Desarrollo
Santiago, Chile
eduardo.graells@bsc.es

Ricardo Baeza-Yates
NTENT & Northeastern University
California, USA
rbaeza@acm.org

Mounia Lalmas
Spotify
London, UK
mounia@acm.org

ABSTRACT

The role of the Web in political exchange has been crucial for society. Its platforms have connected people and allowed manifestation, organization, and access to information; however, they have also produced negative outcomes, such as increased polarization and fast disinformation spreading. These types of phenomena are not completely understood in the context of continuous technological change. Here we propose to grow knowledge in these issues by focusing on representativeness, through the following question: How demographic groups are represented in the discussion on micro-blogging platforms? Our aim is to answer this question on the discussion about a specific topic, abortion, as observed on one of the most popular micro-blogging platforms. As a case study, we followed the abortion discussion on Twitter in two Spanish-speaking countries from 2015 to 2018. Our results indicate differences in representativeness with respect to country, stance, and time of publication, a process that affects to on-going legislation. These findings show that demographic groups differ in how they generate content, and that under- and over-represented groups are not the same between countries, implying that single-country outcomes are not generalizable.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in collaborative and social computing**; • **Information systems** → **Social networks**.

KEYWORDS

Social Networks, Stance Prediction, Data Bias

ACM Reference Format:

Eduardo Graells-Garrido, Ricardo Baeza-Yates, and Mounia Lalmas. 2020. Representativeness of Abortion Legislation Debate on Twitter: A Case Study in Argentina and Chile. In *Companion Proceedings of the Web Conference 2020 (WWW '20 Companion)*, April 20–24, 2020, Taipei, Taiwan. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3366424.3383561>

This paper is published under the Creative Commons Attribution 4.0 International (CC-BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW '20 Companion, April 20–24, 2020, Taipei, Taiwan

© 2020 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC-BY 4.0 License.

ACM ISBN 978-1-4503-7024-0/20/04.

<https://doi.org/10.1145/3366424.3383561>

1 INTRODUCTION

The Web has been crucial for exchanging points of view. Its several platforms have connected people and allowed manifestation, organization, and access to information. However, they have also produced negative outcomes, such as increased polarization [19] and fast disinformation spreading [41]. These phenomena are not completely understood, arguably due to how demographic patterns change on the Web, and the difficulty of performing longitudinal studies around these topics. The lack of population-representative information source can lead to biased analysis and algorithms [1].

With this context in mind, we aim to improve understanding in how people participate and take a stance on the debate around a controversial issue, *abortion*. Abortion is a hard topic to talk about, as it is not only about political stances, but also about deep private matters [43]. Its debate on micro-blogging platforms has been studied before [21, 44, 49, 50], delivering insights on how the different stances on the issue relate and are discussed. Even though this analysis has been mostly static, longitudinal studies covering abortion debate exist [15, 18], however, previous longitudinal analyses have not focused on *who* discusses *what*, and *with whom*; additionally, they have focused on a single country — mainly, the U.S.A. To the extent of our knowledge, there are no studies about abortion debate with a focus on understanding how representative the discussion is in longitudinal terms. This paper aims to provide answers to such research question, stated as follows: *What is the representativeness of the abortion debate on micro-blogging platforms?*

Here we develop a pipeline of methods to study opinions expressed on micro-blogging platforms. We infer demographic attributes and political stance of users using a state-of-the-art classification method. Then, we use census data to establish differences in how demographic and stance groups are represented at two levels. On the one hand, we measure the demographic representativeness of each group. On the other hand, we measure how content in debate is distributed, with a focus on geography, stance, and time.

With this methodology, we study the Twitter debate about abortion in Argentina and Chile during 2015–2018, two neighboring Spanish-speaking countries in South America. Although they share many cultural similarities, they have several differences in terms of abortion legislation. On the one hand, Chile is a country that, until the approval of its current abortion law, had one of the most severe abortion laws in the world [45]. Its current abortion bill was sent to congress in January 2015 and approved as law in September 2017,

eliciting discussion on social media until today. The law allows abortion under three grounds: “endangerment of a woman’s life; embryonic anomaly or malformation incompatible with life; and pregnancy arising from sexual violence” [34]. On the other hand, Argentina does not have an abortion law, although its current Sanitary Code from 1921 allows abortion on two grounds (endangerment of life and pregnancy from sexual violence). In 2018, a free abortion law was proposed in Congress, but was rejected after two months of legislation [7]. These events were relevant for abortion debate in Latin American countries, as they had an impact on how abortion activism was performed on the physical world [36].

Our main results indicate that both countries, regardless their similarities, have completely different population distributions on Twitter, implying that single-country outcomes may not be generalizable without controlling for this fact. Additionally, we find that content distributions varies by country, stance, and time of publication, implying that studies on controversial issues need to include these variables in their analysis.

The contributions of this research are as follows: (i) a methodology to measure the stance representativeness around a controversial issue; and (ii) a case study of the methodology applied to the debate on abortion in Argentina and Chile. In conclusion, even though Twitter has severe bias in population representation [1, 5], when carefully considering demographic factors to debias the data, it provides valuable insights on public opinion, and allows to understand how people behaves in political debate while embracing web-based platforms.

2 RELATED WORK AND MOTIVATION

There are two main areas related to our work: *Political Debate on Twitter* and *Representativeness of Findings*.

2.1 Political Debate on Twitter

Twitter has been a platform that has enabled the study of controversial discussion at scale [17]. In the context of abortion, different perspectives have been studied: how people from each stance interacts with others [49]; the linguistic characteristics of ideological discourse [44]; and the spread of anti-abortion policy [50]. To characterize stances on these types of controversial issues, stances must be predicted, as they are not always explicit, *e.g.*, self-reported. Two main approaches have been proposed: network analysis [16, 17] and lexical analysis [13, 31]. Both have also be used together [42]. Regardless of the approach, the main assumption is that homophily, *i.e.*, the tendency to relate to like-minded people, plays a big role in political classification [4]. Still, stance inference is not a fully solved problem. Recent developments have differentiated between participating in the debate and taking a stance [51].

In our work, we infer stance using a state-of-the-art classifier, XGBoost [9]. Under the absence of ground truth to classify user demographic attributes and political stance, we bootstrapped self-reported data [38]. In addition to defense and opposition to abortion rights, we consider an *undisclosed* category [51]. While our motivation is similar to other works, our analysis differs in focus and coverage, as we put emphasis on demographics and stances in two countries.

2.2 Representativeness of Findings

Twitter is a non-representative platform *per se* as it has several biases with respect to gender, age, education, location, *etc.* [1, 5]. Thus, any analysis based on Twitter data may not be representative, due to the lack of a stratified sample of the general population, and the fact that the platform is not representative of the whole population. For instance, it has severe over-representation of white males from urban areas [5, 30, 32], as well as people in their 20s [46]. In geographical terms, central regions dominate the content generation and distribution more than expected from their population sizes [22]. These biases do not imply that any obtained insights are restricted to these over-represented groups. A study about the #BlackLivesMatter movement detected that, regardless of under-representation, African-Americans were more active than others group in the debate [37]. A study about commuter populations revealed that the infra-representation of females did not hinder the study outcomes with respect to gender [29]. In the case of abortion, domain experts in health have pointed out that, even though the population is biased, if done carefully, Twitter allows to understand what “the public is *actually* seeing,” an understanding that would help to design communication strategies regarding abortion legislation and medical practice [25]. Finally, a recent study about abortion on Twitter found that insights from Twitter match those from nationally representative surveys [20]. Then, while Twitter is not a representative sample, some Twitter-based findings are informative and relevant when the methods and results are properly debiased.

In our work, we analyze representativeness from two aspects. On the one hand, we measure how representative of the population the debate is, by predicting demographic features and comparing them to general population of each country. All our analyses are done at the demographic group level. On the other hand, we measure how each stance, and each country, participates on the debate, with a focus on the content distribution. We find that different countries have different demographic representativeness, and that content distributions vary according to the legislative events in each country. As such, our work adds evidence to the pile of previous work on representativeness of Twitter debate.

Here we established the differences and contributions with the prior research that our work draws from. In the next section we describe the data set where this work is performed.

3 DATA SETS

We describe the data sets used to study abortion legislation debate. The analysis has a temporal coverage of four years of Twitter debate (2015–2018), a geographical coverage of two countries (Argentina and Chile), and a political coverage of two abortion laws (one approved, one rejected). Chronologically, the debate starts with the abortion bill proposed in 2015 by President Michelle Bachelet in her second presidential regime, and approved as law in 2017. In Argentina, the legislative debate was held in 2018, during the regime of President Mauricio Macri. In total, we used three data sets: a corpus of Twitter debate, a census (2017) from Chile, and a census (2010) from Argentina.

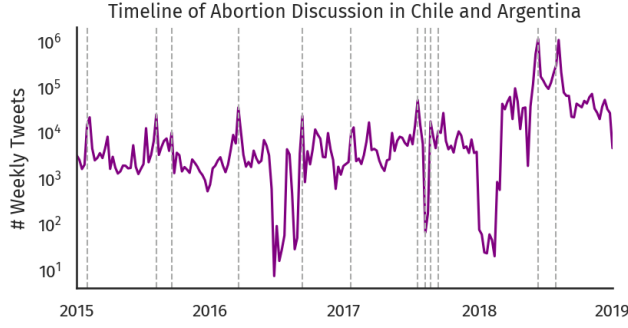


Figure 1: Weekly tweet volume in the period of study. Dotted lines represent relevant events in legislation. Events in 2015–2017 happened in the Chilean Congress. Events in 2018 happened in the Argentinian Congress.



Figure 2: Word cloud of term frequency in the corpus.

3.1 Twitter Corpus

On Twitter, users have a profile and publish micro-posts, usually with limited number of characters allowed (the limit is currently 280). Each micro-post (a *tweet*) may contain multimedia items, hash-tags (or topic indicators, e.g., #abortion), mentions of others users, and links to websites. A micro-post may also be published again by someone else than its original author (in Twitter, this is known as *retweeting*), cited or quoted by others, and replied to.

User profiles in Twitter contain the following features: a *screen-name* or alias, a full name (which may not be validated), an optional location in free text form (eventually fictional [26]), an optional self-description or biography, an optional URL, the number of published micro-posts, the number of followers or subscribers, and the number of friends or subscriptions to other profiles.

Between January 1, 2015, and December 31, 2018, we crawled tweets using the Twitter Streaming API. The query parameters were keywords related to abortion, composed into a query using the OR operator, and apply to the tweet content. The keyword set included general abortion vocabulary (e.g., aborto(s), abortista(s), abortados(as), ..., tenses of *to abort* in Spanish), hashtags, both, general (e.g., #aborto –abortion–, #aborto3causales –abortion three grounds–, #noalaborto –no to abortion–) and contextual (e.g., #marchaabortolegal –protest for legal abortion–), mentions to relevant accounts involved in the debate (e.g., public health institutions,

Table 1: Population with age 13 or older, per gender, reported in the census from 2010 (Argentina) and 2017 (Chile).

Age Cohort	Argentina		Chile	
	Female	Male	Female	Male
< 18	1,743,257	1,783,368	612,585	586,351
18–29	3,942,142	3,919,153	1,704,901	1,666,363
30–39	2,942,278	2,834,870	1,240,967	1,260,447
≥ 40	7,636,729	6,508,177	3,481,952	3,960,218
Total per Gender	16,264,406	15,045,568	7,040,405	7,473,379

NGOs), and phrases (e.g., “pregnancy interruption”). Initially, the data set contained 31.4M tweets from 1.8M users.

We identified users who self-reported their gender (male, female)¹ and country (Argentina, Chile) on their profiles. Then, we proceeded to propagate these labels to the rest of the data set using a classifier (Section 4 details this process). To focus on the debate, we kept only those profiles with valid demographic attributes (either self-reported or predicted with high confidence), and that belonged to the largest connected component of the full interaction network, comprised of retweets, mentions, replies, and quotes. As result, we kept 6M tweets from 663K users.

Figure 1 shows the weekly volume of tweets. One can see that the volume of debate presents several peaks, most of them co-occurring with legislative events. Additionally, one can see weeks periods with low content volume (July 2016, September 2016, December 2017, and January 2018). In those weeks, the crawler presented problems gathering data. In terms of content, Figure 2 shows a word cloud of the most frequent terms in the corpus.

3.2 Census Data

To estimate representativeness of our data set, we use census data from both countries.

The Argentinian government held a population census on October 2010 [28]. The reported population was 40.1M inhabitants (31.3M were 13 years old or older). The Chilean government held a population census on April 2017 [27]. The reported population was 17.5M inhabitants (14.5M were 13 years old or older). Table 1 shows the population in each demographic group under consideration in this work. Note that the underage cohort (< 18) contains from 13 to 17 years, as 13 years old is the minimum Twitter user age according to the Terms of Service. These age cohorts are arguably coarse, but they are common in demographic studies of Twitter profiles [48]. This granularity aids in understanding Twitter demographics, due to over-sampling of people in their 20s [46].

We seek to analyze the abortion debate with a focus on the representativeness of the political debate on micro-blogging platforms. In the next section we describe the methodology applied to these data sets.

4 METHODOLOGY

This section details how we analyze the abortion debate in this work. We describe how we classify users into demographic groups

¹Note that Spanish is a gendered language. In contrast to English, neutral pronouns do not exist.

and stances using a state-of-the-art classifier. Then, from this classification results we can analyze representativeness of several content and behavioral analysis supported by the data.

4.1 Classification Framework

The classification task is, given a user u , predict its *gender* (male, female), *location* (a country), *age cohort* (such as 30–39 years old), and an *abortion stance* (such as in favor or against it). We first define a feature matrix, where each user is a row, and each feature is a column. The feature matrix itself is a horizontal concatenation of multiple matrices:

- (1) A user-term matrix, where a cell at (i, j) contains the number of times user i has used the term j in her/his tweets. Terms include words, hashtags, user names, URLs, and emoji. We only consider terms that have a minimum frequency of 50 in the data set.
- (2) A second user-term matrix, where a cell at (i, j) contains the number of times user i has used the term j in her/his biographical description. We only consider terms that have a minimum frequency of 10 in the data set.
- (3) A matrix of user and meta-features, where we have included profile features such as the domain of the user self-reported URL (e.g., a home page), a time-zone (if available), the number of emoji in the profile description, the use of emoji in the account’s reported name, and any other user-specific metric, such as lexicons applied to user data.
- (4) An adjacency matrix of retweets, where a cell at (i, j) contains the number of times user i has retweeted user j .
- (5) An adjacency matrix of mentions, replies, and quotes, where a cell at (i, j) contains the number of times user i has written tweets directed at user j . This matrix contains only users in the data set. For instance, an account that is mentioned but does not participate in the debate is implicitly contained as a mention term (*@account_name*) in the user-term matrix.

For each attribute we need to predict, we instantiate a single classifier, using the feature matrix as input. We base our work in the XGBoost classifier [9], a state-of-the-art gradient boosting [14] framework based on decision trees. It works by training a number of weak learners for the classification task, and then by returning the fraction of learners that predict each possible category. During training, each learner is instructed to correct its mistakes, focusing each time on the observations that were harder to classify.

Our parameters are the following: 300 *estimators* (the number of trees or weak learners), a *learning rate* of 0.1 (how much the feature weights decrease after each iteration to avoid over-fitting), and a *max delta step* of 1 (the maximum change in a tree leaf from one iteration to the next). These parameters are fairly conservative for a XGBoost classifier. This is due to potential class imbalance in our data set, given that the Twitter population is biased [1], and that we rely on self-reported information, as described next. Additionally, to avoid over-fitting, we train with early stopping, extracting a validation set of 10% of the training observations.

4.2 User Labeling

Having a classification task, we need labeled users to train each classifier. In this context, we rely on self-reported user information,

Table 2: Seed patterns for each abortion stance.

Stance	Patterns in Biography	Patterns in Tweets
<i>defense</i>	#abortolegal, #abortolibre, #abortoseguro, #abortogratis, #feminista, a favor del aborto, #proeleccion, #prochoice	@abortolegalcl, #nobas-tantrescausales, #seraley, @Cam-pAbortoLegal
<i>opposition</i>	@siemprexlavida, derecho a la vida, #antiaborto, contrari[oa] al aborto, #stopaborto, las dos vidas, cristiano/a, #salvoslasdosvidas, aborto no es la solución, #siempreporlavida, #porlasdosvidas, #profamilia, #provida, #noaborto	#salvoslasdosvidas, #sialavida, #abortolegalno, [#@]noaborto, #noesley, @mmreivindica, @noaborto_arg

both in explicit and implicit disclosures of the relevant attributes. We follow a bootstrapped approach, where we learn using the self-reported attributes, and then propagate them to the rest of the data set [38]. We resort to different strategies in labeling users and potentially filtering the feature matrix.

To label gender, we first check her/his first name in the reported full name, by checking from a list of known names. For users without an identifiable name, we checked their biographies for typical expressions disclosing gender (e.g., “Mother of two kids”). To label a user location, we check her/his self-reported location name against a manually built gazetteer. And to label age, we followed a similar process, by matching common phrases in biographies that contained age or date of birth (e.g., “25 years old”). Users were binned into age cohorts (< 18 , $18-29$, $30-39$, ≥ 40), also used in other works [48]. We removed columns from the feature matrix that would over-fit the classifier, as they were used for labeling (e.g., numbers with two or four digits that may represent an age or a date). To aid classification, the meta-data features included a custom built lexicon of biography vocabulary, with categories such as *celebrities* (both, local and global), *sports* (soccer teams and players), *politics* (political parties), *social media* (names of other sites, such as Instagram), *etc.*

Finally, we define the probability of user u belonging to a class c as the fraction of weak learners that vote for c given u . We only accept classification outcomes that are above a specific threshold (0.7 for gender and location, 0.65 for age, both manually determined by experimentation).

4.3 Stance Prediction

To predict stance we followed the same approach as for demographic features. However, labeling users is not as straight-forward. First, in abortion there are two main stances, colloquially denoted *pro-life* and *pro-choice*. Although commonly used, these terms are semantically overloaded, *i.e.*, they carry an implicit false leaning on behalf of the opposite stance [23]. We therefore classify users into two groups or stances: *opposition* (instead of pro-life), and *defense* (instead of pro-choice). Since these stances are known, we have *a priori* information about the words, organizational accounts, and hashtags characterizing them. We use this domain knowledge to build a list of *seed patterns*, to be applied either to profile descriptions or tweet content. Users that match those patterns are labeled with the corresponding stance. As keyword usage is not exclusive

to each stance, for instance, due to content injection [12] or hashtag hijacking [24], we only keep the labels where users match patterns from one stance but not from the other. In a similar way as we do with age, to avoid over-fitting we remove features that match the seed patterns from the feature matrix, as they perfectly separate users from both groups. Table 2 shows a subset of the seed patterns used to label users to each stance. As examples, the *defense* patterns include #abortolibre (unrestricted abortion) and the campaign @CampAbortoLegal. The *opposition* patterns include #sialavida (yes to life) and #salvemoslas2vidas (let’s save the two lives).

Given that weak learners perform slightly better than random, and under the assumption that some users cannot be classified into two stances [51], we bin users according to the fraction of learners that voted *defense*. A threshold of (0.49, 0.51) defines *undisclosed* users, in which we group both undecided people and those who decide to not disclose their stance, but to disseminate information.

With this classification approach we are able to infer gender, location, and stance for all users. Age is an optional attribute, being the hardest attribute to measure due to the lack of general information about users (for instance, interaction with accounts in other topics) and population bias.

4.4 Analysis of Representativeness

Social media platforms have population biases, and Twitter is no exception. Such biases are not inherently bad, however, it is important to measure them to know how generalizable a social-media analysis is. We measure two types of representativeness: *demographic* and *in content*.

Demographic Representativeness. We aim to understand how much of the national population of a country is represented in the debate, and how these representations vary by country, as not all countries discuss in the same way [39]. To measure this, we estimate the user rate per 1,000 inhabitants for each pairwise combination of gender and age cohorts in a country. Then, we compare the similarity of this distribution with the national distribution from a census, using the Kendall-Tau rank-correlation coefficient between matrices:

$$\tau_B = \frac{n_c - n_d}{\sqrt{(n_0 - n_1)(n_0 - n_2)}},$$

where n_c is the number of concordant pairs, n_d is the number of discordant pairs, $n_0 = n(n - 1)/2$, $n_1 = \sum_i t_i(t_i - 1)/2$, $n_2 = \sum_j u_j(u_j - 1)/2$, t_i is the number of tied values in the i^{th} group of ties for the first matrix, and u_j is the number of tied values in the j^{th} group of ties in the second matrix. The coefficient tells us how much similar (or dissimilar) is the Twitter population with the national population.

Content Representativeness. Regardless of the demographic representativeness of the data set, sometimes a small fraction of users produces most of the content [3]. We aim to understand the distributions of content with respect to country, stance, and time, to find whether the content distributions imply equal representations between groups in the data set, or, in other words, how different the content generation process is in each group under analysis.

This type of phenomena has been modeled before through power laws [10]. For instance, power laws have been used to compare national Web domains [2]. They are characterized by a single exponent α , as seen on the following formula:

$$Pr(\text{tweets published by group}) \propto k^{-\alpha}.$$

By fitting a power law to tweets per user in a country, in a stance, and in every year under study, we quantify the content representativeness of the data, and characterize how these content distributions change with respect to legislative events in every year.

The consideration of both types of representativeness allow us to answer our second research question, providing evidence on whether political debate generates comparable patterns of engagement and participation in different population groups.

The series of methods defined in this section, while drawing from previous work in the literature, provide a way to measure representativeness of stances and demographic attributes in abortion as seen on micro-blogging platforms. In the next section we describe the results of applying these methods to the data set from Section 3.

5 RESULTS

Our research question is: *what is the representativeness of the abortion debate?* We aim to understand the representativeness taking into account two aspects: demographics and content distribution. In this section we present the results of applying our methodology (Section 4) to the abortion debate in Argentina and Chile from 2015 to 2018 (Section 3). We describe the demographic and classification results of the 663,340 users in the data set, and then proceed to answer our research question.

5.1 Classification Results

Table 3 contains performance metrics for each predicted attribute, estimated with 5-fold cross validation. Location and stance exhibit the best performance, both with a precision of 0.93 and a high recall (0.93 and 0.92). In terms of stance, this performance exceeds typical precision, which lies around 85% [11]. This shows that our choice of classifier was correct. Gender and age cohort exhibit acceptable performances; however, recall that we considered these predictions only in cases where confidence was above a specific threshold. In our preprocessing stage, we discarded users for whom gender was not predicted. For age, we kept users without a predicted cohort, as they may have publish relevant information. The number of profiles with predicted age is 163K. The table also mentions the top features for each classifier. One can see that colored heart emojis plays an important role in predicting gender and stance, and that words associated with abortion debate are relevant for gender (such as *morir* –to die–, and *cuerpo* –body–). To predict country, the most important feature is time-zone, as well as local hashtags (e.g., #aborto3causales refers to Chilean legislation), and mentioning other social-media sites on the biography (*biography-lexicon: social media*). To predict age, important features include having a link to an Instagram account, using social-media slang (e.g., *fav*), using emojis, and having the word *estudiante* –student– in the biography.

Table 3: Classification metrics of demographic attributes and stance. Metrics were estimated through 5-fold cross validation.

	Precision	Recall	Top Features
Gender	0.71 ± 0.02	0.68 ± 0.03	♥, morir, profile:ingeniero, cuerpo, profile:n_emojis, pibas, name:♥, profile:abogada, biolex:sports
Location	0.93 ± 0.04	0.93 ± 0.04	tz:Santiago, causales, #aborto3causales, chile, legal, #abortolibre, libre, biolex:socialmedia, favor, profile:n_emojis
Age Cohort	0.72 ± 0.01	0.64 ± 0.04	dominio, d:instagram.com, objetores, fav, profile:n_emojis, director, #aborto3causales, profile:estudiante, cambiar, pichetto
Stance	0.93 ± 0.02	0.92 ± 0.03	♥, rt:175273213, ♥, nacer, decidir, existio, amenazada, torcuato, #noalabortoenargentina, matar

Table 4: Discriminating terms for each demographic group (AR is Argentina, CL is Chile; M is male, F is female).

Country	Gender	Age	Terms in Tweets	Terms in Biographies
AR	F	18–29	clandestino, provida, legal, sexual, #abortolegal, educacion, esi, vos, #yoaborte, traumatico	estudiante, futura, 📚, 19, 20, psicologia, lic, 18, 21, medicina
		30–39	#abortolegal, #abortolegal, #quesealey, @campabortolegal, #niunamemos, #senadoresquesealey, #activaelcongreso, salud, punible, pagina12.com.ar	abogada, docente, periodista, mama, madre, militante, investigadora, escribo, mujeres, editora
		< 18	♥, legal, fav, 18, gt, 🌟, vidas, dale, quiero, tenes	♥, ig, 📷, snap, 🌟, 17, ♥, 16, 🍰, 🌸
		≥ 40	macri, #salvoslas2vidas, dominio, extincion, debate, cfk, @mauriciomacri, cristina, #aborto, infobae.com	madre, abuela, esposa, hijos, mama, docente, anti, argentina, casada, nietos
	M	18–29	—, criticas, @agustinlaje, #salvoslas2vidas, normalizacion, #abortolegalesmuerte, destruye, 📺, choriplanero	estudiante, hincha, futuro, ingenieria, boca, river, fanatico, anos, futbol, deportivo
		30–39	#abortolegal, «, », #abortolegal, bit.ly, debate, #aborto, @teresita_ok, pagina12.com.ar, @lavozcomar	periodista, abogado, docente, diseñador, economista, grafico, padre, escribo, investigador, papa
		< 18	♥, twitter, podrido, tw, cansaron, re, fav, tema, legal, bolas	📷, 📷, snap, insta, instagram, ig, 📺, futbol, ♥, jugador
		≥ 40	macri, debate, dominio, #salvoslas2vidas, infobae.com, cfk, extincion, #aborto, cristina, @mauriciomacri	casado, padre, argentino, hijos, peronista, nacido, abogado, papa, anti, esposo
		18–29	#abortolibre, #abortolegalchile, chile, libre, causales, #aborto3causales, #chilesepintadeverde, #nobastan3causales, weon, 🎵	estudiante, chile, udec, chilena, udp, uc, universidad, pedagogia, 🇨🇱, 24
		30–39	#aborto3causales, chile, causales, #abortolibre, #nobastan3causales, libre, objecion, conciencia, 📢, contraloria	mama, abogada, madre, periodista, mujer, hija, chile, feliz, amante, chilena
		< 18	#abortolegalchile, @andrea_provida, chile, @andre_balnes, interesade, cuantes, @xpolilla, pl, #chilesepintadeverde, @yamecanseaborto	🇨🇱, account, chile, 📷, fan, 🌟, raja, 🍷
		≥ 40	chile, #aborto3causales, causales, #porlas2vidas, bit.ly, libre, @mileschile, dc, tres, comision	madre, hijos, abuela, casada, esposa, mama, derecha, mujer, chilena, chile
	M	18–29	causales, #aborto3causales, chile, #abortolibre, libre, udi, #porlas2vidas, derecha, pinera, kast	estudiante, chile, ingenieria, universidad, comercial, uc, civil, derecho, colo, anos
		30–39	#aborto3causales, causales, chile, #abortolibre, objecion, contraloria, conciencia, libre, udi, ajusta	ingeniero, colocolino, ing, padre, administrador, chile, comercial, papa, civil, diseñador
		< 18	inculcar, afiche, docente, hicimos, profesores, expresion, rompio, pretende, pertenece, #abortolegalchile	📖, colegio, 17, 🇨🇱, medio, seminarista, 📷, 📷, colo-colo, deserve
		≥ 40	chile, causales, libre, #aborto3causales, dc, #porlas2vidas, bachelet, pinera, @sebastianpinera, bit.ly	casado, padre, hijos, ingeniero, chile, chileno, catolico, papa, derecha, esposo

To characterize the words most associated with each demographic group (country, gender, age cohort), Table 4 shows the most discriminating terms in both, tweets and biographies, weighted using log-odds ratio with uninformative Dirichlet prior [33]. We observe that females discuss about abortion (*clandestino* –clandestine–, *sexual*, *derecho* –right to–) and other, related phenomena (*#niunamemos*, a hashtag about female homicides in Latin America), while males tend to discuss about politics, and to mention other males (e.g., @mauriciomacri, @sebastianpinera, @agustinlaje, kast). Young people tend to use more emojis than other cohorts, both in their tweets and in their biographies. In general, the table shows coherent phenomena with current expectations about these demographic groups.

5.2 Demographic Representativeness

Using census data [27, 28], we estimated Twitter user rates per 1,000 inhabitants for Argentina and Chile, according to each age cohort and gender in the data set. Table 5 shows the results. Engagement in the debate is different not only between countries, but also within demographic groups. The most represented group in Chile are 18–29 years old females (2.71 rate), followed by men in

Table 5: Twitter user rates per 1,000 inhabitants for each demographic group. Total per gender includes users without predicted age.

	Argentina		Chile	
	female	male	female	male
< 18	42.57	6.63	1.62	0.30
18–29	17.89	5.32	2.71	2.20
30–39	0.79	0.64	1.93	2.35
≥ 40	0.68	0.87	0.63	0.86
Total per gender	24.99	10.95	6.08	6.60

their thirties (2.35). This contrasts with Argentina, where females between 13 and 17 years old have a much higher rate (42.57), followed by 18–29 years old females (17.89). This rate decays abruptly for the other age cohorts, which exposes similar numbers to Chile. A Kendall rank-coefficient of $\tau = 0.12$ ($p = 0.64$) confirms that the two matrices are not similar. In contrast, despite Argentina having a larger population than Chile, their national census distributions are similar ($\tau = 0.79$, $p < 0.01$). The female representativeness of Argentina contrasts previous studies where males are predominant on

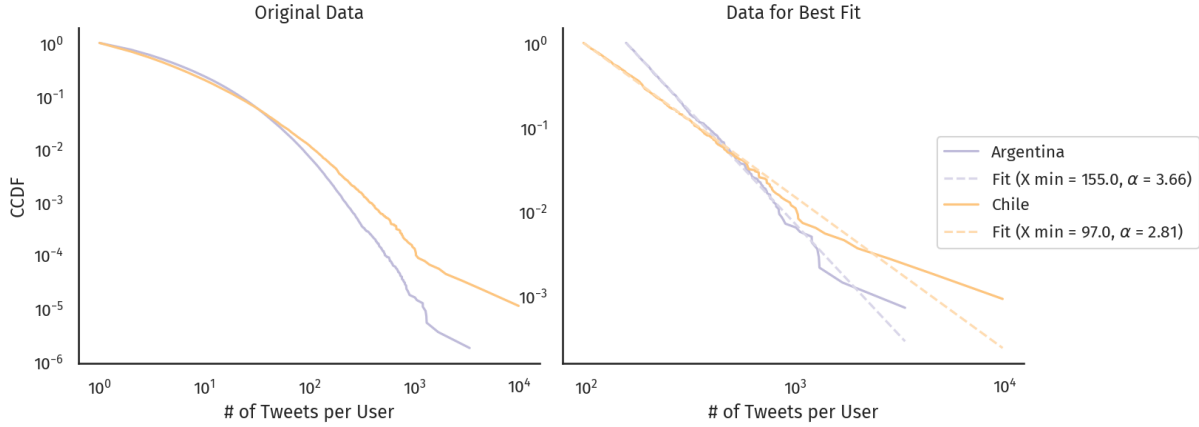


Figure 3: Distribution of number of tweets per user from each country.

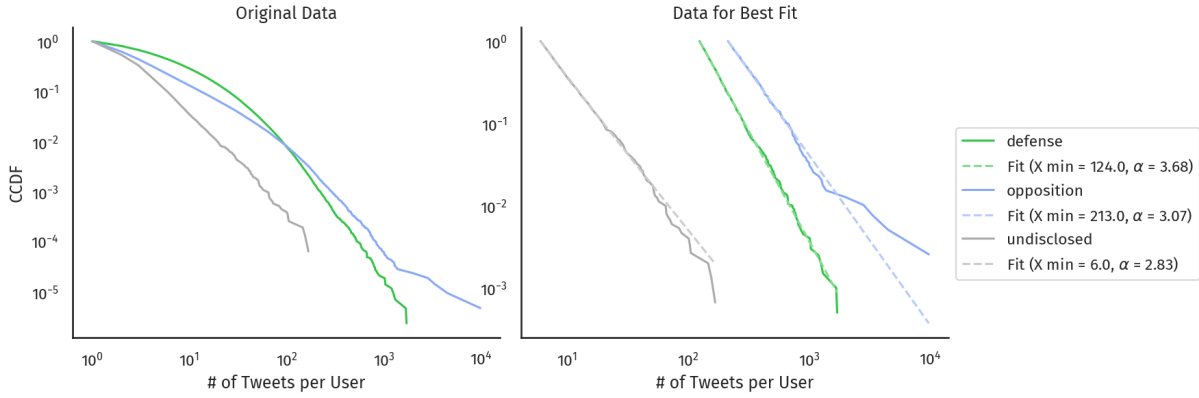


Figure 4: Distribution of number of tweets per user from each stance.

the social platform [5, 30, 32], perhaps due to the gendered nature of the topic.

5.3 Content Representativeness

A relevant aspect of representativeness lies on who produces the content, as it is not uniformly distributed [3]. For instance, only 8.23% of users generated 50% of the tweets in the data set. This analysis can be formalized by modeling the distribution of tweets per user as power laws, following the definition from Section 4.

Figure 3 shows the power law distribution of content per country. The power law exponents are $\alpha_C = 2.81$, and $\alpha_A = 3.66$, indicating that content in Argentina is more concentrated than in Chile, in terms of who generated the data. Power law exponents usually lie in the range [2, 3] [10], which means that these values are higher than what one would expect in regular conditions, for instance, on non-specific debate on Twitter.

The same analysis can be done at the stance level, as shown on Figure 4. This time we include in the analysis the *undisclosed* group, which has the lowest exponent ($\alpha_u = 2.83$) – this may be explained due to this group having people discussing the issue without getting deeply involved and without an agenda. Next is *opposition* ($\alpha_o =$

3.07) and *defense* ($\alpha_d = 3.68$), with higher exponents, signaling that content is concentrated in a more systematic way in *defense*. We hypothesize that this reflects how activism in defense of abortion rights is pushing the issue, given that the legislative efforts in place were aligned with their objectives.

If we take time into account, the exponents for Chile every year go from $\alpha_{2017C} = 2.46$ to $\alpha_{2015C} = 2.69$. The yearly exponents are fairly similar, signaling that the structure of the debate inside the country was stable in terms of content generation (Figure omitted due to space). Figure 5 depicts a different phenomenon in Argentina, with the preceding years before legislation showing similar exponents ($\alpha_{2015A} = 2.97$, $\alpha_{2016A} = 2.64$, $\alpha_{17A} = 3.06$), and the last year with a much higher one ($\alpha_{2018A} = 3.63$). This signals how disruptive the debate was in Argentina in terms of content generation during a year with legislation.

In the stance content generation per year, *opposition* exhibits a stable behavior in time, with exponents ranging from $\alpha_{2018o} = 1.95$ to $\alpha_{2017o} = 2.14$, suggesting that people opposed to abortion rights tend to behave in similar ways regardless of the progress in legislative debate (Figure omitted due to space). In contrast, Figure 6 shows how different is the distribution for *defense* in the aftermath

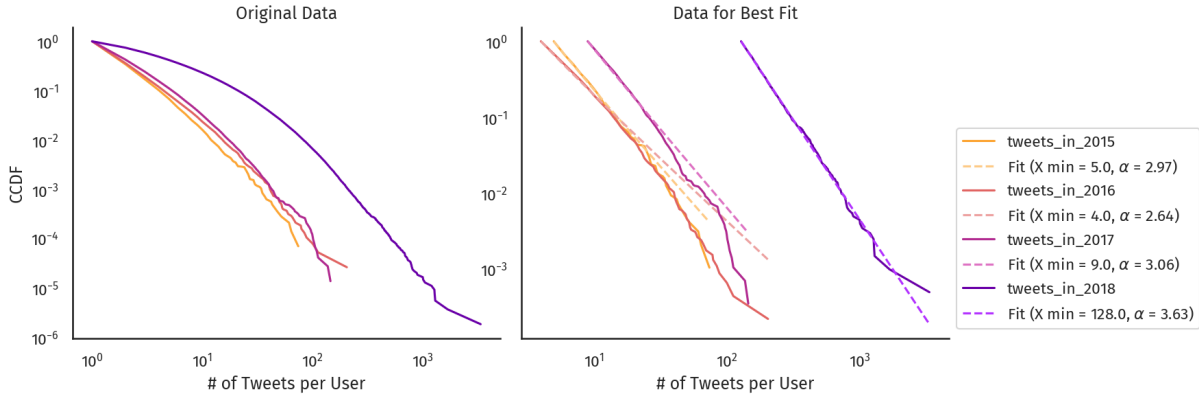


Figure 5: Distribution of number of tweets per user from Argentina, per year.

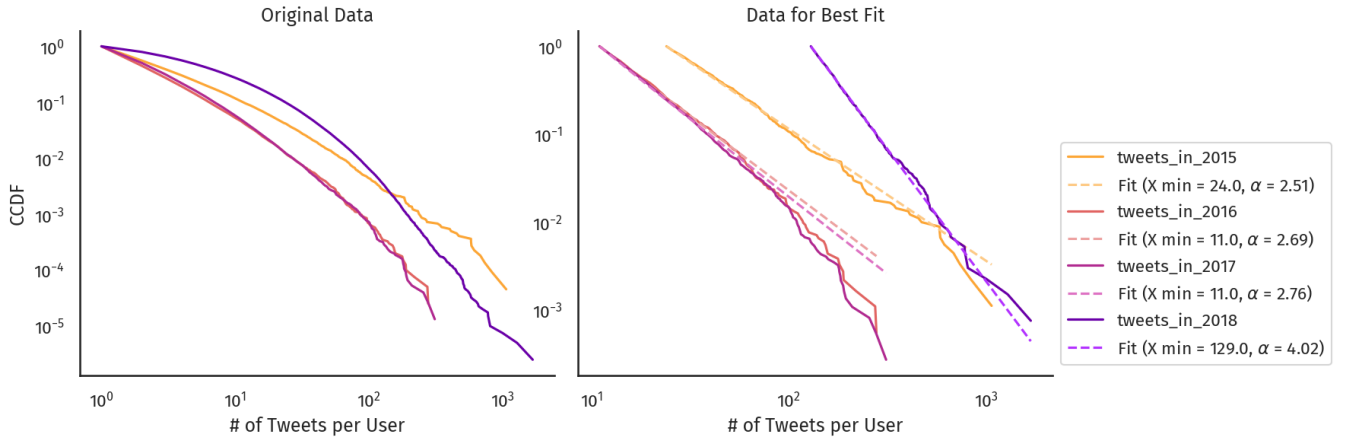


Figure 6: Distribution of the number of tweets per user in the *defense* stance, per year.

of legislation in Chile, and during legislation in Argentina. Throughout the years, the exponents monotonically increase, from $\alpha_{2015d} = 2.51$, to $\alpha_{2018d} = 4.02$. This unusually high exponent suggests that people in defense to abortion rights incrementally concentrated and organized content generation. Note that the minimum exponent in *defense* is still greater than the maximum in *opposition*, suggesting that the former group follows a specific strategy as legislative events unfold.

These results show that Argentina and Chile, even though similar in population distribution, culture and geography, have different demographic representations of Twitter users, and different content distributions. Moreover, different stances also exhibit different strategies in content generation. Thus, in this section we have quantified two types of representativeness in the abortion debate under analysis, effectively answering our research question.

6 DISCUSSION

There are plenty of arguments in the debate around abortion and taking a stance, which may be grounded on rational discussion [23]

or deep private feelings [43]. Most of the reviewed work in Section 2 refers to understanding these factors in the context of the Web. However, to the extent of our knowledge, there is a gap in knowledge regarding the adoption of new technologies, the differences in how debate reflects the population (or not) in different countries, and whether there are changes of opinion. Our results have shown that, indeed, there is quantitative evidence that can shed light on these knowledge gaps.

A crucial aspect of political debate is its representativeness. Different countries may have dissimilar population distributions in the discussion, regardless of how similar they may be culturally. This implies that what can be inferred from web-based analysis if only one country is considered, may apply partially to other countries, regardless of how similar they may be. Moreover, as we found in our content-based analysis, changes in time may change the distributions within same groups of users. Thus, given that political debate is a dynamic phenomenon, our results suggest that longitudinal studies should consider demographic and content factors. Otherwise, their findings may be unrepresentative.

There are two main limitations in our work. Firstly, we measured debate during legislation, but did not consider the effect of the associated political events. For instance, the on-going events may have enticed people from unrepresented groups to participate. Second, the lack of ground truth makes our evaluation strictly based on the data we have self-labeled or inferred from self-reported information. A line of future work is the study of how to further evaluate our results and their sensibility to, for instance, the set of seeded words, the definition of stance boundaries, and user weighting [8]. Given hashtag injection, relying on them may make the training groups not as separable as assumed [40]. Moreover, there is still work to do on how to generalize the method to other issues and other populations. Including additional demographic factors is another important direction for future work [37].

7 CONCLUSIONS

In this paper we characterized the representativeness of abortion debate on Twitter. We did so by performing a longitudinal analysis of the debate in two Latin-American countries, Argentina and Chile. We empirically quantified the phenomena using established methods from the literature, finding that representativeness, both in terms of demographics and of content, changes with geography, time, and stance. The use of gradient-boosted trees showed potential in the context of abortion debate, enabling the propagation of self-reported attributes and our own knowledge about the issue at scale. The use of traditional techniques to study the Web, namely, modeling of attributes using power laws, and the usage of discrete choice models to understand user stances, provided insights that could be analyzed dynamically.

In the long term, political debate around abortion does not end, even if legislation does. For instance, in the last decade, the U.S.A. have seen hundreds of new restrictions to state-level abortion support [6], such as clinical programs and insurance coverage, and the introduction of a new act in the state of Alabama to impose almost a complete ban on abortion starting November 2019. In Spain, several restrictive reforms and budget cuts have been proposed by recent governments [35]. In Mexico, in September 2019, the state of Oaxaca approved an abortion law, a debate that featured the green and blue handkerchiefs that we hypothesize inspired the usage of colored heart emojis [47]. Hence, in a growing political context regarding abortion, understanding the dynamics of public debate is crucial to bridge gaps between people of opposing views in these controversial issues.

To conclude, an important line of future work is the study of these dynamics in a global context, including a more fine-grained demographic analysis on abortion and other controversial issues. In these polarizing times, the Web has the potential to act as a positive influence on how people relate to others.

ACKNOWLEDGMENTS

We thank the authors of the twemoji package, available under a CC-BY 4.0 license.

REFERENCES

- [1] Ricardo Baeza-Yates. 2018. Bias on the Web. *Commun. ACM* 61, 6 (May 2018), 54–61.
- [2] Ricardo Baeza-Yates, Carlos Castillo, and Efthimis N Efthimiadis. 2007. Characterization of national web domains. *ACM Transactions on Internet Technology (TOIT)* 7, 2 (2007), 9.
- [3] Ricardo Baeza-Yates and Diego Saez-Trumper. 2015. Wisdom of the Crowd or Wisdom of a Few?: An Analysis of Users' Content Generation. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media* (Guzelyurt, Northern Cyprus). ACM, ACM, New York, NY, USA, 69–74.
- [4] Pablo Barberá. 2015. Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political Analysis* 23, 1 (2015), 76–91.
- [5] Pablo Barberá and Gonzalo Rivero. 2015. Understanding the political representativeness of Twitter users. *Social Science Computer Review* 33, 6 (2015), 712–729.
- [6] Heather D Boonstra and Elizabeth Nash. 2014. A surge of state abortion restrictions puts providers—and the women they serve—in the crosshairs. *Guttmacher Policy Review* 17, 1 (2014), 9–15.
- [7] Amy Booth. 2018. Argentina votes on bill to legalise abortion up to 14 weeks. *The Lancet* 391, 10138 (2018), e21 – e22. [https://doi.org/10.1016/S0140-6736\(18\)31321-7](https://doi.org/10.1016/S0140-6736(18)31321-7)
- [8] Dallas Card and Noah A Smith. 2018. The Importance of Calibration for Estimating Proportions from Annotations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, Vol. 1. ACL Anthology, New Orleans, Louisiana, 1636–1646.
- [9] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, New York, NY, USA, 785–794.
- [10] Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. 2009. Power-law distributions in empirical data. *SIAM review* 51, 4 (2009), 661–703.
- [11] Raviv Cohen and Derek Ruths. 2013. Classifying political orientation on Twitter: It's not easy!. In *Seventh International AAAI Conference on Weblogs and Social Media*.
- [12] Michael Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. 2011. Political polarization on Twitter. In *International Conference on Weblogs and Social Media*.
- [13] Michael D Conover, Bruno Gonçalves, Jacob Ratkiewicz, Alessandro Flammini, and Filippo Menczer. 2011. Predicting the political alignment of Twitter users. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom)*. IEEE, Boston, MA, USA, 192–199.
- [14] Jerome H Friedman. 2001. Greedy function approximation: A gradient boosting machine. *Annals of Statistics* (2001), 1189–1232.
- [15] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2017. The effect of collective attention on controversial debates on social media. In *Proceedings of the 2017 ACM on Web Science Conference*. ACM, ACM, Troy, New York, USA, 43–52.
- [16] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2018. Political Discourse on Social Media: Echo Chambers, Gatekeepers, and the Price of Bipartisanship. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, Lyon, France, 913–922.
- [17] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2018. Quantifying controversy on social media. *ACM Transactions on Social Computing* 1, 1 (2018), 3.
- [18] Venkata Rama Kiran Garimella and Ingmar Weber. 2017. A long-term analysis of polarization on Twitter. In *Eleventh International AAAI Conference on Web and Social Media*. AAAI, Montréal, Québec, Canada, 528–531.
- [19] Nabeel Gillani, Ann Yuan, Martin Saveski, Sorous Vosoughi, and Deb Roy. 2018. Me, My Echo Chamber, and I: Introspection on Social Media Polarization. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, Lyon, France, 823–831.
- [20] Eduardo Graells-Garrido, Ricardo Baeza-Yates, and Mounia Lalmas. 2019. How Representative is an Abortion Debate on Twitter?. In *Proceedings of the 10th ACM Conference on Web Science*. ACM, 133–134.
- [21] Eduardo Graells-Garrido, Mounia Lalmas, and Ricardo Baeza-Yates. 2015. Finding intermediary topics between people of opposing views: a case study. In *Social Personalisation & Search*, Christoph Trattner, Denis Parra, Peter Brusilovsky, and Leandro Balby Marinho (Eds.). CEUR, Santiago, Chile.
- [22] Eduardo Graells-Garrido, Mounia Lalmas, and Ricardo Baeza-Yates. 2016. Encouraging diversity-and representation-awareness in geographically centralized content. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*. ACM, ACM, Sonoma, California, USA, 7–18.
- [23] Kate Greasley. 2017. *Arguments about abortion: Personhood, morality, and law*. Oxford University Press, Oxford, UK.
- [24] Asmelash Tekla Hadgu, Kiran Garimella, and Ingmar Weber. 2013. Political hashtag hijacking in the US. In *Proceedings of the 22nd International Conference on World Wide Web*. ACM, Rio de Janeiro, Brazil, 55–56.
- [25] Leo Han, Lisa Han, Blair Darney, and Maria I Rodriguez. 2017. Tweeting PP: An analysis of the 2015–2016 Planned Parenthood controversy on Twitter. *Contraception* 96, 6 (2017), 388–394.

- [26] Brent Hecht, Lichan Hong, Bongwon Suh, and Ed H Chi. 2011. Tweets from Justin Bieber's heart: the dynamics of the location field in user profiles. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, Vancouver, Canada, 237–246.
- [27] Instituto Nacional de Estadísticas. 2018. Censo 2017. <http://www.censo2017.cl/microdatos/>. [Online; accessed February 15th, 2020].
- [28] Instituto Nacional de Estadísticas y Censos. 2018. Censo 2010. <https://www.indec.gov.ar/indec/web/Nivel4-Tema-2-41-135>. [Online; accessed February 15th, 2020].
- [29] Wendy Liu, Faiyaz Al Zamal, and Derek Ruths. 2012. Using social media to infer gender composition of commuter populations. In *Proceedings of the When the City meets the Citizen Workshop, the International Conference on Weblogs and Social Media*.
- [30] Paul A Longley, Muhammad Adnan, and Guy Lansley. 2015. The geotemporal demographics of Twitter usage. *Environment and Planning A* 47, 2 (2015), 465–484.
- [31] Haokai Lu, James Caverlee, and Wei Niu. 2015. Biaswatch: A lightweight system for discovering and tracking topic-sensitive opinion bias in social media. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, ACM, Melbourne, Australia, 213–222.
- [32] Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and J Niels Rosenquist. 2011. Understanding the Demographics of Twitter Users. *ICWSM* 11, 5th (2011), 25.
- [33] Burt L Monroe, Michael P Colaresi, and Kevin M Quinn. 2008. Fightin' words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis* 16, 4 (2008), 372–403.
- [34] Adela Montero and Raúl Villarreal. 2018. A critical review of conscientious objection and decriminalisation of abortion in Chile. *Journal of medical ethics* 44, 4 (2018), 279–283.
- [35] Melania Moscoso and R Lucas Platero. 2017. Cripwashing: The abortion debates at the crossroads of gender and disability in the Spanish media. *Continuum* 31, 3 (2017), 470–481.
- [36] Ciara Nugent. 2018. Argentina Votes Today on Legalizing Abortion. Here's What That Means for Women's Rights Across Latin America. <http://time.com/5358823/argentina-abortion-vote-latin-america/> [Online; accessed February 15th, 2020].
- [37] Alexandra Olteanu, Ingmar Weber, and Daniel Gatica-Perez. 2016. Characterizing the Demographics Behind the #BlackLivesMatter Movement. In *2016 AAAI Spring Symposium Series*.
- [38] Marco Pennacchiotti and Ana-Maria Popescu. 2011. A machine learning approach to Twitter user classification. In *Fifth International AAAI Conference on Weblogs and Social Media*.
- [39] Barbara Poblete, Ruth Garcia, Marcelo Mendoza, and Alejandro Jaimes. 2011. Do all birds tweet the same?: Characterizing twitter around the world. In *Proceedings of the 20th ACM international conference on information and knowledge management*. ACM, Glasgow, Scotland, UK, 1025–1030.
- [40] Arti Ramesh, Shachi H Kumar, James Foulds, and Lise Getoor. 2015. Weakly supervised models of aspect-sentiment for online course discussion forums. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Vol. 1. ACL Anthology, Beijing, China, 74–83.
- [41] Gustavo Resende, Philippe Melo, Hugo Sousa, Johnatan Messias, Marisa Vasconcelos, Jussara Almeida, and Fabrício Benevenuto. 2019. (Mis) Information Dissemination in WhatsApp: Gathering, Analyzing and Countermeasures. In *The World Wide Web Conference*. ACM, 818–828.
- [42] Mrinmaya Sachan, Avinava Dubey, Shashank Srivastava, Eric P Xing, and Eduard Hovy. 2014. Spatial compactness meets topical consistency: jointly modeling links and content for community detection. In *Proceedings of the 7th ACM international conference on Web search and data mining*. ACM, New York City, NY, USA, 503–512.
- [43] Carol Sanger. 2016. Talking About Abortion. *Social & Legal Studies* 25, 6 (2016), 651–666.
- [44] Eva Sharma, Koustuv Saha, Sindhu Kiranmai Ernala, Sucheta Ghoshal, and Munmun De Choudhury. 2017. Analyzing Ideological Discourse on Social Media: A Case Study of the Abortion Debate. In *Proceedings of the 2017 International Conference of The Computational Social Science Society of the Americas*. ACM, Santa Fe, New Mexico, USA, 3.
- [45] Bonnie L Shepard and Lidia Casas Becerra. 2007. Abortion policies and practices in Chile: Ambiguities and dilemmas. *Reproductive Health Matters* 15, 30 (2007), 202–210.
- [46] Luke Sloan, Jeffrey Morgan, Pete Burnap, and Matthew Williams. 2015. Who tweets? Deriving the demographic characteristics of age, occupation and social class from Twitter user meta-data. *PloS one* 10, 3 (2015), e0115545.
- [47] VERNE. 2019. La despenalización del aborto en Oaxaca se convierte en tendencia mundial. https://verne.elpais.com/verne/2019/09/26/mexico/1569455628_115209.html. [In Spanish; Online; accessed February 15th, 2020].
- [48] Zijian Wang, Scott Hale, David Ifeoluwa Adelani, Przemyslaw Grabowicz, Timo Hartman, David Jurgens, et al. 2019. Demographic Inference and Representative Population Estimates from Multilingual Social Media Data. In *The World Wide Web Conference*. ACM, 2056–2067.
- [49] Sarita Yardi and Danah Boyd. 2010. Dynamic debates: An analysis of group polarization over time on Twitter. *Bulletin of Science, Technology & Society* 30, 5 (2010), 316–327.
- [50] Amy X Zhang and Scott Counts. 2016. Gender and Ideology in the Spread of Anti-Abortion Policy. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, San Jose, CA, USA, 3378–3389.
- [51] Qiang Zhang, Shangsong Liang, Aldo Lipani, Zhaochun Ren, and Emine Yilmaz. 2019. From Stances' Imbalance to Their Hierarchical Representation and Detection. In *The World Wide Web Conference*. ACM, 2323–2332.