

# Every Colour You Are: Stance Prediction and Turnaround in Controversial Issues

Eduardo Graells-Garrido\*  
Barcelona Supercomputing Center  
Barcelona, Spain  
eduardo.graells@bsc.es

Ricardo Baeza-Yates  
Northeastern University at SV  
California, USA  
rbaeza@acm.org

Mounia Lalmas  
Spotify  
London, UK  
mounia@acm.org

## ABSTRACT

Web platforms have allowed political manifestation and debate for decades. Technology changes have brought new opportunities for expression, and the availability of longitudinal data of these debates entice new questions regarding who participates, and who updates their opinion. The aim of this work is to provide a methodology to measure these phenomena, and to test this methodology on a specific topic, abortion, as observed on one of the most popular micro-blogging platforms. To do so, we followed the discussion on Twitter about abortion in two Spanish-speaking countries from 2015 to 2018. Our main insights are two fold. On the one hand, people adopted new technologies to express their stances, particularly colored variations of heart emojis (❤️ & 💙) in a way that mirrored physical manifestations on abortion. On the other hand, even on issues with strong opinions, opinions can change, and these changes show differences in demographic groups. These findings imply that debate on the Web embraces new ways of stance adherence, and that changes of opinion can be measured and characterized.

## CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in collaborative and social computing.**

### ACM Reference Format:

Eduardo Graells-Garrido, Ricardo Baeza-Yates, and Mounia Lalmas. 2020. Every Colour You Are: Stance Prediction and Turnaround in Controversial Issues. In *12th ACM Conference on Web Science (WebSci '20)*, July 6–10, 2020, Southampton, United Kingdom. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3394231.3397907>

## 1 INTRODUCTION

The Web is an important medium to exchange points of view. Its several platforms have connected people and allowed manifestation, organization, and access to information. Longitudinal studies covering political debates exist [13, 16], although they have not looked at who discusses what, and with whom. As our case study, we focus on the debate about *abortion*. Abortion is a hard topic to talk about, as it is not only about political stances, but also about

deep private matters [32]. Debates on micro-blogging platforms on this topic have been studied before, describing how the different stances relate and are discussed [20, 34, 40, 41].

However, the aforementioned work has focused on a single country (USA), with English as main language, and textual content as main unit of analysis. In contrast, here we study the Twitter debate on abortion in Argentina and Chile, two neighboring Spanish-speaking countries in South America, during 2015–2018. Although both countries share many cultural similarities, they have several differences in terms of abortion legislation and Twitter population [19]. Chile is a country that, until the approval of its current abortion law, had one of the most severe abortion laws in the world [35]. 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. 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 the Argentinian Congress, but was rejected after two months of legislation [3]. What characterizes this context is how physical manifestations and protests have connected activist movements from several countries. In the last twenty years, abortion activists in Argentina have expressed themselves through green and purple kerchiefs [38]. The movement, called *green wave*, has influenced how abortion activism was (is) carried out in Latin America [5].

With this context in mind, we aim to better understand how people take a stance on debates around abortion specifically, but also in controversial issues in general. More precisely, we look into how these on-going movements have characteristics that go beyond written language, and how people react to on-going legislation while these movements are active. We aim to answer two research questions: RQ1) *How people make use of new technologies to express their positions on controversial issues?*; and RQ2) *Which demographic and profile factors characterize opinion change?*

To answer these questions, we designed a methodology to study opinions expressed on micro-blogging platforms. We detail how we label demographic attributes and stances in profiles taking advantage of information already available on user profiles. Then, we describe methods that allowed us to answer our research questions, by means of inspecting how a stance classifier predicts stance, and by the linear regression over a metric of stance turnaround.

The main contribution of this work are the novel insights regarding the abortion discussion derived from applying our methodology. Regarding RQ1, we observed that people adopted a new technology to express their stance: emojis, *i.e.*, pictures represented as

\*Also with Data Science Institute, Universidad del Desarrollo.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*WebSci '20*, July 6–10, 2020, Southampton, United Kingdom

© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-7989-2/20/07...\$15.00

<https://doi.org/10.1145/3394231.3397907>

encoded characters in text, part of mainstream text-input user interfaces. Particularly, heart colored emojis (♥ & 💙) are strong predictors of stance, both in tweet content and user profiles. Regarding RQ2, we observed that turnarounds in abortion stance can be inferred, identifying how demographic and profile characteristics explain the variations in stance after important events on the issue, such as legislative actions in Congress.

In conclusion, this paper showcases how the analysis of a longitudinal discussion can support the analysis of socially relevant phenomena: how people express their stances, mirroring physical manifestations, and what are the characteristics of people that change their views in time, enabling the measurement of how different events impact stance of people. In an ever-changing world, embracing Web platforms brings also new ways of expression.

## 2 RELATED WORK

Twitter has been a platform that has enabled the study of controversial discussion at scale [15]. In relation to abortion, different perspectives have been studied: how people from each stance interacts with others [40]; the linguistic characteristics of ideological discourse [34]; and the spread of anti-abortion policy [41]. To characterize stances on these types of controversial issues, stances must be predicted, as they are not always explicit. Two types of approaches are common. Stance can be predicted using network interactions, based on the assumption that like-minded people are more likely to interact [14, 15]. In addition, lexical analysis have shown to allow predicting stance as vocabulary within stances tend to have strongly associated words [10, 26]. However, stance prediction is not a fully solved problem. For instance, participating in the debate and taking a stance are two actions that are assumed to be the equivalent, but are not [42]. In our work we apply a mixed approach, and define an *undisclosed* stance to account for participation in the debate without disclosing stance.

In the lexical approach, a common feature is the usage of hashtags in micro-post content (e.g., #freeabortion, #notoabortion). Hashtags are a form of expression native to the Web that is increasingly associated with physical manifestations of political debates and protests [37]. We believe that current-generation web-technologies, such as emojis, a focus of this paper, are missing in these studies. Doing so poses challenges, for instance, as there are cultural differences in emoji usage [2], and there can be several interpretations of the same emoji [27]. Nevertheless, emojis are so popular that people constantly request new ones [12]. In addition, their usage patterns allow to predict user characteristics, such as gender [7]. In our context, emoji has been analyzed in political discussion on social networks [25], but only with respect to what representation is best for predicting the emoji to be used in text, regardless of its actual meaning, *i.e.*, the work would be similar in a non-political discussion. Although they are typically understood as emotional or sentiment cues [11], we find that emojis are a powerful predictor of abortion stance in Argentina and Chile, not only in micro-post content, but also in profile elements—mirroring how colored kerchiefs are used in physical manifestations [5, 38]. To the extent of our knowledge, this association between stance on controversial issues and emojis has not been described before.

Longitudinal studies on political debate on Twitter exist. Eight years of debate on Twitter in the USA showed that polarization gradually increased in time, in terms of people supporting democrat/republican politicians [16]. Five years of debate on the same country provided insights on how physical-world events had impact on virtual discussion [13]. On a smaller scale, four months of Twitter debate on judicial decisions on same-sex marriage showed that Supreme Court decisions polarized the public and generated emotional shifts in public opinion [8]. While our motivation is similar to longitudinal works such as Garimella et al. [13], our analysis differs in focus. Our attention is not on how polarization evolves in time, instead, we identify users who have changed their views before and after an important event for the discussion, and find which demographics and profile characteristics relate to stance turnaround (or the absence of it).

## 3 DATASET

We describe the dataset used to study the abortion legislation debate on Twitter, the case study in this work. 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 started with the abortion bill proposed in 2015 by Chilean former President Bachelet, and approved as law in 2017. In Argentina, the legislative debate was held in 2018, during President Macri’s period.

On Twitter, users have a profile and publish micro-posts, usually with limited number of characters allowed (currently 280). Each micro-post (a *tweet*) may contain multimedia items, hashtags (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, 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 [24]), 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.

We crawled tweets using the Twitter Streaming API between January 1, 2015, and December 31, 2018. The query parameters were keywords related to abortion, composed into a query using the OR operator, and applied to the tweet content. The keyword set included general abortion vocabulary (e.g., aborto(s), tenses of *to abort* in Spanish), hashtags, both general (e.g., #aborto3causales –abortion three grounds–, #noalaborto –no to abortion–)<sup>1</sup> and contextual (e.g., #marchaabortolegal –protest for legal abortion–), mentions to accounts involved in the debate (e.g., public health institutions, NGOs), and phrases (e.g., “pregnancy interruption”).

Initially, the dataset contained 31.4M tweets from 1.8M users. This dataset was pre-processed and filtered to ensure that we analyzed discussion and debate. Firstly, we identified users who self-reported their gender (male, female) and country (Argentina,

<sup>1</sup>In Chile, 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” [29].



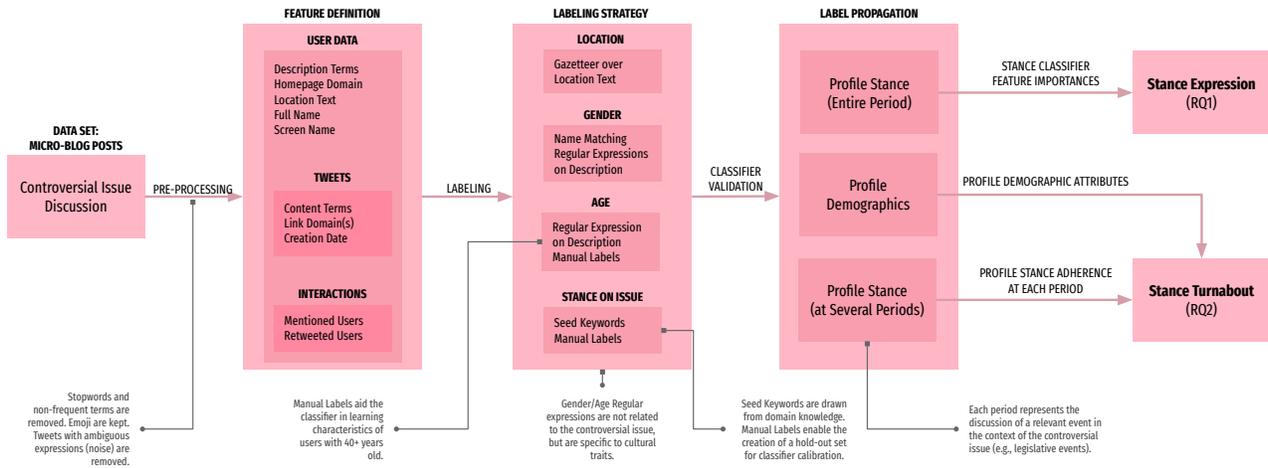


Figure 2: Schematic diagram of the proposed methodology.

page, a time-zone (if available), the number of emojis in the profile description, and the use of each emoji in the account’s reported name.

- (4) Two adjacency matrices. One of retweets, where a cell  $(i, j)$  contains the number of times user  $i$  has retweeted user  $j$ ; and one of mentions, replies, and quotes, where a cell  $(i, j)$  contains the number of times user  $i$  has written tweets directed at user  $j$ .

This feature matrix is used in the following two steps, user profile labeling, and classifier validation and application.

## 4.2 User Profile Labeling

To answer our research questions, we need to know various characteristics of user profiles. We predict them using a classifier, but first we need labeled data. This step in our methodology takes care of generating labeled subsets of users for each needed profile characteristic. The overall approach is simple, but non trivial: we rely on self-reported information, both in explicit and implicit disclosures of the relevant attributes.

**4.2.1 Location (binary).** To label a user location, we check their self-reported location name against a manually built gazetteer for all locations within the countries of interest.

**4.2.2 Gender (binary).** To label a user gender, we first check their first name in the reported full name, by checking from a list of known names. For users without an identifiable name, we check their biographies for typical expressions disclosing gender (e.g., “Mother of two kids”). This list of expressions is built manually, as it is dependant on the main language of the dataset (but not on the specific issue under analysis).<sup>3</sup>

**4.2.3 Age (cohorts).** To label age we match common phrases in biographies that contained age or date of birth (e.g., “25 years old” [36] in Spanish). Instead of focusing on the exact age, users are grouped into four age cohorts ( $< 18$ ,  $18-29$ ,  $30-39$ ,  $\geq 40$ ). Identifying people in the fourth cohort is arguably harder, as they are

<sup>3</sup>Spanish is a gendered language. In contrast to English, neutral pronouns do not exist.

less likely to report their age (based on our experience). Thus, manual labels of these profiles may be required.

**4.2.4 Stance on a Controversial Issue (abortion, binary).** To label stance we follow two strategies. The first one is similar to labeling demographic attributes, with the exception that instead of patterns in phrases, we look for seed patterns and keywords associated with each stance. The second one is to build a set of manually labeled users that do not comply with these patterns.

Before applying seed patterns and keywords, we need to define the actual stances under study. 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 [21]. Therefore, we adopt these two stances: *opposition* (instead of pro-life), and *defense* (instead of pro-choice). We use domain knowledge from our previous work analyzing the abortion discussion [19] to build a list of seed patterns and keywords. Table 1 shows a subset of the seed patterns and keywords used to label users to each stance in Argentina and Chile, including words, expressions, hashtags, activist accounts, and campaign accounts associated with each stance (the table includes translations and explanations). Given that keyword usage is not exclusive to each stance (for instance, due to hashtag hijacking [22]), we only label users that match these patterns for one stance but not for the other.

As result, we have a labeled subset of the dataset with demographic and stance labels. Next, we describe how to propagate these labels to the rest of the dataset.

## 4.3 Classifier Validation and Application

We follow a bootstrapped approach to predict profile characteristics for all users: we learn using the labels obtained from the previous step, and then propagate them to the rest of the dataset [31].

**4.3.1 Classifier Training and Validation.** For each profile characteristic we want to predict, we train a XGBoost (XGB) classifier [6]. XGB is a gradient boosting algorithm based on decision trees. It

**Table 1: Seed patterns for each abortion stance.**

Stance	Patterns in Biography	Patterns in Tweets
<i>defense</i>	#abortolegal ( <i>#legalabortion</i> ), #abortolibre ( <i>#freeabortion</i> ), #abortoseguro ( <i>#safeabortion</i> ), #abortogratis ( <i>#freeabortion</i> ), feminista ( <i>#feminist</i> ), a favor del aborto ( <i>supports abortion</i> ), #proeleccion ( <i>#prochoice</i> ), #prochoice	@abortolegalcl (activist account), #nobastantrescausales ( <i>#threegroundsarenote- nough</i> ), #seraley ( <i>#itwillbelaw</i> ), @CampAbortoLegal (campaign)
<i>opposition</i>	@siemprexlavida ( <i>@alwaysxlife</i> ), derecho a la vida ( <i>right to live</i> ), #antiaborto ( <i>#antiabortion</i> ), contrario/a al aborto ( <i>against abortion</i> ), #stopaborto, las dos vidas ( <i>the two lives</i> ), cristiano/a ( <i>christian</i> ), #salvemosalasdosvidas ( <i>#savethetwolives</i> ), aborto no es la solución ( <i>abortion is not the solution</i> ), #siempreporlavida ( <i>#alwaysforlife</i> ), #porlasdosvidas ( <i>#forthetwolives</i> ), #profamilia ( <i>#profamily</i> ), #provida ( <i>#prolife</i> ), #noalaborto ( <i>#notoabortion</i> )	#salvemosalasdosvidas ( <i>#letssavethetwolives</i> ), #sialavida ( <i>#yestolive</i> ), #abortolegalno ( <i>#notolegalabortion</i> ), #noalaborto ( <i>#notoabortion</i> ), #noesley ( <i>#itsnotlaw</i> ), @mmreivindica (activist account), @noalaborto_arg (campaign)

works by training a number of weak learners using randomly built small subsets of the data, and then returning the fraction of learners that predict each possible category.<sup>4</sup> To avoid over-fitting we train with early stopping, by using a validation set of 20% of the training observations. We also remove columns from the feature matrix that were used for labeling. This includes numbers with two or four digits, used in age cohort labeling, and seed keywords for each stance, as they perfectly separate users from both groups—and our aim is to predict a stance for users that do not use these terms in their content. Finally, to validate the predictive performance of each classifier, we measure precision and recall using 5-fold cross-validation.

**4.3.2 Predicting Demographic Characteristics.** In tree-based models, the classification output is the proportion of weak learners that vote for each stance in a prediction, which we interpret as the prediction confidence of a given profile characteristic for a specific user. We only accept classification outcomes above the following thresholds: 0.7 for gender and location, and 0.65 for age (selected manually through experimentation).

**4.3.3 Stance Adherence Probabilities.** Since in abortion there are two stances, the prediction confidence of one stance is symmetrical with respect to the other, which allows us to work using one stance as unit of analysis without losing generality. However, the complexities of the issue under analysis pose two challenges. First, using seed keywords to label user stance means that only users with extreme positions on the stance spectrum are labeled. Hence, the XGB classifier makes the implicit assumption that all users have extreme views on the issue. This may not be realistic, as not all people has their views positioned in those extremes, and some people may have an undecided position [42]. Second, it is not clear if the proportion of stances in the labeled set of users is the same

<sup>4</sup>The parameters for each classifier are as follows: 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).

as in the non-labeled set [4]. For instance, a prediction confidence of 0.5 may not imply an undecided profile, because these values are not real probabilities.

We address both challenges by converting stance prediction confidence into probabilities by calibrating the classifier [4]. For calibration to work effectively, the dataset used for this purpose should be disjoint from the one used for training. Here we use the set of manually labeled users with stance from the previous step to calibrate the XGB prediction outcomes with Platt’s calibration method [30], converting them into stance adherence probabilities. This allows us to classify users into three classes: *opposition* (probability of *defense* label  $p(\text{defense})$  within  $[0, 0.4]$ ), *undisclosed* ( $p(\text{defense})$  within  $(0.4, 0.6]$ ), and *defense* ( $p(\text{defense}) \geq 0.6$ ).

At this point, we have a dataset with labels either through user reporting, or through prediction with high confidence thresholds. The next steps provide answers to our research questions.

## 4.4 Measuring Stance Expression

Our first research question is concerned with how people express their stance. It is known that specific traits of language and linguistic focus are associated with abortion stances [34, 41]. In terms of the Web, this is translated to hashtags and links. However, as new technologies enable other forms of communication, such as emojis [12], we also ask ourselves whether these new ways of expression can be associated with stance.

A first glimpse into the potential answer of stance expression can be obtained by looking at the predictive power of features. Recall that gradient boosted trees are weak learners that are trained on a subset of the data, both in terms of observations and features. In XGB, the feature importance, defined as *total gain*, quantifies the importance of a feature through the whole set of learners. There, an important feature not only helps in classification (tree building), but also in improving the accuracy of trees (boosting).

Therefore, we quantify how features relate to stance expression by grouping them into types (emojis, terms, links, *etc.*), and estimating whether the mean feature importance of each type is statistically different from the rest. We apply the Tukey’s HSD test, which compares all possible pairs of means, and corrects for family error rates, effectively answering our research question.

## 4.5 Inferring Stance Turnarounds

Usually, the stances related to controversial issues are analyzed as a static phenomenon. Indeed, in our first question, we infer a single predominant stance for users in the whole period under analysis. However, a feature matrix can be built for any period of time, enabling the prediction of stance probabilities at different time windows within the dataset. As a result, for each time window, the difference in predicted probabilities may indicate whether users changed their leaning toward a stance or remained in their previous position. A change in stance is known as a *turnaround*.

To analyze turnarounds, we compare two time windows separated by an event, in a similar way to how natural experiments work. For instance, legislation events and court decisions do not depend on the discussion in social media, but they may stimulate such discussion (see Figure 1). We therefore select users who were

**Table 2: Classification metrics of demographic attributes and stance (5-fold cross validation, with the exception of the hold-out set for stance classification).**

Attribute	Labeled N	Precision	Recall
Gender	212,302	0.71 ± 0.02	0.68 ± 0.03
Location	122,105	0.93 ± 0.04	0.93 ± 0.04
Age Cohort	10,902	0.72 ± 0.01	0.64 ± 0.04
Stance	26,994	0.93 ± 0.02	0.92 ± 0.03
Stance (Hold-out manually labeled)	1,019	0.88	0.84

active in both time windows, and build the corresponding feature matrices for each period. By applying the XGB stance classifier to each matrix, we obtain stance adherence probabilities for each user and period. The difference between two periods is defined as:

$$\Delta(t_0, t_1, u) = p(\text{defense}_{t_1} | u) - p(\text{defense}_{t_0} | u),$$

where  $\Delta(t_0, t_1, u)$  lies in the range  $[-1, 1]$ . When  $\Delta = 1$ , the turnaround is completely toward *defense*. When  $\Delta = -1$ , the turnaround is completely toward *opposition* (recall that  $p(\text{defense})$  is symmetrical with the probability of being in *opposition*). A value of  $\Delta = 0$  implies no change in opinion between periods for the corresponding user  $u$ .

To answer our second research question, we infer the relationship between profile characteristics and turnaround categories (or its absence, *i.e.*, to *remain* in a stance). As profile characteristics, we consider demographic features, as well as attributes easily obtained from each profile, including the number of followers, the number of followees, an activity ratio (number of tweets divided by account age in days), account age (in years), the stance predicted in the first period, and the usage of important terms in expressing opinion (a potential result from RQ1). This would allow to test social hypotheses, for instance, if popular accounts are less likely to change opinion due to their exposure, or whether as people get older, they become more conservative.

We measure the relationship between the aforementioned variables and the value of  $\Delta$  through adjusting a linear regression:

$$\Delta = \beta_0 + \beta \cdot X + \epsilon,$$

where  $X$  is a user feature vector,  $\beta$  is the regression coefficients vector,  $\beta_0$  is the intercept, and  $\epsilon$  is the error term. The value of each element in  $\beta$  describes whether the corresponding feature is significant for stance turnaround.

The series of methods defined in this section, while drawing from previous work in the literature, provide a way to measure new aspects of stances in abortion as seen on micro-blogging platforms, namely, expressiveness and turnarounds. Next, we describe the results of applying these methods to our use case.

## 5 RESULTS

We present the results of applying our methodology (Section 4) to the abortion debate in Argentina and Chile (Section 3).

We start with describing the performance of the demographic attributes (gender, location, age cohort) and stance classifiers applied to the 663,340 users in our dataset. Table 2 reports the performance metrics, estimated using 5-fold cross validation. We obtain the best

**Table 3: Distribution of users per demographic group.**

Country	Gender	< 18	18–29	30–39	≥ 40	n/a
Argentina	Female	15.10	12.32	0.48	1.06	71.04
	Male	5.48	8.98	0.87	2.77	81.91
Chile	Female	1.75	7.68	4.67	4.14	81.76
	Male	0.27	5.49	4.99	5.23	84.02

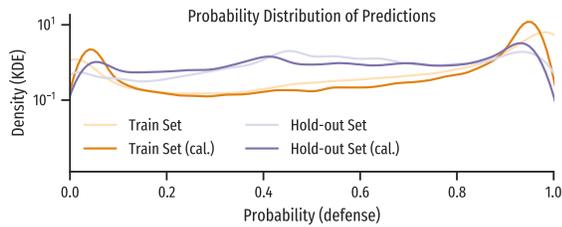
**Table 4: Top features of profile demographic characteristics.**

Attribute	Top Features
Gender	♥, morir (to die), profile:ingeniero (male engineer), cuerpo (body), profile:n_emojis, pibas (Argentinian word for girls), name:♥, profile:♥, profile:abogada (female lawyer), [sport words in the profile]
Location	timezone:Santiago, causales (grounds), #aborto3causales (#abortion3grounds), Chile, legal, #abortolibre (#freeabortion), libre (free), [social media keywords on the profile], favor, profile:n_emojis
Age Cohort	dominio (domain), home page: instagram.com, objetores (objects), fav (social media slang), profile:n_emojis, director, #aborto3causales (#abortion3grounds), profile:estudiante (student), cambiar (to change), Pichetto (Argentinian politician)

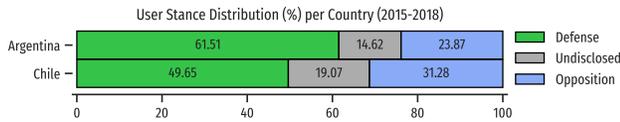
performance for location and stance, precision of 0.93 for both and high recall (0.93 and 0.92, respectively). The performance of the gender prediction is below recent work using deep learning [39] by 0.18, and the performance of the age cohort prediction is above the same approach [39] by 0.23. Arguably the scenarios are not directly comparable, due to differences in using content-specific and general datasets. For instance, age prediction performance is good, but the size of the validation dataset is small.

To account for potential prediction errors, we only consider predictions where the confidence is above specific thresholds. This particularly affects the age attribute, as 75.4% of user age predictions were below the acceptance threshold (see Table 3 for the distribution of predicted demographic characteristics). We kept those users, as they may have posted relevant information. For location, Argentinian users makes 86% of the dataset. An imbalance was expected, as Argentina has around three times the population of Chile. This pronounced imbalance suggests that the population in Argentina has different Twitter adoption patterns [19].

Table 4 shows the most important features in predicting demographic characteristics. Colored heart emojis play an important role in predicting gender (note that previous analysis of emoji usage only identified ♥ as associated with gender [7]), and words associated with the abortion debate are relevant to gender too (e.g. *morir* –to die–, and *cuerpo* –body–). When predicting location, the most important features include timezone, local hashtags (e.g., #*aborto3causales* refers to Chilean legislation), and the mentioning of other social-media sites on the biography, which may be related to different usage of social-media in the two countries [19]. For age, important features include having a link to an Instagram account, using social-media slang (e.g., *fav*), the total number of emojis in the description, and having the word *estudiante* –student– in the biography. These observations are in general not surprising, but they show insights on the usage of emojis, in terms of which ones (gender) and how many (age) help in classifying users.



**Figure 3: Probability distribution of predictions for the training set (self-reported profiles) and the hold-out set used for calibration.**



**Figure 4: Stance distribution per country.**

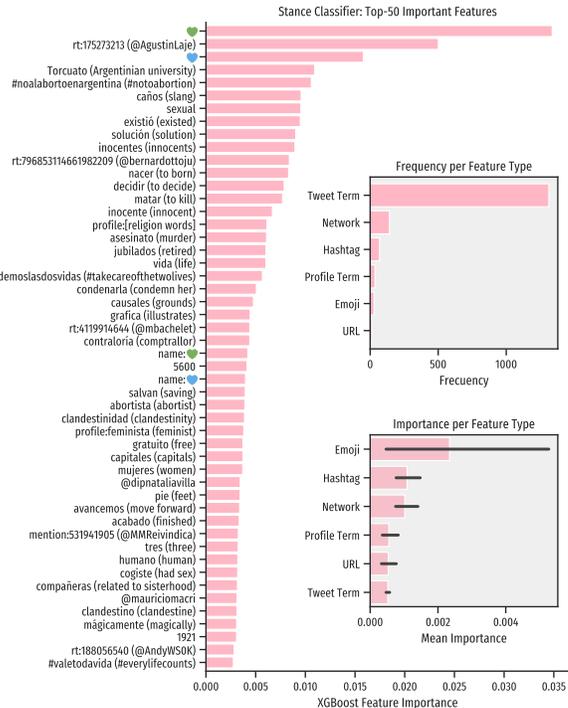
With respect to stance prediction, the performance exceeds typical values, where precision lies around 85% [9]. In the labeled set, all users had a self-reported stance, which implied an explicit position on abortion. However, the performance achieved on the hold-out set is slightly worse (0.88 precision, 0.84 recall), but is still within a good performance range. This set was comprised with 1,019 manually labeled users, randomly selected from the unlabeled group. Its proportion of labels is slightly more balanced than in the self-reported set: 80% *defense*, 20% *opposition*; compared with 85% and 15%, respectively. Figure 3 shows the distribution of calibrated probability outcomes from the classifier on both datasets, automatically labeled (training set) and hold-out (manually labeled), where calibration was performed using Platt scaling. The training set exhibits two probability peaks, one at each extreme, whereas the hold-out set does not exhibit this behavior, and shows a third peak around 0.4. If we consider an *undisclosed* category, this could signal that people who do not express explicitly their position are more likely to be conservative (*opposition*). To account for this uncertainty, we considered the *undisclosed* category in our analyses. In total, 59.86% of users are in *defense*, 15.24% are in *undisclosed*, and 24.90% are in *opposition* (see Figure 4 for the distribution per country).

In summary, the prediction of profile characteristics yielded results that are on-par with the state-of-the-art in terms of stance and location, which we need to answer RQ1. To answer RQ2, we also need gender and age, so in the regression analysis we only considered profiles with predictions generated with high confidence.

### 5.1 Stance Expressiveness

Our first research question is: *how people make use of new technologies to express their positions on controversial issues?* In other words, we want to understand how new forms of communication on the Web enable both, prediction of stance, but also expression, in the sense of explicit adherence to a stance.

Figure 5 shows the top-50 most important features for stance classification, from a total of 1,602 features. Two different emojis appear on the top-50 (♥ & ♥), in four different features, as they



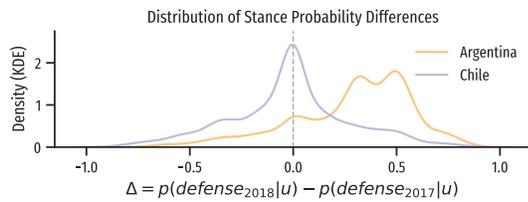
**Figure 5: Relevant features for stance prediction. The larger bar chart includes the top-50 features, the embedded bar charts show the frequency of each feature type, and each type’s mean and standard deviation of the corresponding importance.**

can appear as terms in tweet content, and as terms in a profile’s name. Since feature importance alone does not indicate association with a stance, Table 5 shows the top-30 features associated with each country and stance (including *undisclosed*). The table confirms that the green heart is associated with *defense*, whereas the blue heart is associated with *opposition*, which are the colors adopted by activists to symbolize their position on abortion stance in physical manifestations in Latin America [38]. Other terms show expected associations, such as a focus on *women* (*mujeres*) and the *right to choose* (*decidir*) in *defense*, or the focus on *life/death* (*vida/muerte/matar*), and the *right to be born* (*nacer*) in *opposition*. Features associated with *undisclosed* are mixed, with a diverse set of generic profile features.

Figure 5 contains two additional charts, with frequency and mean importance per type of feature: emojis, network features (interactions), hashtags, terms in tweets and profile biography, and URLs. The frequency of emojis is smaller than most of the other features (29), with exception of URLs. However, on average, it is one of the most important (mean XGB importance: 0.00234). To validate whether there are differences in the main features, we applied the multiple comparison Tukey’s HSD test. The differences between emojis and the rest of the feature types are all significant (rejecting the null hypothesis of no differences, with  $p < 0.001$ ), while the difference between other types is not statistically different, with the exception of tweet terms with network interactions and hashtags. Hence, even though few emojis are related to stance, their

**Table 5: Top-30 features for stance classification weighted by association to each country.**

Country	Defense	Opposition	Undisclosed
Argentina	♥, legal, clandestino (clandestine), seguro (safe), gratuito (free), sexual, morir (to die), decidir (to decide), anticonceptivos (contraceptive), educación (education), abortar (to abort), clandestinidad (clandestinity), mujeres (women), años (years), pibas (girls)	💙, matar (to kill), #aborto (#abortion), profile:n_emojis, vida (life), solución (solution), @mauriciomacri, Macri (Argentinian President), nacer (to born), humano (human), niño (kid), @marianoobarrio, abortistas (abortists), asesinato (murder), muerte (death)	profile:n_emojis, [profession words in the profile], [social media words in the profile], [education words in the profile], [family words in the profile], tema (topic), católico (catholic), profile:siempre (always), profile:💙, timezone:Santiago, aborto (abortion), profile:periodista (journalist), ateo (atheist), Cristina (former Argentinian President), favor (favor)
Chile	mujeres (women), legal, causales (grounds), ♥, decidir (to decide), tres (three), despenalización (decriminalization), seguro (safe), gratuito (free), derechos (rights), UDI (far-right Chilean political party), años (years), clandestino (clandestine), protocolo (protocol), bit.ly (link sharing URL)	vida (life), 💙, muerte (death), izquierda (left), niños (kids), asesinato (murder), matar (to kill), ideología (ideology), nacer (to born), negocio (biz), humano (human), aborto (abortion), niño (kid), inocentes (innocent), crimen (crime)	timezone:Santiago, profile:n_emojis, [profession words in the profile], [education words in the profile], estaría (would be), [family words in the profile], profile:música (music), profile:periodista (journalist), profile:siempre (always), provida (profile), profile:gusta (like), [social media words in the profile], profile:fanático (male fan of), gente (people), profile:ex



**Figure 6: Probability distribution of turnaround.**

predictive power is higher than language associated with abortion stances, and higher than other types of features.

As answer to RQ1, we observe that emojis significantly express stance in a way that mirrors physical manifestations, either as a symbol of adherence (in the name), or as a symbol of support in discussion (in tweet content). We discuss this further in Section 6.

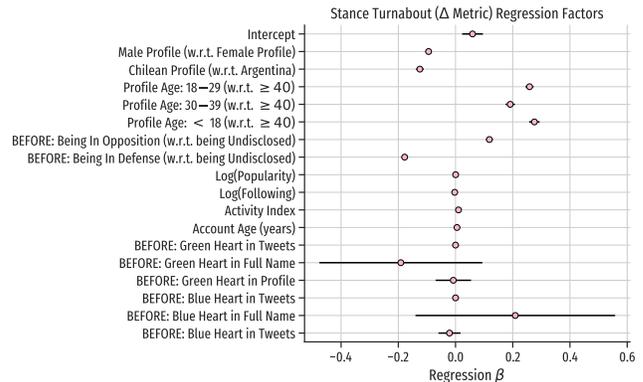
### 5.2 Stance Turnarounds

Our second research question is: *which demographic and profile factors characterize opinion change?* To answer it, we study two time windows of three months in two consecutive years: May, June, and July, in 2017 and 2018. During these months in 2017, the Chilean Senate approved the abortion bill proposed by President Bachelet (see Figure 1). One year later, a free abortion bill was being discussed in Argentina, during the government of President Macri. Given that the new Chilean law allowed abortion on three grounds, the possible approval of a free abortion in Argentina was likely to influence the following discussion in Chile. Thus, we hypothesize that the discussion in Argentina during 2018 was a moment to express a potentially updated view on abortion, with respect to the previous year, particularly for Chileans, but also for Argentinians who participated in the 2017 debate.

In our dataset, 11,931 users participated in the discussion during both periods, and for whom we have a predicted age and gender with high confidence. In this subset, the proportion of Chilean accounts is greater than in the full dataset (31% instead of 14%), as Argentinian users tended to participate in the discussion only in 2018. Figure 6 shows the distribution of probability differences for the users appearing in both periods. Chilean users present a symmetric distribution, suggesting that the majority of users maintained their

Stance in 2017	Argentina			Chile		
	Opposition	Undisclosed	Defense	Opposition	Undisclosed	Defense
Opposition	26	11	63	64	15	21
Undisclosed	14	12	73	32	26	42
Defense	8.3	8.6	83	15	19	66
	Opposition	Undisclosed	Defense	Opposition	Undisclosed	Defense
	Stance in 2018			Stance in 2018		

**Figure 7: Stance turnarounds per country.**



**Figure 8: Regression factors and their confidence intervals from stance turnaround regression.**

views (see Figure 7), whereas Argentinian users exhibit a clear tendency to change *toward defense*.

To infer which demographic and content factors were important in determining a turnaround, we performed a linear regression over the turnaround metric  $\Delta$ . The independent variables of the regression included the demographic factors predicted for each user and profile characteristics, such as number of followers, followees, activity ratio, account age, and the usage of colored emojis. The number of followers, followees and activity ratio were log-transformed due to their skewed distributions.

The model exhibited a good fit of the data (adjusted  $R^2 = 0.397$ ,  $MSE = 0.1011$ ,  $F = 524.2$ ,  $p < 0.001$ , Log-Likelihood  $LL = -458.70$ ). According to the adjusted  $R^2$  coefficient, the model explains 40% of the variance in stance turnaround. Figure 8 shows the regression

factors with their confidence intervals. As some variables were categorical, the regression used dummy variable coding, and thus, some categories are reference values that are encoded in the intercept. Keeping in mind that positive factors indicate leaning toward *defense*, and negative factors indicate leaning toward *opposition*, one can see that males tend to turnaround toward *opposition* (male  $\beta = -0.0942$ ) in comparison to females. Also, as people gets older, their turnaround difference changes from toward *defense* to *opposition* (from  $\beta = 0.2757$  for  $< 18$ ,  $\beta = 0.2587$  for  $18-29$ , and  $\beta = 0.1912$  for  $30-39$ , with  $\geq 40$  as reference).

Even though the distribution of differences in Chile is symmetrical, in comparison to Argentina, Chileans lean toward *opposition* ( $\beta = -0.1624$ ). Network metrics (popularity and following) are not significant; conversely, the amount of activity as well as the age of the account are positively related with turnaround toward *defense* ( $\beta = 0.0103$  and  $\beta = 0.0051$ , respectively), meaning that active and old accounts are more likely to be in *defense*. Surprisingly, the usage of emojis was only significant for  in tweet content during the first period ( $\beta = 0.0051$ ).

To conclude, stance turnaround happens as shown in the probability distribution in Figure 6. Moreover, publicly available features (profile characteristics and emoji usage) plus other predictable attributes (demographics) provide a solid baseline to analyze stance turnaround in quantitative terms as shown in Figure 8, effectively answering our research question.

## 6 DISCUSSION

We discuss the main implications of this work, in terms of *technological means of expression* and *stance turnarounds and demographic composition*. We also discuss the limitations of our work.

**Technological Means of Expression.** The identification of political stance through color is not new. Women’s causes in the last century have been associated with green, white, and purple, expressed through clothing accessories [33]. However, in non-Spanish speaking countries these causes did not include abortion rights. In Latin America, the usage of colors also emerged from feminist movements [38], and, as our results indicate, this phenomenon is manifested in web platforms, not only for feminist positions regarding abortion () , but also for opposition to abortion rights (). In the last decade, emoji input has been a common feature in modern user interfaces, although they have been commonly used as a emotional device rather than a political one. Compared to other web-based means of expression, emojis have emerged in the abortion discussion using out-of-domain symbolism. For instance, hashtags tend to include concepts, dates, and mottos relevant to the issue under discussion, and are indeed related with protests [37]. Conversely, these emojis have an arguably neutral appearance (a metaphorical heart) and a color that is not necessarily associated with a political party [38]. Their current prevalence, and the adoption of younger generations of emojis as a natural device to engage in conversation, imply that future studies should reconsider whether text-only insights still apply, and whether removing non-domain/non-textual context eliminates powerful signals.

**Stance, Turnarounds and Demographic Composition.** Abortion is also a public health issue, and domain experts in health have pointed out that, even though the Twitter population is biased [1],

if done carefully, Twitter allows to measure what “the public is *actually* seeing,” an understanding that would help to design communication strategies regarding abortion legislation and medical practice [23]. Previous work has improved the understanding of *who* is this public, by measuring its representativeness of the general population [19]; here we observed that demographic factors and profile characteristics explain a significant amount of the variance in stance turnaround between two specific periods. A main implication of this result is that the proposed methodology can be applied to measure how people react to specific events, which may be political, but may also be communication strategies to improve views on public health interventions.

**Limitations.** Our work is scoped in several aspects. First, the lack of ground truth makes our evaluation strictly based on the data labeled or inferred from self-reported information. Moreover, as the Twitter Streaming API is a sample, the dataset does not cover the entire discussion on the platform. However, our previous work showed that insights derived from this dataset match those from nationally representative surveys [18]. Second, our methodology needs further validation. Individual steps can be evaluated, such as a formal comparison of the classifier with other approaches [39], and measuring the classifier performance in turnaround scenarios with labeled data. Finally, our main insight is the identification of previously unseen features, but we cannot rule out the possibility that other important features were excluded due to our seed keyword identification processes. A sensitivity analysis of seed keywords should be performed.

## 7 CONCLUSIONS

In this paper we characterized the expressiveness and turnarounds of stances in controversial issues. We did so by performing a longitudinal analysis of the abortion debate in two Latin-American countries, Argentina and Chile. We empirically quantified these phenomena using established methods from the literature, finding that, even though abortion is not a new topic, the resources being employed to take a stance and express it are evolving, as the Web does; and that it is possible to characterize changes in opinion according to demographic factors and profile characteristics. 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, but also the identification of novel ways in which people express their stance. The inclusion of demographic and profile characteristics on turnaround analysis show that it is possible to evaluate the effect on stance of legislative events, and potentially other types of interventions, such as communication strategies for public health awareness.

In 2010, in the rise of Twitter and other social platforms, a famous article from The New Yorker stated that the revolution will not be tweeted [17]. Times have changed: our results show that these platforms reflect important aspects of physical manifestation in events that span multiple years, and that people adopt new means to express themselves and communicate their positions. The Web has the potential to act as a positive influence on how people relate to others, and future work should look into how new & old ways of communication (such as emojis, memes, and short videos)

are related to people's political beliefs, and how people of opposing views could be brought together through them.

**Acknowledgments.** We thank the authors of the twemoji package (CC-BY4), Ana Freire for insightful discussion, and the anonymous reviewers for their insightful comments.

## REFERENCES

- [1] Ricardo Baeza-Yates. 2018. Bias on the Web. *Commun. ACM* 61, 6 (May 2018), 54–61.
- [2] Francesco Barbieri, German Kruszewski, Francesco Ronzano, and Horacio Sag-gion. 2016. How cosmopolitan are emojis?: Exploring emojis usage and mean-ing over different languages with distributional semantics. In *Proceedings of the 24th ACM international conference on Multimedia*. ACM, Amsterdam, The Nether-lands, 531–535.
- [3] Amy Booth. 2018. Argentina votes on bill to legalise abortion up to 14 weeks. *The Lancet* 391, 10138 (2018), e21–e22.
- [4] Dallas Card and Noah K Smith. 2018. The Importance of Calibration for Esti-mating 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.
- [5] Z e Carpenter. 2019. This Was the Decade of Feminist Uprisings in Latin Amer-ica. TheNation.com [Online; accessed May/15/2020].
- [6] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A scalable tree boosting sys-tem. In *Proceedings of the 22nd ACM SIGKDD international conference on knowl-edge discovery and data mining*. ACM, San Francisco, USA, 785–794.
- [7] Zhenpeng Chen, Xuan Lu, Wei Ai, Huoran Li, Qiaozhu Mei, and Xuanzhe Liu. 2018. Through a gender lens: Learning usage patterns of emojis from large-scale Android users. In *Proceedings of the 2018 World Wide Web Conference*. ACM, Lyon, France, 763–772.
- [8] Tom S Clark, Jeffrey K Staton, Yu Wang, and Eugene Agichtein. 2018. Using Twitter to Study Public Discourse in the Wake of Judicial Decisions: Public Re-actions to the Supreme Court's Same-Sex-Marriage Cases. *Journal of Law and Courts* 6, 1 (2018), 93–126.
- [9] Raviv Cohen and Derek Ruths. 2013. Classifying political orientation on Twitter: It's not easy!. In *Seventh International Conference on Weblogs and Social Media*. AAAI, Cambridge, USA, 91–99.
- [10] Michael D Conover, Bruno Goncalves, 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.
- [11] Bjarke Felbo, Alan Mislove, Anders Sogaard, Iyad Rahwan, and Sune Lehmann. 2017. Using millions of Emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In *Proceedings of the 2017 Con-ference on Empirical Methods in Natural Language Processing*. ACL, Copenhagen, Denmark, 1615–1625.
- [12] Yunhe Feng, Wenjun Zhou, Zheng Lu, Zhibo Wang, and Qing Cao. 2019. The World Wants Mangoes and Kangaroos: A Study of New Emoji Requests Based on Thirty Million Tweets. In *The Web Conference*. ACM, San Francisco, USA, 2722–2728.
- [13] 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, Troy, USA, 43–52.
- [14] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2018. Political Discourse on Social Media: Echo Chambers, Gate-keepers, 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.
- [15] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2018. Quantifying controversy on social media. *ACM Transac-tions on Social Computing* 1, 1 (2018), 3.
- [16] 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, Montr al, Qu bec, Canada, 528–531.
- [17] Malcolm Gladwell. 2010. Why the revolution will not be tweeted. *The New Yorker* 4 (2010), 1–9.
- [18] 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, Boston, USA, 133–134.
- [19] Eduardo Graells-Garrido, Ricardo Baeza-Yates, and Mounia Lalmas. 2020. Rep-resentativeness of Abortion Legislation Debate on Twitter: A Case Study in Ar-gentina and Chile. In *Companion Proceedings of the Web Conference 2020*. ACM, Taipei, Taiwan, 765–774.
- [20] Eduardo Graells-Garrido, Mounia Lalmas, and Ricardo Baeza-Yates. 2015. Find-ing 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.
- [21] Kate Greasley. 2017. *Arguments about abortion: Personhood, morality, and law*. Oxford University Press, Oxford, UK.
- [22] Asmelash Teka Hadgu, Kiran Garimella, and Ingmar Weber. 2013. Political hash-tag hijacking in the US. In *Proceedings of the 22nd International Conference on World Wide Web*. ACM, Rio de Janeiro, Brazil, 55–56.
- [23] Leo Han, Lisa Han, Blair Darney, and Maria I Rodriguez. 2017. Tweeting PP: An analysis of the 2015–2016 Planned Parenthood controversy on Twitter. *Contra-ception* 96, 6 (2017), 388–394.
- [24] 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 *Pro-ceedings of the SIGCHI conference on human factors in computing systems*. ACM, Vancouver, Canada, 237–246.
- [25] Chaya Liebeskind and Shmuel Liebeskind. 2019. Emoji prediction for Hebrew political domain. In *Companion Proceedings of The 2019 World Wide Web Confer-ence*. ACM, San Francisco, USA, 468–477.
- [26] Haokai Lu, James Caverlee, and Wei Niu. 2015. Biaswatch: A lightweight sys-tem for discovering and tracking topic-sensitive opinion bias in social media. In *Proceedings of the 24th ACM International Conference on Information and Knowledge Management*. ACM, Melbourne, Australia, 213–222.
- [27] Hannah Jean Miller, Jacob Thebault-Spieker, Shuo Chang, Isaac Johnson, Loren Terveen, and Brent Hecht. 2016. "Blissfully Happy" or "Ready to Fight": Varying Interpretations of Emoji. In *Tenth International Conference on Web and Social Media*. AAAI, Cologne, Germany, 259–268.
- [28] 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.
- [29] Adela Montero and Ra l Villarreal. 2018. A critical review of conscientious ob-jection and decriminalisation of abortion in Chile. *Journal of medical ethics* 44, 4 (2018), 279–283.
- [30] Alexandru Niculescu-Mizil and Rich Caruana. 2005. Predicting good probabili-ties with supervised learning. In *Proceedings of the 22nd International conference on Machine Learning*. ACM, Bonn, Germany, 625–632.
- [31] Marco Pennacchiotti and Ana-Maria Popescu. 2011. A machine learning ap-proach to Twitter user classification. In *Fifth International Conference on Weblogs and Social Media*. AAAI, Barcelona, Spain, 281–288.
- [32] Carol Sanger. 2016. Talking About Abortion. *Social & Legal Studies* 25, 6 (2016), 651–666.
- [33] Marian Sawer. 2007. Wearing your politics on your sleeve: The role of political colours in social movements. *Social Movement Studies* 6, 1 (2007), 39–56.
- [34] Eva Sharma, Koustuv Saha, Sindhu Kiranmai Ernala, Sucheta Ghoshal, and Mun-mun De Choudhury. 2017. Analyzing Ideological Discourse on Social Media: A Case Study of the Abortion Debate. In *Proceedings of the 2017 International Con-ference of The Computational Social Science Society of the Americas*. ACM, Santa Fe, USA, 3.
- [35] 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.
- [36] 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.
- [37] Zachary C Steiner-Trelkeld, Delia Mocanu, Alessandro Vespinani, and James Fowler. 2015. Online social networks and offline protest. *EPJ Data Science* 4, 1 (2015), 19.
- [38] Barbara Sutton and Nayla Luz Vacarezza. 2020. Abortion Rights in Images: Vi-sual Interventions by Activist Organizations in Argentina. *Signs: Journal of Women in Culture and Society* 45, 3 (2020), 731–757.
- [39] 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, San Francisco, USA, 2056–2067.
- [40] Sarita Yardi and danah boyd. 2010. Dynamic debates: An analysis of group po-larization over time on Twitter. *Bulletin of Science, Technology & Society* 30, 5 (2010), 316–327.
- [41] 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.
- [42] Qiang Zhang, Shangsong Liang, Aldo Lipani, Zhaochun Ren, and Emine Yilmaz. 2019. From Stances' Imbalance to Their Hierarchical Representation and Detec-tion. In *The World Wide Web Conference*. ACM, San Francisco, USA, 2323–2332.