Analyzing Ideological Discourse on Social Media: A Case Study of the Abortion Debate

Eva Sharma*, Koustuv Saha*, Sindhu Kiranmai Ernala*, Sucheta Ghoshal*, Munmun De Choudhury
Georgia Institute of Technology
Atlanta, Georgia 30308, United States
(evasharma,koustuv.saha,sernala3,sucheta,munmund)@gatech.edu

ABSTRACT

Social media provides a unique platform enabling public discourse around cross-cutting ideologies. In this paper, we provide a methodological lens for studying the discourses around the controversial topic of abortion on social media. Drawing from the theoretical framework of "Critical Discourse Analysis", we study discourse around abortion on Twitter through analysis of language and the manifested socio-cultural practices. First, employing a large dataset of over 700 thousand posts, we find that abortion discourse can be classified into three ideologies: For, Against, and Neutral to Abortion. We observe these ideological categories to be characterized by distinctive textual and psycholinguistic cues. Finally, we analyze the nature of discourse across ideologies against the backdrop of socio-cultural practices associated with abortion. Our findings reveal how the hegemonic nature of the rhetoric that has historically shaped the abortion debate in society is reconceptualized on Twitter. We discuss the role of social media as a public sphere that shapes critical discourse around controversial topics.

CCS CONCEPTS

• **Human-centered computing** → *Empirical studies in collaborative and social computing*; • **Social and professional topics** → *Cultural characteristics*;

KEYWORDS

social media, twitter, abortion debate, public sphere, critical discourse analysis

ACM Reference format:
Eva Sharma*, Koustuv Saha*, Sindhu Kiranmai Ernala*, Sucheta Ghoshal*, Munmun De Choudhury. 2017. Analyzing Ideological Discourse on Social Media: A Case Study of the Abortion Debate. In *Proceedings of The Computational Social Science Society of the Americas, Santa Fe, NM, October 2017 (CSS'2017),* 8 pages.
https://doi.org/10.475/123_4

1 INTRODUCTION

An active "public sphere" is a crucial element of social, political, and cultural change. Since its emergence in the ancient Greek agoras, public sphere has served as a facilitator of informed and logical discussion around a variety of societal topics [23]. The seminal work of communication scholar Gerard A. Hauser notes that, by coming together to freely discuss and identify contentious problems, individuals can reach a common judgment, form public opinion, and influence collective action, policy, and decision-making [25].

Social media platforms such as Twitter have emerged as prominent forums promoting open and democratic exchanges around many controversial topics. Many have argued these platforms to be extending the public sphere due to their ability to facilitate exchange of opinions [29, 42, 44]. Research has examined controversial topics on social media around policy change and activism, such as abortion [52], gun control [3], climate change [43], Lesbian Gay Bisexual Transgender (LGBT) rights [40, 51], and racial inequality [11]. However, to our knowledge, few empirical studies have examined the intricacies engendering the expression of diverse, politically charged, socio-culturally complex, and often stigmatized viewpoints in social media discourse like these, and how they relate to the societal context at large.

In this paper, we advance prior work by presenting computational methods to analyze how public discourse around controversial topics is being re-conceptualized in social media. Specifically, we examine the socio-cultural practices around the controversial topic of abortion on Twitter. We choose the abortion debate given its historical significance and recent resurgence with the ruling of U.S Supreme court striking down Texas abortion restrictions (June 2016) [37], the Poland protests (October 2016) [38] and the latest GOP health bill defunding Planned Parenthood for a year [39]. The abortion debate has contributed to polarization of ideologies in public discourse through the years, around issues ranging from the personhood of a fetus, to moralities around motherhood [41]. These viewpoints contained in the moral discourse of abortion contribute to significant ideological differences about the rightness and wrongness of abortion. Moreover, abortion has multiple facets in its debate, such as political, religious, medical, legal and so on [36]. These multitude of facets bring together diverse viewpoints in public discourse from varied groups. Twitter caters to such diverse audience enabling them to participate in the abortion discourse, setting the stage for the study presented in this paper.

To study abortion discourse on Twitter, we draw from social theories of critical discourse, specifically the theoretical framework of Critical Discourse Analysis (CDA) proposed by Norman Fairclough [14]. This theory allows us to understand how the controversial topic of abortion is discussed on social media, and thereby understand how existing offline world hegemonic discourses around this topic are manifested online. We address the following two research questions:

* These authors contributed equally.

CSS'2017, October 2017, Santa Fe, NM

ACM ISBN 123-4567-24-567/08/06…$15.00
https://doi.org/10.475/123_4

RQ1: What linguistic attributes characterize different ideological perspectives around the abortion debate on Twitter?

RQ2: How do the different ideological perspectives on Twitter re-conceptualize and reproduce the offline socio-cultural practices associated with abortion?

Towards our research goals, we use a large dataset of over 700 thousand public posts shared on Twitter around the topic of abortion. We observe that the abortion discourse on Twitter manifests via three contrasting ideological perspectives: *For Abortion*, *Against Abortion*, and *Neutral to Abortion*, which can be automatically and accurately identified with a machine learning based classification framework. However, we find notable imbalance in the expression of these ideologies: the discourse on *Against Abortion* is thrice as much as *For Abortion*. Analyzing linguistic cues of these three ideological stance, we observe significant differences: for instance, *Against Abortion* expresses greater death and familial concerns compared to *For Abortion*, while the latter manifests a more prominent collective identity. Finally, studying these ideological discourses within the context of socio-cultural processes centering around abortion in the society, our approach discovers several political, ethical and institutional facets of the debate on Twitter.

Our work introduces a methodological "lens" bridging the theoretical framework of CDA and computational large-scale data analysis to study discourse on controversial topics, such as abortion, on social media. Thereby, we discuss how our insights can improve our understanding of the role of social media as a public sphere that shapes discourse on contentious issues of the society.

2 BACKGROUND AND PRIOR WORK

2.1 Ideological Discourse on Abortion

Controversial topics of societal significance are likely to spark many ideological discourses around them. The socio-cognitive theory of ideology, describes ideologies as shared mental representations of social groups [49]. Relevant to the case of abortion, the moral, ethical, and cultural values around this topic have given rise to two prominent ideologies: the 'pro-choice' advocating for abortion rights and the 'pro-life' condemning abortion. According to a 2016 Gallup report [19], 47% and 46% respondents in the US reported "prochoice" and "prolife" stances respectively. Discourses around abortion have been emphatically hegemonic shaped by the mainstream, dominant cultural ideas about female sexuality [26]. The ethical and moral dilemmas with the status of the fetus, and the notion of taking a life complementarily contribute to this culturally hegemonic discourse, triggering the contrasting ideologies around it. For instance, pro-life activists adopt the rhetoric of fetal personhood, and advocate that abortion ends a human life. This ideology has consistently advanced the principle that the practice should be stopped, just like any other form of unjustified killing [24]. Challenging the mainstream pro-life rhetoric, feminist theorist Adrienne Rich states, "Arguments against abortion have in common a valuing of the unborn fetus over the living woman" [41]. Given this polarizing context, increasing enforcement of legal constraints and regulations on abortion led the women's rights movements of 1960s to advocate more for reproductive rights, and to this day, women's right to safe affordable abortion remains as a key subject of discourse around it [10]. Seeking to understand how these regulations and legal decision making on abortion impact the societal practices around abortion, prior work has also studied discourses on sex, motherhood, and abortion [26]. We extend this line of work by analyzing the abortion discourse on Twitter and how it shapes the ideological divide.

2.2 Understanding Public Discourse: Critical Discourse Analysis

Researchers have adopted a number of theories and methods to study and understand public discourse around controversial topics [12, 48]. Among the theories related to power and ideology, Michel Foucault's formulations of "order of discourse" and "power-knowledge" [18] and Antonio Gramsci's notion of "cultural hegemony" [21] have been widely adopted. Notably, in the late 1980s, Fairclough, Wodak, and van Dijk contributed to the development of the "Critical Discourse Analysis" (CDA) framework [15], building on social science theories to examine ideologies and power relations in public discourse [5]. Chouliaraki and Fairclough state that, "It is an important characteristic of the economic, social and cultural changes of late modernity that they exist as discourses as well as processes that are taking place outside discourse, and that the processes that are taking place outside discourse are substantively shaped by these discourses" [8].

CDA thus draws from critical theory of language, which considers use of language as a form of social practice and regards the context of language usage crucial for discourse [14]. More formally, it provides a framework for analyzing discourse which consists of three inter-related processes of analysis tied to the respective dimensions of discourse—discourse as text, discourse as a discursive practice, and discourse as a socio-cultural practice.

Consequently, CDA lends itself as a suitable theoretical framework to understand online discourse. However, there is limited work that applies CDA to online discourse studies, except the works of Tornberg et al. and Lidskog et al. [27, 47]. Since it is known that the historical, political, and social significance as well as the ethical and moral debates associated with the topic of abortion have shaped its discourse through centuries [31], in our work, we adopt the theoretical framework of CDA as a way to analyze these dimensions and their complexities around the abortion debate on Twitter.

2.3 Ideology, Controversial Topics, and Social Media

In his 2011 book, "The Political Power of Social Media" [44], Clay Shirky asks the question: "Do digital tools enhance democracy?". Whether a specific digital tool like social media extends the public sphere around societal topics has piqued the interest of many scholars in recent years. Although empirical work on this subject is hard to come by, Shirky notes that social media platforms are making the landscape of public discourse denser and more participatory, and that the networked populations are gaining greater access to information, more opportunities to engage in public speech, and an enhanced ability to undertake collective action. In another study, Liu and Weber adopt quantitative techniques to find that social media platforms (especially Twitter) are not an ideal public sphere for democratic conversations and demonstrate hierarchical levels of communication [29].

In an early work, Schneider examined computer mediated communications as a democratic public sphere, through a case study on abortion [42]. In fact, social media discourse on controversial topics like issues of racial profiling and same-sex marriage have been extensively studied [1, 11, 30, 45]. Such discourse demonstrates two contrasting facets. On the one hand, it is affected by social stigma and hate speech where people often find it difficult to express their views, especially if their opinion is not popularly accepted [34]. On the other hand, it still provides a platform for marginalized views to be expressed, broadening public discourse and adding new perspectives to everyday discussion of contentious issues. This dichotomy makes examining the discourse on controversial topics of special interest.

Prior work studying ideological content around controversial topics on social media, such as Twitter, have also looked at analyzing the prevalence and nature of public opinion around policy change. Several studies have found evidence for moral culture wars between ideologies during policy changes in controversial topics like gun control, same-sex marriage, abortion [3, 51, 52]. Our work goes beyond identifying these ideological differences into studying how historical, socio-cultural processes around controversial topics like abortion are being re-conceptualized on social media. Further, to identify political ideologies on Twitter, some studies have combined network and content analyses [2, 9]. While we do not analyze networks, we draw on these observations by incorporating the richer context of the abortion discourse, beyond text, via analysis of its socio-cultural practices.

3 DATA AND METHODS

3.1 Abortion Data Acquisition

The source of data for this paper is Twitter. We utilized several manually curated seed hashtags in an iterative manner to obtain a sample of Twitter posts and associated metadata; these hashtags captured a variety of discussions on Twitter around abortion. We started with the hashtag *#abortion*, and then identified frequently co-occurring or trending hashtags related to it through a website called Hashtagify [1], which calls itself a hashtag search engine. Out of the 10 retrieved hashtags (ref: Table 1), we decided to use *#prochoice* and *#prolife* – the two hashtags with the highest correlation with #abortion, which in turn inspired the choice to append *#antilife* and *#antichoice* to our seed list of hashtags to capture contrary ideologies on the abortion debate. Thus, our final set of seed hashtags included: *#abortion*, *#prochoice*, *#prolife*, *#antichoice*, and *#antilife*.

Next, using these hashtags as search query terms, we collected Twitter data consisting of *tweet id*, *text*, *username*, *date*, *hashtags*, *geo-location*, *mentions*, and number of *re-tweets* and *favorites*. Our final dataset contains 731,080 tweets posted between January 2015 and September 2016 by 104,433 unique users (mean 7 tweets per user). The overall descriptive statistics are reported in Table 2 and Figure 2 shows distribution of tweets over unique users. Additionally, we found that 93,374 unique hashtags co-occurred with the seed hashtags, and Figure 1 shows the distribution of the top 10 of these hashtags.

3.2 Classifying Ideologies on Abortion

[1] http://hashtagify.me/hashtag/abortion Accessed: 2017-02-16

Table 1: Top 10 hashtags and correlation with *#abortion.*

Hashtag	% Corr.
#prolife	24.0
#prochoice	11.7
#tcot	9.4
#PlannedParenthood	3.9
#UniteBlue	3.0
#PJNET	2.5
#WarOnWomen	2.3
#Gosnell	2.2
#fem2	2.0
#p2	1.9

Table 2: Descriptive statistics of the Twitter dataset.

Statistic	Value
# Unique Hashtags	93,374
# Tweets	731,080
# Unique Users	104,433
μ Tweets per User	7.00
Mdn Tweets per User	1.00
σ Tweets per User	121.24
# Unique Hashtags	93,374
# Retweets	1,242,099
# Favorites	1,230,056
# @-Mention Tweets	241,316

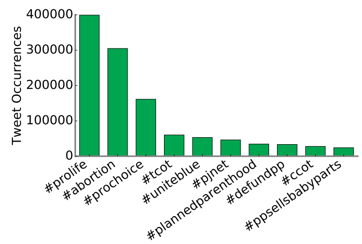

Figure 1: Top 10 hashtags by number of occurrences.

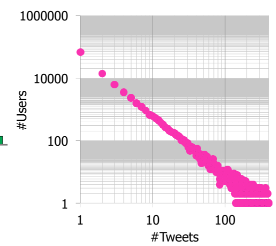

Figure 2: Distribution of number of users by number of tweets.

Table 3: Ideological categories and two example memos per qualitative coding.

Code	Example Memo
For	Occurrence of "Horrible Anti-Abortion"; Co-occurrence of #shoutyourabortion and #StandWithPP
Against	#UnbornLivesMatter when expressed in Positive Sentiment; Co-occurrence of #ppsellbabyparts and #planned-butcherhood
Neutral	Occurrence of #mustread; Co-occurrence of #abortion and #TopNews

3.2.1 Qualitative Coding. Addressing our two RQs necessitates identifying the various ideological perspectives manifested in abortion discourse in our collected Twitter data. Since the tweets in our dataset did not include ground truth information on their ideological stance, we used qualitative coding. One human rater familiar with the abortion debate first examined a random sample of 200 tweets from our dataset using an open coding approach [20], followed by employing an iterative process to categorize different tweets into "codes" relevant to our study [28]. We generated a set of hand-coded rules, that we refer to as "memos", to create a codebook; this codebook contained the definitions of the codes, their correlations, and specific examples. Table 3 shows a sample of the memos from a total of 73 memos in our codebook. We then applied the codebook to code a second sample of 200 tweets into: *For Abortion* (tweets that voice support for abortion), *Against Abortion* (tweets that argue against the practice of abortion), and *Neutral to Abortion* (tweets that do not express an explicit stance on the issue). Since these codes align with three distinctive stances on the abortion debate, henceforth we refer to them as ideological categories.

3.2.2 Classification Framework. Using all of the above hand-labeled tweets (400 in total) as training data, we build a classification model to classify tweets into the three ideological categories around abortion—*For Abortion*, *Against Abortion*, and *Neutral to Abortion*. The classifier uses the following features:

(1) *n*-gram language model, that includes the top 5000 unigrams and bigrams extracted from the tweet text data. We represent each tweet as a feature vector of the normalized frequency counts of these *n*-grams.
(2) Memos from the above qualitative coding task, where, for a given tweet, each memo is featurized based on the following rules: a) if the tweet contains a particular phrase or hashtag noted in the memo; b) if the tweet contains co-occurring phrases or hashtags characteristic of the memo; and c) sentiment [2] of the tweet in combination with the aforementioned rules. This way, we include the memos (ref: Table 3) as binary features for each tweet.

With these features, we train different classifiers—Support Vector Machine (SVM) with a linear kernel, and a Random Forest classifier. For our analysis, we use the best performing model. We employ balanced class weights (to handle disproportionate distribution of the three ideological categories in training data). For parameter tuning we employ k-fold cross validation ($k = 10$) and then apply the trained model to an unseen held out dataset. With this classifier, we then machine label the remaining 730,680 tweets in our dataset to belong to one of the three categories—*For Abortion*, *Against Abortion*, and *Neutral to Abortion*.

3.3 Linguistic Characterization of Ideological Categories

Recall that our first research question (RQ1) characterizes the linguistic attributes in abortion discourse on Twitter. To do so, we use two approaches:

3.3.1 Psycholinguistic Characterization. First, we seek to characterize the text of the ideological categories from a psycholinguistic perspective. For this, we employ Linguistic Inquiry and Word Count, or LIWC [46]. Borrowing from prior work [11], we use the following LIWC categories that we deemed most relevant to understanding the language of the abortion debate: (1) *affective attributes* (categories: anger, anxiety, sadness, swear), (2) *cognitive attributes* (categories: cognitive mech, discrepancies, inhibition, negation, causation, certainty, and tentativeness), (3) *linguistic style attributes* (categories: past, present, future tense, first, second, third person, indefinite pronoun, article, adverb, verb, aux verb, preposition), and (4) *social/personal concerns* (categories: family, friends, social, health, bio, body, death, humans, religion, sexual).

3.3.2 Hashtag Usage. Hashtags are often used on Twitter for signaling and discoverability purposes [6]: they are therefore of prime importance in content production and distribution. Hence, we examine the uniqueness of hashtags to examine their usage in the abortion debate. For this purpose, we use a language differentiation technique known as Sparse Additive Generative Models of Text, or SAGE [13]. SAGE is a generative model of text where each class label or latent topic (here *For Abortion*, *Against Abortion*, or *Neutral to Abortion*) is endowed with a model of the deviation in log-frequency from a constant background distribution. With a language model of vocabulary size 1000, we utilize SAGE to identify the highly used distinctive hashtags from the three ideological categories.

3.4 Socio-Cultural Practices of Ideological Categories

Our second research question (RQ2) studies the socio-cultural practices of the three ideological categories. For the purpose, we utilize an unsupervised language modeling approach, specifically a topic model [4] to identify topics corresponding to each category, and then an iterative inductive open coding method on top of these topics to extract broad themes that capture the socio-cultural practices associated with abortion.

3.4.1 Topic Modeling. We employed MALLET[3] to build a topic model on our tweets. The topic models extracted, include latent semantically coherent clusters of words. Using the default parameters of MALLET for building topic models for each ideological category, we obtained 40 topics per ideology.

3.4.2 Theme Extraction. Following topic modeling, we sought to identify semantically interpretable broader themes that describe the abortion discourse. For this purpose, we employed a human rater familiar with the abortion debate. Using inductive open coding, the rater went through each topic to come up with a set of topical descriptors while consulting various external information and news sources on abortion. Next, these descriptors were iteratively revised to develop a coherent descriptor vocabulary. Finally, the rater combined these descriptions for each ideological category, identifying 16 overall themes across *For Abortion*, *Against Abortion*, and *Neutral to Abortion* ideological categories.

4 RESULTS

4.1 Classifying Ideologies on Abortion

We begin by presenting the performance metrics of our ideology classifier. Our best performing classifier, an SVM model is trained on labeled data of 400 tweets as described in the previous section. Table 4 reports the performance metrics of our classifier, as evaluated using a k-fold cross-validation ($k = 10$) (ref: Fig. 3). Our classifier achieves a mean accuracy of 67% and best accuracy of 81% on the training data, which is better than the baseline accuracy of 41% (a baseline model is one in which all tweets are labeled as the majority class). This improvement in accuracy over the baseline demonstrates that the content of tweets has a meaningful contribution in predicting ideological stance. The rich domain relevant tokens from language model (ref: table 6) and the hand-labeled memo based features capture high level concepts characterizing the ideological categories, thus serving as good predictive features.

[2] We used Stanford sentiment analysis model (nlp.stanford.edu/sentiment) to classify a tweet's sentiment into positive, negative, or neutral.

[3] MAchine Learning for LanguagE Toolkit: mallet.cs.umass.edu

Table 4: Performance metrics of ideology classification based on k-fold cross-validation (k=10).

Metric	μ	σ	max.
Accuracy	0.67	0.18	0.81
Precision	0.67	0.15	0.78
Recall	0.65	0.16	0.77
F1-score	0.64	0.17	0.77

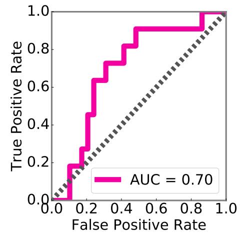

Figure 3: ROC curve of ideology classification.

Table 5: Distribution of tweets and users per ideology.

	Manual Labeled		Machine Labeled	
	# Tweets	# Users	# Tweets	# Users
For Abortion	131	97	144,824	38,908
Against Abortion	168	146	442,988	60,310
Neutral to Abortion	101	79	142,868	31,188
Total	400	322	730,680	130,406

Next, we use this classifier to machine label the ideological stance of the held-out 730,680 tweets. A random sample of 120 of these tweets were manually cross-verified by one human rater, which led to a classification accuracy of 76%, demonstrating consistent performance of the classifier in unseen data. We find that, out of 730,680 tweets (Table 5):

- *For Abortion* is expressed in 19.8% of the tweets. An example tweet in this category says, *"Antis can't decide bodily autonomy doesnt apply to pregnant people #prochoice=#prolife #ocra #abortion"*.
- *Against Abortion* occurs in 60.6% of the tweets. One such example tweet is, *"Sorry we r not in business of murdering unborn children. #DefundPP #ProLife"*.
- The remaining 19.5% are *Neutral to Abortion*, such as '*'Former DPP encouraged #abortion on demand: Britains abortion laws will be challenged"*.

We notice that the proportion of *Against Abortion* is significantly higher than that of other ideologies, indicating an imbalance in the abortion ideologies on Twitter. With these classified tweets, we now present the results for our RQs.

4.2 RQ1: Linguistic Characterization

In this subsection, we present the results of linguistic characterization of abortion discourse through psycholinguistic and hashtag analysis.

Psycholinguistic Analysis. Table 8 reports the mean values of different LIWC measures in the three ideological stances, including the outcome of Kruskal-Wallis significance tests comparing their mutual differences.

Starting with the LIWC measures under affective attributes, we observe that, tweets in *For Abortion* show higher occurrences of *anger* words (H = 547.4), as compared to the other ideological categories. These categories have tweets like, "I seriously hate people who r prolife. Not ur body. Not ur decision.", "#antichoice want to block #prochoice women's health care" expressing dissatisfaction

Table 6: Top features of ideology classifier (* p < .0001, ** .001 < p < .01, * .01 < p < .05). Scores in [1.3, 8.2].**

Feature	Score (p)	Feature	Score (p)
standwithpp	***	praytoendabortion	**
defundpp	***	unbornlivesmatter pjnet	**
supreme court	**	standwithpp prochoice	**
feminism	**	trust women	**
abortion uniteblue	**	domestic terrorism	**

Table 7: Highly used distinctive hashtags per ideology obtained using SAGE [13].

Ideology	Top Hashtags
For	prochoice, women, choice, standwithpp, rights, feminism, waronwomen, reprorights, support, sign
Against	prolife, defundpp, unbornlivesmatter, god, babies, parenthood, ccot, catholic, unborn, human
Neutral	abortion, court, bill, reproductivehealth, texas, scotus, trump, clinic, access, health, mustread

with other ideologies, likely because they perceive abortion as their right to choose.

Next, for the measures grouped under cognitive attributes, tweets in *For Abortion* include the highest occurrences of *cognitive mech.*, *negation*, *causation*, and *inhibition* words. One possible explanation could be that through *For Abortion* tweets, individuals bring in life histories, social interactions, and psychological predispositions in expressing their viewpoints. Personal accounts of life and social experiences are known to be associated with greater cognitive processing [35], e.g., as expressed in the tweet "It's so easy to say what a woman can do with her body when you're not a woman".

Under the different linguistic style attributes, the ideologies show distinctive interpersonal focus. *Against Abortion* demonstrates the highest social orientation through the use of second person pronouns (H = 1388.8), including a collective attentional focus indicated in the greater use of first person plural pronouns (H = 462.1). Both of these characteristics may be attributed to *Against Abortion* being the largest and dominant ideological category per our earlier observations (Table 5). In contrast, *first person singular pronoun* (H = 701.6) occurs the most in *For Abortion.* This indicates that users show high self-attention focus within *For Abortion* such as in tweets like, "Tell #antichoice politicians: you don't speak for me"; "I stand in #solidarity with #Polish women against total #abortion ban". The *For Abortion* tweets focus on the here and now, as observed through the use of present tense words. Moreover, high occurrence of *lexical density* words in *For Abortion* (adverbs, verbs and auxiliary verbs) indicates greater linguistic intricacies within this ideological expression on Twitter. This aligns with prior findings that argues the *For Abortion* dimension of the abortion debate to adopt a more complex narrative stance [12].

Finally, "social/personal concerns" measures demonstrate significant domain specific relevance to our work. For instance, several categories like *sexual*, *health* and *bio* are pertinent to the overall discourse on abortion and hence show high H-statistic values. Further, the measures of *religion* (H = 9024.9) and *death* (H = 1960.4)

Table 8: Results of Kruskal-Wallis tests comparing ideological categories *For Abortion* (*F*), *Against Abortion* (*A*) and *Neutral to Abortion* (*N*) for LIWC measures. Statistical significance reported after Bonferroni correction (α = .05/33).

Category	F	A	N	H-stat	p
Affective Attributes					
Anger	0.0118	0.0105	0.0086	547.43	***
Anxiety	0.0029	0.0019	0.0020	153.71	***
Cognitive Attributes					
Causation	0.0130	0.0101	0.0104	893.22	***
Certainty	0.0077	0.0081	0.0056	481.48	***
Cognitive Mech	0.0868	0.0756	0.0766	1683.7	***
Inhibition	0.0083	0.0069	0.0085	477.00	***
Negation	0.0134	0.0110	0.0075	1840.6	***
Interpersonal Focus					
1st P. Plural	0.0066	0.0067	0.0042	462.17	***
1st P. Singular	0.0101	0.0094	0.0065	701.65	***
2nd P.	0.0076	0.0087	0.0040	1388.8	***
3rd P.	0.0044	0.0036	0.0037	32.60	***
Temporal References					
Future Tense	0.0045	0.0055	0.0045	193.31	***
Past Tense	0.0070	0.0078	0.0067	112.23	***
Present Tense	0.0571	0.0472	0.0361	9821.8	***
Lexical Density and Awareness					
Adverbs	0.0188	0.0166	0.0165	240.04	***
Article	0.0233	0.0260	0.0235	841.59	***
Verbs	0.0744	0.0659	0.0518	8568.9	***
Auxiliary Verbs	0.0493	0.0417	0.0326	7532.5	***
Social/Personal Concerns					
Bio	0.0466	0.0421	0.0756	56434	***
Body	0.0045	0.0045	0.0023	485.92	***
Death	0.0036	0.0070	0.0033	1960.4	***
Family	0.0034	0.0074	0.0049	1711.6	***
Friends	0.0004	0.0009	0.0004	31.86	***
Health	0.0376	0.0340	0.0679	72910	***
Home	0.0018	0.0016	0.0018	15.94	***
Humans	0.0227	0.0197	0.0091	10509	***
Religion	0.0039	0.0177	0.0089	9024.9	***
Sexual	0.0329	0.0264	0.0619	93114	***
Social	0.0725	0.0707	0.0493	13111	***
Work	0.0403	0.0449	0.0407	2716.6	***

which are significantly high for *Against Abortion*, emphasize the pro-life stance adopted by these tweets around condemning abortion and arguing for lives of the "unborn": "millions of innocent babies slaughtered"; "life is sacred".

Hashtag Usage. Table 7 presents the top 10 most distinctive hashtags per ideological category as given by the SAGE technique. While there are some hashtags relating to the debate on abortion directly, such as *#prochoice* and *#prolife*; several other hashtags are unique to each of the three ideological perspectives, demonstrating their use as a mechanism of content production and distribution across the ideological spectra. For example, while on one hand, *For Abortion* talks mainly about feminism and women's rights through hashtags like *#waronwomen* and *#reprorights*, *Against Abortion* uses *#unbornlivesmatter*, *#defundpp* to disseminate opinions against abortion. Notably, the hashtags used in *Neutral to Abortion* stand out to show that this content is produced mainly by "informers" [33] who share perspectives, news, and external and popular opinions around the abortion debate (note hashtags *#mustread*, and *#scotus*). Taken together, we conjecture that tweets in both these ideological categories tend to use hashtags to advance their respective viewpoints and consolidate support for their own position, at the same time discrediting competing ones [32]. Further, content distributed through *#catholic*, and *#god* adds to our earlier observation of the presence of a religious rhetoric within the *Against Abortion*.

4.3 RQ2: Socio-Cultural Practices

Table 9 shows five major themes corresponding to each ideology that appeared from the human annotation based theme extraction task we applied on the results of topic modeling. To understand the relevance of identifying these major themes, we also report the top contributing topic words per theme. We observe distinctive differences in the thematic discourse of the three ideologies (see Figure 4). Our results suggest that the major themes revolve around socio-cultural practices ("Religious views", "Abortion is murder") or time sensitive news and offline events ("State bans") related to abortion. Further, we observe the presence of prominent themes like "PlannedParenthood" (PP) which reveal the institutional and organizational circumstances around abortion.

However, we find the existence of focused themes, which are used to propagate the socio-cultural arguments for a single ideological category. For instance, "Women's Rights" is a major theme in *For Abortion*, that has no occurrence in the *Against Abortion* category. For example, a *For Abortion* tweet related to this theme says: "*#antichoice wants to block #prochoice women's health care. Violation of reproductive rights hurts people.*". In contrast, "Abortion is murder" is recurrent in *Against Abortion* but has no occurrence in *For Abortion*—here is an example of *Against Abortion* tweet around this theme: "*All abortions are murder. No one should have the legal right to choose abortion.*".

These findings align with prior results regarding how tweets in the two ideological categories *For Abortion* and *Against Abortion* present their arguments using contrasting concepts and disparate viewpoints. While those in *Against Abortion* back their argument in the context of religion and notion of abortion as an act of killing, *For Abortion* tweets put forth their views in terms of feminism, reproductive health and women's right to choose. Along similar lines, on the one hand, *Against Abortion* tweets discuss topics like unborn lives matter and defund Planned Parenthood, on the other, those in *For Abortion* show support with topics like Stand with PP and Violence against abortion clinics.

5 DISCUSSION

5.1 Theoretical and Practical Implications

In our work, we observe how the discourse on abortion is reconceptualized on Twitter. The concept of Critical Discourse Analysis (CDA) helps to understand the notion of ideological discourse as a crucial social practice revealing the manifestation of the power dynamics inherent in the society [14]. CDA postulates that diverse "ideological-discursive formations" (IDFs) are associated with social

Table 9: Example themes from Topic Modeling, their description and top words

Theme	Description	Top words
State bans	State level regulations on abortion	ban, state, govt, bill, ohio
Women's rights	Abortion as women's fundamental right	women's, rights, pills, reproductive, healthcare
Religious views	Church's stance on abortion	jesus, religion, bible, god, faith
Abortion is murder	Perceiving abortion as an act of killing	kill, murder, wrong, life, baby
Planned Parenthood	organization for reproductive health services	planned, parenthood, defund, pp, clinics

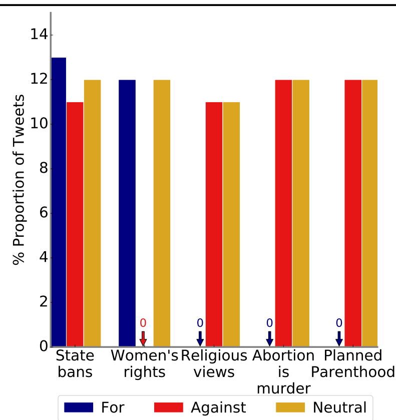

Figure 4: Proportion of tweets across the ideological categories of For Abortion (For), Against Abortion (Against), and Neutral to Abortion (Neutral) per theme.

institutions and there is usually one IDF which is clearly dominant [16]. With political, social, and cultural institutions shaping the morals around abortion, in our study we find similar dynamics being reflected on Twitter. By analyzing 700 thousand tweets, we observe the *Against Abortion* stance poses as the dominant ideology in contrast with the ones advocating for abortion rights (i.e., the *For Abortion* ideology). Our methods adopt Fairclough's notion of examining language as an amalgamation of text and its associated context, and reveals that Twitter can indeed facilitate a comprehensive understanding of the discourse around the topic of abortion. In the following paragraphs we situate our findings and observations within CDA's theoretical framework.

Discourse as Text. Texts as elements of discursive events have the power to bring change in participants' knowledge, beliefs, values, and much else [14]. A structured and systematic analysis of texts therefore helps in effectively interpreting the language of an ideological discourse and its impact on the immediate environment. In RQ1, we considered the language from the tweets as discursive units and material manifestations of discourse [7]. On these texts, we conducted psycholinguistic analysis to investigate the differential expressions of the ideological stances embedded in the content.

Abortion being an inherently sensitive topic, has been known to trigger heavily polarized ideological stances in the offline world [17]. Our linguistic analysis in RQ1 relating to cognitive processing, interpersonal focus, or lexical density attributes of the ideological categories, suggest that Twitter is no exception. That said, our findings also indicated a significant volume of tweets, which instead of reflecting a specific ideology echoed information about events and practices around abortion, therefore appearing to us as ideologically neutral. Prior work also found the usage of neutral hashtags in the Twitter discourse on abortion [50]; although our work validates that trend, the nuances in the ideological stance communicated through the language needs further investigation.

Discourse as Socio-Cultural Practice. Analyzing the public discourse around abortion thus inevitably leads to an understanding of the relationship of *social structure* and agency of the practice itself. Relatedly, through our results from RQ2, we find several ethical, institutional, and organizational circumstances from the offline world being reconceptualized on the Twitter platform. The common practice we observe in the Twitter rhetoric *Against Abortion* predominantly holds the perception of abortion as an act of murder, which only reflects the classic argument of fetal personhood and abortion as an act of violence against the life of fetus [22]. On the other hand, our findings also reveal that "Women's Rights" is a major theme in *For Abortion* and is never a subject of discussion in *Against Abortion*. This echoes the popular counter-discourses around abortion, the crux of which is women claiming the decision power over their reproductive life, and the right to safe and affordable abortion.

Re-conceptualization on Twitter. A thorough analysis of the Twitter discourse comprising the markers of texts or language as discursive practices and socio-cultural practices has provided us with an intricate and thorough understanding of the public opinions around abortion, and its inherent deep-seated ideological complexities. In many ways, we have characterized and analyzed how this debate is being reconceptualized in the online context, albeit a specific social media platform. The insights we gleaned can be particularly valuable to gauge the collective vibe of the abortion debate—information which, in turn, can be useful for policymakers and activists toward social, political, and collective action. The methodological lens we provide in this paper can also be utilized to study the dynamically changing perspectives around the debate during periods of significant socio-political events. Broadly, beyond abortion, with Twitter facilitating public discourse surrounding socially contested problems, our work provides a quantitative, methodological lens to enable future researchers examine the platform's performance as the public sphere of the twenty-first century.

5.2 Limitations

There are limitations to our approach and findings. We cannot claim to have captured the complete discourse around abortion on Twitter: recall, we used the most popular hashtags related to the topic of abortion as our search terms, which also have socio-cultural and political significance. Hence our work did not capture personal experiences related to abortion and the insights gathered from our results do not span the topics of social stigma that engenders abortion and several awareness campaigns. We also note that there could be a self-selection bias in the Twitter demographic group we studied, which means our findings might not be directly generalizable to the offline world or other social media. In particular, *Against Abortion* surfaces as a dominant group in the abortion discourse in our study—further research is needed to understand to what extent this ideological imbalance is a reflection of the rhetoric in the offline world.

6 CONCLUSION

Our study has provided a methodological lens to study the ideologically diverse discourse around the controversial topic of abortion on social media, specifically Twitter. Linguistic analysis results of over 700 thousand tweets revealed an ideological imbalance, where the *Against Abortion* category surfaced as a dominant ideology. We also observed offline socio-cultural practices around abortion being reconceptualized on Twitter. To the best of our knowledge, we provided some of the first empirical insights into the texts and social practices around abortion on Twitter, incorporating the theoretical framework of Critical Discourse Analysis.

REFERENCES

[1] Aseel A Addawood and Masooda N Bashir. 2016. âĂIJWhat is Your Evidence?âĂİ A Study of Controversial Topics on Social Media. *ACL 2016* (2016), 1.
[2] Pablo Barberá. 2015. Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political Analysis* 23, 1 (2015), 76–91.
[3] Adrian Benton, Braden Hancock, Glen Coppersmith, John W Ayers, and Mark Dredze. 2016. After Sandy Hook Elementary: A year in the gun control debate on Twitter. *arXiv preprint arXiv:1610.02060* (2016).
[4] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *JMLR* 3, Jan (2003), 993–1022.
[5] Jan Blommaert and Chris Bulcaen. 2000. Critical discourse analysis. *Annual review of Anthropology* (2000), 447–466.
[6] Axel Bruns and Jean E Burgess. 2011. The use of Twitter hashtags in the formation of ad hoc publics. In *Proceedings of ECPR*.
[7] Jean K Chalaby. 1998. Beyond the prison-house of language: Discourse as a sociological concept. In *The Invention of Journalism*. Springer, 57–67.
[8] Lilie Chouliaraki and Norman Fairclough. 1999. *Discourse in late modernity: Rethinking critical discourse analysis*. Edinburgh University Press.
[9] Michael Conover, Jacob Ratkiewicz, Matthew R Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. 2011. Political polarization on twitter. *ICWSM* 133 (2011), 89–96.
[10] Flora Davis. 1999. *Moving the mountain: The women's movement in America since 1960*. University of Illinois Press.
[11] Munmun De Choudhury, Shagun Jhaver, Benjamin Sugar, and Ingmar Weber. 2016. Social Media Participation in an Activist Movement for Racial Equality. In *ICWSM*.
[12] Michele Dillon. 1993. Argumentative complexity of abortion discourse. *Public Opinion Quarterly* 57, 3 (1993), 305–314.
[13] Jacob Eisenstein, Amr Ahmed, and Eric P Xing. 2011. Sparse additive generative models of text. *ICML* (2011).
[14] Norman Fairclough. 1992. *Discourse and social change*. Polity press.
[15] Norman Fairclough. 1997. Discourse across disciplines: discourse analysis in researching social change. *AILA Review* 12, 1 (1997), 3–17.
[16] Norman Fairclough, Jane Mulderrig, and Ruth Wodak. 2011. Critical discourse analysis. *Discourse studies: A multidisciplinary introduction* (2011), 357–378.
[17] Myra Marx Ferree. 2002. *Shaping abortion discourse: Democracy and the public sphere in Germany and the United States*. Cambridge University Press.
[18] Michel Foucault. 1971. Orders of discourse. *Social science information* 10, 2 (1971), 7–30.
[19] Gallup. 2016. Gallup Report on Abortion. http://www.gallup.com/poll/1576/abortion.aspx. (2016). Accessed: 2017-01-02.
[20] Barney G Glaser and Anselm L Strauss. 2009. *The discovery of grounded theory: Strategies for qualitative research*. Transaction publishers.
[21] Antonio Gramsci, Geoffrey Nowell-Smith, and Quintin Hoare. 1971. *Selections from the Prison Notebooks of Antonio Gramsci: Ed. and Transl. by Quintin Hoare and Geoffrey Nowell Smith*. International Publishers.
[22] Kate Greasley. 2017. *Arguments about Abortion: Personhood, Morality, and Law*. Oxford University Press.
[23] Jürgen Habermas. 1991. *The structural transformation of the public sphere: An inquiry into a category of bourgeois society*. MIT press.
[24] Glen A Halva-Neubauer and Sara L Zeigler. 2010. Promoting fetal personhood: The rhetorical and legislative strategies of the pro-life movement after planned parenthood v. casey. *Feminist Formations* 22, 2 (2010), 101–123.
[25] Gerard A Hauser. 1999. *Vernacular voices: The rhetoric of publics and public spheres*. Univ of South Carolina Press.
[26] Maxine Lattimer. 1998. Dominant ideas versus womenâĂŹs reality: Hegemonic discourse in British abortion law. In *Abortion Law and Politics Today*. Springer, 59–75.
[27] Rolf Lidskog and Ulrika Olausson. 2013. To spray or not to spray: The discursive construction of contested environmental issues in the news media. *Discourse, Context & Media* 2, 3 (2013), 123–130.
[28] Thomas R Lindlof and Bryan C Taylor. 2010. *Qualitative communication research methods*. Sage.
[29] Zhe Liu and Ingmar Weber. 2014. Is Twitter a public sphere for online conflicts? A cross-ideological and cross-hierarchical look. In *International Conference on Social Informatics*. Springer, 336–347.
[30] Gilad Lotan, Erhardt Graeff, Mike Ananny, Devin Gaffney, Ian Pearce, et al. 2011. The Arab Spring| the revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions. *IJOC* 5 (2011), 31.
[31] Kristin Luker. 1984. *Abortion and the Politics of Motherhood*. Vol. 3. Duke Univ Press.
[32] Dawn McCaffrey et al. 2000. Competitive framing processes in the abortion debate: Polarization-vilification, frame saving, and frame debunking. *The Sociological Quarterly* 41, 1 (2000), 41–61.
[33] Mor Naaman, Jeffrey Boase, and Chih-Hui Lai. 2010. Is it really about me?: message content in social awareness streams. In *Proceedings of CSCW*. ACM, 189–192.
[34] Elisabeth Noelle-Neumann. 1974. The spiral of silence a theory of public opinion. *Journal of communication* 24, 2 (1974), 43–51.
[35] James W Pennebaker, Matthias R Mehl, and Kate G Niederhoffer. 2003. Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology* 54, 1 (2003), 547–577.
[36] Pew. 2016. 5 Facts about Abortion. http://www.pewresearch.org/fact-tank/2016/06/27/5-facts-about-abortion/. (2016). Accessed: 2017-01-02.
[37] Politico. 2016. Supreme Court strikes restrictive Texas abortion law. http://www.politico.com/story/2016/06/supreme-court-strikes-restrictive-texas-abortion-law-224831. (2016). Accessed: 2017-06-12.
[38] Politico. 2016. Women in black protest PolandâĂŹs new abortion proposal. http://www.politico.eu/article/women-in-black-protest-polands-new-abortion-proposal/. (2016). Accessed: 2017-06-12.
[39] Politico. 2017. What's actually in the GOP health care bill. http://www.politico.com/story/2017/05/04/gop-health-care-bill-details-explained-237987. (2017). Accessed: 2017-06-12.
[40] Christopher Pullen and Margaret Cooper. 2010. *LGBT identity and online new media*. Routledge.
[41] Adrienne Rich. 1995. *Of woman born: Motherhood as experience and institution*. WW Norton & Company.
[42] Steven M Schneider. 1996. Creating a democratic public sphere through political discussion: A case study of abortion conversation on the Internet. *Social Science Computer Review* 14, 4 (1996), 373–393.
[43] Alexandra Segerberg and W Lance Bennett. 2011. Social media and the organization of collective action: Using Twitter to explore the ecologies of two climate change protests. *The Communication Review* 14, 3 (2011), 197–215.
[44] Clay Shirky. 2011. The political power of social media: Technology, the public sphere, and political change. *Foreign affairs* (2011), 28–41.
[45] Laura M Smith, Linhong Zhu, Kristina Lerman, and Zornitsa Kozareva. 2013. The role of social media in the discussion of controversial topics. In *SocialCom*. IEEE, 236–243.
[46] Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology* 29, 1 (2010), 24–54.
[47] Anton Törnberg and Petter Törnberg. 2016. Muslims in social media discourse: Combining topic modeling and critical discourse analysis. *Discourse, Context & Media* (2016).
[48] Ralph Turner. 1996. The moral issue in collective behavior and collective action. *Mobilization: An International Quarterly* 1, 1 (1996), 1–15.
[49] Teun A Van Dijk. 2006. Politics, ideology, and discourse. In *Encyclopaedia of Language & Linguistics (Second Edition)*. Elsevier Ltd, 728–740.
[50] 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.
[51] Amy X Zhang and Scott Counts. 2015. Modeling ideology and predicting policy change with social media: Case of same-sex marriage. In *Proceedings of CHI*. ACM, 2603–2612.
[52] Amy X Zhang and Scott Counts. 2016. Gender and Ideology in the Spread of Anti-Abortion Policy. In *Proceedings of CHI*. ACM, 3378–3389.