Examining Regional Abortion Discourse in America Via Twitter

Zachary Horvitz
Anthropology 2300: Anthropological Demography
Dr. David Kertzer
May 10, 2019

Introduction

Demographers brandish regressions and data to describe fluctuations in diverse populations. Such tools are incredibly powerful; by translating complex, heterogeneous, social realities into quantitative approximations built of variables and trends, these methods offer academics a means of both generalization and prediction. Through them, communities worlds apart can be described on the same chart, and a population scientist needs to only increment an independent variable to simulate how reality could unfold. The seductive power of these quantifiable methods exerts a strong pull on the field's focus. Susan Greenhalgh describes the resulting understanding in the discipline: "if you can't measure it, it isn't important" (Greenhalgh 1996:48). However, there are critical aspects of social reality that are ostensibly impossible to measure, yet heavily influence demographic outcomes. "Culture," is one of those nebulous phenomena: whether considered in terms of behaviors emerging from social practice (Carter 1998), evaluative networks (Hammel 1990), or beliefs promoted by influential institutions (Kertzer 1998).

Christine Bachrach offers a very different model of culture. She presents cultural ideas as emerging from networks of meanings. In this cognitive anthropological framework, cultural beliefs are made up of multiple discrete components. These basic elements, which Bachrach terms schemas, connect different concepts (Bachrach 2014), like love and marriage, or birth and life. Critically, these basic connections can be explicitly represented and *measured* through approaches like textual analysis. Bachrach writes, "Data mining techniques for textual data...draw out the relationships among ideas, the distribution of ideas across social space, and the affective meanings associated with people and events..." (Bachrach 2014: 20). Such

techniques reveal cultural schemas by elucidating the relationships between concepts that appear together in paragraphs and sentences (Bachrach 2014; Carley 1994). When extracted from a textual source, schemas can be examined and counted. In turn, they can be related to demographic transitions, and even, Bachrach ventures, accounted for in models (Bachrach 2014). However, for such an approach to be viable, an appropriate data set must exist.

Over the last decade, a collaborative cloud of textual discourse has taken the world by storm. Across the planet, humans spend billions of hours consuming and posting on social media. Now, conversations, arguments, and opinions, one ephemeral, are now fixed in public posts, messages, and tweets. Viewed through Bachrach's formulation of culture, this corpus provides a bonanza of data from which cultural schema, and their associations, can be drawn, analyzed, and compared. Thus, following Bachrach's mandate, this paper will explore several ways in which social media text data can be mined for beliefs related to the demographic context. In particular, I will use Twitter data to characterize geo-localized differences in discourse surrounding abortion. In addition, I will explore the extent to which those social media exchanges capture the political and social realities of the practice. To do this, I examine whether regional discursive patterns correlate with public opinion and policy.

Twitter for Demographic Research

Social scientists have begun exploring the viability of Twitter data for use in their disciplines. For example, public health researchers have found that markers of healthy lifestyles in tweets correlate with lower mortality statistics in corresponding counties (Ngyuyen et al. 2016). Taking a complementary approach, another paper found a strong negative correlation between future tense usage in Tweets and the spread of HIV at the county-level (Ireland et al.

2015). Related work by demographers has shown that the word frequencies used on Twitter are highly predictive of demographic categories that include ethnicity and religion (Bokányi et al. 2015). These publications all emphasize the possibilities for Twitter data, and the ability of geo-tagged tweets to encode data on regional norms, identities and behaviors.

Several papers have also examined abortion discourse on Twitter across the United States. Sharma et al. classified abortion related tweets as for, neutral, and against abortion to examine how "offline hegemonic discourses" materialize online (Sharma et al. 2017). These researchers examined frequent hashtags associated with the Pro-life and Pro-choice movements, and identified sets of terms and themes that correlated with each group. They found tweets in support of abortion tended to feature references to reproductive rights (e.g. feminism, rights, prochoice, and women), while the tweets against abortion frequently connected abortion with the murder of children and invoked religious themes (Sharma et al. 2017). Other work, like that of Han et al, examined a year of rhetoric around a leaked video the purported to incriminate Planned Parenthood. In their paper, the authors documented the changing nature of discourse as time passed, which included a shift in style from "sensational" terminology to calls for mobilization, that men were much more likely to mention #defundpp hashtag than women, and that the #defundpp tag was much more common in the southern United States (Han et al. 2017).

All together, the research on county-level health outcomes and abortion discourse display the potential for Twitter to serve as a tool that can bridge anthropological and demographic perspectives. The work on abortion rhetoric reveals that charged ideologies can embed themselves in tweeted text, while the county-level work reveals that these messages can simultaneously be associated with demographic behavior on the ground. This project seeks to

connect these two branches of study by relating abortion discourse to belief and policy at the regional level.

Unfortunately, Twitter data has its limitations. Studies have concluded that the Twitter population is neither random nor representative; it predominantly comprises young adults and male users, and is concentrated in urban areas (Yildiz et al. 2017). In addition, only a small percentage of Tweets are geotagged, and these geotags are more accurate among some demographics than others (Pavalanathan and Eisenstein, 2015). Future research could explore the influence of these biases on the following results.

The Abortion Twitter Dataset

To construct my data set, I used the Python Tweepy API, which allows scripts to collect recent tweet data. Using Tweepy, I queried the term "abortion" on Twitter, and collected all English posts from Thursday, April 25th, to Monday, April 29th. In total, I collected 200,071 posts from over 120,010 users worldwide. There were two possibilities for mapping tweets to a region. If the tweet was tagged with a location (i.e. "New York City"), I used that as the identifier. If this was not the case, I used the location associated with the user. User were only associated with a state if their Twitter location ended in a valid state name or state acronym. Users were associated with a city if their location was of the form: City, State. Using this method, I identified the states of 30,929 users, and the cities of 26,412.

For every state and major city, I collected the location's corresponding messages. These posts are of three primary forms: original posts, replies, and retweets. The following is an example post from someone in California:

> What's missing from the conversation about late abortions, explained by a doctor https://t.co/ajrpitZa4i?

Other data are retweets. The following is a retweet from Texas:

> "RT @LiveAction: If this isn't torture, nothing is.
> An eyewitness says babies born alive after abortions were "turned over with their face…"

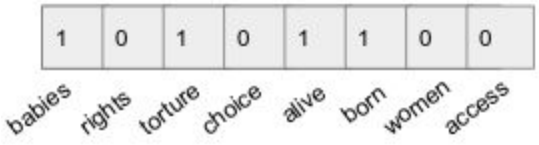
In contrast to posts, retweets contain a prepended "RT," followed by the Twitter handle of the original poster, and, at times, a line of additional commentary.

Bag of Words:

In computational linguistics, a baseline method for representing segments of text is the Bag of Words (BoW) approach (Harris 1954). The approach evaluates a text as an unordered set of its words. For example, the aforementioned retweet could be represented as:

[abortions:1, alive:1, babies:1,born:1,face:1,eyewitness:1, nothing:1, says:1, torture:1, turned:1]

Using the following method, we can construct vectors that have a slot for each common word in our corpus. For each tweet, we can tally the number of times each word appears. Rewritten in this format, the previous tweet could be represented as:



The tweet contains 1 count of "babies," "torture," "alive," and "born," and is missing "rights," "choice," "women" and access.

With a BoW approach, representing the tweets from a region in aggregate is a trivial task: we can sum each tweet's list of word counts elementwise to produce one set of counts. Beyond scalability, these "document" vectors are easily comparable. To juxtapose the texts of two regions, one can simply compare each of their respective word counts. Unfortunately, the benefits of BoW come with a fundamental disadvantage: The representation of text does not

capture the relationship between words in a sentence. For example, while the sentences “I am prolife not prochoice” and “I am prochoice not prolife” have opposite meanings, their BoW representations are identical. However, there is a partial solution to this problem: We can instead treat adjacent pairs of words as single features. In the previous case, one tweet would have the bigram “not+prochoice” while the other would have “not+prolife,” and the difference would be preserved. Another similar approach is to construct bigrams from pairs of words that occur together in a single tweet.

Cosine Similarity: Comparing Different Regions

Once we have produced aggregate summary vectors for states, we can investigate their similarities. One method for doing this is to compare states via their vectors’ cosine similarity, calculated as follows:

$$\textbf{Similarity} = cos(\theta) = \frac{A \cdot B}{||A||||B||}$$

Here, A is the vector of words associated with one region, and B corresponds to the other. The more the two word vectors depart, the larger the angle between them, θ. The cosine of θ is equivalent to the dot product of the two vectors normalized by their magnitudes.

States

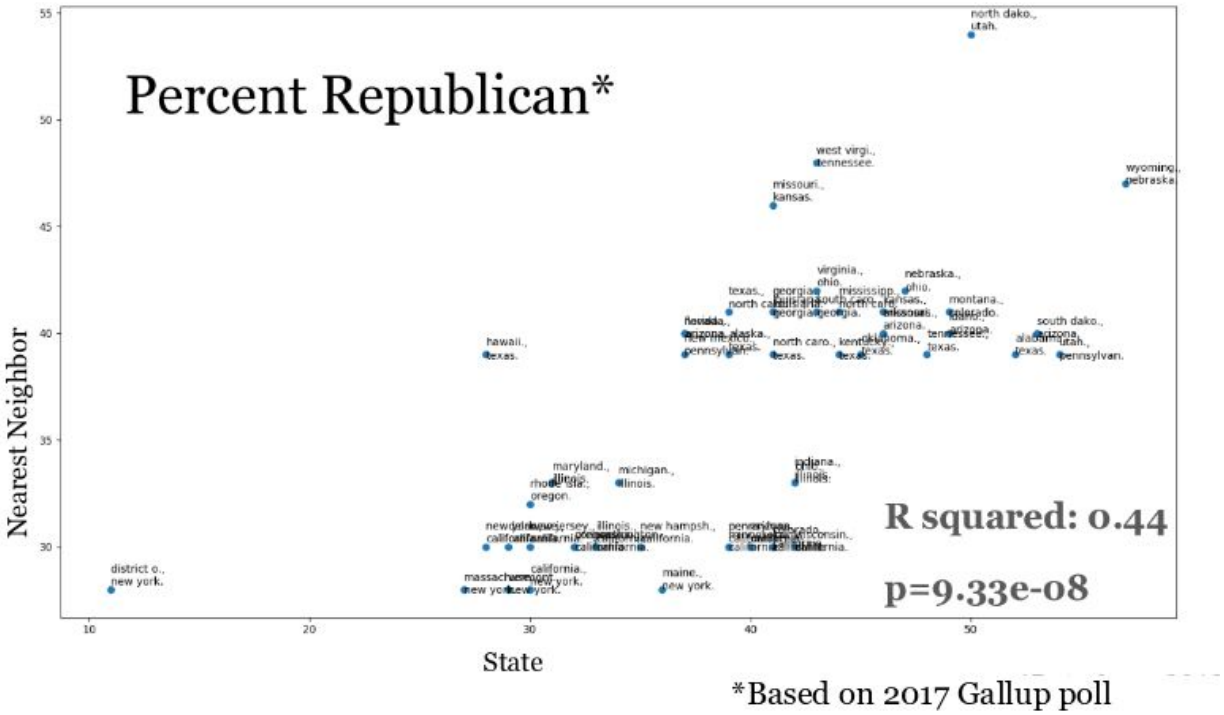
Taking the cosine similarity of each pair of state vectors (built from the top 20,000 most common bigrams) yields the following results:

States and Their Nearest Neighbors

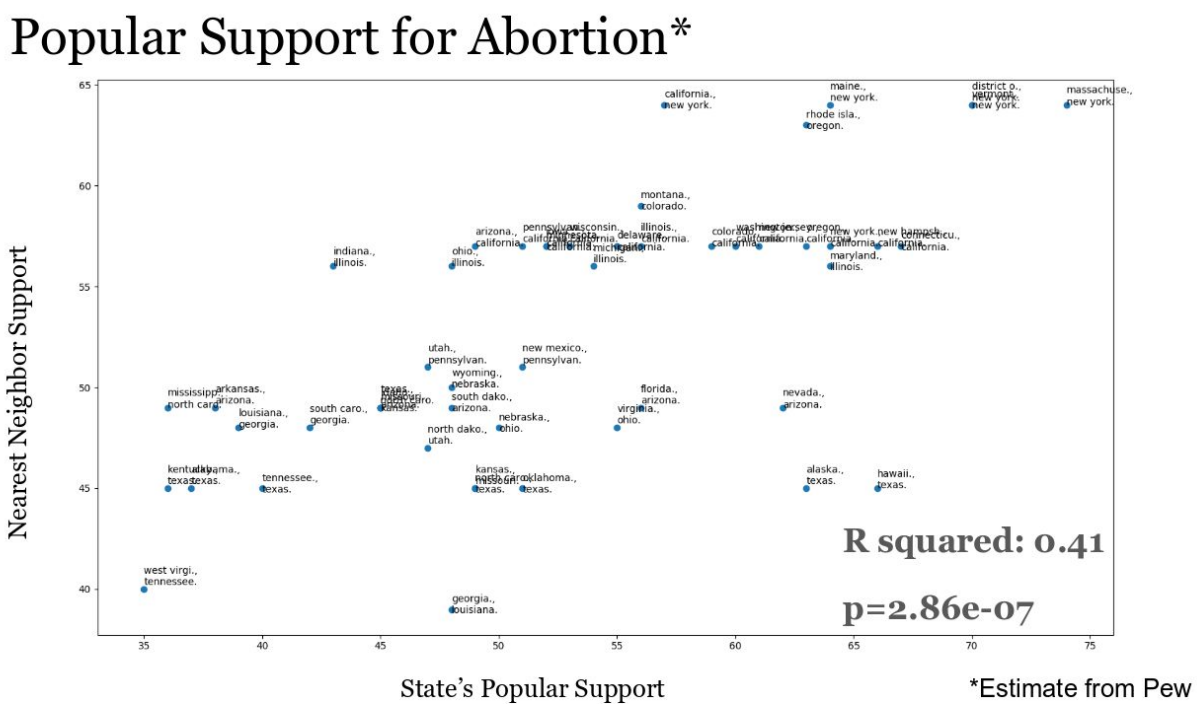
State	Most Similar To:	Difference
Rhode Island	Oregon	0.2446
Washington	California	0.0985
Kansas	Missouri	0.13
DC	New York	0.1026
Tennessee	Texas	0.0946

Broadly, these associations make intuitive sense, and map onto regional similarities. Washington and California are both on the west coast, while both the District of Columbia and New York are on the east coast. In addition, states in the southern United States appear similar to other southern states. An interesting exception to the geospatial rule is Rhode Island and Oregon. However, these two regions have political similarities; both have large constituencies that support democratic candidates.

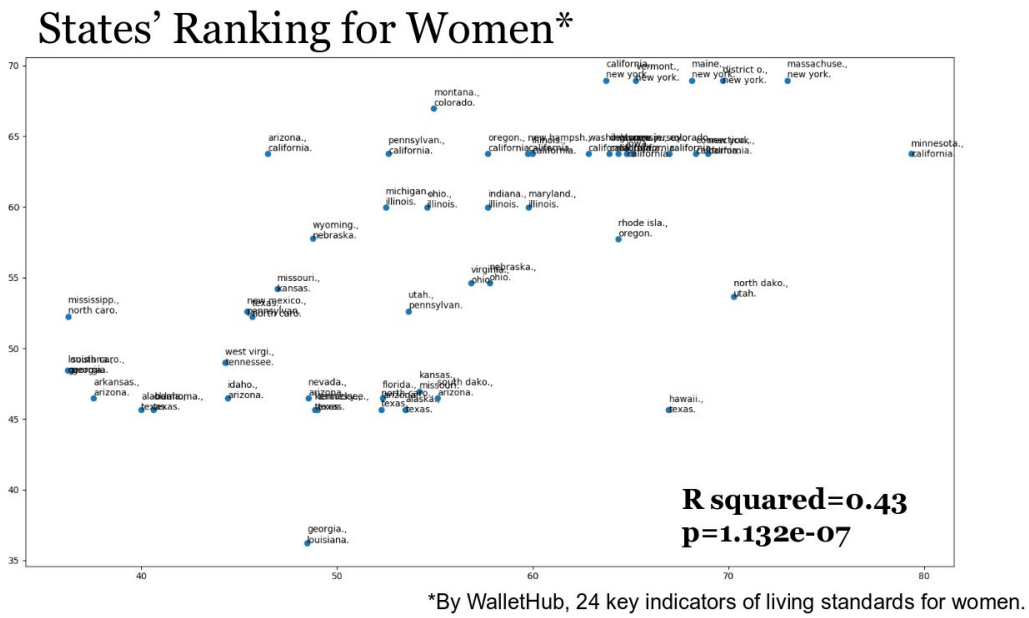
In fact, the political compositions of these neighboring states strongly correlate in general:



Based on Pew survey results, similar states also tend to have comparable levels of support for legalizing abortion:



States with similar Twitters also agree on a third metric: WalletHub's rankings of states based on opportunities for women, which comprises an aggregate metric of factors ranging from Women's median income to rates of sexual harassment in the workplace. Again, there is a strong positive correlation between Twitter neighbors:

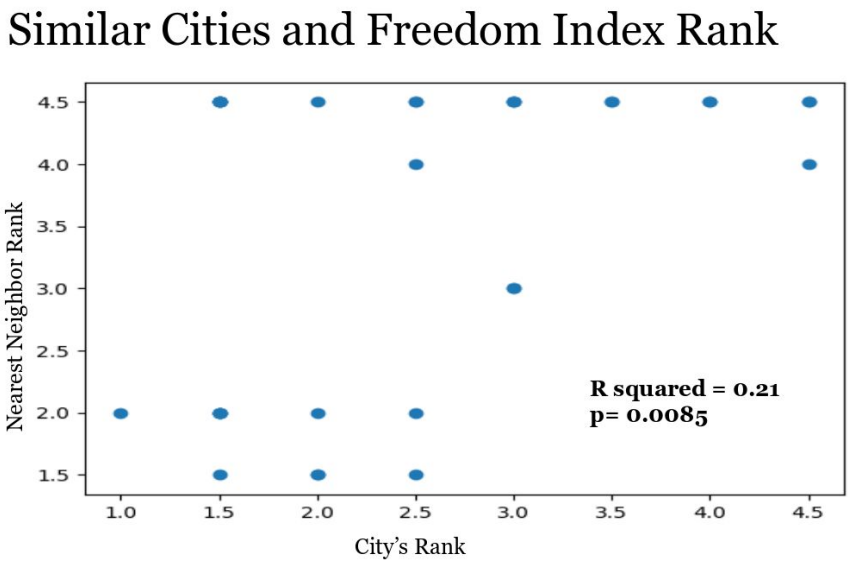


In contrast to these political metrics, the ACS median income of similar Twitters is weakly correlated (With an R squared of 0.24). Therefore, across these three strong correlations, income is not being used as a proxy.

Cities

Unfortunately, few metrics related to popular support for contraceptives and political party are available at the municipal level. However, the National Institute for Reproductive Health (NIRH), a reproductive freedom advocacy group, published a report in 2017 that ranked 40 major cities by their Local Reproductive Freedom Index. The report was part of a broader push by the organization to promote pro-choice activism at the local level. To evaluate an urban center, the index takes into account local policies on "health, rights, and justice policies"(NIRH 2017). Ranked by this metric, Los Angeles New York City, and San Francisco topped the list of progressive cities with scores of 4.5. In contrast, Jacksonville, Florida received the worst scorecard, and was given a rank of 1.

Limiting our data to tweets from cities ranked by the NIRH allows us to compare the ranks of similar locales.These results are moderately correlated: Cities with similar Twitter clouds tend to also share similar ranks.



Thus, similarities in Twitter posts between regions correlate with similar metrics of policy at the level of cities and popular support for abortion at the level of states. These data evidence that there is some correspondence between political reality on the ground, and the discourses that unfold online. However, we have not yet examined what the nature of that difference is. Thus far, we have only compared the broad contours of regional discourses in relation to each other. However, our methodology allows us to magnify specific cases, and examine their differences.

Zooming In: Juxtaposing Mississippi and New York Bigrams:

We can compare the vectors Mississippi and New York by subtracting their respective brigram vectors to determine which word counts most diverge. The following lists are constructed from bigrams that occur significantly more in Mississippi than in New York, and vice-versa:

Mississippi

'yall+idea', 'idea+type', 'type+horrible', 'horrible+shit', 'shit+see', 'see+keep', 'keep+banning', 'banning+abortion', 'abortionrights+activists', 'babies+born', 'born+alive', 'carolina+governor', 'governor+democrat', 'democrat+roy', 'cooper+vetoed', 'vetoed+legislation', 'legislation+protect', 'protect+babies', 'alive+result', 'result+bot', 'roy+cooper', 'north+carolina', 'slave+owners', '6week+abortion', 'wrong+certain', 'certain+exceptions', 'kansas+highest', 'highest+court', 'owners+slaverys', 'slaverys+legal', 'legal+dare', 'dare+oppose', 'oppose+slavery', 'slavery+natural', 'part+society', 'society+slave', 'natural+part', 'protects+abortion', 'bot+north', 'margaret+sanger', 'joe+biden', 'abortion+kills', 'human+right', 'never+understood', 'understood+anyone', 'anyone+would', 'would+need', 'need+lateterm', 'life+induce', 'induce+labor'

New York

'supreme+court', 'kansas+supreme', 'right+abortion', 'last+night', 'ban+abortion', 'night+indiana', 'signed+ban', 'indiana+governor', 'governor+holcomb', 'holcomb+signed', 'abortion+weve', 'weve+already', 'already+filed', 'filed+suit', 'ruled+right', 'baby+born', 'sounds+like', 'completely+false', 'false+pathological', 'pathological+liar', 'liar+misogynist', 'misogynist+sounds', 'like+talks', 'talks+abortion', 'abortion+together', 'state+constitution', 'court+ruled', 'abortion+embedded', 'court+rules', 'president+trump', 'texas+wants', 'wants+put', 'put+women', 'women+death', 'death+abortion', 'abortion+ohio', 'ohio+others', 'others+trying', 'trying+restrict', 'restrict+much', 'kansas+constitution', 'abortion+bans', 'rules+state', 'states+constitution', 'embedded+state', 'going+get', 'title+x', 'states+passed', 'abortion+providers', 'good+news'

In Mississippi, common bigrams include “protect+babies,” “born+alive,” and “abortion+kills,” and others reference to ‘democrat’ statesmen. Given the high-occurrence of these terms, we can take an additional step, and query these words in our corpus to find the actual tweets where they occur. “Abortion+kills” occurs in several tweets retweeted in Missouri:

> RT @Liz_Wheeler: Actually abortion kills a human baby. https://t.co/ydJ1EqvSKw
> RT @nrlc: Abortion kills an innocent person. Every time. https://t.co/MwgyVvADJ2

Each of these tweets implies a strong association between abortion and the death of a baby, and frame stopping a pregnancy as ending a life. Similarly, querying “born+alive” yields the common retweet:

> RT @charliekirk11: Evil: North Carolina Governor, Democrat Roy Cooper vetoed legislation to protect babies born alive as a result of a bot…

The text above criticizes a democratic politician for his stance on abortion. Political support is voiced in many other tweets, including:

> @kikilutefisk @HoodForGovernor He likes the 6 week abortion ban, "working families" and Jesus. That's all I have.

Additionally, several common bigrams that appear more in Mississippi tweets related to specific policy announcements. While some of the related tweets strike a more neutral tone:

> RT @MSNBC: JUST IN: Kansas' highest court rules that the state constitution protects abortion rights and blocks a 1st-in-the-nation ban on…

Others are politically charged:

> Democrats, Activists 'Infuriated' as Jim Hood Defends Six-Week Abortion Ban | Jackson Free Press | https://t.co/3gPlo2UuCT

Of particular significance, the discourse in Mississippi also contains divergent voices, many of which express their support for reproductive rights. This is frequently done via retweets, which are often from out of state. Many Twitter users in Mississippi retweeted @goddessjimmi, who posted the following in response to the passage of a fetal heartbeat bill in South Carolina:

> @goddessjimmi: Y'all have no idea the type of horrible shit we are about to see if they keep on banning abortion. https://t.co/loDthkkKiM

Here, the user with the moniker @goddessjimmi deplores legislation that makes abortion illegal, and intimates that women will go to unsafe lengths to end pregnancies. Her post was retweeted thousands of times, and across multiple states.

In summary, the bigrams that are significantly more prevalent in Mississippi than New York point our attention to tweets that associate abortion with murder and lampoon democratic representatives. However, other bigrams come from news circulated around policy, and many came from a viral pro-abortion retweet that originated from outside the state.

A large percentage of bigrams more common in New York than Mississippi come from tweets celebrating or circulating the April 26th ruling of the Kansas Supreme court to uphold protections for abortion rights:

> A win for Women, individual freedom, advocates of small government: Kansas Supreme Court rules state constitution… https://t.co/lMmRrtH8g3
>
> RT @JuanSaaa: Kansas Supreme Court rules state constitution protects abortion rights https://t.co/qU0LlSh5Hu
>
> Kansas Supreme Court rules state constitution protects abortion rights - The Washington Post https://t.co/1B55IURCXg

Other New York tweets with common bigrams (like "women+death" and "trying+restrict") refer to policies outside of state, and their threat to women and women's reproductive rights:

> @jennmalinchalk: Texas wants to put women to death for having an abortion, Ohio and others are trying to restrict it so much that by the…
>
> RT @NPWF: Abortion restrictions and sexual violence are rooted in the same sexism. @NPWF & @endsxlviolence are joining together during #SA…
>
> Most of the abortion restrictions on the books are based on junk science https://t.co/Po6ZC2n6of
>
> Delaying Trump's Latest Abortion Coverage Restriction Shows That When Women Speak Out, We Win https://t.co/WrBITEEFxC

The tweets above frame the abortion bans as having adverse effects on women, encourage mobilization, and even strive to debunk anti-abortion perspectives as "junk" science. Such themes have historically pervaded pro-choice rhetoric.

The Abortion Debate at a Glance

Anthropologists, historians, and sociologists have collaborated to build up a vast literature on the discourses around abortion in the United States, and to characterize the fierce

conflict between the "pro-life" and "pro-choice" movements. The pro-life movement "encmposses a broad range of ideological positions" including both "liberal Catholics and Protestants" and more "fundamentalist" Christians and Evangelicals (Ginsburg 1989: 8; Hoffmann et al. 2005). Faye Ginsburg writes that these activists, "not only share the goal of decriminalizing abortion, but also see abortion as symptomatic of other social problems," including "irresponsible sexual behavior" among women, the devaluation of people like "the 'unborn' child," and the influence of profit and "market rationality" over human relations (Ginsburg 1989: 9). In contrast, pro-choice activists view the issue as one of women's rights, self-determination, and choice (Fried 2013; Tribe 1992; Ginsburg 1989).

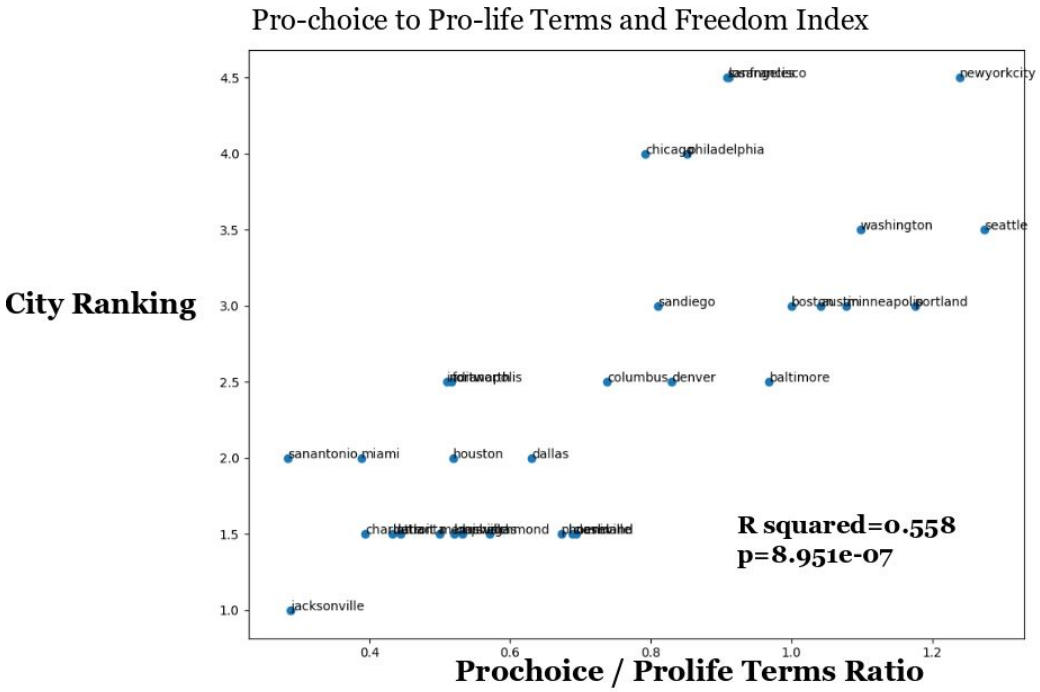
Both sides of the abortion debate continually compete to reframe the issue within their terms by using their own rhetoric, debunking each other's framings ,and vilifying the opposing movement (McCaffrey and Keys 2000). To support its campaign, the pro-life movement deploys language "highly resonant with religious conservatives," by expressing that position that "all life is sacred" and associating women with maternal care (McCaffrey and Keys 2000: 47; Halva-Neubauer and Zeigler 2010). On the other side of the debate, pro-choice activists champion civil rights rhetoric, and paint the pro-life activists as religious fanatics (McCaffrey and Keys 2000).

Relative Term Frequency: Viewing Discursive Dominance

Sharma et al. uncovered strong evidence of one of these moral oppositions in their work studying Abortion discourse on Twitter. The anti-abortion tweets they surveyed tended to invoke a frame of religious morality and "fetal personhood" while the pro-abortion tweets framed the issue as one of "women's rights" and choice (Sharma et al. 2017). Given this finding, I explored

the relative frequencies of a selection of keywords related to these two frames. For “fetal personhood,” I selected the associated terms “life,” “god,” “baby,” ,"human","unborn","murder","born", and "heartbeat," which frequently appeared both in the literature and in the Twitter corpus. I used the terms “rights,” “feminism,” “women,” “choice,” and “freedom” to index the reproductive rights moral framework.

At the city level, the ratio of reproductive rights to life to fetal personhood terms strongly correlated with NIRH ranking:



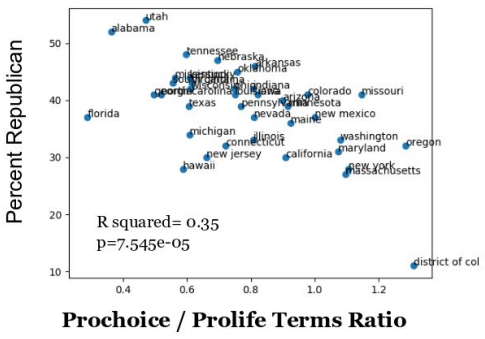
This result indicates that where pro-choice terminology dominates, the corresponding municipalities maintain more progressive reproductive policies. One reasonable explanation for this phenomenon is that local public opinion is a latent variable, influencing both Twitter discourse and local policy. An examination of specific cities’ discourses supports this hypothesis. Of all data points, Seattle has the highest ratio of pro-choice to pro-life language: 1.27. Tweets from the

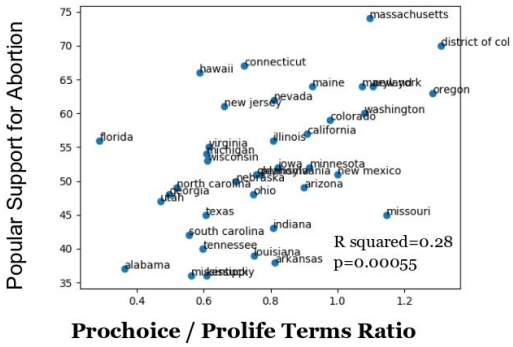
Pacific Northwest city use pro-choice language to heavily criticize legislation that attempts to ban abortion, and reveal strong opinions in support of reproductive access:

> RT @emrazz: 1 in 4 US women under the age of 45 have had an abortion.
> Abortion is healthcare. Healthcare is a right. That doesn't change j…
>
> @johngocee @PattyMurray I know what happens to women who are forced to carry non-viable fetuses to term because th… https://t.co/JYNVxuxvsy
>
> The worst and pathetic thing about Trump and his Pence dictated "stand" on abortion and privacy rights is HE DOES N… https://t.co/HRI8o8DUWR
>
> US: Kansas court bolsters abortion rights, blocks ban @AJENews
> https://t.co/6OrhVm6zNN
>
> RT @laurenduca: Equality is impossible without reproductive rights.
> There is a war on abortion access, and the siege on women's bodily aut…

Such a conclusion is also generally supported by state-level data. States with more pro-choice language on Twitter have smaller Republican constituencies Additionally, increased reproductive rights language on Twitter correlates with support for legalizing abortion:

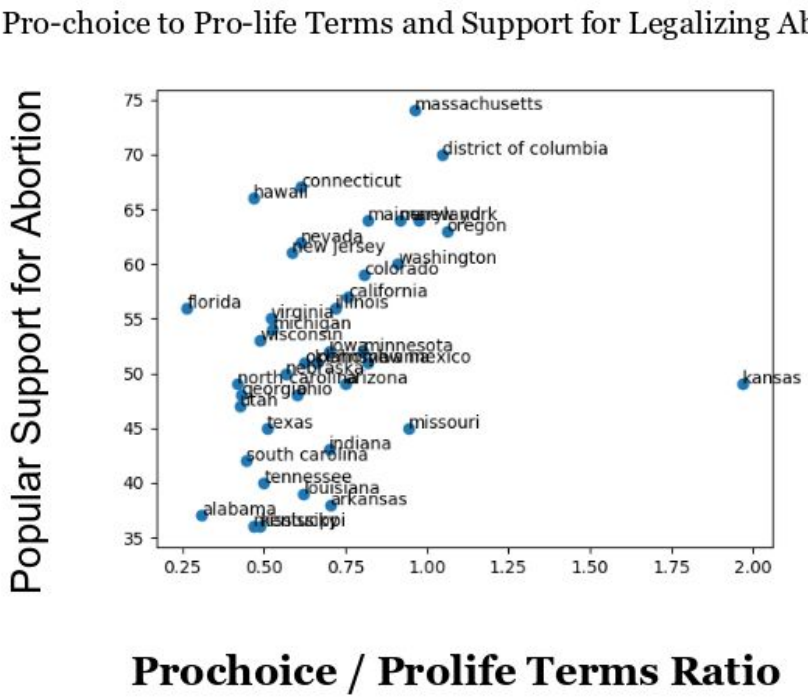
Pro-choice to Pro-life Terms and State* Demographics





*Kansas is excluded

However, there is a critical exception to these trends: Kansas. When Kansas is added back to the regression, the R squared value comparing the ratio belief precipitously falls to 0.10. According to the Pew research data, less than 50 percent of Kansas residents support abortion legalization in all or most cases, and the Republican bloc dominates in the state. However, in Kansas, pro-choice terminology occurs 1.96 times more than pro-life terms:



An explanation lies within Kansas' tweets. Close inspection reveals that many of the state's tweets and retweets circulate the news on April 26th that the Kansas Supreme Court ruled in favor of abortion rights:

> RT @sherman_news: Developing story: Kansas Supreme Court says state constitution guarantees abortion rights #ksleg https://t.co/Y909DLtM8z
>
> Kansas Supreme Court upholds abortion rights, blocks ban: https://t.co/PTrSfajqrG #KAKEnews

> Kansas you make me proud. Who woulda thought this conservative state would take this stand! #abortion #womenrights. https://t.co/KlL7tquSrV

> Woohoo! Kansas leading the way on state protected abortion rights (yes I said Kansas!) https://t.co/67Js4n4hVu

Thus, the Kansas Twitter cloud captures a disruptive political event: a landmark court case. As previously mentioned, this ruling reverberated across the Twitter network; users shared the news across the United States. The embedding of such political news has several implications. For one, it indicates that any analysis of regional differences should ideally examine and control for specific events, and likely draw tweet data from a wider time frame than the one available in this paper. In addition, the Kansas tweet examples indicate that these stories are not just passed on through a neutral medium. As users retweet and share about breaking news, they can insert their own voice and commentary. In the last two examples above, we are offered a glimpse of this. In both, Twitter users from Kansas place their own positive spin on the event, portraying the ruling as an encouraging, albeit surprising, sign of progressive reform to come. Future work could examine the prospect of such seismic political events drawing out otherwise marginalized voices, and reshaping Twitter discourses for the long term.

Twitter As Mobilization

In the above examples, users have used their tweets and retweets to opine or share news surrounding legislation. However, Twitter also provides constituents with a novel means of organizing and mobilizing (Theocharis et al. 2014). In our corpus, this was highly visible in Jacksonville, Florida, where users united around the Prolife and StandForLife hashtags. The latter was included in 25% of Jacksonville tweets. This hashtag was used multiple times by multiple different users, many of whom posted the same formulaic message to their governor:

> Gov. @RonDeSantisFL, as one of your #ProLife constituents, I urge you to #StandForLife and take bold leadership to… https://t.co/9GEYqXj07Z

A similar message was also shared to a Florida senator:

> Sen. @BillGalvano, as a #ProLife Floridian, I urge you to stop stalling and move the Parental Consent for Abortion… https://t.co/FR1l84lnyZ

These tweets were largely facilitated by a handle @SBAList, which circulated the following message:

> **#StandForLife** Action Alert: Florida minors need parental consent before they get a tattoo… but under current law, they can get an abortion without their parents' consent. Click button below to tweet Gov **@RonDeSantisFL** to tell him: Pass the Parental Consent for Abortion Act!

Such tweets capture a concerted effort to lobby the Floridian government through Twitter; an effort that dramatically contributed to the ratio of pro-life terminology in the corpus. As with the Kansas Supreme Court case, this discernable act of mobilization raises questions about the appropriate timespan for a regional twitter study such that the data in aggregate is unperturbed by specific political acts. However, even when we remove tweets with #StandForLife and #ProLife hashtags, the ratio of pro-choice to pro-life keywords is still one of the lowest in the data set, at around 40 percent. The robust number indicates that pro-life language is still dominant. This severe asymmetry is evidence that Jacksonville's reproductive policies, rated as regressive by the NIRH, may have strong grassroots support from local constituents, and will not be overturned by pro-choice advocates in the reasonable future.

Retweets

When similar word patternings appear in tweets from very different parts of the country, there are several possible causes. Some portion of this shared linguistic practice is the result of tweets being written with the same language, making similar connections, or responding to

overlapping events. However, because of the format of the dataset, a significant portion of shared verbiage comes from identical retweets. 74 percent of all the statuses in the corpus are retweets. Of the unique retweets, 25 percent appeared in in multiple different states. Thus, retweets are a huge portion of our dataset, and represent streams of content that link users from multiple distinctive regions.

Retweets are a social feature unique to Twitter. Stolee and Caton write, "Rather than simply replying to another user, this function actually em-beds the content of the original user's post within the body of the retweet along with their own commentary…. A single feed might actually contain many voices, creating a dialogical narrative curated by one user" (Stolee and Caton 2018). However, in our corpus, many of the retweets lack commentary. In them, users are simply propagating the content of the original tweet through their networks. When viewed through Erving Goffman's Participation Framework, these specific posts take on an added subtext. In Goffman's model, there are many different forms of speaker: An "author" composes the words, while the "animator" provides the 'sounding box' for them by making the utterance, and a "principal" is the entity whose beliefs are being represented (Goffman 1981). In the case of a retweet with little commentary, the original user is a relatively obvious author, and the retweeter is a clear animator. However, in this case, the principal becomes blurred. When a pro-choice or pro-life post is reshared, the retweeter makes a conscious choice to disseminate the information. In the process, they take on the role of an implicit endorser, and appear to represent their own beliefs. Therefore, they morph into a principal. Because this is the case, tracking the text from retweets provides just as valuable a window into the beliefs of Twitter users in a region as original content.

Networks of Influence

The substantial role of retweets in the abortion twitter dataset necessitates an exploration of their networks. Of the retweeted handles, several dominated the abortion discourse across cities. The top 10 retweeted handles among major cities are displayed below, along with the cities they appeared in, sorted by count. All of these, with the exception of @LiveAction, are pro-choice leaning:

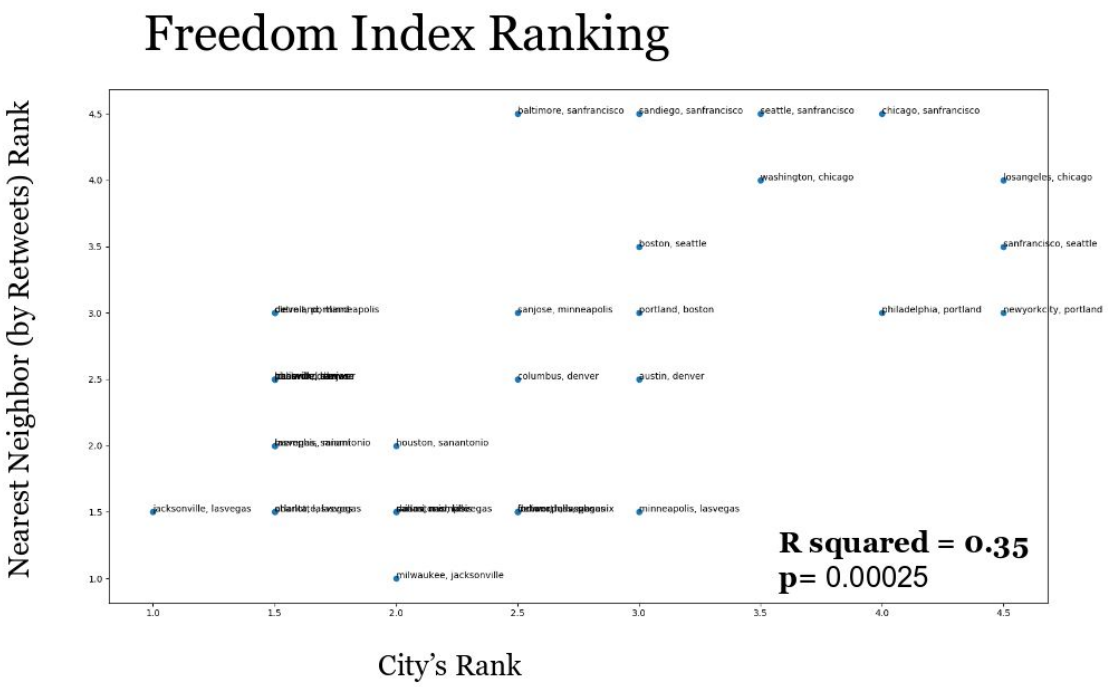
On the one hand, the interconnectedness of geographically disparate locations via retweets seemingly confounds the intention of this research project, which was to characterize specific bounded regions. However, on the other, these networks of retweets provide an additional means to examine regional differences. Instead of tracking words or bigrams, we can

count which handles are retweeted within a city, and leverage these patterns to build our comparison. This approach provides promising results:

Similar Cities (By Retweets)

City	Nearest Neighbor	Distance
Seattle	San Francisco	0.527
Houston	San Antonio	0.535
Dallas	Nashville	0.579
Louisville	San Jose	0.61
Miami	Memphis	0.56

As before, we can graph the NIRH rankings of cities associated with similar twitters to examine general trends:



When compared by retweeted handles, there is a substantially stronger correlation between the Freedom Index ranking of similar Twitters. What could explain this? The topology of these retweet networks may actually be capturing regional differences more robustly than word or bigram counts. There is strong evidence for this in the literature. Research has found that retweets, especially concerning controversial issues like gun control, same-sex marriage, and climate change, primarily spread within liberal and conservative group boundaries (Brady et al. 2017). The wider sociological field has long recognized this phenomenon, of "similarity breeding connection," or "homophily," and its ability to shape social relationships. (MchPhearson et al 2001).

The fact that twitter retweet patterns correlate with Freedom Index is further evidence that advocacy groups like the NIRH are not merely contending with antiquated municipal policy. To the contrary, these similarities provide evidence that constituents in cities are embedded in politically charged social networks that span the country, and that the beliefs that are circulated through these influential networks may also influence policy.

Classification

Once real-world information is transposed into quantitative features, it can serve as the basis for regression and classification tasks. The performance of such classifiers can offer insight into how useful that information is for predicting a given label. Additionally, classifiers can be reverse-engineered, and provide a window into which available features do actually matter. In the case of the NIRH Freedom index, a city-tweets-to-Freedom-Index classifier allows us to examine the extent to which rank can be predicted from Twitter, *and* which Twitter language matters for such a prediction. Thus, as an extension of my project, I trained a simple Naive Bayes

classifier to predict the rank of a city based on the words that co-occur together in regional tweets. Naive bayes methods are family of simple classifiers; they assume that different features (here, the count of a word pair), are independent, and compute the probability of a class given the appearance of each feature. Because of their simplicity, such bayesian classifiers perform well on tasks with limited data sets, like email spam filtering (Sahami et al 1998). Here, our dataset is incredibly small; we only have Freedom Index labels for 40 city vectors. Additionally, to evaluate our classifier, we need to hold out a set of testing examples, making our training sample smaller still. However, even with these constraints, preliminary results are promising. A multinomial naive bayes, with 5-folds (i.e, train on 4/5ths of the dataset) yields:

Task	Accuracy
Predict If City Rank >=3	76%
Predict If City Rank (1,2,3,4)	45%

On a binary classification task, where cities are labeled based on whether their ranking is above or below 3, the classifier scored around 75%. While not astoundingly powerful, this model outperforms a random baseline. Similarly, a 45% accuracy across the four classes is significantly higher than random chance. These results indicate that the Twitter data bigram counts provide a signal about a city's reproductive policies. Because of the simplicity of the model, unpacking the relevant features is a trivial task: We can just inspect the likelihood of each word pair for a given Freedom Index rank.

The word pairs that were much more likely to occur in cities given scores of 1 or 1.5 were:

[born+rt, babies+born, protect+rt,alive+born,alive+babies,@liveaction:+rt, abortions+rt]

These results offer strong evidence that Twitter language discussing abortion in terms of "babies," "birth," and "life" are predictive of a locality's reproductive policy. As identified earlier, these associations are deeply linked to the pro-life belief system and rhetoric. Another feature much more likely to occur in lowly ranked cities is "@Liveaction:+rt." The @Liveaction moniker is associated with a widely retweeted pro-life profile based in Los Angeles. Thus, the model is not just picking up on pro-life language, but the social networks through which that language circulates. Therefore, the classifier provides additional evidence that the city policies that limit abortion reproductive access correspond with strong anti-abortion sentiment.

Public Opinion and Abortion Policy

The aforementioned results indicate that Twitter discourses correlate with both public support for legalizing abortion at the state level and reproductive policy at the city level. These discourses appear to capture the circulation of tweets through ideological networks, and messages that articulate opinions through pro-life and pro-choice frames.

A reasonable inference from these results is that popular beliefs about the morality of abortion strongly correlate with abortion policy. At the city level, the ratio of pro-choice to pro-life terms strongly correlated with Freedom Index rankings, and cities that had similar Twitters also tended to have similar rankings. At the state-level, where data is available, this ratio was positively correlated with belief, and similar Twitters also correspond to similar proportions of support for legalization. Additionally, a review of tweets from select cities showcased morally opposed positions on abortion, indicating that these perspectives were finding their way into the Twitter corpus. For example, Jacksonville, which was ranked last by the NIRH, had a Twitter

flooded with pro-life messaging. Together, these results make a compelling case that city policy tends to have grassroots support, which manifests on Twitter. For the advocacy groups like the NIRH, this conclusion has critical implications: Any attempt to change existing city policy could face staunch resistance in locations like Jacksonville. Instead, to be most effective advocacy groups should focus their resources on raising awareness about city policy in locations where there is dissonance between popular support and law. From the Twitter dataset, it appears that Seattle and Portland are two such cases. While both have incredibly high rates of pro-choice terminology, their Freedom Indices lag behind, at 3.5 and 3, respectively.

This paper corroborates a chorus of voices that associate grassroots advocacy with reproductive laws. A study by Marshall Medoff found that state policy is largely determined by factors including membership in the National Abortions Rights Action League, and that when a state's Roman Catholic population was larger, abortion laws were more restrictive (Medoff 2003). Additionally, a study by Matthew Wetstein demonstrated a correlation between public opinion and abortion policy at the state level, and offers a causal hypothesis (Wetstein 1996). Wetstein concludes his report by calling for future research on "the analysis of aggregate public opinion variables and their relationship with abortion policy," and bemoaning the limited data available (Wetstein 1996: 130). This paper extends Wetstein's research in at least three critical ways. First, it provides further evidence of the connection between opinion and abortion policy. Secondly, it extends his conclusions to the municipal level. Thirdly, and perhaps most importantly, this project further establishes the ability of social media corpora to encode regional information and beliefs, even where data is otherwise unavailable.

Conclusion

In Pierre Bourdieu's *Outline of a Theory of Practice*, he terms unchallenged knowledge as "doxa," dominant but contested knowledge as "orthodoxy," and competing, subversive discourses as "heterodoxy" (Bourdieu 1977). In the United States, abortion has been pulled from doxa; the fight over abortion access has emerged as a contentious issue that has cut through the national conversation. However, this polarization is far from uniform across the country. From one state or city to another, different beliefs prevail, and local conceptions of heterodoxy and orthodoxy are shifted or inverted. Geotagged tweets provide can provide a glimpse into how such beliefs are contested, and the fault lines that cut through a region's discourse and demarcate heterodoxy.

However, when mining Twitter data, researchers are not limited to just estimating the broad portions of people who hold differing beliefs. Textual analysis can also map out the specific associations that characterize those different regions. First, this can be done with broad strokes that compromise depth for breadth and simplify whole dialogues to sets of vectorized bigrams. These bigrams are meaningful in their own right; they can draw out relations between concepts like "abortion" and "child," or "women" and "rights," and the prevalence of these connections. Yet, research does not have to end there. Those preliminary, quantifiable insights can serve as launchpads into queries and comparisons that delve into specific messages and debates. Conversely, those deeper inquiries can fuel new statistical tests and classifiers. This paper has danced between those two poles, the broad and the narrow, to examine the regional differences on abortion discourse, and to tie those characteristics to political and social reality.

Bibliography

Bachrach, C.A.. "Culture and demography: From reluctant bedfellows to committed partners." Demography (2014).: 3-25.https://www.jstor.org/stable/42919986.

Bokányi, Eszter, Dániel Kondor, Laszlo Dobos, Tamas Sebok, József Stéger, István Csabai and Gábor Vattay. “Race, Religion and the City: Twitter Word Frequency Patterns Reveal Dominant Demographic Dimensions in the United States.” CoRR abs/1605.02951 (2016): n. pag.

Bourdieu, Pierre. Outline of a Theory of Practice. Cambridge: Cambridge University Press, 2001.

Brady, William J., Julian A. Wills, John T. Jost, Joshua A. Tucker, and Jay J. Van Bavel. "Emotion Shapes the Diffusion of Moralized Content in Social Networks." Proceedings of the National Academy of Sciences 114, no. 28 (2017): 7313-318. doi:10.1073/pnas.1618923114.

Carley, K. M. (1994). Extracting culture through text analysis. Poetics, 22, 291-312.

Carter, Anthony. "Cultural models and demographic behaviour." The Methods and Uses of Anthropological Demography (1998).: 246-267.

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 CSSSA’s Annual Conference on Computational Social Science, Santa Fe, NM, USA, October 19–22, 2017. https://doi.org/10.1145/3145574.3145577.

Fried, Marlene Gerber. "Reproductive Rights Activism in the Post-Roe Era." American Journal of Public Health 103, no. 1 (2013): 10-14. doi:10.2105/ajph.2012.301125.

Ginsburg, Faye D. Contested Lives: The Abortion Debate in an American Community. Berkeley, CA: University of California Press, 2006.

Goffman, Erving. Forms of Talk. Oxford: Basil Blackwell, 1981.

Greenhalgh, Susan. ""The Social constructioni of Population Science"." Comparative Studies in Society and History . 38, no. 1 (Jan 1996).: 26-66.

Halva-Neubauer, Glen A., and Sara L. Zeigler. "Promoting Fetal Personhood: The Rhetorical and Legislative Strategies of the Pro-Life Movement after Planned Parenthood v. Casey." Feminist Formations 22, no. 2 (2010): 101-23. doi:10.1353/ff.2010.0011.

Hammel, E.. ""A Theory of Culture for Demography"." Population and Development Review . 16, no. 3 (Jan 1990).: 455-485.

Han, L., Han, L., Darney, B., & Rodriguez, M. I. (2017). Tweeting PP: an analysis of the 2015-2016 Planned Parenthood controversy on Twitter. *Contraception*, *96*(6), 388–394. doi:10.1016/j.contraception.2017.08.011

Harris, Zellig S. "Distributional Structure." Papers on Syntax, 1981, 3-22. doi:10.1007/978-94-009-8467-7_1.

Hoffmann, John P., and Sherrie Mills Johnson. "Attitudes toward Abortion among Religious Traditions in the United States: Change or Continuity?" Sociology of Religion 66, no. 2 (2005): 161. doi:10.2307/4153084.

Ireland, M. E., Schwartz, H. A., Chen, Q., Ungar, L. H., & Albarracín, D. (2015). Future-oriented tweets predict lower county-level HIV prevalence in the United States. Health Psychology, 34(Suppl), 1252-1260. http://dx.doi.org/10.1037/hea0000279

John P. Hoffmann, Sherrie Mills Johnson, Attitudes Toward Abortion among Religious Traditions in the United States: Change or Continuity?, Sociology of Religion, Volume 66, Issue 2, Summer 2005, Pages 161–182, https://doi.org/10.2307/4153084

Kertzer, David. "The proper role of culture in demographic explanation." The Continuing Demographic Transition (1998).: 137-157.

Mccaffrey, Dawn, and Jennifer Keys*. "COMPETITIVE FRAMING PROCESSES IN THE ABORTION DEBATE: Polarization-vilification, Frame Saving, and Frame Debunking." The Sociological Quarterly 41, no. 1 (1999): 41-61. doi:10.1111/j.1533-8525.1999.tb02019.x.

Mcpherson, Miller, Lynn Smith-Lovin, and James M. Cook. "Birds of a Feather: Homophily in Social Networks." Annual Review of Sociology 27, no. 1 (2001): 415-44. doi:10.1146/annurev.soc.27.1.415.

Medoff, Marshall H. "The Determinants and Impact of State Abortion Restrictions." American Journal of Economics and Sociology 61, no. 2 (2002): 481-93. doi:10.1111/1536-7150.00169.

NIRH. Local Reproductive Freedom Index. New York: 2017.

Nguyen, Q. C., McCullough, M., Meng, H. W., Paul, D., Li, D., Kath, S., … Li, F. (2017). Geotagged US Tweets as Predictors of County-Level Health Outcomes, 2015-2016. American journal of public health, 107(11), 1776–1782. doi:10.2105/AJPH.2017.303993

Pavalanathan, Umashanthi, and Jacob Eisenstein. "Confounds and Consequences in Geotagged Twitter Data." Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015. doi:10.18653/v1/d15-1256.

Sahami, Mehran, Susan T. Dumais, David Heckerman and Eric Horvitz. “A Bayesian Approach to Filtering Junk E-Mail.” (1998).

Stolee, Galen, and Steve Caton. "Twitter, Trump, and the Base: A Shift to a New Form of Presidential Talk?" Signs and Society 6, no. 1 (2018): 147-65. doi:10.1086/694755.

Theocharis, Yannis, Will Lowe, Jan W van Deth, und García. „Using Twitter to mobilize protest action: online mobilization patterns and action repertoires in the Occupy Wall Street, Indignados, and Aganaktismenoi movements.“ Information, Communication & Society, 2015: 220.

Tribe, Laurence H. Abortion: The Clash of the Absolutes. New York: W.W. Norton, 1992.

Wetstein, Matthew E. Abortion Rates in the United States: The Influence of Opinion and Policy. Boulder, CO: NetLibrary, 1999.

Yildiz, Dilek, Jo Munson, Agnese Vitali, Ramine Tinati, and Jennifer A. Holland. "Using Twitter Data for Demographic Research." Demographic Research 37 (2017): 1477-514. doi:10.4054/demres.2017.37.46.