Gab, Does it Ever Change?
Final Project for CSC 2611

Reid McIlroy-Young

April 5, 2019

Abstract

Gab, an online social network with lax moderation and history of issues caused by its community. I studied how the deadly attack by one of its members. on October 27 2018, on a synagogue and resulting temporary shutdown of the site affected the language used on the site. I found there were substantial changes in the rate of usage of certain words, particularly those related to anti-Semitism, but that more nuanced measures were difficult to draw conclusions from and will require a longer time window to generate results.

1 Introduction

Gab is social media platform with similar functionality to Twitter [4]. The site is deeply tied to the alt-right and other right wing extremists, there has been some work characterizing *Gab*'s community[11, 20], but it has mostly focused on characterizing the extremism on the platform. The dynamics of *Gab* have yet to be explored beyond the most cursory of examination [14].

The site has a complex history of extremism, which peaked in October of 2018 when a Gunman entered a synagogue in Pittsburgh and shot 18 people. The perpetrator was an active member of and had even announced his attack on the site [16]. This attack lead to many service providers breaking ties with the site, including the payment processor and registrar, which caused a service outage lasting about two weeks. This outage creates an natural experiment, and even if a member was not aware of the site's shutdown immediately the minor design changes and domain change (*gab.ai* to *gab.com*) are additional hints of change. *Gab* its self is also very interesting site to study the dynamics of due to its small, by social network standards, user base of 800 thousand, but highly connected community. The social network has low levels of clustering when compared to Twitter, which means it is likely that novel ideas and language will spread quickly.

In this report I first examine hows the frequencies of word usage have changed across the site, both throughout the life and before and after the shutdown. What does the regular pattern of speech look like on the site and how was it effected by the shutdown? I found that the pattern speech shows is that of focus on contemporary events, frequently those related to contemporary US politics and right wing conspiracy theories. This pattern was not effect much by the shutdown.

I also examine the semantic change of words, via the use of word embeddings [15, 7]. This shows some change in the dynamics, but limitations of the sample size and the unregulated nature of spelling and grammar on the site make drawing conclusions difficult.

2 Related Work

The previous work on Gab has mostly focused on characterizing the site's hate speech [11] and its social network [20]. The work on hate speech is particular relevant to this work as it does show considerable amounts of hate speech, in excessive of other site such as Twitter. The social network information is less relevant to this analysis but does suggest that the community is more homogenous than places like *Reddit* which means changes in the sites language will be spread across the community faster than places like *Reddit* or *Twitter*.

There is a substantial body of work dealing with online communities and how different behaviours flourish and change as a result of the community and moderation procedures [17, 1]. The work of Chandrasekharan [3] on the effects of *Reddit* banning sub-forums due to their hateful content and frequent violations of of anti-harassment the policy shows the site's moderation had a considerable effect on the community. Their work shows that even small measures such as banning the most extreme violators can lead to measurable improvements on a site. Thus if *Gab* wanted to change things after the shooting, even small measures could be used to 'calm down' the community. Of course when asked about changing things the chief executive and co-founder of the site, Andrew Torba, saide: 'Absolutely not.' [19]

The anti-Semitic nature of the attack means studying the anti-Semitic language on the site is an obvious starting point. Some work has already been done on this by Finkelstein, et al. [5], although the focus was across multiple domains. The results of Finkelstein's work show that *Gab* has unusually high amounts of ant-Semitic language, high comparable to */pol/*, and that it tends to increase after major events.

3 Methodology

Gab is a website with mostly user generated content, analogous to twitter [4]. Most posts are broadcasts, one user to their followers and are public to even unregistered users. There are in text annotations such as hashtags and @ mentions, and many users include urls in their posts. The posts can also have images and videos (hosted both by *Gab* and third parties such as *Youtube*) associated with them. The site's moderation policy is intentionally lax, to the point of it being an advertised benefit, although terrorism is one of the few explicitly disallowed activities, the others are spam and illegal pornography (pornography is generally allowed although it is rare and tagged as not safe for work). The posts can be upvoted, down voted, commented on or reposted by users, although all but the latter are infrequent.

The users on *Gab* each have unique usernames with their followers and those they are following being publicly visible. These following revelations form a directed graph, which is the basis for the data collection. The users also have a 300 or fewe word bio, which defaults to one of a small set of free-thinker quotations, e.g. *"I might disagree with your opinion, but I'm willing to give my life for your right to express it." - Voltaire* is the bio given to my account.

The site is primarily English, but their are small amounts (less than 1%) of posts in both Portuguese and German, along with very rare posts in other languages. Some posts are tagged correctly by language although many are not so there is some corruption of the data.

3.1 Data Collection

The data collection was done by a custom crawler using the standard snowball sampling procedure [6]. I started with a list of users, the *WHO TO FOLLOW* list, and then using an undocumented API downlaoded all the information on those users and created a list of all people following them, and all people they follow. The went to that new list and repeated the process. This repeated until no more users could be found. This procured found a total of 835,695 people which is close, but lower, to the estimates given by Gab (those estimates are part of an ongoing controversy as *Gab*'s financials are currently being questioned).

The scraping of the site was done every few weeks for the last two months, with the second and later samplers being started with the user lists of the previous run, instead of the much smaller *WHO TO FOLLOW* list. The each scraping takes about three days and was done with 128 crawlers running in parallel. The complete API responses were collected and stored. The last scraping was started March 27, 2019 and it's data are the basis for this report Table 1 shows an overview of the data. Figure 1 shows the usage of *Gab* over time with the shutdown marked with a line.

3.2 Data Pre-prepossessing

The data collected on *Gab* are very powerful in their completeness, but this makes converting them into a usable form difficult. Initially I tried using the standard *NLTK* [12] tokenizing tools. But I found these struggled with parsing the heavily non-standard spelling, and spacing conventions on *Gab*, which with its longer than *Twitter* character limit has many more posts with newlines, indented content and usage of emojis as separators or punctuation. Thus to parse and tokenize I ended up using a regex instead.

	Before	After	Total
Number Registered Users	712,093	123,602	835,695
Users with Posts	265,900	93,684	307,623
Number of Posts	22,567,328	9,078,113	31,645,441
Mean posts per user	16.8	18.4	111.3
Outlier corrected mean posts per user	16.8	18.4	17.4
Median posts per user	3	3	3
Words			423,539

Table 1: Overview of data on *Gab*, note that when calculating the Outlier corrected means users with over 100 posts had their rounds rounded to 100

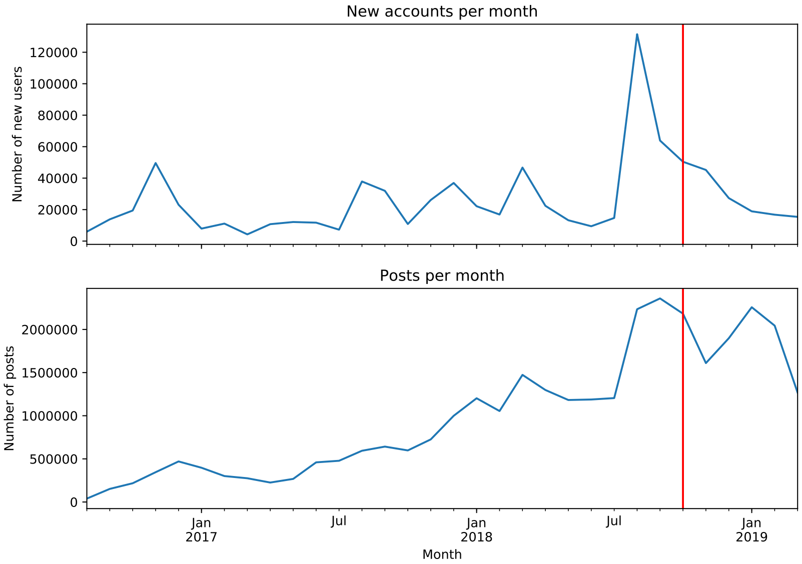

Figure 1: Site usage, with the shooting indicated with the red line

3.3 Attempted methods

I was hoping to use dynamic topic models in this work [2], but was never able to get them to yield useful results. I attempted with the two implementations first the pure Python one by *gensim* [18] which was far to slow on even $\frac{1}{32}$ of the data. Then the Java version created by Ble, et al [2], which was able to analysis my test set in only three few days, but when run on the full dataset did not complete within eight days. I also did train a couple vanilla LDA models, but the resulting topics were not very good and they still took a couple days to complete. The large run times of the topic models I believe is due to the shape of the data, topic models are traditionally run on corpora with many words per document, and relatively few documents. While my data have many documents (posts) and few words per document. One avenue of further work is to see how if combining documents, such as by grouping all posts by author per month, would improve performance and interpretability.

I was also hoping to use the change point detention model [10], but found it had already been done on *Gab* to look at hate speech and anti-Semitism [5]. This combined with me running low on time meant I did not do it.

4 Results

The code used for this project can be found ar github.com/reidmcy/csc-2611-final-project, there is also a much larger private repo that was for previous work, along with some continuing projects and shows the commit history, access can be provide to this upon request.

4.1 Ngrams

The first line of inquire was counting word occurrence an co-occurrence over time. This ngram analysis was done by binning the posts by month, then counting all unigrams, bigrams, trigrams and tetragrams. These were then filtered to remove, ones with hashtags and @ mentions. I also discard the tetragrams from my analysis as they proved to be very noisy and trigrams are the more common tool. Figure 2 shows the number of unique ngrams overtime of each type.

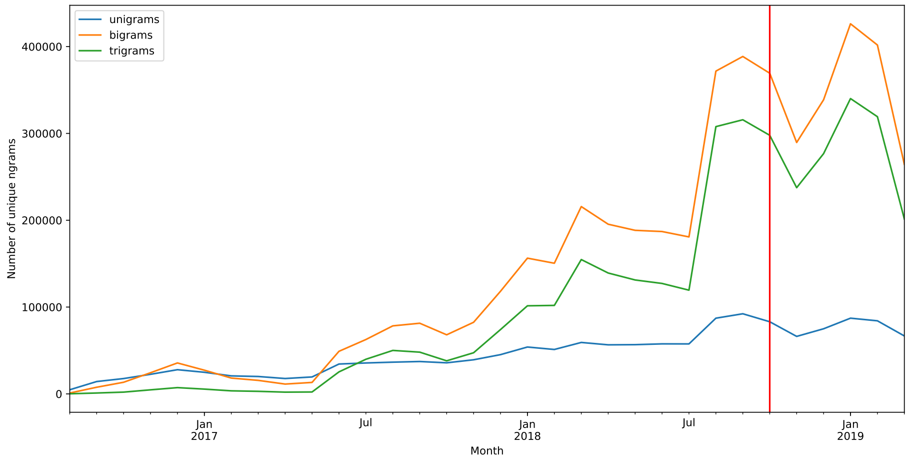

Figure 2: Number of unique ngrams over time, shooting date marked with a line

The top new trigrams each month, with new meaning not being top in a previous month, are shown in figure 3. The spiky nature of the occurrences shows that the discussion on the site would change quickly month to month. This makes the work usage generally more variable month to month.

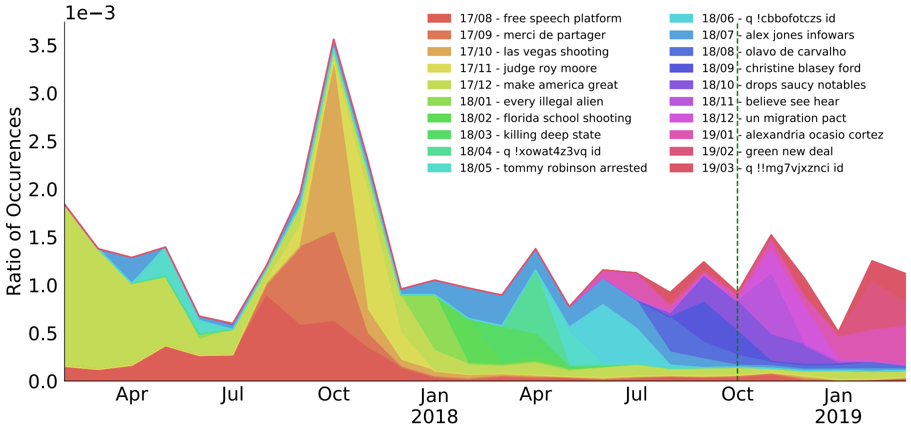

Figure 3: By month top new trigrams, with the shooting indicated with the line

To establish if the shooting had an effect on ant-Semtic langauge used on *Gab*, I looked at the rate of usage of the top 5 (by count the anti-Semitic terms. The terms are from *hatebase.org* [8]. Figure 4 shows that usage of some terms increased quickly after the shooting, but this did not last. Interestingly the word ‘jew’ became much more frequent on *Gab*, with table 2 showing the top trigrams containing ‘jew’, the before corpus is more than twice as large so the null-hypostasis would be for the counts to be half as large as before, which is not the case at all.

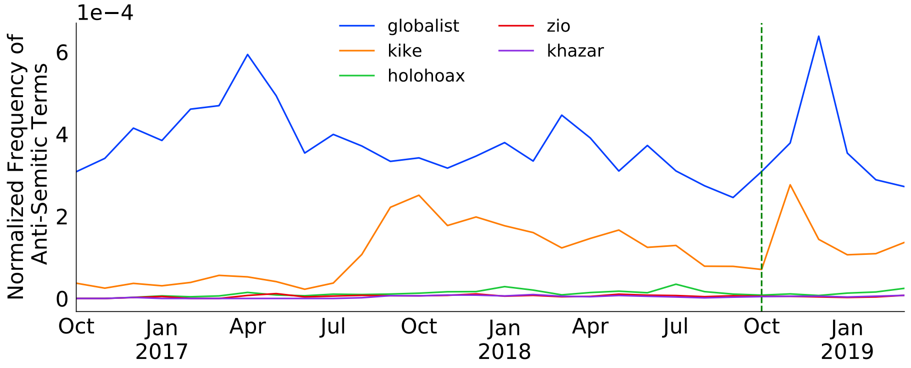

Figure 4: By month ratio of occurrence of the top 5 anti-Semitic words via hatebase.org, with the shooting indicated with the line

Before words	count	After words	count
jew,world,order	582	jewish,conservative,journalist	2701
anti,white,jewish	531	speaking,jew,hatred	1349
white,jewish,activist	493	jew,hatred,sharia	1349
already,taken,jewtube	331	simply,american,jewish	1348
serious,question,jewish	322	american,jewish,conservative	1348
question,jewish,groups	322	genetic,jewesses,jews	487
six,million,jews	252	nit,jew,islamist	426
000,000,jews	239	jew,haters,gab	402
jews,support,country	236	majority,jews,support	395
jew,jew,jew	234	vast,majority,jews	367
genetic,jews,jewesses	214	jew,world,order	339
false,system,jews	211	jews,rape,kids	273

Table 2: Before and after counts of top trigrams containing ‘jew’

Addition to anti-Semitic language I examined if language associated with the shooting had been affected. For some words, such as those shown if figure 5 there was a temporary increase in frequency. But when the whole of the corpus was considered there is not a clearer pattern. Table 3 shows the words with the largest log odds change before and after, with some filtering for spam already applied. The new and old words do not show a clear pattern and mostly seem to be affected by other events.

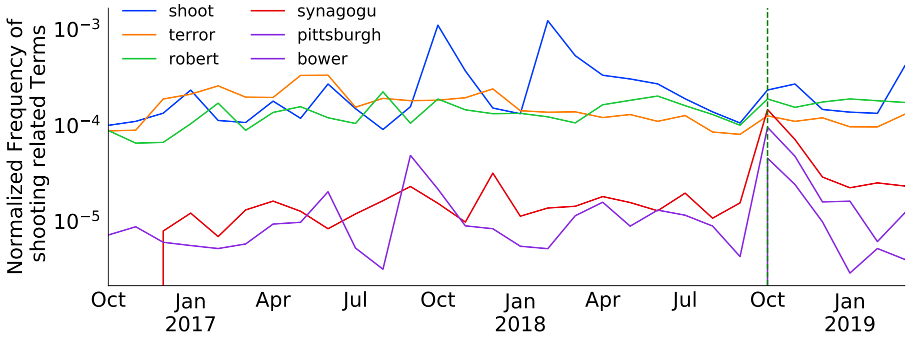

Figure 5: By month ratio of occurrence of some shooting related words, with the shooting indicated with the line. Note that the words were stemmed for the counting

4.2 Word2Vec

To see if shifts in meaning could be detected I created a series of word2vec [15] models using *gensim*[18]. These were trained on the posts before the shutdown,

Word	Log Odds	Word	Log Odds
ballots	2.82517	gillette	-4.32011
dec	2.68070	bonner	-4.12487
broward	2.50657	christchurch	-3.93367
oven	2.26191	alckmin	-3.83449
nov	2.21719	fords	-3.64577
compact	2.12065	candidatos	-3.49373
thanksgiving	2.07243	covington	-3.48308
acosta	2.01614	1d	-3.27645
christmas	1.97001	candidato	-3.20228
merry	1.85361	omarosa	-2.63009
december	1.77640	uol	-2.60407
macron	1.76386	marina	-2.59409

Table 3: Words with the lowest and highest log odds of occurring in After from the top 10,000 words by count in After

those after the shutdown and on each month separately. To tun the models I conducted the standard analogies tasks on, finding a window size of 10, using hierarchical softmax and 512 dimensions performed best after a basic grid search. For this analysis though I used 300 dimensions as performance difference is small and it allows me to use the Google News word2vec pre-trained embeddings. I also trained a word2vec model on all *reddit* posts from October 2018 to use as a baseline in comparisons.

My first attempt at seeing the semantic change of words was to use the procrustes aliment procedure [7]. Where each embedding is aligned to the previous one. When done monthly the results proved to be very noisy on the words I considered, such as 'jew' or 'liberal'. I believe this is mostly due to the size of the training data. Some months would have a cosine difference of 0, and the next month would be 1, with no pattern and when the order of alignments was changed (I tried aligning sequentiality starting at each month, and wrapping around to the first when necessary) the difference would be completely different.

To combat the noise I used the Google News word2vec pre-trained embeddings as a stable basis and aligned all my models to it, so it was unchanged after each alignment. This proved to lead to much less noise, although it means the measurements are now blind to shifts within a hypersphere defined by equal cosine difference from a point in Google News word2vec pre-trained embeddings space.

The results of the newly aligned word2vecs were also quite noisy, figrue 6 shows the cosine distance of words from their values in the Google News word2vec. You can see that month to month they shift by up to .2 with very little pattern and that all are very far away from the Google News word2vec originals. I also looked at some specifically sensitive terms to see if they also had the same pattern, figure 7. It is interesting that the terms such as 'women' or ‘black’ are more similar than ‘computer’ to the Google News word2vec originals, suggesting that they are weighted higher in the alignment and that less sensitive words may be more divergent.

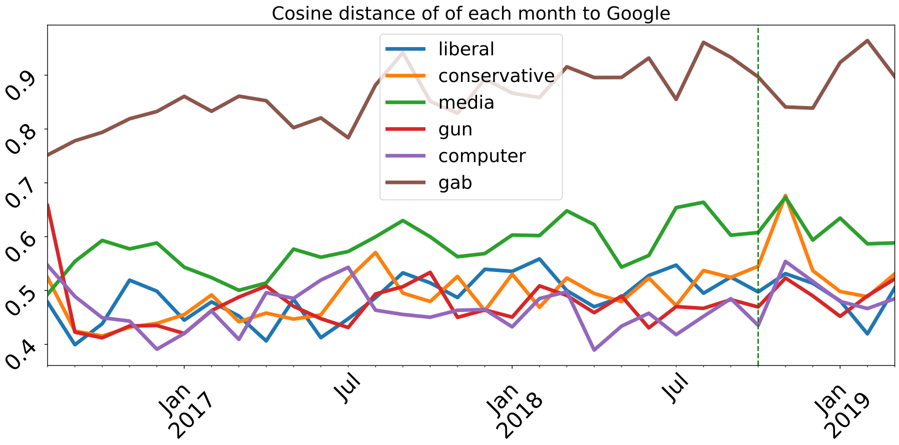

Figure 6: By month cosine distance of select words from their values in the Google News word2vec

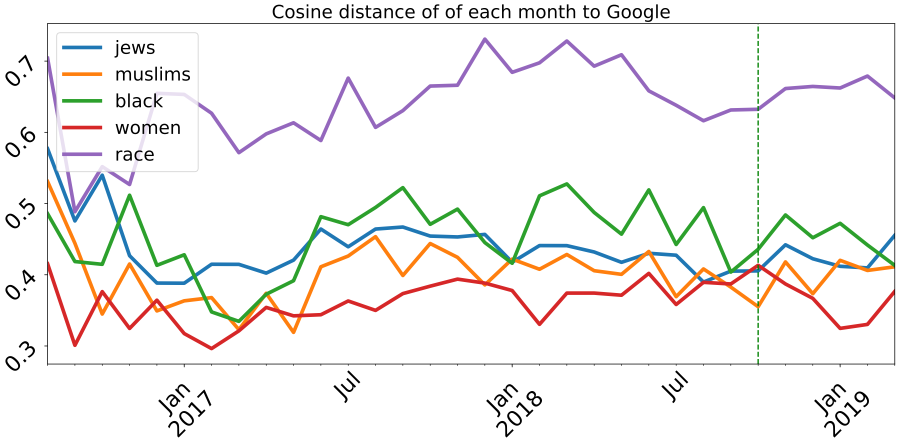

Figure 7: By month cosine distance of select ethnic words from their values in the Google News word2vec

To looking at specifically words related to anti-Semitism I found the nearest 3000 neighbours in Google News and looked at the top words also share by the Before. After and *Reddit* word2vecs. Figure 8 shows their cosine distances to the Google News. An alternative way to view this is by doing a t-SNE (on the top 10,000 words) and reducing to two dimensions the space, this is then a two dimensional measure that will not have the issue of not being sensitive to rotations in a hypersphere that the previous measures do, but it does have the much large downside of using information destroying dimension reduction, figure 9 shows how the words shift. That the blue and green lines have nearly perfect overlap shows that the before and after semantics of those words did not shift much.

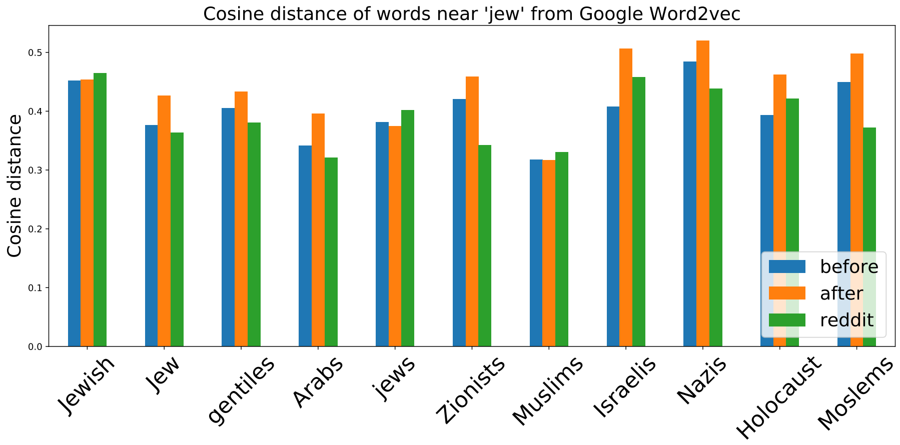

Figure 8: Cosine distance of different words from Google to aligned word2vecs

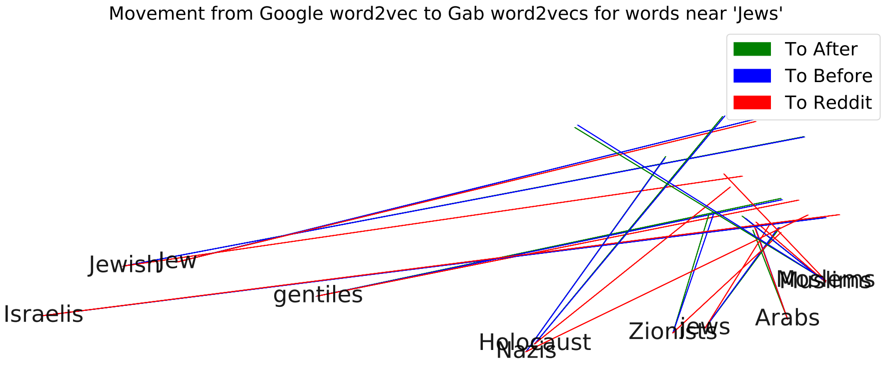

Figure 9: t-SNE embedding of all 4 aligned word2vecs, lines show movement from Google to to other Word2Vecs

5 Discussion

In this report I showed that there was a measurable shift in the language usage on *Gab* immediately after the shooting, but that it was likely temporary. The shift was also mostly in word frequencies and I was not able to provably identify any real semantic change. This lack of provability is due to two main factors, one the lack of a control and two the small size of the data. Future work could be done with holdout sets and randomization to start getting error bars along with simply having more data. The word usage change I was able to find maths other work [5] on the matter and suggests that long term affects on the communities language will be minor. Thus more nuanced measure will have to be used to detect real semantic change.

The shifts in usage patterns to *Gab* also suggest a substantial change in the community, the more ancillary members are leaving and only the true believers remain. The community also is more aware of it's observation by outsiders, and one thing I would like to see in the future is an analysis *Gab*'s non-standard English. Some popular topics, such as conspiracy theories like Qanon [9], lead to the creation of their own languages. These new words have bee used for a long time by political groups to communicate [13] and under harsher scrutiny they may be getting more intense.

References

[1] Kelly Bergstrom. dont feed the troll: Shutting down debate about community expectations on reddit. com. *First Monday*, 16(8), 2011.

[2] David M Blei and John D Lafferty. Dynamic topic models. In *Proceedings of the 23rd international conference on Machine learning*, pages 113–120. ACM, 2006.

[3] Eshwar Chandrasekharan, Umashanthi Pavalanathan, Anirudh Srinivasan, Adam Glynn, Jacob Eisenstein, and Eric Gilbert. You can't stay here: The efficacy of reddit's 2015 ban examined through hate speech. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW):31, 2017.

[4] Gordon Darroch. Gab alt-right's social media alternative attracts users banned from twitter. *The Guardian*, 2016.

[5] Joel Finkelstein, Savvas Zannettou, Barry Bradlyn, and Jeremy Blackburn. A quantitative approach to understanding online antisemitism. *arXiv preprint arXiv:1809.01644*, 2018.

[6] Leo A Goodman. Snowball sampling. *The annals of mathematical statistics*, pages 148–170, 1961.

[7] William L Hamilton, Jure Leskovec, and Dan Jurafsky. Diachronic word embeddings reveal statistical laws of semantic change. *arXiv preprint arXiv:1605.09096*, 2016.

[8] Hatebase. https://hatebase.org, accessed 2019-03-27, Mar 2019.

[9] Liam Stack Justin Bank and Daniel Victor. What is qanon: Explaining the internet conspiracy theory that showed up at a trump rally, August 2018.

[10] Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. Statistically significant detection of linguistic change. In *Proceedings of the 24th International Conference on World Wide Web*, pages 625–635. International World Wide Web Conferences Steering Committee, 2015.

[11] Lucas Lima, Julio CS Reis, Philipe Melo, Fabricio Murai, Leandro Araujo, Pantelis Vikatos, and Fabricio Benevenuto. Inside the right-leaning echo chambers: Characterizing gab, an unmoderated social system. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 515–522. IEEE, 2018.

[12] Edward Loper and Steven Bird. Nltk: the natural language toolkit. *arXiv preprint cs/0205028*, 2002.

[13] Ian Haney López. *Dog whistle politics: How coded racial appeals have reinvented racism and wrecked the middle class.* Oxford University Press, 2015.

[14] Reid McIlroy-Young and Ashton Anderson. From welcome new gabbers to the pittsburgh synagogue shooting: The evolution of gab. In *Proceedings of the Annual Conference on Weblogs and Social Media (ICWSM 2019)*, 20019.

[15] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[16] Abby Ohlheiser and Ian Shapira. Google's app store has banned gab - a social network popular with the far-right - for 'hate speech', October 2018.

[17] Whitney Phillips. The house that fox built: Anonymous, spectacle, and cycles of amplification. *Television & New Media*, 14(6):494–509, 2013.

[18] Radim Řehůřek and Petr Sojka. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta, May 2010. ELRA. `http://is.muni.cz/publication/884893/en`.

[19] Kevin Roose. On gab, an extremist-friendly site, pittsburgh shooting suspect aired his hatred in full, Oct 2018.

[20] Savvas Zannettou, Barry Bradlyn, Emiliano De Cristofaro, Michael Sirivianos, Gianluca Stringhini, Haewoon Kwak, and Jeremy Blackburn. What is gab? a bastion of free speech or an alt-right echo chamber? *arXiv preprint arXiv:1802.05287*, 2018.