

Gab, Does it Ever Change?

Reid McIlroy-Young

March 21, 2019

Questions

- Did the site shutdown measurably effect the site's speech?
- Are there other underlying trends in the site?
- Can we find trends in the hate speech?

Gab

Snowball sampling of Gab, last run March 27

	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
Words			423,539

Other Data

- Reddit
 - All comments made during October 2018
 - 112,573,001 Posts
 - 271,400 words in vocabulary (above 50 occurrences)
- Google News Word2vec - pretrained
 - Baseline word2vec for comparisons
 - 3,000,000 words in vocabulary (case sensitive)

Gab Usage, Before and After

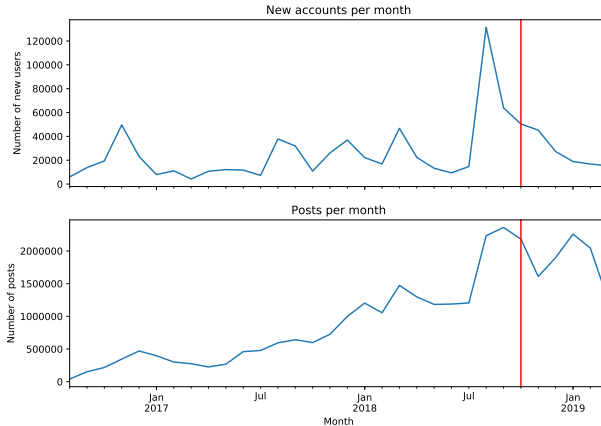


Figure: Site usage, with the shooting indicated with the red line

Top trigrams from After, by Month

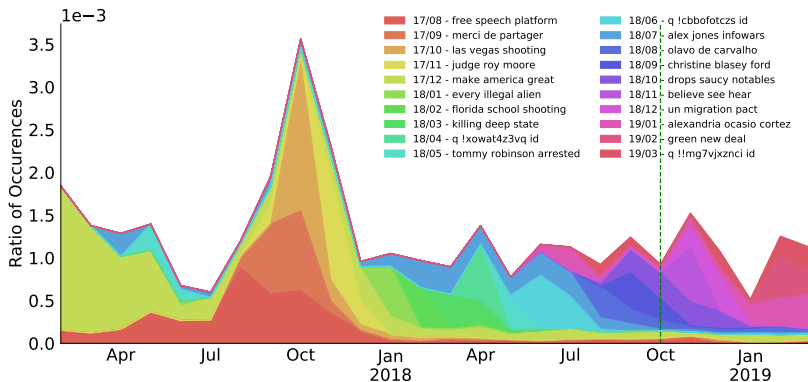


Figure: By month top new trigrams, with the shooting indicated with the line

Anti-Semitic words

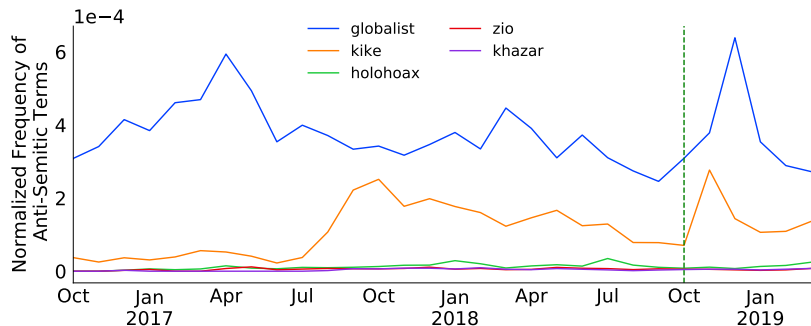


Figure: By month ratio of occurrence of the top 5 anti-Semitic words via hatebase.org, with the shooting indicated with the line

Trained Word2vecs

Before All posts before on Gab the shutdown

After All posts on Gab after the shutdown

Monthly Models trained on posts from each month since Gab's creation, 32 in total

Reddit All posts on Reddit in October 2018

Google The Google news pre trained word2vec

The Word2Vecs were all trained with a windows size of 10, and 300 dimensions, and a few other hyper parameters that were tuned with an analogies task

Differences between words

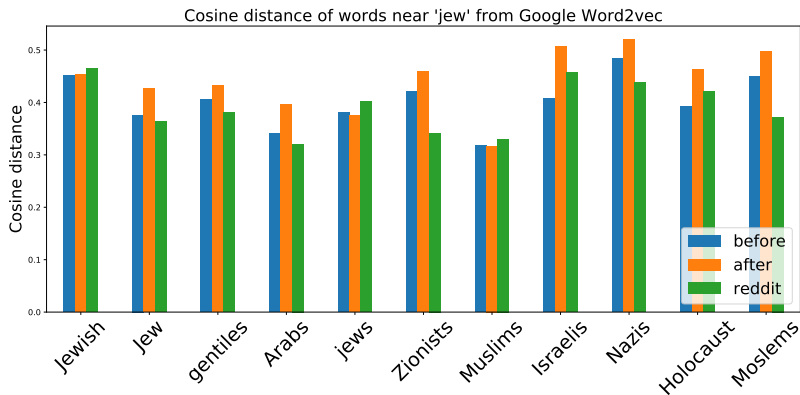


Figure: Cosine distance of different words from Google to aligned word2vecs

Differences between words

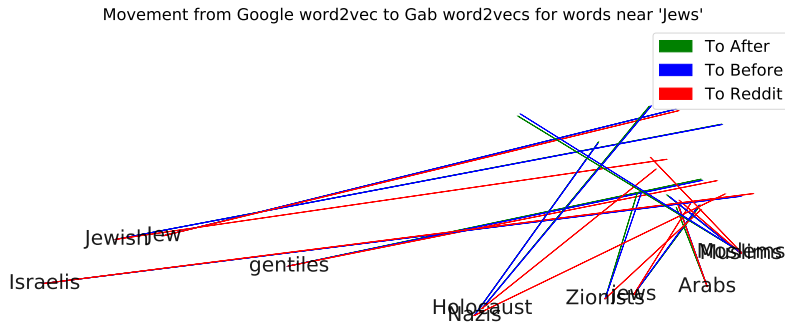


Figure: t-SNE embedding of all 4 aligned word2vecs, lines show movement from Google to to other Word2Vecs

Monthly Cosine Distances

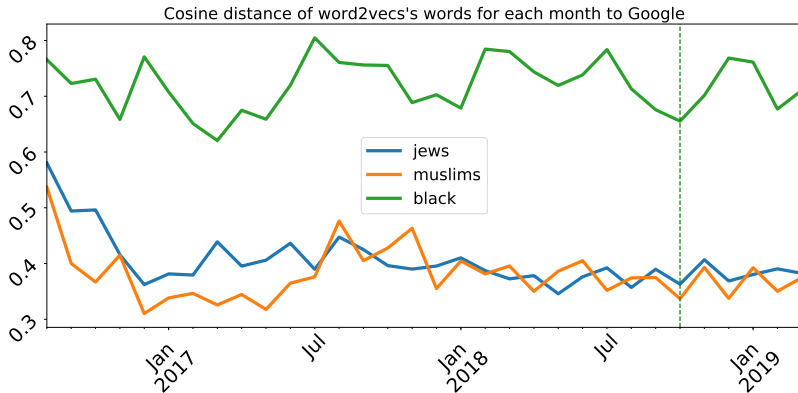


Figure: Distance from Google for select words by month

Monthly Cosine Distances

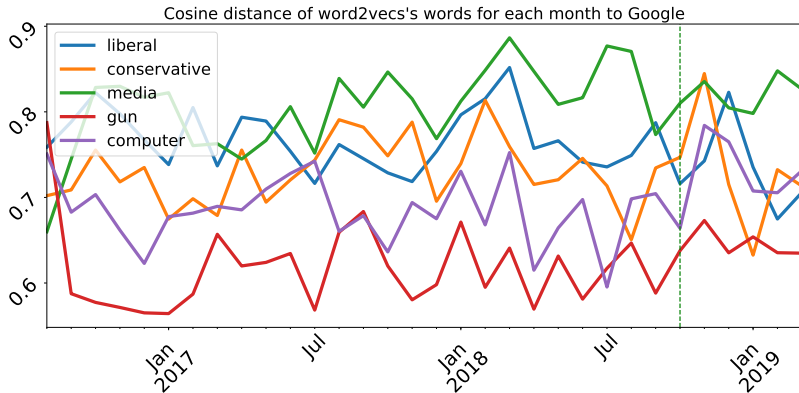


Figure: Distance from Google for select words by month

Least and Most Likely Words

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	ford's	-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: Words with the lowest and highest log odds of occurring in After from the top 10,000 words by count in After

Questions?