Identification And Categorization Of Toxic Twitter Posts Via Clustering

Foreword

This research was conducted between 1st November 2019 and 22nd January 2020 by Alexandros Kornilakis (University of Crete, FORTH-ICS institute) and Andrew Patel (F-Secure Corporation) as part of EU Horizon 2020 projects PROTASIS and SHERPA. SHERPA is an EU-funded project which analyses how AI and big data analytics impact ethics and human rights. PROTASIS is a project that aims to expand the reach of systems security to the international community via joint research efforts. The PROTASIS project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement, No. 690972.

Summary

Due to the complex nature of human language, automated detection of negativity and toxicity in content posted on forums, comments sections, and social networks is a difficult task. We posit that an accurate method to cluster textual content is a necessary precursor to any system that may eventually be capable of detecting abusive content, especially on platforms that limit the length of messages that can be authored (such as Twitter). Clustering can be used to find similar phrases, such as those found in regular spam, reply-spam-based propaganda, and content artificially amplified by organized disinformation groups. It is also useful for identifying topics of conversation (*what people think about something or someone, what people are talking about*), and may also be used to measure sentiment around those topics (*how strongly people agree or disagree with something or someone*). As this article will illustrate, accurate clustering can also be used to identify other interesting phenomena, such as users who attempt to "hide" spam by slightly altering each of their tweets, groups of accounts spreading hatred towards specific demographics, and groups of accounts spreading disinformation, hoaxes, and fake news.

In this article, we detail our own novel clustering methodology, based on meta embeddings and community detection, and the results of applying that methodology a number of different datasets collected from Twitter, including replies to US politicians, tweets captured against hashtags pertaining to the 2019 UK general elections, and content gathered from UK far-right activists. We present several examples of the output of our clustering methodology, including analysis and interpretation of the results we obtained, and an interactive site for readers to explore. We also discuss some future directions for this line of research.

Introduction

Anyone who's read comments sections on news sites, looked at replies to social media posts authored by politicians, or read comments on YouTube will appreciate that there's a great deal of toxicity on the internet. Some female and minority high-profile Twitter users are the target of constant, serious harassment, including death threats (https://www.youtube.com/watch?v=A3MopLxgvLc) from both individuals and coordinated groups of users. Social media posts authored by politicians, journalists, and news organizations

often receive large numbers of angry or downright toxic replies from people who don't support their statements or opinions. Some of these replies originate from fake accounts that have been created for the express purpose of trolling - the process of posting controversial comments designed to provoke emotional reactions and start fights. Trolling is a highly efficient way to spread rumors and disinformation, alter public opinion, and disrupt otherwise meaningful conversation, and, as such, is a tool often used by organized groups of political activists, commercial troll farms, and nation state disinformation campaigns.

On Twitter, troll accounts sometimes use a technique called reply-spamming to fish for engagement. This technique involves replying to a large number of high-profile accounts with the same or similar messages. This achieves two goals. The first is organic visibility - many people read replies to posts from politicians, and thus may read the post from the troll account. The second is social engineering – people get angry and reply to the troll's posts, and occasionally the owner of the high-profile account may be tricked into engaging with the post themselves. Although high-profile accounts are rarely engaged by such tactics, there are examples of it happening. Recently, a parody account named Shaniqua O'Toole, claiming to be a Guardian journalist, managed to gain engagement from a few high-profile verified Twitter accounts by posting replies to their tweets that contained fake screenshots that looked like headlines on The Guardian's website.

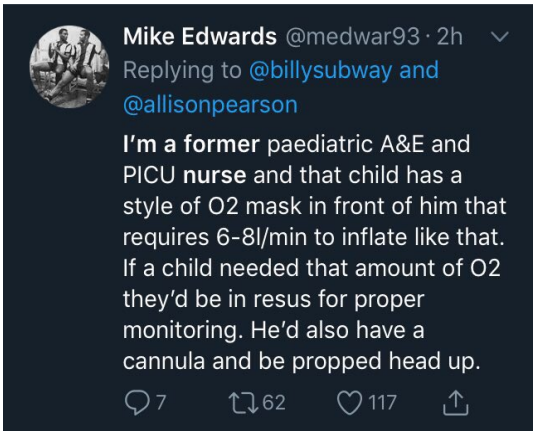
Above: an example of a fake Guardian headline containing parody columnist "Shaniqua O'Toole"

Successful reply-spam-based disinformation in the lead-up to the 2019 UK General Election

Reply-spam was also used to successfully propagate disinformation during the run-up to the December 2019 UK general election. One such occasion involved a situation where a journalist attempted to show a picture of a child sleeping on the floor of an overcrowded hospital to Boris Johnson during a television interview. Instead of looking at the picture, Johnson pocketed the reporter's phone and attempted to change the subject of their conversation. A clip of the interview went viral on social media, and shortly after, a large number of accounts published posts on various social networks, including Facebook and Twitter, claiming to be an acquaintance of one of the senior nurses at the hospital, and that the aforementioned nurse could verify that the picture was faked (https://twitter.com/marcowenjones/status/1204183081009262592).

Above: some of the original reply spam tweets regarding the Leeds Hospital incident. Note how they are all replies to politicians and journalists.

Above: tory activists on Twitter reinforced the original campaign with more copy-paste reply spam



Above: this was shortly followed by a second campaign containing a different tweet that was also copy-pasted across social networks (by the same group of tory activists)

Many of the accounts that posted this content on Twitter were created specifically for that purpose, and deleted shortly afterwards (https://twitter.com/r0zetta/status/1204519439640801280). The picture of the child sleeping on the floor of the hospital had appeared a week prior to the interview with Johnson in a local newspaper, and at that time, both the story and picture had been verified with personnel at the hospital. However, the fake social media posts were amplified to such a degree that voters, including those living in Leeds, believed that the picture had been faked. At least on Twitter, this

disinformation was spread using reply-spam aimed at posts authored by politicians and journalists.

During the run-up to the 2019 UK general elections, posts on social networks were enough to propagate false information. Very few traditional "fake news" sites were uncovered, and it is unlikely that those that were found had any significant impact. Fake news sites are traditionally created in order to give legitimacy to fabricated, "clickbait" headlines. However, people are often inclined to share a headline without even visiting the original article. As such, fake news sites are rarely necessary. Nowadays, it is often enough to simply post something emotionally appealing on a social network, promote it enough to reach a handful of people, and then sit back and watch as it is organically disseminated by proxy. Once a rumor or lie has been spread in this manner, it enters the public's consciousness, and can be difficult to later refute, even if the initial claim is debunked (https://twitter.com/r0zetta/status/1210499949064052737).

Dealing with social media posts on a large scale

Anyone who runs a prominent social media account is unlikely to be able to find relevant or interesting replies to content they've posted due to the fact that they must wade through hundreds or even thousands of replies, many of which are toxic. This essentially amounts to an informational denial of service for both the account owner, and anyone with a genuine need to contact them. Well-established anti-spam systems exist to assist users with this problem for email, but no such systems exist for social networks. Since notification interfaces on most social networks don't scale well for highly engaged accounts, an automated filtering system would be a more than welcome feature.

Detection of unwanted textual content such as email spam and hate speech is a much easier task than detecting nuances in language indicative of negativity or toxicity. Spam messages typically follow patterns that can be accurately separated with clustering techniques or even regular expressions. Hate speech often contains words that are rarely used outside of their context, and hence can be successfully detected with string matches and other relatively simple techniques. One might assume that sentiment analysis techniques could be used to find toxic content, but they are, unfortunately, still rather inaccurate on real-world data. They often fail to understand the fact that the context of a word can drastically alter its meaning (e.g. "You're a rotten crook" versus "You'll beat that crook in the next election"). Although accurate sentiment analysis techniques may eventually be of use in this area, software designed to filter toxic comments may require more metadata (such as the subject matter, or topic of the message) in order to perform accurately, or to provide a better explanation as to why certain messages were filtered.

Motivation for using a clustering / topic modelling approach

In the context of our work, clustering (or topic modelling) is the process of grouping phrases or passages (or, in this case, tweets) into "buckets" based on their topic or subject matter. Clustering of textual content is useful for finding similar phrases, such as those found in regular spam (e.g. porn bots), reply-spam-based propaganda, and content artificially amplified by organized disinformation groups. It is also useful for identifying topics of conversation (*what people think about something or someone, what people are talking about*), and may also be used to measure sentiment around those topics (*how strongly people agree or disagree with something or*

someone). As this article will illustrate, accurate clustering can also be used to identify other interesting phenomena, such as users who attempt to "hide" spam by slightly altering each of their tweets (something that is cumbersome to detect via regular expressions), groups of accounts spreading hatred towards specific demographics, and groups of accounts spreading disinformation, hoaxes, and fake news. Furthermore, the results of accurate clustering and topic modeling can be fed into downstream tasks such as:

- systems designed to fact-check posts and comments
- systems designed to detect and track rumors and the spread of disinformation, hoaxes, scams, and fake news
- systems designed to identify the political stance of content published by one or more accounts or conversations
- systems designed to quantify public opinion and assess the impact of social media on public opinion
- trust analysis tasks (including those used to determine the quality of accounts on social networks)
- the creation of disinformation knowledge bases and datasets
- detection of bots or spam publishers

To this end, we have attempted to build a system that is capable of clustering the type of written content typically encountered on social networks (or more specifically, on Twitter). Our experiments focus on tweets posted in reply to content authored by prominent US politicians and presidential candidates.

Experiments

We started by collecting two datasets:

Set 1: US Democrats

The first set captured direct replies to tweets published by a number of highly engaged democrat-affiliated Twitter accounts - @JoeBiden, @SenSanders, @BernieSanders, @SenWarren, @ewarren, @PeteButtigieg, @MikeBloomberg, @amyklobuchar, @AndrewYang and @AOC - between Sun Dec 15 2019 and Mon Jan 13 2020. A total of 978,721 tweets were collected during this period. After preprocessing, a total of 719,617 tweets remained.

Set 2: Donald Trump

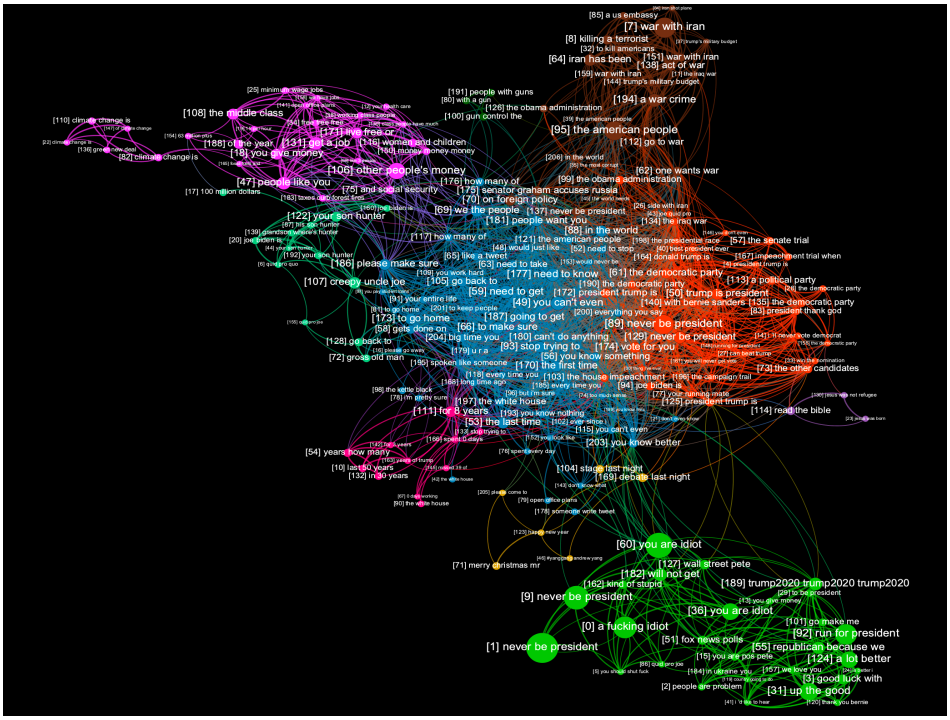
The second set captured direct replies to tweets published by @realDonaldTrump between Sun Dec 15 2019 and Wed Jan 08 2020. A total of 4,940,317 tweets were collected during this period. Due to the discrepancy between the sizes of the two collected datasets, we opted to utilize a portion of this set containing 1,022,824 tweets. After preprocessing, a total of 747,232 tweets remained.

We developed our own clustering methodology for this research, which involved preprocessing of captured data, converting tweets into sentence vectors (using different techniques), combining those vectors into meta embeddings, and then creating node-edge graphs using similarities between calculated meta embeddings. Clusters were then derived by performing community detection on the resulting graphs. A detailed description of our methodology can be found in appendix 1 of this article.

Experiment 1: US Democrats

Our first experiment involved clustering of a subset of data in set 1 (US democrats). We clustered a batch of 34,003 tweets, resulting in 209 clusters. We created an interactive demo using results of this clustering experiment that can be found here: https://twitter-clustering.web.app/ Note that this interactive demo will not display correctly on mobile browsers, so we encourage you to visit it from a desktop computer. Use the scroll wheel to zoom in and out of the visualization space, left-click and drag to move the nodes around, and click on nodes or communities themselves to see details. Details include names of accounts that were replied to the most in tweets assigned to that cluster, subject-verb-object triplets and overall sentiment extracted from those tweets, and the two most relevant tweets, loaded on the right of the screen, as examples. Different communities related to different topics (e.g. Community 2 contains clusters relevant to recent events in Iran).

A image below is a static graph visualization of the discovered clusters. Labels were derived by matching commonly occurring words, and bigram combinations of those words with ngrams and subject-verb-object triplets found in the tweets contained within each cluster. The code for doing this can be found at https://github.com/r0zetta/meta_embedding_clustering under the code directory.



We ran sentiment analysis on each cluster by taking the average sentiment calculated across all tweets contained in the cluster. Sentiment analysis was performed with TextBlob's lexical sentiment analyzer. We then summarized negative and positive groups of clusters by counting words, ngrams, and which account was replied to. We also extracted subject-verb-object triplets from clusters using the textacy python module.

```
Positive clusters: 51 - 9498 tweets.
svo
(you, have, vote) / (you, will never be, president) / (i, love, you) / (we, love, you) / (bho, sent, pallets) /
(you, want, get rid) / (it, has, to be) / (you, remove, %) / (you, would pardon, trump) / (we, need, you) /
(you, don't have, chance) / (jesus, was not, refugee) / (trump, is kicking, people) / (i, 'm gon, na keep) /
(you, 're, moron) /

ngram
thank you for / you and your / you want to / the stock market / running for president / we need to /
keep up the / the middle class / of the united / to pay for / is going to / a lot of / vote for you /
why don't you / you are the /

word
like / get / people / president / good / thank / would / one / joe / pay / trump / better /
bernie / money / vote /

replied_to
1996    JoeBiden
1773    AndrewYang
1359    AOC
1111    BernieSanders
857     PeteButtigieg
698     ewarren
644     SenSanders
534     amyklobuchar
440     MikeBloomberg
86      SenWarren
```

Note how, in the above, sentiment analysis has incorrectly categorized a few statements such as "you will never be president" and "you're a moron" as positive.

```
Negative clusters: 151 - 22867 tweets.
svo
(you, will never be, president) / (you, are, liar) / (you, open, mouth) / (you, missed, %) /
(you, have, vote) / (you, want, to be) / (you, have, idea) / (you, know, nothing) / (we, need, president) /
(you, are, traitor) / (you, read, bible) / (you, spent, days) / (people, are, problem) / (you, lost, friend) /
(you, are, joke) /

ngram
you are a / you and your / you need to / you will never / quid pro quo / you want to / will never be /
quid pro joe / the white house / go back to / never be president / the united states / the most corrupt /
a war crime / in the last /

word
people / joe / trump / like / president / know / iran / war / one / us / get / never / need /
son / obama /

replied_to
6249    JoeBiden
4001    AOC
2777    PeteButtigieg
2376    BernieSanders
1962    ewarren
1875    AndrewYang
1280    SenSanders
1107    amyklobuchar
1020    MikeBloomberg
220     SenWarren
```

As you can see in the above, negative clusters outnumbered positive clusters by a factor of two.

```
Toxic clusters: 7 - 1638 tweets.
svo
(you, are, idiot) / (you, need, to go) / (you, are, traitor) / (iran, shot, plane) / (you, were, vice president) /
(they, don't know, better) /

ngram
you are a / this is a / the most corrupt / most corrupt administration / you are an / are an idiot /
you really are / go back to / back to bartending / please go away / and go back / you need to /
tit for tat / joe biden was / a war crime /

word
joe / corrupt / idiot / stupid / shut / like / shit / fucking / liar / please / go / stop /
back / away / mouth /

replied_to
467     JoeBiden
393     AOC
225     PeteButtigieg
132     ewarren
120     BernieSanders
84      AndrewYang
76      MikeBloomberg
66      SenSanders
54      amyklobuchar
21      SenWarren
```

Above are clusters designated toxic by virtue of their average sentiment score.

```
Positive tweets: 9498   Negative tweets: 22867  Toxic tweets: 1638
Positive clusters: 51   Negative clusters: 151  Toxic clusters: 7

JoeBiden        POS: 22.91%     NEG: 71.73%     TOX: 5.36%
AOC             POS: 23.62%     NEG: 69.55%     TOX: 6.83%
amyklobuchar    POS: 31.50%     NEG: 65.31%     TOX: 3.19%
PeteButtigieg   POS: 22.21%     NEG: 71.96%     TOX: 5.83%
MikeBloomberg   POS: 28.65%     NEG: 66.41%     TOX: 4.95%
ewarren         POS: 25.00%     NEG: 70.27%     TOX: 4.73%
SenSanders      POS: 32.36%     NEG: 64.32%     TOX: 3.32%
AndrewYang      POS: 47.51%     NEG: 50.24%     TOX: 2.25%
BernieSanders   POS: 30.80%     NEG: 65.87%     TOX: 3.33%
SenWarren       POS: 26.30%     NEG: 67.28%     TOX: 6.42%
```

Above is a breakdown of replies by verdict for each candidate. Percentage-wise, @AndrewYang received by far the most positive replies, and @AOC and @SenWarren received the largest ratio of toxic replies.

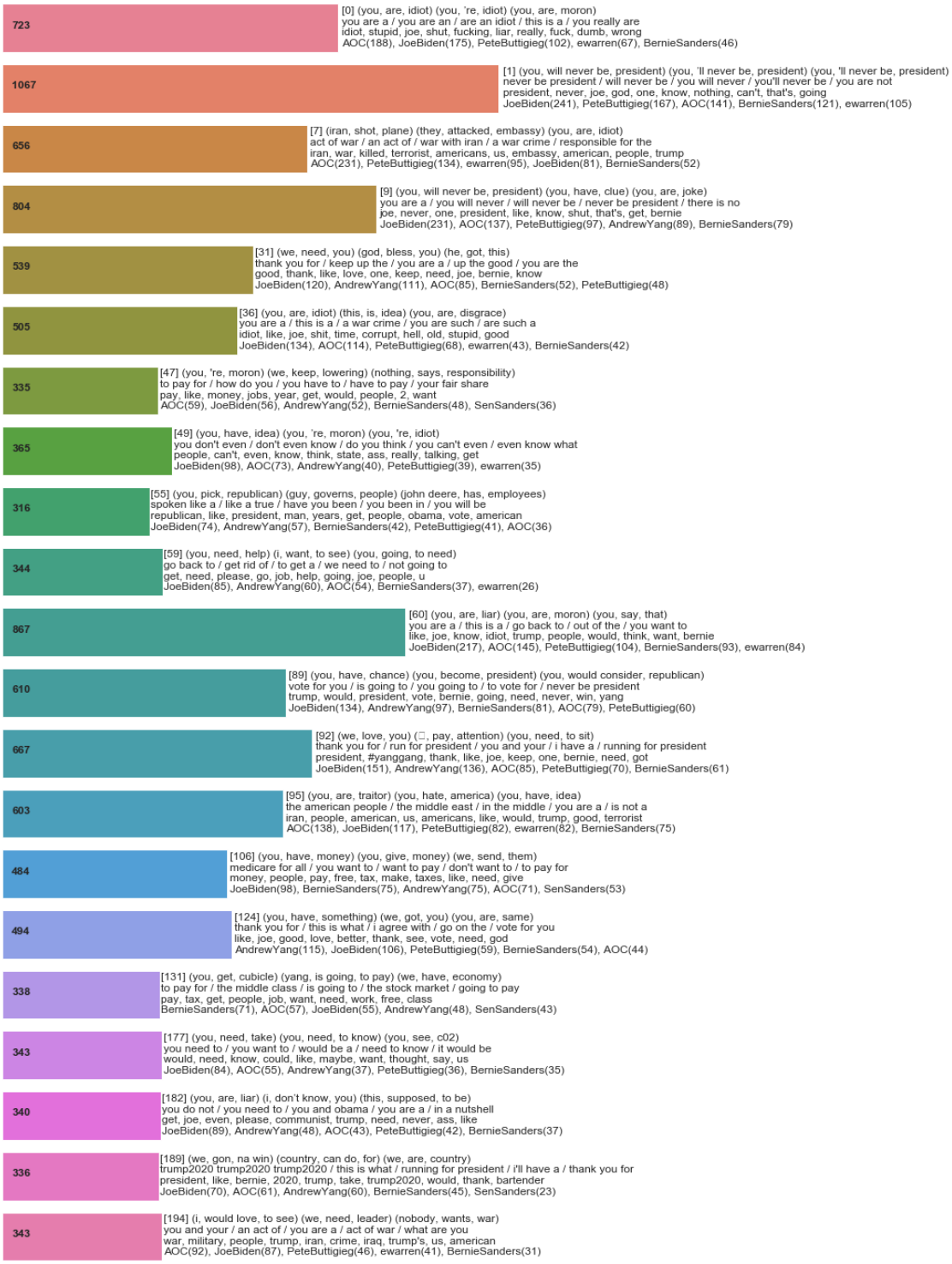
This simple analysis isn't, unfortunately, all that accurate, due to deficiencies in the sentiment analysis library used.

The following chart contains summaries of some of the larger clusters identified. Most of the larger clusters contained negative replies, including common themes such as:

- you are an idiot/moron/liar/traitor (or similar)
- you will never be president
- Trump will win the next election

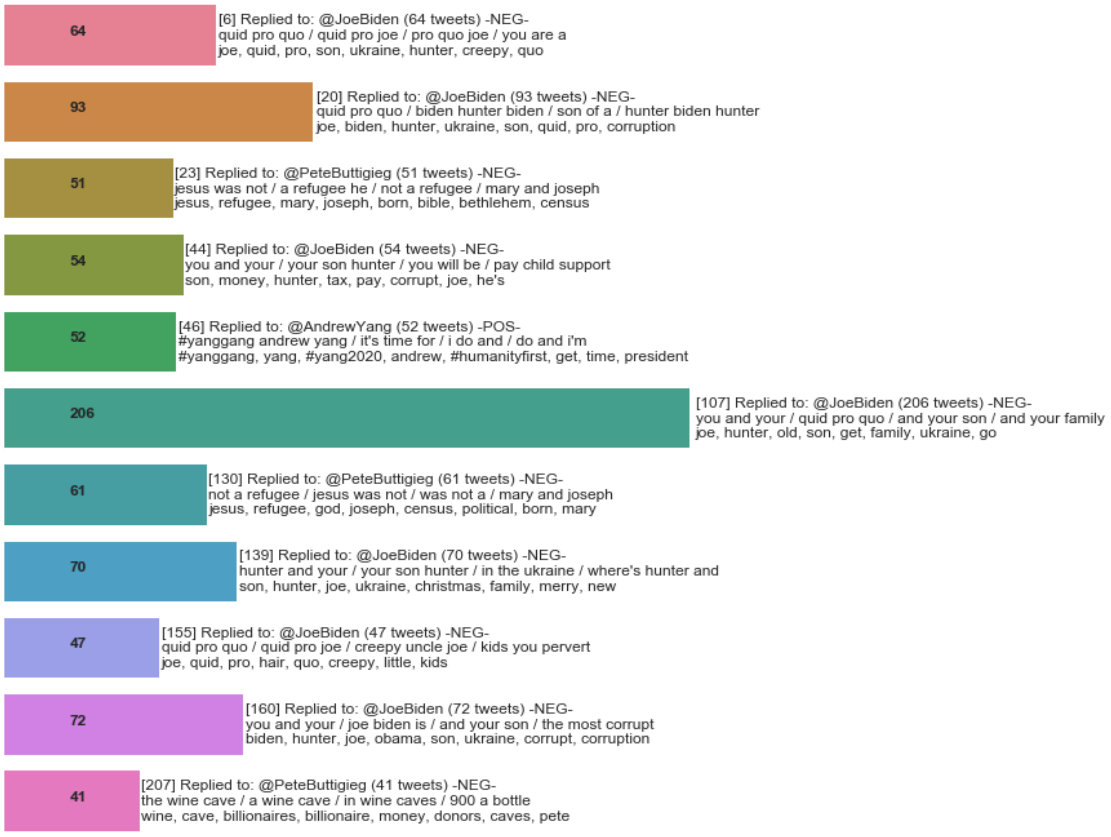
Positive themes included:

- We love you
- You got this
- You have my vote


723
[0] (you, are, idiot) (you, 're, idiot) (you, are, moron)
you are a / you are an / are an idiot / this is a / you really are
idiot, stupid, joe, shut, fucking, liar, really, fuck, dumb, wrong
AOC(188), JoeBiden(175), PeteButtigieg(102), ewarren(67), BernieSanders(46)
1067
[1] (you, will never be, president) (you, 'll never be, president) (you, 'll never be, president)
never be president / will never be / you will never / you'll never be / you are not
president, never, joe, god, one, know, nothing, can't, that's, going
JoeBiden(241), PeteButtigieg(167), AOC(141), BernieSanders(121), ewarren(105)
656
[7] (iran, shot, plane) (they, attacked, embassy) (you, are, idiot)
act of war / an act of / war with iran / a war crime / responsible for the
iran, war, killed, terrorist, americans, us, embassy, american, people, trump
AOC(231), PeteButtigieg(134), ewarren(95), JoeBiden(81), BernieSanders(52)
804
[9] (you, will never be, president) (you, have, clue) (you, are, joke)
you are a / you will never / will never be / never be president / there is no
joe, never, one, president, like, know, shut, that's, get, bernie
JoeBiden(231), AOC(137), PeteButtigieg(97), AndrewYang(89), BernieSanders(79)
539
[31] (we, need, you) (god, bless, you) (he, got, this)
thank you for / keep up the / you are a / up the good / you are the
good, thank, like, love, one, keep, need, joe, bernie, know
JoeBiden(120), AndrewYang(111), AOC(85), BernieSanders(52), PeteButtigieg(48)
505
[36] (you, are, idiot) (this, is, idea) (you, are, disgrace)
you are a / this is a / a war crime / you are such / are such a
idiot, like, joe, shit, time, corrupt, hell, old, stupid, good
JoeBiden(134), AOC(114), PeteButtigieg(68), ewarren(43), BernieSanders(42)
335
[47] (you, 're, moron) (we, keep, lowering) (nothing, says, responsibility)
to pay for / how do you / you have to / have to pay / your fair share
pay, like, money, jobs, year, get, would, people, 2, want
AOC(59), JoeBiden(56), AndrewYang(52), BernieSanders(48), SenSanders(36)
365
[49] (you, have, idea) (you, 're, moron) (you, 're, idiot)
you don't even / don't even know / do you think / you can't even / even know what
people, can't, even, know, think, state, ass, really, talking, get
JoeBiden(98), AOC(73), AndrewYang(40), PeteButtigieg(39), ewarren(35)
316
[55] (you, pick, republican) (guy, governs, people) (john deere, has, employees)
spoken like a / like a true / have you been / you been in / you will be
republican, like, president, man, years, get, people, obama, vote, american
JoeBiden(74), AndrewYang(57), BernieSanders(42), PeteButtigieg(41), AOC(36)
344
[59] (you, need, help) (i, want, to see) (you, going, to need)
go back to / get rid of / to get a / we need to / not going to
get, need, please, go, job, help, going, joe, people, u
JoeBiden(85), AndrewYang(60), AOC(54), BernieSanders(37), ewarren(26)
867
[60] (you, are, liar) (you, are, moron) (you, say, that)
you are a / this is a / go back to / out of the / you want to
like, joe, know, idiot, trump, people, would, think, want, bernie
JoeBiden(217), AOC(145), PeteButtigieg(104), BernieSanders(93), ewarren(84)
610
[89] (you, have, chance) (you, become, president) (you, would consider, republican)
vote for you / is going to / you going to / to vote for / never be president
trump, would, president, vote, bernie, going, need, never, win, yang
JoeBiden(134), AndrewYang(97), BernieSanders(81), AOC(79), PeteButtigieg(60)
667
[92] (we, love, you) (□, pay, attention) (you, need, to sit)
thank you for / run for president / you and your / i have a / running for president
president, #yanggang, thank, like, joe, keep, one, bernie, need, got
JoeBiden(151), AndrewYang(136), AOC(85), PeteButtigieg(70), BernieSanders(61)
603
[95] (you, are, traitor) (you, hate, america) (you, have, idea)
the american people / the middle east / in the middle / you are a / is not a
iran, people, american, us, americans, like, would, trump, good, terrorist
AOC(138), JoeBiden(117), PeteButtigieg(82), ewarren(82), BernieSanders(75)
484
[106] (you, have, money) (you, give, money) (we, send, them)
medicare for all / you want to / want to pay / don't want to / to pay for
money, people, pay, free, tax, make, taxes, like, need, give
JoeBiden(98), BernieSanders(75), AndrewYang(75), AOC(71), SenSanders(53)
494
[124] (you, have, something) (we, got, you) (you, are, same)
thank you for / this is what / i agree with / go on the / vote for you
like, joe, good, love, better, thank, see, vote, need, god
AndrewYang(115), JoeBiden(106), PeteButtigieg(59), BernieSanders(54), AOC(44)
338
[131] (you, get, cubicle) (yang, is going, to pay) (we, have, economy)
to pay for / the middle class / is going to / the stock market / going to pay
pay, tax, get, people, job, want, need, work, free, class
BernieSanders(71), AOC(57), JoeBiden(55), AndrewYang(48), SenSanders(43)
343
[177] (you, need, take) (you, need, to know) (you, see, c02)
you need to / you want to / would be a / need to know / it would be
would, need, know, could, like, maybe, want, thought, say, us
JoeBiden(84), AOC(55), AndrewYang(37), PeteButtigieg(36), BernieSanders(35)
340
[182] (you, are, liar) (i, don't know, you) (this, supposed, to be)
you do not / you need to / you and obama / you are a / in a nutshell
get, joe, even, please, communist, trump, need, never, ass, like
JoeBiden(89), AndrewYang(48), AOC(43), PeteButtigieg(42), BernieSanders(37)
336
[189] (we, gon, na win) (country, can do, for) (we, are, country)
trump2020 trump2020 trump2020 / this is what / running for president / i'll have a / thank you for
president, like, bernie, 2020, trump, take, trump2020, would, thank, bartender
JoeBiden(70), AOC(61), AndrewYang(60), BernieSanders(45), SenSanders(23)
343
[194] (i, would love, to see) (we, need, leader) (nobody, wants, war)
you and your / an act of / you are a / act of war / what are you
war, military, people, trump, iran, crime, iraq, trump's, us, american
AOC(92), JoeBiden(87), PeteButtigieg(46), ewarren(41), BernieSanders(31)

Several clusters contained replies directed at just one account. They contained either replies to specific content posted by that account, or comments specifically directed at the politician's history or personal life, including the following:

- Comments about Joe Biden's son
- Replies to Pete Buttigieg correcting him on a tweet about Jesus being a refugee
- Comments about Joe Biden's involvement in the Ukraine
- Comments about Pete Buttigieg's net worth, and something about expensive wine
- Highly positive replies to Andrew Yang's posts



Noteworthy clusters

@AOC: you're a fucking idiot aholec.
@MikeBloomberg: you are a fuc$ing idiot.
@JoeBiden: you sir, are an idiot.
@PeteButtigieg: and you sir, are an idiot.
@AOC: and you aoc are an idiot!
@JoeBiden: you sir are an idiot.
@AOC: aoc you are a fucking idiot.
@ewarren: you are.... such an idiot
@JoeBiden: biden, you are an idiot.
@PeteButtigieg: you sir are a idiot.
@AOC: what an idiot you are.
@AOC: aoc you are an idiot.
@AOC: you are such an idiot.
@SenSanders: you are a blithering idiot.
@AOC: aoc, you are such an idiot
@ewarren: you are an ill informed idiot.
@AOC: you're an idiot. that is all 🤦‍♀️
@AOC: yea, well you're an idiot
@AOC: hell naw bisch... you're an idiot.
@PeteButtigieg: you are a friggin idiot.....
@AOC: aoc, you are an idiot
@AndrewYang: you are such a fucking idiot.
@AOC: lol. you're such an idiot
@AOC: aoc you are an idiot!
@MikeBloomberg: so you really are an idiot.
@PeteButtigieg: you sir are an idiot...
@ewarren: you are an idiot. stfu!
@PeteButtigieg: good god you're an idiot
@AOC: aoc you are an idiot
@AOC: you're such a fucking idiot
@AOC: lol you are an idiot
@PeteButtigieg: you are an idiot sir.
@AndrewYang: yeah... you're a fucking idiot.

@AndrewYang: yang gang we can do it!!!
@JoeBiden: at least he's open - i like #va02
@AndrewYang: the west coast supports you. you know what to do.
@AndrewYang: you can do it boss! #yang2020 #yanggang
@PeteButtigieg: keep going all the way through
@JoeBiden: sorry joe. you hold that honor
@JoeBiden: you've got my vote, joe!
@JoeBiden: ... i really hope you get the nomination...
@PeteButtigieg: i'm with you, pete! don't get discouraged.
@MikeBloomberg: i know you are " rinpoche"
@JoeBiden: i'm with you joe. #enoughisenough
@MikeBloomberg: you will have my vote, sir.
@amyklobuchar: chats right the law is egal for everyone
@AndrewYang: keep doing you!; #yanggang ❤️
@amyklobuchar: and they voted for him.
@JoeBiden: you have my support joe!💪🏾
@JoeBiden: so agree with you! i stand with joe!!
@amyklobuchar: but this is what they voted for.
@JoeBiden: yeah joe! keep this up!!!
@JoeBiden: nope! pretty sure that was you and daddy o!
@JoeBiden: you now have my vote! love, love, love this!
@JoeBiden: joe malarkey biden. perfect definition of you
@JoeBiden: whoomp there it is. well played joe biden.
@JoeBiden: and bernie by similar amounts.
@AndrewYang: i will do my part
@MikeBloomberg: you've got my vote, mr. bloomberg!
@JoeBiden: sure joe, whatever you say! 😂😂😂
@JoeBiden: keep it up joe..we got your back👍
@JoeBiden: thanks joe. we know you have our backs. and i hope you know,,we have yours too
@AndrewYang: get comfortable up there. we are here to stay. 🧢🤠👍
@JoeBiden: if you like your dr you can keep your dr
@JoeBiden: this is the joe i am supporting!!
@JoeBiden: oh uncle joe you're my favorite
@JoeBiden: ya, hillary was saying the same shit.
@AndrewYang: we have your back! be strong!

Above: two discovered clusters – one containing toxic replies, and another containing praise

@BernieSanders: well stop using california and australia's fires to claim climate change when you know they were arso
n !!!
@JoeBiden: are you crazy?? you sound like you have alzheimer's fires are not caused by global change or climate chan…
@AOC: people started those wild fires !! not climate change stop lying
@AOC: the australian fires were caused by climate change fanatics like you. people lit the fires on purpose.
@SenSanders: just because you are probably right about climate change contributing to fires in australia doesn't mak…
@SenSanders: tell me exactly how probably lit fires are a consequence of global climate change.
@SenSanders: bernie is flat-out lying about climate change being the cause of the horrific wildfires in australia an…
@AndrewYang: democrats starting fires in australia to prove climate change 🤣
@SenSanders: if 184 arrests for the fires in australia is your proof of climate change, your spot on......time for g…
@SenSanders: you moronic tool, wild fires were not caused by climate change! weather has done cylindrical patterns f…
@SenSanders: hey bernie, 79% of the australian fires are set by humans. how is this caused by climate change ??? ge
t real, guy!!!
@AOC: any fire is horrible, but how can you blame climate change when 50% of australian fires are deliberately or ha…
@AOC: a good question would be if any of the fires were started by climate change people
@ewarren: thank you, for calling attention to climate change & subsequent fires in australia. we are a na…
@JoeBiden: so remind me again exactly what fires in australia have to do with climate change and exactly how you're…
@AOC: arson is the cause of the fires in australia. not climate change. put that in your pipe.
@JoeBiden: it's coming out that a lot of those fires are being set. probably by climate change anarchists who are t…
@AOC: is it really climate change, or is it un-managed forests that are causing the widespread fires? i have a feeli…
@BernieSanders: i'd like to start with this fire is not climate change based.these fires were lit by some stupid peo…
@AOC: you blithering idiot, climate change has *no relationship* to the fires in australia. okay? none. only pe…
@AOC: just like you thought climate change started the fires in australia when it turns out to be crazy hoaxers just…
@AOC: yeah. doesn't help your cause when the fires were arson set by climate change protesters. stop trying to gaslig
ht us.
@AOC: folks. she's not saying that climate change started the fires. she's saying that climate change is the reason…
@AOC: they are charging people with arson so that's climate change? just like the fires in cali and everywhere else.…
@ewarren: i'm betting you think climate change is responsible for the fires in australia
@AOC: it doesn't even matter how those fires started. if climate change didnt't increase heat and drought in austral…
@BernieSanders: leftist propaganda. you set fires and then claim climate change. i'm so g…
@AOC: and the australian pm says that the fires are not connected to climate change. this is what is wrong all over
the world!
@SenSanders: bernie believes that climate change starts forest fires...
@BernieSanders: moron... i guess the fires in 1851 and after are all because of climate change...
@BernieSanders: so the chief scientist of australia says the fires are not related to climate change (wh…
@AOC: for anyone saying it's fake news without climate change and global warming the fires in australia wouldnt be t…
@ewarren: you still think the fires in australia are a result of climate change. you'll never be president of anythi…
@MikeBloomberg: you can now work for the fake news you idiot. climate change did not create the australian fires, b…
@SenSanders: stop blaming every disaster on #climatechange. forest fires aren't created from climate change you fear…
@ewarren: so fires never happened before??? arsonists are being driven by climate change??? one flew over the cucko
o's nest.

The above discovered cluster contains accounts propagating a hoax that the 2019 bushfires in Australia were caused by arsonists

@PeteButtigieg: you moronic propagandist. jesus was not a refugee!! mary and joseph returned to their hometown tempo…
@PeteButtigieg: nice try but you should read your bible. jesus was born in bethlehem because his parents, joseph a…
@PeteButtigieg: you do realize what you have said is not accurate. jesus was not a refugee, joseph and mary went to…
@PeteButtigieg: pete, you need to read the bible. he was not a refugee. mary and joseph were returning to bethlehem…
@PeteButtigieg: you should read the bible this is not the story! jesus, mary and joseph were not refugees! pathetic…
@PeteButtigieg: i'm sorry but christ was not a refugee before his birth. mary and joseph were there in bethlehem for…
@PeteButtigieg: too bad you don't know your bible. mary and joseph were not refugees. they had just returned to thei…
@PeteButtigieg: a man that claims to be a christian doesn't the bible. joseph and mary arrived in bethlehem from naz…
@PeteButtigieg: /dems stop lying! jesus was not a refugee; they went to bethlehem for the roman census and when they…
@PeteButtigieg: no where in the bible is mary, joseph and jesus referred to as refugees! they were travelers on a m…
@PeteButtigieg: for those of you discounting jesus as a refugee: jesus, mary, and joseph certainly were refugees wh…
@PeteButtigieg: pete, according to bible mary and joseph were not refugees why did you have to say that?? hhmmm
@PeteButtigieg: you clearly are clueless. jesus wasn't a refugee nor was joseph and mary. they were merely traveling…
@PeteButtigieg: liar. read the bible. joseph wasn't a refugee or an immigrant.
@PeteButtigieg: jesus christ was never a refugee nor were mary and joseph. pick up the bible sometime and when you'r…
@PeteButtigieg: what a silly comment. jesus was not a refugee. in fact, joseph was returning to his hometown to com…
@PeteButtigieg: are you a liar or are you really this dumb? jesus was not a refugee. mary and joseph were following r
oman law.
@PeteButtigieg: hey pete, read the bible, mary and joseph went to bethlehem for a census, not because they were refu…
@PeteButtigieg: actually read the bible pete. jesus mary and joseph were not refugees. they followed the law and tra…
@PeteButtigieg: jesus christ was not a refugee   st. joseph had to do a census   really, the bible passages aren't t
hat long
@PeteButtigieg: dude, you need to learn about the bible. jesus wasn't a refugee, his parents were traveling for the…
@PeteButtigieg: you just proved that you haven't read the bible - joseph and mary were not refugees.. they made a te…
@PeteButtigieg: gee pete, i thought joseph and mary went to bethlehem for a census...nothing in the bible i read eve…
@PeteButtigieg: politician in all senses? lying even about the bible 😡 jesus wasn't a refugee, joseph and mary went…
@PeteButtigieg: dude, jesus, mary and joseph were not refugees. get a bible will ya?
@PeteButtigieg: the arrival of jesus christ & mary & joseph were not refugees.
@PeteButtigieg: try actually reading the bible, he was not born a refugee. mary went into labor on their way to pay…
@PeteButtigieg: get the story straight. joseph, mary and jesus were not refugees. it's all in a book. the bible. chec
k it out.
@PeteButtigieg: again you show ignorance of holy scripture. jesus mary nor joseph were refugees, scripture is clear…
@PeteButtigieg: i think you need to re-read your bible. he did not come as a refugee. joseph and mary had to go to c…
@PeteButtigieg: what bible do you read that depicts joseph, mary and jesus as refugees? that's just plain false. no…
@PeteButtigieg: first you get the bible wrong, then jesus mary and joseph... now the founding fathers and the consti…
@PeteButtigieg: all these people complaining that jesus, mary and joseph weren't refugees apparently never read the…
@PeteButtigieg: your pandering is pathetic. jesus was not a refugee, mary and joseph were not immigrants and they we
re not poor.
@PeteButtigieg: for all those criticizing pete, and claiming jesus and his parents were not refugees, read matthew 2…
@PeteButtigieg: obviously you are not familiar with the holy bible. joseph & mary went from nazareth…
@PeteButtigieg: need to maybe reread the bible again, nowhere in it does it say joseph, mary, or jesus were refugees…
@PeteButtigieg: it might help to actually read the story of the birth of jesus! mary & joseph were required to tr

Above is one of a few clusters containing replies only to Pete Buttigieg, where Twitter users state that Jesus wasn't a refugee

@amyklobuchar: excellent job last night! donated to your campaign this morning. as i looked at the group you outsh…
@AndrewYang: hey andrew. i donated $10 to the campaign last night to enter the star wars sweepstakes. unfortunately,…
@amyklobuchar: you were my first 2020 donation. last night was not it, amy. you were in my top two. good luck.
@amyklobuchar: you did a great job last night. i was truly impressed. will be donating to your campaign.
@AndrewYang: i didn't even realize when i donated last night that i could get this chance with you....love it!
@AndrewYang: we donated $1000 last night. joined the #yanggang2020 after listening to the podcast.
@AndrewYang: yes president andrew yang, i donated last night. forward!!!
@amyklobuchar: i've been a fan since the beginning. i just donated $25 to your campaign. i wish i could give more
!!!!
@amyklobuchar: you had a fantastic debate last night. i have not made my mind up yet and have donated to a couple di…
@AndrewYang: you better win! i donated my last $5 till payday friday 😬
@AndrewYang: bought two last night (i've already heard the audible version) hoping it goes towards and counts as a p…
@ewarren: was i the last person in the last decade to donate to your campaign?
@AndrewYang: and a good year for you to become #potus. my son is 5. last night we talked about 2020 goals. he said h…
@AndrewYang: i'd buy a hoodie and a bumper sticker but i donated my last $20 to the cause. all in as some would say.
@amyklobuchar: watched the debate in southern oregon. donated to your campaign the next day. good luck!
@SenSanders: blew away the competition during the last quarter of fundraising. he just reached equal footing in the…
@AndrewYang: help us keep up this momentum ahead of the primaries by donating to the link below! today is the last d…
@AndrewYang: we gotcha! and to the awesome #yanggang textbanker who just texted me to get one last donation in for 2…
@AndrewYang: gifted his fellow candidates with a copy of his book at last night's #demdebate. his campaign sent me…
@AndrewYang: you are amazing! your performance last week at my alma mater (lmu) was revelatory! say y…
@BernieSanders: i swung a trump supporter last night to support bernie. i did it while volunteer texting for bernie.…
@amyklobuchar: just made my first campaign donation! amy, you were outstanding tonight and we need a president to br…
@SenSanders: first time i have ever donated to a campaign
@AndrewYang: there were 10 candidates last time, and there're only 7 candidates. you spoke for 6.4 minutes…
@AndrewYang: thank you andrew, i have depression for last couple years, following your campaign and watch yang gang…
@AndrewYang: i've been wanting a yang/wiliamsom ticket for the last year! she's awesome! and so are you!
@PeteButtigieg: i donated twice to the campaign this past week. join in all the fun by contributing here:
@AndrewYang: got an email from the yang campaign today saying how i've donated 18 times this year. dang! time to mak…
@AndrewYang: first time ever donated to any campaign
@amyklobuchar: great debate. you're the real thing. i donated tonight. 💪🏻
@BernieSanders: go ask hillary....she gave you $600k during the last campaign, right?
@ewarren: you filled out a questioner, don't get too excited. last time you were that excited, you we…
@AndrewYang: firdt candidate i've ever donated to and just gave my third donation! let's do this #yanggang
@PeteButtigieg: first presidential campaign my husband and i have ever donated to! my husband even…
@AndrewYang: monthly donor for months now plus i give after every major event! i've never donated to a campaign befo…
@AndrewYang: remember how jumped up in iowa at the last minute on a shoestring budget? i feel you could easily do t
he same.

The cluster shown above contains positive comments to democratic presidential candidates that were posted after a debate

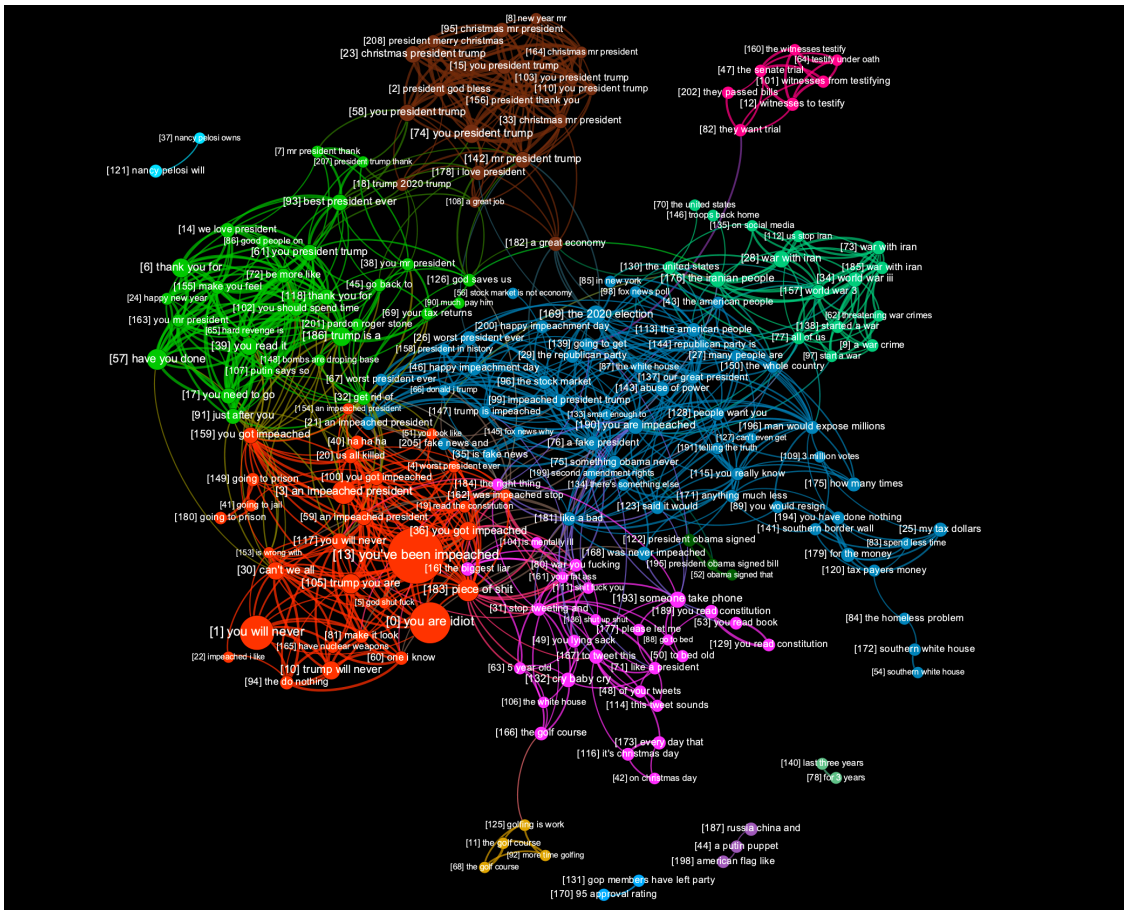
Example output from this dataset can be found here: https://github.com/r0zetta/meta_embedding_clustering/blob/master/example_output/tweet_graph_analysis_dems.txt

---

Experiment 2: realDonaldTrump

Our second experiment involved clustering of a subset of data in set 2 (@realDonaldTrump). We processed a batch of 30,044 tweets, resulting in 209 clusters.

A image below is a static graph visualization of the discovered clusters:



Using the same methodology as in our first experiment, we separated the clusters into positive, negative, and toxic, and then summarized them. Positive clusters included both statements of thanks and wishes of Merry Christmas and a Happy New Year, but also included the incorrectly categorized phrase “you are a puppet”. A summarization of negative clusters didn’t find any obvious false-positives, and included themes such as recent impeachment hearings, and comments on the amount of time the president has spent playing golf. Clusters deemed toxic contained, as expected, a lot of profanity.

```
Positive clusters: 52 - 7260 tweets.
svo
(god, bless, you) / (we, love, you) / (you, are, president) / (you, should spend, time) / (this, is, president) /
(i, love, you) / (we, have, back) / (you, are, assualt) / (you, are, puppet) / (they, want, trump) /
(you, got, votes) / (stock market, is not, economy) / (we, will end, it) / (it, going, to be) /
(you, do, it) /

ngram
happy new year / you mr president / god bless you / merry christmas to / you and your / thank you for /
thank you mr / to you and / and your family / you are a / christmas to you / me me me / the stock market /
you president trump / all is well /

word
president / mr / trump / thank / god / love / great / christmas / merry / new / happy / 2020 /
like / year / time /

Negative clusters: 131 - 16364 tweets.
svo
(you, are, president) / (you, started, war) / (you, are, criminal) / (you, should spend, time) /
(you, blocked, witnesses) / (you, played, golf) / (you, are, assault) / (it, 's, christmas) /
(you, did, nothing) / (you, lost, vote) / (you, have, idea) / (there, is, house) / (you, need, to go) /
(you, are embarrassing, yourself) / (i, read, transcript) /

ngram
you are the / you are a / in the senate / donald j trump / the united states / of the united /
president of the / you are not / to be impeached / the white house / out of the / you are impeached /
you have been / you were impeached / start a war /

word
trump / president / impeached / get / like / people / never / war / nothing / know / read /
impeachment / iran / one / go /

Toxic clusters: 26 - 6420 tweets.
svo
(you, are, idiot) / (you, are, president) / (you, are, man) / (you, are, disgrace) / (this, is, news) /
(he, 's, idiot) / (you, lying, sack) / (you, fucking, idiot) / (dems, are, joke) / (jesus, add, 2) /
(you, need, to update) / (you, are, liar) / (you, failed, us) / (it, is, to believe) / (you, are, embarrassment) /

ngram
you are a / you are the / shut the fuck / piece of shit / you are an / worst president ever /
the worst president / just shut up / you are so / in the history / you got impeached / abuse of power /
you are such / are such a / trump is a /

word
fuck / trump / impeached / fucking / president / ever / like / stupid / man / old / shit /
idiot / history / liar / piece /
```

Final values for this set were as follows:

Positive tweets: 7260 (**24.16%)** Negative tweets: 16364 (**54.47%)** Toxic tweets: 6420 (**21.37%)**

Note how @realDonaldTrump received a great deal more toxic replies than any of the accounts studied in the previous dataset. Note also that tweets contained in negative and toxic clusters totalled roughly three times that of tweets in positive clusters.

Here are some details from the largest identified clusters. They include the following negative themes:

- You are an idiot/liar/disgrace/criminal/#impotus
- You are not our president
- You have no idea / you know nothing
- You should just shut up
- You can’t stop lying
- References to Vladimir Putin

Here are some of the positive themes identified in these larger clusters:

- God bless you, Mr. President
- We love you
- You are the best president

Noteworthy clusters

god bless you and your family!!! thank you for all you do for america!!! ❤️❤️🙏🙏🎁🇺🇸🇺🇸🇺🇸🇺🇸🇺🇸
thank you mr. president for all you have done for our country. god bless america and you and your family. ❤️🇺🇸
we're all praying mr. president. for you, your family and our great nation.
always praying, president trump, for you and this great country we live in! god bless and merry christmas!
drain the swamp love you president trump 🇺🇸 🇺🇸 🇺🇸 🙏 truth will prevail 🇺🇸🇺🇸🇺🇸🇺🇸🇺🇸 godbless merry christmas 🎄
you're the best president ever! stay strong prayers for you mr. president 4 more years!!
thank you president trump for the great job you are doing. god bless you, and your family.
merry christmas 🎄 to you and your family! god bless the 🇺🇸 !! let's keep america great! vote trump 2020!!
always praying for you mr president! you are wonderful and amazing. my church loves trump!
happy christmas season all over the world. keep us in your prayers. god bless usa.
🙏🙌❤️🙌🙏 patriots all over are praying for our potus and his family. thank you mr. president 🙏🙌 #wwg1wga #kaga2020
i will. dear whatever god is listening. thank you for bringing me this glorious day. thank you. amen.
we love you president trump!! keep helping our country! god bless you and your beautiful family!!!
god bless trump, his family, and all that dare to stand as a fortress against the evil who plot against our great nation.
dang i'm 15 sand the differences you pulled out with and changing my family helped thank you mr 45th president
q sent me! #sheepnomore #wwg1wga love you president trump. god bless you more and more ❤ merry christmas!
great job at yesterday's rally in michigan, sir. god bless you on your day today. #kag #staythecourse
i have said many prayers for you & your wonderful family. please be strong mr president. i believe in you!
i love your patriotism. thank you for all that you have done for america. may god bless you and your family.
praying! god bless america and president trump and his family!
mr president i am grateful for you every day! god bless you and your family. thank you for everything you do!
my prayers for you,your family and america. love u mr president
you got this! i along with many others are praying over you & your beautiful family! stay strong. #trump2020 #maga #45
love to our president trump, i thank god for him and family.
love you president trump and your whole family! thank you all for everything you are doing for america!!
that's correct! we love 💕 president trump! god loves 💕 president trump! #trumpeconomy is great! merry christmas. #trump2020
i pray you see my tweet to you president trump. we love you and pray for you everyday. job well done. god bless you.
thank you so much 4 protecting america! god bless you and your family!!!! #keepamericagreat #trumpcard #trump2020
great question pres. trump. #maga #psalm91 prayers for you and family continue everyday. #winning2020
thanq president trump! god bless you and your family, and thank you for taking care of the forgotten..we love you!❤❤🙏🙏🇺🇸🇺🇸
❤️ american patriots love and respect our president trump and his beautiful family!!! ❤️
god will bless you and your family mr. president. thank you
god bless you president trump! may he keep you and the first family safe!
you mr. president are awesome. god bless you thank you & keep being the awesome man u sir are.
took to long to fund it!!! glad you did mr. president!! thank you and god bless for putting our military first. ✌️👏👏

Above and below are Christmas-themed clusters, but with quite different messages. The one above contains mostly season's greetings, whilst the one below contains some questions to Trump about his plans for the holidays.

you should recreate your call in a christmas themed fireside chat.
#godisincontrol ##christmasislove check for any gifts. #linkinbio
you gonna have the #viktorleonov swing by and pick you up for some holiday defection, or nah?
a christmas gift for daddy vlad?
he tweeted this on christmas eve folks. how presidential.
im campaigning for you while i christmas shop! 🎄⭐🇺🇸
early christmas present from my aunt.
which church are you going to for christmas service? will you be with melanie and her son?
just keep thinking about this over the holidays, donnie.
are all congressmen returning from the holiday recess? #patriots are just wondering
holiday game night at the wh
give them a new home for christmas.
i wonder if he will text thru christmas
it's christmas. spend time with your son.
relax...it's hanukkah...it's christmas week...why don't you chill out for a few days and celebrate your savior's birth?
have you been visited by the ghost of christmas future yet?
are you on holidays chubs?
you should celebrate christmas with your family for a couple days.
christmas eve wilding brought to you by the stable genius.
this is your chosen message on christmas eve?
where are your buddies spending christmas and hanukkah? #magajudeochristianvalues
what are you doing with your family for christmas?
hey shithead...go hang with your family...it's christmas.
so what are you getting daddy putin for christmas?
forget about this for tonight and tomorrow..enjoy christmas with your family..merry christmas!
go to church, it's christmas eve.
shouldn't you be going to christmas eve service
what did you get your wife's son for christmas?
don't you have a family? it's christmas eve! go be with them
which one of your minions informed you it was christmas?
marry christmas to you and you family president
#christmas for thousands of #childrenincages
did i have to pay for your christmas party?
did you give us the gift of your resignation this christmas?🎁
did anyone get you wind for dummies for christmas?
these children deserve a christmas with their parents! #impotus45

Below is a cluster that found a bunch of "pot calls the kettle black" phraseology. Note how it captures quite different phrases such as "name is pot and he says you're black", "kettle meet black", "pot and kettle situation" and so on. It did fail on that one tweet that references blackface.

sure. still the pot calling the kettle black
pot calling the kettle black. it would be laughable if it wasn't so sad.
isn't that the pot calling the kettle black! watch this:
woah! pot, kettle, black springs to mind!!!
hi kettle, this guy's 👆 name is pot and he says you're black.
classic pot calling the skillet black"
you're a dirty potus. pot calling the kettle black.
and we have this. the pot calling the kettle black lesson from trump the liar!!!
says the pot calling the kettle black!
you calling someone corrupt is a pot and kettle situation.
hahahahahaha the gigantic pot calling the kettle black . wow you need a jacket
pot calling the kettle black..this is rich
now there's the pot calling the kettle black.
morally??? really. kettle meet black
well that's the pot calling the kettle black!
pot calling kettle black much
the pot calling the kettle black comrade?
news flash: pot calls kettle black
now there's a pot calling the kettle black! joke of the day!!!!
ain't that rich! the pot calling the kettle black!😂
kind of a pot calling the kettle black there cheeto dust!
no that's about right....see it's about the pot calling the kettle black.
kind of the pot calling the kettle black don't you think dirty donnie!
if that ain't the pot calling the kettle black....
hmm, is this the pot calling the kettle black?
another guy that wore blackface and you're calling him highly respected?
shish! you are the pot calling the kettle black!
is this your version of the "pot calling the kettle black"????? #resign #impotus
oh wow you're calling the kettle black again
seriously??? the pot calling the kettle black 😊
well isn't this the coffee pot calling the kettle black

This next one (below) is interesting. It found tweets where people typed words or sentences with spaces between each letter.

😂u r such an insecure liar!!!!!!! u r not popular & majority knows that u r a fraud!!!!!!
f u c k t r u m p ! ! ⁹
we want d o n a l d t r u m p in j a i l!!!! lock him up!!!
have u looked in the mirror lately? as if u r such a prize? u r garbage!
what does mean is trump? t - traitor r- russia u-. unfit m- moscow p- puppet
omg! what reviews r u reading? u r the laughing stock of the country!
oh for godsakes give it up spanky! b o r i n g
who gives a shit about ken starr?? y o u a r e g u i l t y
u r do nothing u scammed america
nah, you're still 3rd potus impeached. that will follow you f o r e v e r.
impeached president is d. i. s. g. r. a. c. e. d.
china no.1 !!!!u r nothing
t r u m p t r a i n 🚂 🇺🇸 got my full support!
you sound a bit n e r v o u s ... something bothering you?
u r right! we r victims of your cruel, immoral and corrupt behavior. u r a vicious liar.
it is not a h o a x or a s c a m. you are a f r a u d.
this american has had it with donald j trump. i dislike u @ ur immoral compass. u r a joke.
u r not politician. u r corrupt. your best friend bribing. get lost
no. not liars. thoughtful and strategic. two things u r not #impeached45 #votebluetosaveamerica
i can't believe u r president and r this shallow. thk g-d i'm already on my 2nd drink. so embarsssing. what a dumbass
so proud to be canadian!🇨🇦 f u c k. t r u m p!
it's over dumb donald. you've been impeached. that can never be changed. e v e r ! !
t a x r e t u r n s
perhaps if they "ask very politely" you would do ur job and help. u r not a king.... u r a childlike court jester.
i really following you because of your boldness and firm decisions. u r the trust politician.
not believe i am n the father and the father is n me words i say r not mine but r fathers dwelling n me doing his wor
ks
he was flat broke until the presidency. w h o r e s
t h a n k y o u d o n a l d t r u m p !!!
k i l l y o u r s e l f
s h u p u p trump ! ! !

Below is a cluster that identified "stfu" phraseology.

```
you need to shut the fuck up, infant.
just stop being a little bitch
shut the fuck up, donny.
tone down the drama there, orange.
shut the fuck up, loser.
just please, please shut the fuck up.
sit down and shut up
shut up just shut the fuck up
ok - i give up..who are you talking about..
shut yo hoe ass up
just shut the fuck up already!
will you pls quit whining!!!!
you've left your caps lock on!
can you just shut the fuck up !!!     loser
shut the fuck up, puto
shut the fuck up boomer
shut the fuck up traitor
all you do is lie. shut up.
leave it in gods hands!
turn off the caps don.
shut up and get out!!
oh do shut the fuck up.
sit down and shut the fuck up.
shut the fuck up already!!!
oh please, just shut up!!!
ahhhh, quit yelling at me!!
why don't you just fucking shut up?
turn off the caps lock bud
shut up you fuckin dick
please please shut the hell up!!!!!!
bro shut the fuck up lmfao
```

Example output from this dataset (and others studies) can be found here: https://github.com/r0zetta/meta_embedding_clustering/tree/master/example_output

---

Content regarding the recent Iranian situation

As mentioned in our methodology section (later in this article), the technique we're using does sometimes identify multiple clusters containing similar subject matter. While looking through the clusters identified from replies to @realDonaldTrump, we found four clusters that all contained high percentages of tweets about a recent situation in Iran. Upon inspection we realized that those clusters contained different takes on the same issue.

Below is a cluster that contains some tweets praising Trump's actions in the region.

```
Cluster 112 contained 77 tweets.
iran(37) let(13) us(8) get(8) president(7) stop(6) mr(6) terrorist(6) trump(6) revenge(5)

0.800: @sm51155422: this is the cultural places of iran. the mourners of martyr soleimani. can you attack them?
0.787: @joshlingard4: plan b , drop   on iran   #worldwarthree #wwlll
0.761: @NRC05847892: well, calculated response sir, let us not escalate unless iran is determined to be wiped out...
0.758: @HoseyniHamed: ithink you should destroy importent places of iran
0.753: @Maryam31765736: can you save the world from the evil iran
0.751: @kittyfrisk: "hey mike... birdie i strike iran, parr i don't, bogey we get kfc"
0.742: @jpnc1987: mr. president, now is the time to stop iran in it's evil tracks.  tighten the screws on iran.
0.732: @jrguerra619: this is what you see in iran. because of your actions. fyi it the fag of revenge.
0.729: @Snowyama10Roth: we seriously need a military base in tehran.  what does iran not understand?  it's a question
of when.
0.727: @leyden28: is this organized terror coming from iran? does the cia know?
0.727: @rrmazi: unless you plan on going out there with your bow and arrow, but the us army will not strike iran unti
l approved by congress
0.720: @amirrezatoa21: you are really stupid nation of iran are more powerful than your army dogs are more beautful t
han you
0.714: @onemangangfan: it is time to turn the tables on iran! keep it up mr president!
0.704: @AlvinMaiorana: stop killing irans general why you do that
0.704: @WinWithWinslow: if iraq threw shoes at w., what will iran throw at trump?
0.697: @JyotiranjanRa17: remove iran from the world map. india is with you.
0.690: @irqmah: unless us stop iran ircg money loundry network  iran will get nuclear weapons soon
0.688: @JCanch: make sure to follow through on threat-52 when iran thinks you're bluffing and they send their quds ah
uffing.
0.685: @darkrob2009: way to go mr. president show iran whos the boss.
0.681: @MichelleSkime: if only we had a solid treaty with iran 👀
```

Below is a cluster that contains some tweets mentioning Iraq and related repercussions of actions against Iran.

```
Cluster 73 contained 148 tweets.
iran(56) us(22) iraq(19) oil(13) war(13) people(10) get(10) take(10) lives(9) embassy(9)

0.779: @runyan50: this tweet makes no sense. iran is in control and iraq is kicking us out.
0.764: @REDAngel2B: can we please just hand u over to iran 🇮🇷 .. ughhh #dumbass
0.762: @dondos1969: alleviate sanctions on iran as an apology and spare us the war .
0.756: @raad198866: i think iran is defeating you in iraq
0.756: @ppark4545: get your war out of iraq deal with iran government in iran
0.754: @LookingoverT: didn't we drone bomb iraq, and iraqis protested?
0.752: @DaveSpearo: guess we are going to war with iran.  #wagthedog
0.741: @andrew94882296: nigeria is supporting iran to attack u.s be beware of buhari and his islamists terrorists
0.737: @sa00gd: but please outside iraq, you can start a war on iran in the iran region, not in iraq please 🙏🙏🙏🙏
🙏🇮🇶🇮🇶
0.733: @mikerardin99: that's pretty ballsy of you to say because iran can now just hit us with a reverse card🤔
0.728: @Mahdi_mr_t: controll by iran!! who have military base in iraq??!
0.728: @huragil: iran could use venezuela to get revenge and harm the us. just fyi
0.720: @Connie_Alberta: is going to crush iran. prayers for us and allied soldiers in the area.
0.720: @tweetheart2020: and why didn't he warn us about the attack on our embassy in iraq?
0.720: @HosseinAnsari: since the us has no cultural site , iran will only take down all your army and assets all over
the world.
0.718: @Sandybisht7: world only want peace, u.s and iran #usirantension
0.717: @CarolDy75984351: sob is starting a war with iran on revenge from more than 30 years ago.
0.717: @aceleeverbrown: you probably talked about the next steps in your illegal war with iran too you saudi bitchsla
ve
0.716: @ignorantpotus: thanks for singlehandedly creating a nuclear iran you irresponsible dumbshit. #ignorantdonald
0.714: @MarilynVercelle: my guess is putin had intel about the plan to bomb iraqi militia.
```

Below is a cluster that contains mostly negative comments about Trump's actions in the region.

```
Cluster 28 contained 264 tweets.
iran(78) us(58) iraq(42) nuclear(31) people(29) trump(28) president(24) want(23) would(19) america(17)

0.802: @ShariAnneHeath1: you are insane.  you just started a war with iran.  can you ensure the security of all ameri
cans in that region?
0.784: @SlytherinXlord: #nowarwithiran fuck you trump iran and us are friends not enemies 🇺🇸🖤🇮🇷❤🇺🇸💙🇮🇷🧡
0.783: @Byk_Mata: you so big mistake about kill of general of iran and tomorrow you out for president of american, al
l people hate you !!!
0.780: @gaurasharma007: neuclor technology given to iran by the "terrorist laboratory pakistan" .... america should b
an on pakistan.
0.770: @tripledot2019: right now iran attack the usa in irak
0.769: @alosh59916224: we reelly want a freedom help us those who are iranian dogs in the iraq
0.765: @holmesj100502: please address the nation on the iraq base attack by iran.
0.762: @tuukkaa2: iran can't bomb bases with the intention of killing us soldiers and except to run free
0.757: @m_omar68: to hell, iran, it is time for the iraqis to lead iraq
0.754: @madar_abd: i continue to kill iranian leaders, i am from iraq 🇮🇶
0.752: @nvChUGJki4y7xl8: america and iran killed the iraqi people
0.751: @Harvey77378671: you mean the security and stability you just trashed in iraq with iran? that security and sta
bility?
0.750: @XRevengerX: terrorists lives in pakistan.... trump bombing iran...common blew up porkistan first
0.750: @FoersterRobert: ...because you want to have war - no peace - with iran? right, warmonger?
0.748: @Moaed70471713: you must cut off iran's arm in iraq, or else america's prestige will be insulted
0.746: @Kansfan2001: nobody knows what they have.... because u pulled us out of the iran nuclear deal..... smooth bra
0.745: @qyHxq6h8o6m10jp: thank you #usa for eliminating those terrorists including #qassemsoleimani , from iran with
0.745: @OkraskaRick: i'm confused now. it's the embassy in iraq and it's the iranians fault?
0.742: @Therock3000000: we as the iranian people want you to support us to overthrow the mullah´s regime  no to negot
iation!
0.742: @proshikkben: one more thing mr. president. demolish and nulify all iran nuclear arsenals immediately.
```

And finally, the cluster below contains a great deal of toxic comments.

```
Cluster 157 contained 144 tweets.
war(40) iran(32) us(26) world(21) people(19) killed(19) fucking(17) trump(13) please(12) get(12)

0.786: @mycousinstevie: it won't work. you are still impeached , and now you are starting a war that will kill americ
ans. disgraceful.
0.785: @mightydeku2: you fucking idiot, you just started a war with iran!
0.766: @SQ190: terrorist america want to kill more people in iran
0.761: @happymartie1: like you're now? taking us to war with iran, just because you wanted a distraction? you're very
weak
0.751: @extreme_miami: just opened a door to by killing the iran general to world war 3 thanks alot djt
0.750: @QLisha13: you fucking idiot iran dose not want fucking war so stop putting us in danger
0.749: @DDub_68: kill one terrorist and liberals think world war iii has started. 🤔🤔🤔
0.749: @jake88336179: starting a world war 3 why killing innocent people
0.730: @hank_viceroy: you're going to get innocent people killed you fucking asshole
0.723: @JonDStrange1: hey asshole!!! you just killed americans for no fucking reason!!!
0.719: @atrahamrepol: stop trying to get us into wwiii you fucking idiot.
0.716: @girlmoon121: please don't hurts i raqi people please kill 🇮🇶 iran not us please
0.716: @AimeeHypatia: dude, do you get whiplash from inciting war to professing regional stability? asking for the wo
rld.
0.715: @BlandafanG: you're going to get us in a war before you're impeached
0.707: @kellykbarksdale: please resign before you get us all killed.
0.699: @Hoscar96397865: yo don't involve us in ur war do it by urself you fucking idiot  we fucking dead
0.698: @Ahmed26768222: i expect that someone will target you and kill you, madman, before you destroy the world
0.698: @HutchesonSteve: so let's bomb iran and cause a distraction.
0.697: @maro757: qais khazali, the terrorist, killed us when you kill him, mr. trump
0.696: @SaudAlsafi: stop your stupid fucking war !
```

Testing our methodology on different data

We tested our topic modeling methodology further by running the same toolchain on a set of tweets collected during the run-up to the UK elections. These were tweets captured on hashtags relevant to those elections (#GE2019, #generalelection2019, etc.). Our methodology turns out to be quite well-suited for finding spam. Here are a few examples:

The output below contains tweets posted by an app called "paper.li", which is a legitimate online service that folks can use to craft their own custom newspaper. It turns out there were a great deal of paper.li links shared on top of the #ge2019 hashtag. Unfortunately, this was one of four clusters identified that contained similar-looking paper.li tweets (which could be found more easily by filtering collected Twitter data by source field).

```
Cluster: 0 contains: 377 tweets.
Sentiment: 180.01
Words: latest(362) #ge2019(343) thanks(326) daily(283) #generalelection2019(39)
svo: (i, love, puppies)(2) (mr, chas, thanks)(2)
ngrams: the latest the(244) daily thanks to(207) thanks to #ge2019(178) #ge2019 the latest(131) #generalelection2019 the latest(27)
tweeted: [ UniLifeQdos (4) guy_lakeman (4) OpenUnionism (3) AndrewKerr4 (3) austenblakemore (3) ]
==================
0.987 @mfavourite: the latest mfavourite daily!  thanks to   #ge2019 #imaceleb
0.985 @tatranky: the latest the tatranky daily!  thanks to    #rt #ge2019
0.982 @MarXPacE: the latest the marxpace daily!  thanks to    #ge2019 #thecrown
0.981 @BringtheFlag: the latest bringtheflag daily!  thanks to    #maga #ge2019
0.981 @Mr_Machado: the latest the mauricio machado daily!  thanks to    #ge2019 #imaceleb
0.979 @wheresmydad: the latest the wheresmydad™ daily!  thanks to    #imaceleb #generalelection2019
0.979 @TechStuffBert: the latest the ukmedia daily!  thanks to  #ge2019 #ajopinion
0.977 @MatthewChattle: the latest the matthew chattle daily!  thanks to    #ge2019 #themaskedsinger
0.977 @michaelcooper: the latest the michaelcooper daily!  thanks to    #ai #ge2019
0.976 @upukcab: the latest the  ok  daily!  thanks to    #ge2019 #boltonfire
0.976 @joaohartley: the latest the hartley's daily!  thanks to    #observador #generalelection2019
0.976 @UniLifeQdos: the latest the unilifeqdos daily!  thanks to    #ge2019 #mondaymotivation
0.976 @Mef0611: the latest the tengku meftah abdul daily!  thanks to  #usrc #generalelection2019
0.975 @MohKhodi: the latest the mendax daily!  thanks to   #ge2019 #generalelection2019
0.975 @welovedavid: the latest the welovedavid daily!  thanks to    #ge2019 #apfactcheck
0.974 @nasmu: the latest the nasir muhammad daily!  thanks to    #writingtowardfreedom #ge2019
0.974 @Hookie62: the latest the hookie daily!  thanks to    #ge2019 #staffing
0.974 @TananariveDue: the latest the tananarive due daily!  thanks to    #ge2019 #themandalorian
0.974 @tayotony: the latest the addiction daily!  thanks to    #ge2019 #imaceleb
0.973 @CCarlutti: the latest the cristiano carlutti daily!  thanks to   #ge2019 #history
0.973 @teh_Dede: the latest teh daily dede!  thanks to    #generalelection2019 #ge2019
0.973 @Nicknowhere: the latest the nicknowhere daily!  thanks to    #euro2020 #ge2019
0.972 @LocalGovNews: the latest localgovnews daily!  thanks to    #connectedchristmas #ge2019
0.972 @LOHADdotcom: the latest the lohad (twice) daily !  thanks to   #ge2019 #cop25
0.972 @crazyhorse2126: the latest the crazyhorse daily!  thanks to    #ge2019 #breaking
0.972 @paulscooking: the latest the paulscooking daily!  thanks to   #ge2019 #ukelection
```

Below we can see some copy-paste disinformation, all shared by the same user. Note that this analysis was run over roughly 30,000 randomly selected tweets from a dataset with millions of entries. As such, I imagine we'd likely find more of the same from this user if we were to process a larger number of tweets.

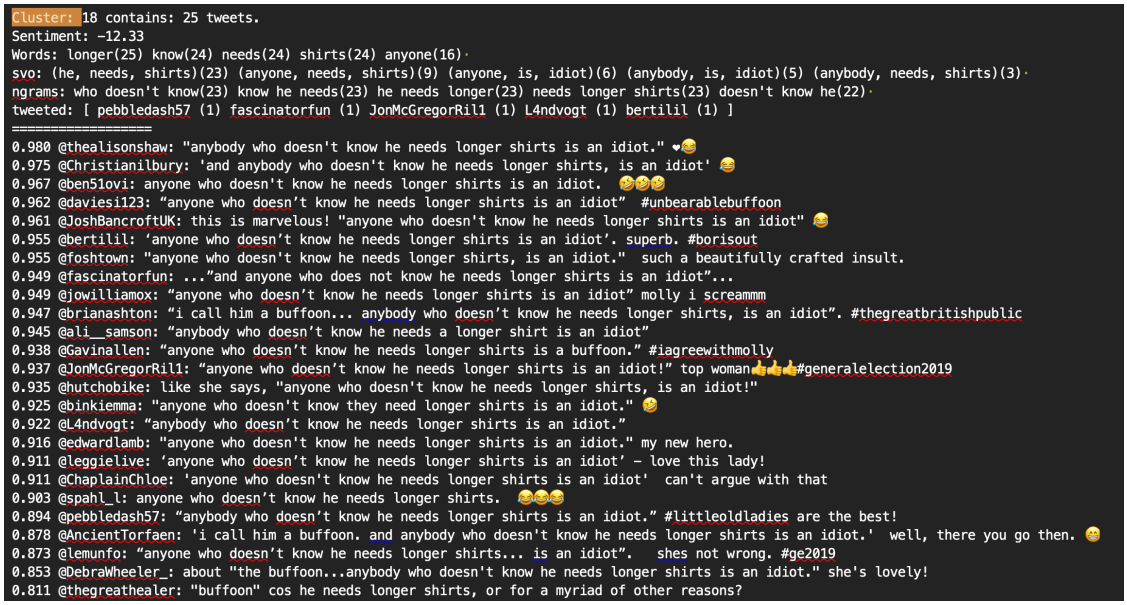
```
@TheMighty_Spurs: fact. jeremy corbyn is a huge threat to our national security. he's a supporter of terrorists 😡😡😡😡😡😡😡😡😡
@TheMighty_Spurs: fact jeremy corbyn is a huge threat to our national security. he's a supporter of terrorists  😡😡😡😡😡😡😡😡😡
@TheMighty_Spurs: fact jeremy corbyn is a huge threat to our national security he's a supporter of terrorists 😡😡😡😡😡😡😡😡😡
@TheMighty_Spurs: fact jeremy corbyn is a supporter of terrorists he is a huge threat to our national security 😡😡😡😡😡😡😡😡😡
```

Below we see some tweets advertising porn, on top of the #ge2019 hashtag. Spam advertisers often piggyback their tweets on trending hashtags, and the ones we captured trended often during the run-up to the 2019 UK general elections.

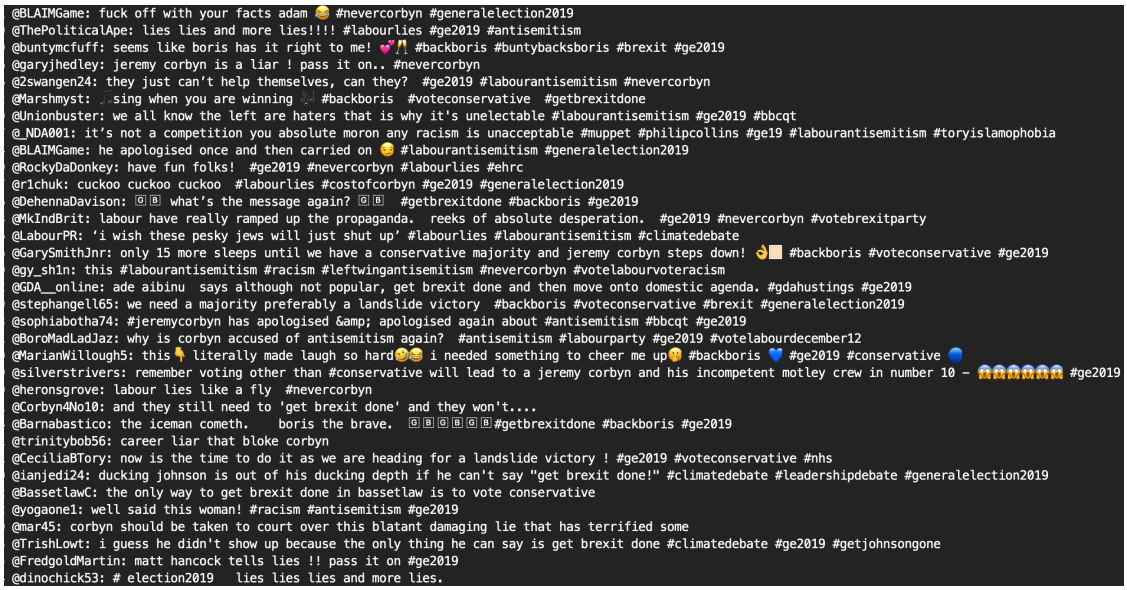
```
@gulmdero: #sexual 👌👌🌶️  #thanksgiving #ge2019 #endomondo #love #sex rt if you wanna a clip in dm #porn #t_co_hb __322__
@mehdirns1980: #horny 🍒❤️🐷  #endorphins #bebaskanluthfi #darbar #ge2019 #sex rt if you wanna a clip in dm #porn #t_co_hb __315__
@jnc6626: #lesbian 👙🐷👢  #exo #blackfriday #bebaskanluthfi #ge2019 #sex kinky teen gobbles dick #porn #t_co_hb __354__
@waltlumuwejjm: #lesbian 🐷👖❤️  #photography #chummakizhi #ge2019 #thanksgiving #sex who wants to see my dm #porn #t_co_hb __256__
@dllookingreat: #bigboobs ❤️👢👌  #endomondo #ge2019 #2019 #obsessedwithexo #sex who wants to see my dm #porn #t_co_hb __231__
@yaanliz: #18👢🍒🐷  #ge2019 #bebaskanluthfi #endomondo #exo #sex i have a shady desire. i need sex #porn #t_co_hb __313__
@gulmdero: #sexy 🐷👢🍒  #ge2019 #exo #love #shindanmaker #sex legal age teenager adores sexy masturbation #porn #t_co_hb __322__
@waltertiewhiph: #horny 🐷🐷🍒  #ge2019 #exo #love #christmas #sex retweet if you want to be inside of me xx #porn #t_co_hb __255__
```

The cluster below identified a certain style of writing also identified tweets coming mostly from one account.

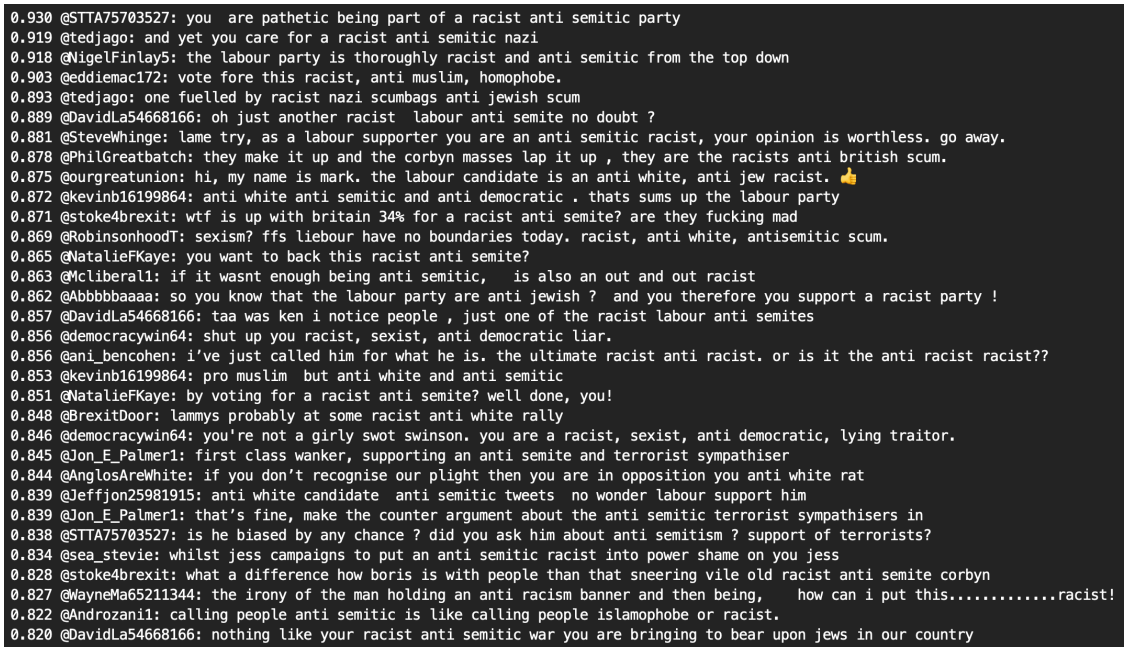
The cluster below picked up on similar phraseology. Not sure what that conversation was about.



Finally, several clusters (shown below) contained a great deal of tweets including the word "antisemitism". Many of the accounts in these clusters could be classified as trolls and/or fake disinformation accounts.



Note that we found similar clusters in data collected by following pro-tory activist accounts and sockpuppets during the same time period (shown below):



Other clusters were discovered in tweets from the same tory accounts, including a few that contained tweets designed to incite hatred towards specific demographics (see below).

```
0.866 @sheriffofknokem: why is it that lefties equate persecution of jews with fearing muslims as being on the same level ....?
0.865 @James84475033: saudi arabia does not import muslims, they export them, to spread their hate of christians ect, axis of evil
0.847 @Joanne62186731: what i can't fathom is this: he loves muslims; muslims hate gays; how do the students square that?
0.845 @miele_rodriguez: islam is a vile &amp;amp; disgusting ideology. muslims are people and should be treated with respect.
0.841 @AirBeanB1: it's almost as if you're directing hatred toward muslims paul. is that your goal?
0.837 @courtgambla: criticism of islam is suddenly enabling hatred of muslims !
0.833 @EFriesl: and muslims hate for christians and jews is sick... why has no one called muslims out for their bigotry???
0.830 @MosmanBarb: wouldn't have to lock up our young girls. but why can muslims scream hate about westerners and jews?
0.828 @ChrisRi85706154: britain and the people have had enough of islam muslims and their backward religion
0.828 @JeaniefaetroonJ: any mention of the hate crimes committed by ..... muslims.
0.826 @Steve_Leave: it's when you are afraid of muslim terrorists! gosh, do i suffer from islamophobia...
0.823 @gofer_1: couldn't agree more, the radical muslims will destroy islam.
0.821 @AlbertSaxon2: well they're behind the importation of the muslims you despise.  not to mention all non white immigration.
0.819 @Joke_Hunt: ok. so are we going to see muslims who openly preach death and genocide doing 18mnths too?
0.817 @PriestleyStacy: meanwhile a radical muslim kills people over an ideology.
0.813 @eelmorts: stop hate crimes by islamists! islamophobia is the result of their hatred towards all but fellow muslims!#outlawun
0.809 @NoelTurner194: have they mass murdered 2m muslims or beaten up strikers then?
0.801 @DavidLo09134020: did they visit the sites in london where muslims slaughtered decent citizens
0.800 @CookieM11559272: this woman tells us why muslims hate everyone.
0.798 @Glynn58660957: there's no such thing i no plenty of muslims . i hate islamist vile barbaric deluded freaks
0.797 @b17vic_Public: i couldn't care less for the savages. . . . . jews, muslims, hindus, all the fucking same with their blood lust.
0.795 @Bateman20191: what experts? tell mama? a muslim group that want to have the non existent 'islamaphobia' classified as a hate crime?
0.794 @The_Duchess__X: i do live in baxterley. where did i say i hate muslims and want them deported?
0.787 @Flic26849250: phobia is a fear of - like spiders islam.... islamophobia- . fear of islam.  that they want to kill you, they are terrorist.
0.784 @riot_rebecca: it's almost like muslims in britain have a problem with jews.
0.784 @cornell_phil: mayor of london is wrong cities with no muslims do not suffer like we do
0.780 @Indigo41105581: criticism &amp;amp; hatred of islam is totally different to bigotry &amp;amp; intolerance of all muslims. behave.
0.780 @abraham_degg: i love reading muslims hate on other muslims for doing something they want to do which is forbidden in there culture haha
0.780 @BrexitBanter52: are you suggesting muslims aren't integrated with wider society, racist?
0.779 @Barrybritish: it's easy to prevent the irrational fear of islam stop murdering people because they are not muslims.
0.778 @BadEvilDick: this folks it what should sharpen your voting. the muslims are stating that this isn't a white country any more
```

It's worth noting that a portion of the accounts identified in our clustered data have been suspended since the data was originally collected. This is a good indication that some of the users who post frequent replies to politicians, and participate in harassment are either fake, or are performing activities that break Twitter's terms of service. Any methodology that allows such accounts to be identified quickly and accurately is of value.

Conclusions and future directions

The methodology developed for our experiments yielded a mechanism for grouping tweets with similar content into reasonably accurate clusters. It did a very efficient job at identifying similar tweets, such as those posted by coordinated disinformation groups, from reply-spammers, and from services that post content on behalf of a user's account (such as paper.li or share buttons on web sites). However, it still suffers from a tradeoff between accuracy and the creation of redundant clusters. Further work is needed to refine the parameters and logic of this methodology such that it is able to assign groups of relatively rare tweets into small clusters, while at the same time creating large clusters of similar content, where appropriate.

In order to fully automate the detection of toxic content and online harassment, additional mechanisms must be researched and added to our toolchain. These include an automated method for creating rich, readable summaries of the contents of a cluster, more accurate sentiment or stance analysis of the contents of a cluster, and better methods for automatically assigning verdicts, labels, or categories to each cluster.

Further research into whether the identified clusters may be used to classify new content is another area worth exploring (initial experiments into this line of research are documented in appendix 2 of this article).

If these future goals can be completed successfully, a whole range of potential applications open up, such as, automated filtering or removal of toxic content, an automated method to assign quality scores to accounts based on how often they post toxic content or harass users, and the ability to track the propagation of toxic or trolling content on social networks (including, perhaps, behind-the-scenes identification of how such activity is coordinated).

The problem of analyzing and detecting abuse, toxicity, and hate speech in online social networks has been widely studied by the academic community. Recent studies made use of word embeddings to recognise and classify hate speech on Twitter (https://arxiv.org/pdf/1809.10644.pdf, https://arxiv.org/pdf/1906.03829.pdf), and Chakrabarty et. al. have used LSTMs to visualize abusive content on Twitter, by highlighting offensive use of language (https://arxiv.org/pdf/1809.08726.pdf).

The challenges involved in detecting online abuse are discussed in this paper published by the Alan Turing Institute (https://www.turing.ac.uk/sites/default/files/2019-07/vidgen-alw2019.pdf) Furthemore, issues surrounding the detection of cyber-bullying and toxicity are discussed in the following publication (https://encase.socialcomputing.eu/wp-content/uploads/2019/05/NicolasTsapatsoulis.pdf). An approach for detecting bullying and aggression on twitter is proposed by Chatzakou et. al at (https://arxiv.org/pdf/1702.06877.pdf). Srivastava et. al have used capsule networks to identify toxic comments (https://www.aclweb.org/anthology/W18-4412.pdf). The challenges of classifying toxic comments are discussed further in the following publication (https://arxiv.org/pdf/1809.07572.pdf).

We note that methods involving the use of word embeddings have been previously used to cluster Twitter textual data (https://ieeexplore.ieee.org/document/7925400), and that community detection has been applied to text classification problems (https://arxiv.org/abs/1909.11706). However, we have not encountered literature referencing the combination of both. To the best of our knowledge, our approach is the most sophisticated method to date for clustering tweets.

Appendix 1: Detailed methodology

This section contains a detail explanation of the methodology we employed to cluster tweets based on their textual content. Since this section is fairly dry and technical, we opted to leave it until the end of this article. Feel free to skip it unless you're interested in replicating it for your own means, are involved in similar research, or are both curious and patient

All the code used to implement this can be found at https://github.com/r0zetta/meta_embedding_clustering under the code/ subdirectory.

1. Data collection, preprocessing, and vectorization

Twitter data was collected using a custom python script leveraging the Twarc module. The script utilized Twarc.filter(follow=*accounts_to_follow*) to follow a list of Twitter user_ids, and only collect tweets that were direct replies to *accounts_to_follow* list provided. Collected data was abbreviated (a subset of all status and user fields were selected) and appended to a file on disk. Once sufficient data had been gathered, the collection was terminated, and subsequent analyses were performed on the collected data.

Collected Twitter data was read from disk and preprocessed in order to form a dataset of relevant tweets. Tweet texts were stripped of urls, @mentions, leading, and trailing whitespace, and then tokenized. If the tweet contained enough tokens, it was recorded, along with information about the account that published the tweet, the account that was replied to, and the tweet's status ID (in order to be able to recreate the original URL). Both the preprocessed tweet texts and tokens were saved during this process.

Three different sentence vectors were then calculated from each saved tweet:

1. A word2vec model was trained on the tokenized tweets. Sentence vectors for each tweet were then calculated by summing the vector representations of each token in the tweet.

2. A doc2vec model was trained on the preprocessed tweet texts. Sentence vectors were then evaluated for each preprocessed tweet text.

3. BERT sentence vectors were calculated for each preprocessed tweet text using the model's encode function. Note that this can be a rather time-consuming process.

Sentence meta embeddings were then calculated by summing the three sentence vectors calculated for each tweet. The resulting sentence meta embeddings were then saved in preparation for the next step.

Traditional methods for clustering textual data (such as Latent Dirichlet Allocation) require text to be stemmed and/or lemmatized (the process of reducing inflected words to their word stem, base, or root form). This process can be cumbersome and inaccurate. Since embeddings capture relationships between similar words in an unsupervised manner, our approach does not require either stemming or lemmatization.

2. Sample clustering

Our clustering methodology involves the following steps:

1. Calculate a cosine similarity matrix between vector representations of the sentences meta embeddings for a batch of samples. This process generates a matrix of similarity values between all possible pairs of vectors in the sample batch.

2. Calculate (or manually set) a threshold value at which we would draw an edge between two nodes in a graph.

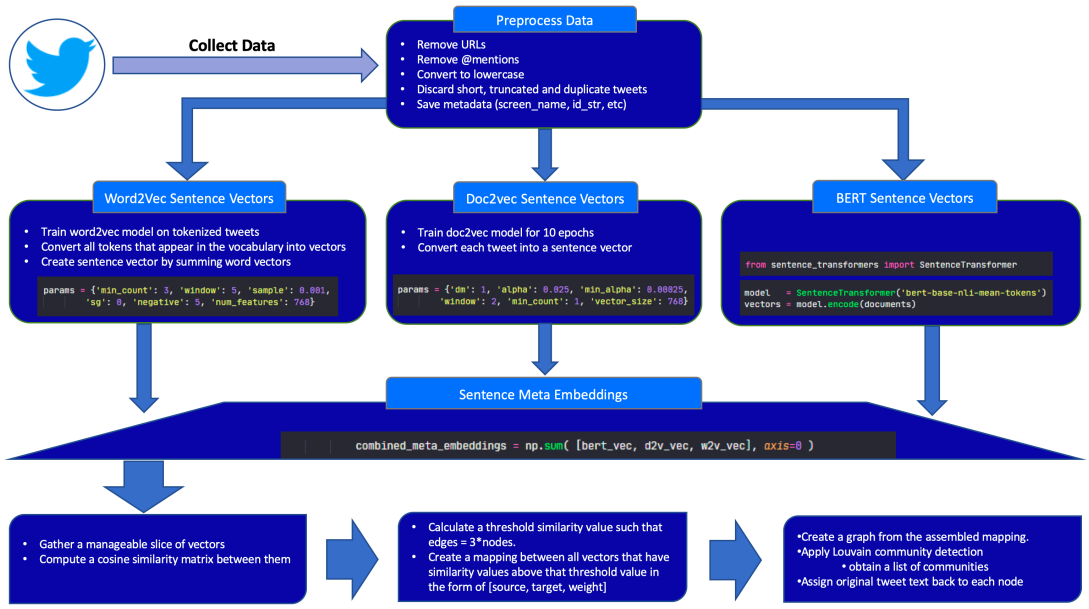
3. Find all vector pairs that have a cosine similarity equal to or greater than the threshold value. Create a node-edge graph from these values, setting the edge weight equal to the cosine similarity between that pair of vectors.

4. Perform Louvain community detection on the resulting graph. This process labels each node based on the community it was assigned to.

5. Process the results of the clustering - for instance, extract common words, n-grams, and subject-object-verb triplets.

6. Perform manual inspection and statistical analysis of the resulting output.

Here is a diagram of the above process:



It is possible to perform reasonably fast (less than 10 seconds) in-memory cosine similarity matrix calculations on small sets (<20000) using, for instance, the sklearn.metrics.pairwise cosine_similarity function. However, larger sets of vectors that don't fit into memory require a calculation loop that can take anywhere between minutes and hours to run. In order to process our large sample sets, we opted to perform processing in batches using the following logic:

1. Start with an array, *current_batch*, populated with small *batch_size* (e.g. 10,000) randomly selected sample vectors from the full set of samples to be clustered. We used randomly sampled vectors during all of our experiments so as to not optimize clustering logic for a deterministic set of inputs.

2. Calculate an in-memory cosine similarity matrix between all vectors in *current_batch*.

3. Calculate a threshold cosine similarity value that will select the top (*batch_size* * *edges_per_node*) samples from *current_batch*.

4. Iterate through the cosine similarity matrix values found for vectors in *current_batch*, adding pairs of nodes to a list, *graph_mapping*, in the form [source, target, cosine_similarity] for each pair whose cosine similarity was equal to or greater than the threshold value calculated in the previous step.

5. Create a node-edge graph using the *graph_mapping* list created in the previous step. Edge weights are assigned to the *cosine_similarity* values obtained during that process. Run the Louvain community detection algorithm on the graph to obtain a list of nodes, labeled by community. This process will not utilize all of the vectors in *current_batch*, so save a list of vectors that were not included in the resulting graph into a new list, *new_batch*.

6. Iterate through the communities found in the previous step, selecting the list of vectors that were assigned to each community.

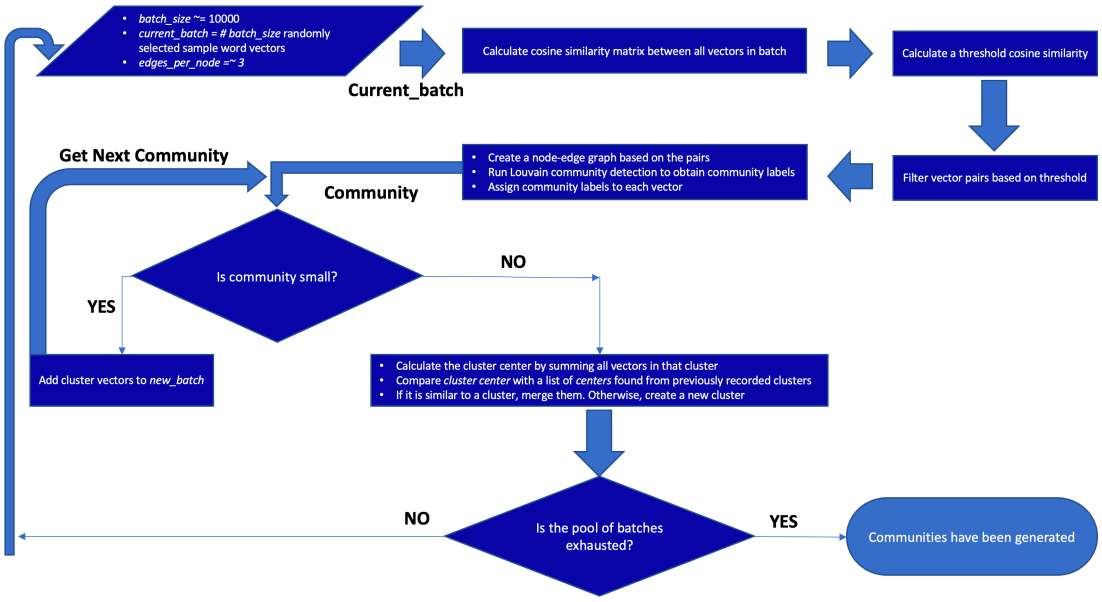
7. If the length of the list of vectors assigned to a community is less than the defined *minimum_cluster_size*, add those vectors to *new_batch* and proceed to the next community.

8. If the length of the list of vectors assigned to a community is equal to or greater than the defined *minimum_cluster_size*, continue processing that cluster.

9. For each cluster that fits the *minimum_cluster_size* requirement, calculate a *cluster_center* vector by summing all vectors in that cluster. Compare *cluster_center* with a list of *cluster_center* values found from previously recorded clusters. If the new cluster center has a cosine similarity value that exceeds a *merge_similarity* value, assign items to the previously recorded cluster. If not, create a new cluster, and assign items to that.

10. Once all communities discovered in step 5 have been processed, add new samples from the pool to be processed to *new_batch* until it reaches size *batch_size*, assign it to *current_batch*, and return to step 1. Once all samples from the pool have been exhausted, or the desired number of samples have been clustered, exit the loop.

Here is a diagram of the above process:



Failsafe

Occasionally, the loop runs without finding any communities that fulfill the *minimim_cluster_size* requirement. This, of course, causes the loop to go infinite. We added logic to detect this (check that the length of *new_batch* is not the same as *batch_size* before proceeding to the next pass). Our fix was to forcefully remove the first 10% of the array and append that many new samples to the end before proceeding to the next pass.

Variable settings

Different batch sizes result in quite different outcomes. If *batch_size* is small, the selection of samples used to create each graph may not contain a wide enough variety of samples from the full set, and hence samples will be missed. If *batch_size* is large, more communities are discovered (and the calculations take longer, require more memory, etc.). We found that setting *batch_size* to 10,000 was optimal in terms of accuracy, speed, and memory efficiency.

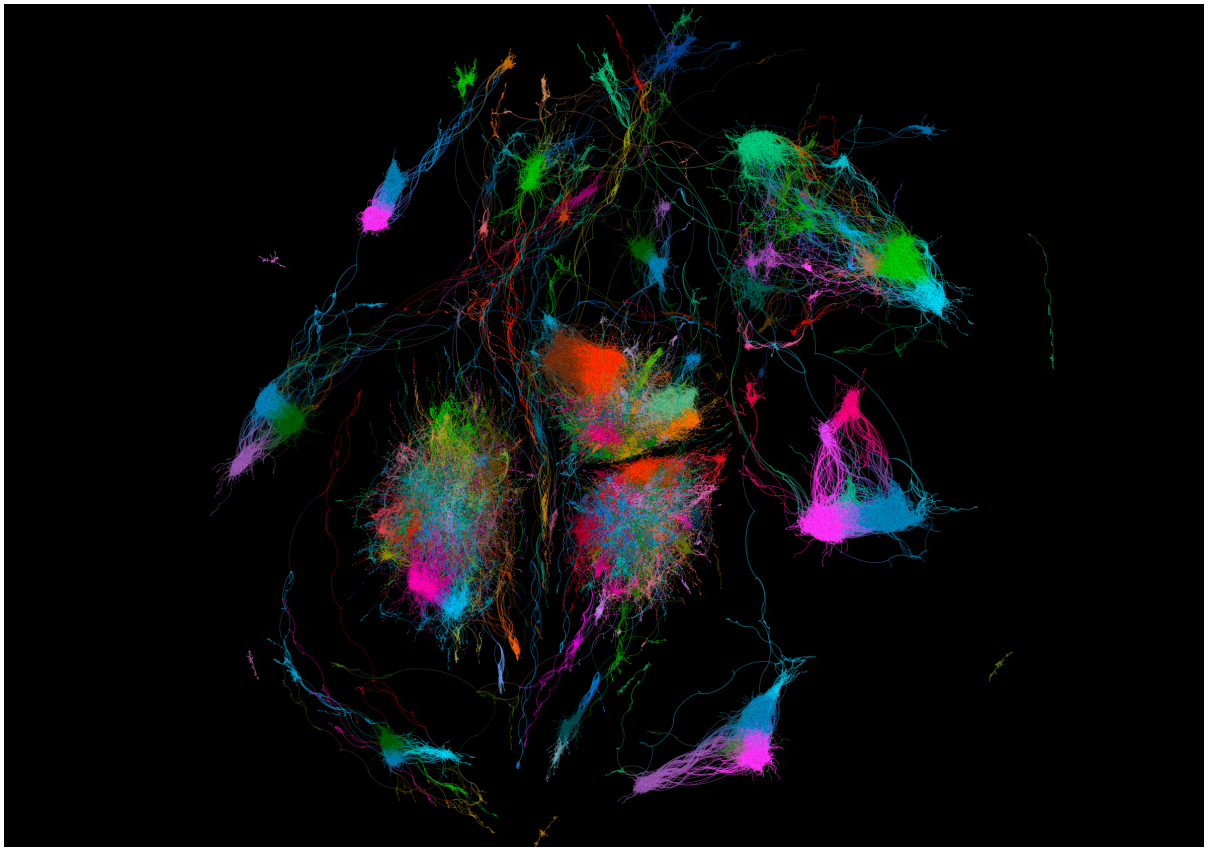
The *edges_per_node* variable has a marked effect on the accuracy of the clustering process. When *edges_per_node* is set to a low value (1-3), less samples are selected from each batch during graph creation, and community detection often finds many very small (e.g. 2-item) communities. However, when *edges_per_node* is set to higher values (>6), a smaller number of larger communities are detected. However, these communities can contain multiple topics (and hence are inaccurate). We found that an *edges_per_node* value of 3 to be optimal for a *batch_size* of 10,000. Increasing *batch_size* often requires also increasing *edges_per_node* to achieve similar looking results.

The *minimum_cluster_size* variable affects the granularity of the final clustering output. If *minimum_cluster_size* is set to a low value, more clusters will be identified, but multiple, redundant clusters may be created (that all contain tweets with similar subject matter). If

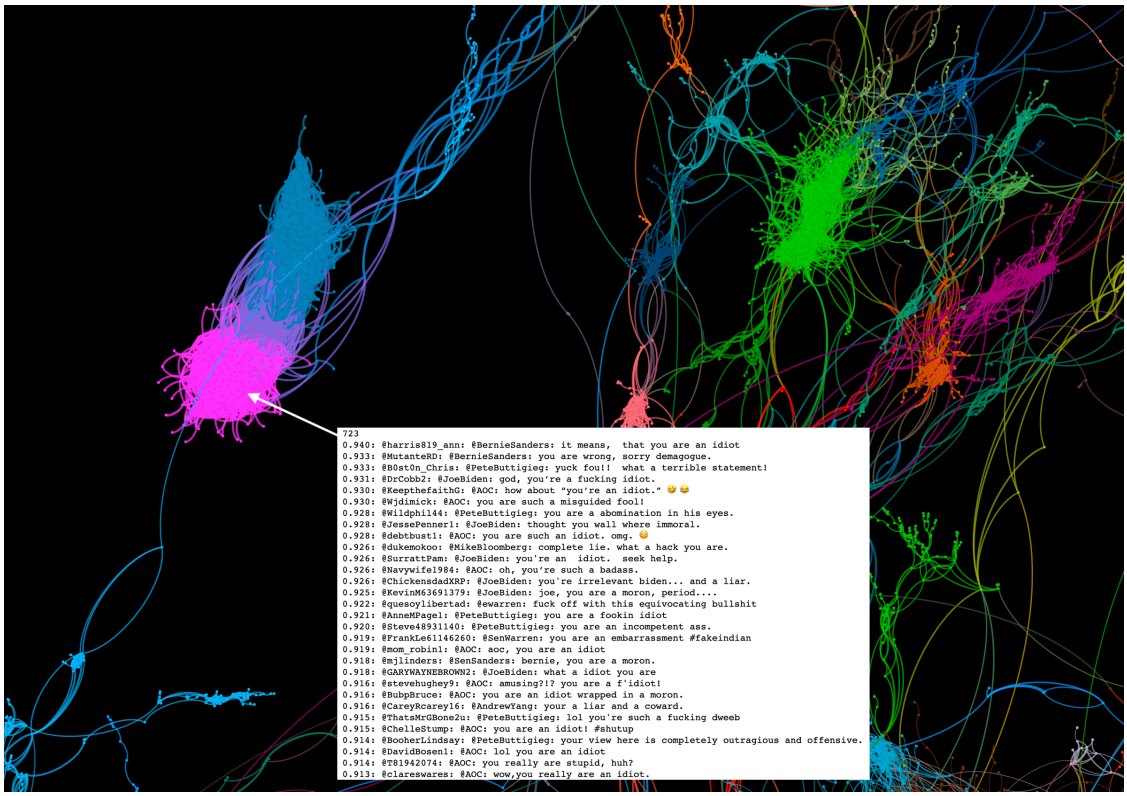
accuracy is not important, setting *minimum_cluster_size* to a higher value will result in less clusters, and less redundancy, but may create clusters containing multiple topics (false positives), and may cause some topics to be lost. In datasets that contain a very wide range of different topics, a high *minimum_cluster_size* value (e.g. 50) may cause the process to not find any relevant communities at all. We found this variable to be very dataset-dependent. We tried values between 5 and 50, but ended up using a value of 50 for our experiments, mostly to allow for aesthetically pleasing visualizations to be created.

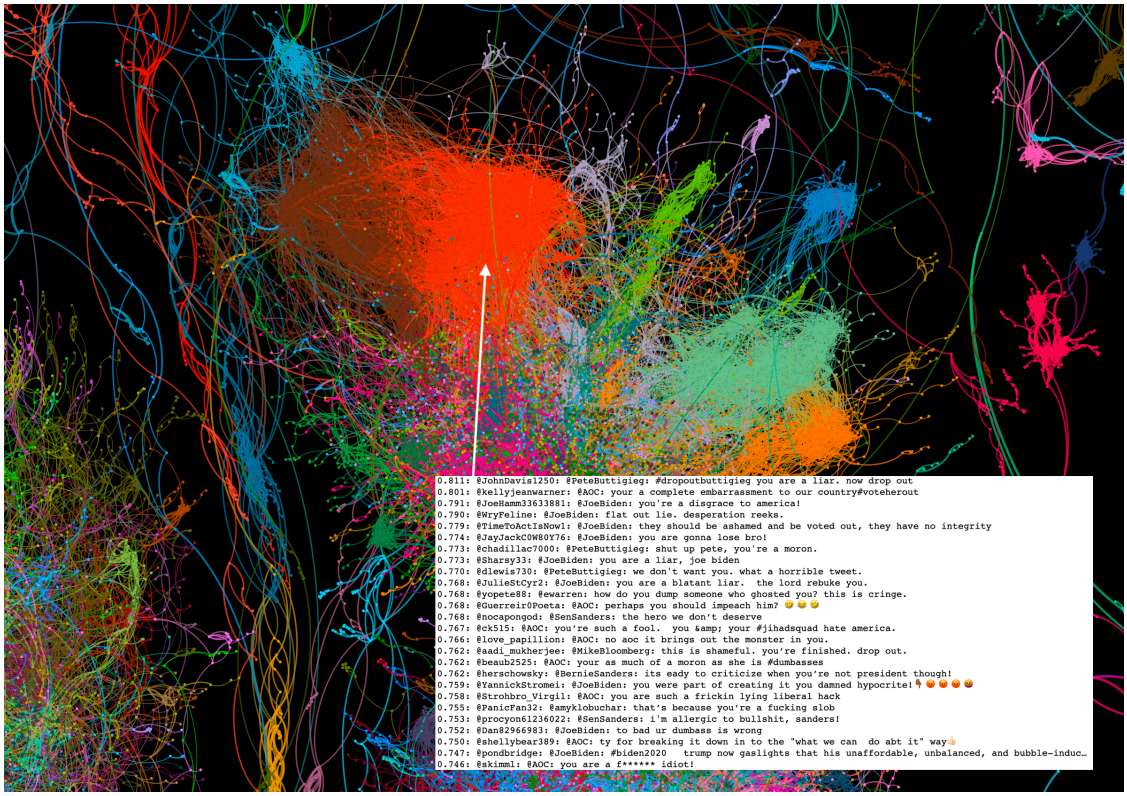
The *merge_similarity* variable has a similar effect on the output as the *edges_per_node* variable discussed earlier. This variable dictates the threshold at which newly identified clusters are merged with previously discovered ones. At lower values, this variable may cause multiple different topics to be merged into the same cluster. At high values, more redundant topic clusters are created. In our setup, we set *merge_similarity* to 0.98.

An example of a visualized graph (the one we generated using 30k tweets from set 1) looks like this:



Below are a few examples of how tweets assigned to identified clusters map onto the visualized graph:





Appendix 2: Experiment: Using identified clusters for new tweet classification

We experimented with the idea that identified clusters might be used to classify new tweets. In order to do this, we clustered approximately 25% of all tweets from each dataset and then attempted to classify the entire captured dataset using the following process:

1. For each tweet in the dataset, calculate meta embeddings using the same models and methods that were used to generate the clusters.

2. Run cosine similarity between the new tweet's meta embedding and all previously identified cluster centers, and find the best match (highest cosine similarity score).

3. If the cosine similarity exceeds a threshold, label that tweet accordingly. If not, discard it. In this case, we used a value of 0.65 as a threshold.

Set 1 (democrats):

184,851 (approximately 25% of the full dataset) tweets were clustered (using a *minimum_cluster_size* of 5) to obtain 3,376 clusters. The full 719,617 set of tweets were then converted into sentence meta embeddings and compared to the clusters found. This process matched 541,812 (75.29%) of the tweets.

Set 2 (realDonaldTrump):

188,010 (approximately 25% of the full dataset) tweets were clustered (using a *minimum_cluster_size* of 5) to obtain 3,894 clusters. The full 747,232 set of tweets were then converted into sentence meta embeddings and compared to the clusters found. This process matched 623,120 (83.39%) of the tweets.

By manually inspecting the resulting output (lists of tweet texts, grouped by cluster) we were able to determine that while some newly classified tweets matched the original cluster topics fairly well, others didn't. As such, identified cluster centers can't reliably be used as a classifier to label new tweets from data captured with similar parameters. When using a threshold value higher than 0.65, a lot less tweets ended up being matched to existing clusters. One possible reason for the failure of this experiment is that some identified clusters contain tweets that only have very high cosine similarity values to the cluster center (above 0.95), whilst others contain tweets with much lower similarities (albeit whilst the content of the tweets match each other). As such, it might be that each cluster must have its own specific threshold value in order to match similar content. We didn't spend a great deal of time exploring this topic, but feel it may be worth researching in the future. Naturally, if this were figured out, cluster centers would likely only be valid for a short duration after they've been created due to the fact that the political and news landscape changes rapidly, and no techniques exist (as of yet) in this area that are able to create models that include a temporal context.