Sending the Trolls to a Subspace: Identifying Polarization through Word Embeddings

Felix Adam[a], Jordi Serra[a], Sebastian Wolf[a]

[a]*Barcelona Graduate School of Economics, Barcelona, Spain*

Keywords: Textmining, Russian Interference, Polarization

1. Introduction

In this report we analyze efforts by the Russian Internet Research Agency (IRA) to polarize the American public on social media / twitter in the context of the 2016 presidential election. The 2016 election featured two candidates that bore starkly different promise for the Russian government. Whereas Hillary Clinton had already built her profile as a strong critic of Russian foreign policy during her tenure as Secretary of State, Donald Trump offered the opportunity for Russia to push American politics into a more favorable space for Russian interests. The topic has caused a lot of controversy in American politics over the last two years. Recently, the finally released, heavily redacted Mueller report has rekindled the debate, and provided evidence that Russia did indeed attempt to aid Trump's election, as had already been alleged by many political observers. Polarization likely played an important role in Russia's strategy to do so. Given Trump's more radical policy positions, a polarized society can be expected to play in his advantage, allowing him to increase his vote at the extremes of the political spectrum. Polarization might also have been a goal of the Russians in and of itself, by weakening American influence in world politics, and lowering faith in democratic vs autocratic political systems.

Any democratic government has a strong interest in counteracting such foreign interference in the national debate that can destabilise the democratic process. Therefore, there is strong policy justification for research into measures that help to identify polarization in

Email addresses: `felix.adam@barcelonagse.eu` (Felix Adam), `jordi.morera@barcelonagse.eu` (Jordi Serra), `sebastian.wolf@barcelonagse.eu` (Sebastian Wolf)

twitter data. In this report, we attempt to do just that. We study a dataset of nearly 3 million tweets by user accounts linked to the IRA and use textmining techniques to uncover targeted efforts to polarize the national debate. We propose to detect polarization through embedding each tweet into a low-dimensional subspace and subsequently clustering them, aiming to find distinct groups. We acknowledge that clusters in our dataset cannot be expected to be perfectly separated, or bunched into a specific number of clusters. To build up credibility among the American public, IRA tweets needed to build a history of realistic tweets to increase followers and be taken seriously. We hence clean the dataset to identify what we describe as "content-creators", IRA agents posing as genuine american twitter users.

Our findings show, that the Russian polarization efforts we're concentrated around distinct topics, such as race, police-brutality, healthcare or even the US position towards Russia itself. Contrary to our inital expectations, we didn't find a clear separation within topics, such as into liberal and conservative clusters. Rather we found that tweets within a cluster are strongly opinionated, trying to push the most extreme side of a given subject. We did not find any indication about aims of polarizing the discussion within topics. Russias polarization strategy can thus be described as trying to amplify the voice of a particular extreme side, be it the complete shutdown of borders or pushing a negative perceptation of the police force in general.

In the second section of this report, we clarify the likely political motivation of the Russian government to engage in a polarization strategy. The third section introduces our methodology. The fourth section describes the data and our pre-processing steps. The fifth section reports the results, and the sixth section discusses our conclusions.

2. The Strategy behind Russian Social Media Efforts

In order to identify efforts to polarize US politics through social media, we need to understand the strategies behind such efforts. In particular, it is crucial to know why Russia would use such strategies, and how they could potentially influence domestic and foreign politics in the US.

Unlike the predicted "End of History", (Fukuyama (1989)) conflicts between liberal democracies in the west and Russia reemerged in the beginning of the 21st century. Following the fall of the Soviet Union, instead of integrating into the West, Russia tried to reestablish it's position as a superpower. Tensions became more pronounced in the beginning of the second decade of the 21st century, with hot conflicts in the middle East, most notably Syria, and the Maidan Revolution in Ukraine. At the heart of these conflicts were and still are diverging interests of the United States and Russia. Be it the integration of the Ukraine into the European Union (EU) and NATO or the Russian support for the Syrian leader Assad. These points of conflict can be seen as a motivator for the Russian interference in the 2016 US presidential elections.

In the aftermath of the 2016 elections, the US intelligence community published a number of assessments of the Russian strategy. Most notably, the Central Intelligence Agency (CIA), the Federal Bureau of Investigation (FBI) and the National Security Agency (NSA) published an assessment of Russian activities, stating that the goals of Russia were to weaken public faith in the US democratic process and denigrate presidential candidate Hillary Clinton (ODNI (2017)). The idea behind these efforts was to undermine the US-led liberal democratic order, posing a threat to Putin's regime. In particular, Putin saw a potential presidency of Hillary Clinton as a threat to his ambitions in Ukraine and Syria, due to Clinton's foreign policy positions as Secretary of State. Trump on the other hand was seen more friendly towards Russia and thus favoured over Clinton. As part of these efforts, Russia used social media networks like Facebook and Twitter, to promote radical discontent with US politics, polarize the discussion and denigrate Hillary Clinton. The report further states, that the Russian strategy changed over time. In the beginning, the goal was to undermine public institutions. However, with Clinton leading over Trump, the Russian efforts shifted towards the defamation of Clinton, trying to harm her electability and potential presidency. Interestingly, the report states that after Trump had won, the efforts to undermine public institutions stopped. The three intelligence agencies pick out the so called Internet Research Agency (IRA), as one of the main sources of these social media accounts. The IRA is described as a private agency with ties to the Russian government, engaging in targeted social influence efforts. The findings of the so-called Mueller investigation (Mueller III (2019)) support this assessment. According to the Mueller Report, the IRA carried out social media campaigns to amplify social discord in the US. In order to do so, Twitter accounts linked to the IRA tweeted on divisive US political and social issues, such as illegal immigration and racial injustice. According to Mueller III (2019), the IRA used social media accounts in two manners. Some accounts were designed around fictitious US personas, posting original content on divisive topics, promoting radical ideas and denigrating or promoting political candidates. Other accounts weren't used for original content, but rather to promote and amplify the impact of the "original" content. Some of the accounts posting such content had a large follower share, such as TEN_GOP, an account pretending to be related to the Tennessee Republican Party. TEN_GOP was clearly used to promote polarization by pushing extreme content and promoting Donald Trump. Some tweets included:

- "Wake up America before it's too late! Europe has already lost its chance! #BanIslam #StopIslam #filibuster"
- "Donald Trump: "I will be the greatest jobs-producing president that God ever created"

The strategy clearly worked, since tweets of the IRA we're picked up by major news outlets in the US and thus shaped at least daily discussions (Mueller III (2019)). We can thus summarize the Russian strategy as follows: Polarize the political discussion in the US, undermine Clintons authority and electability and promote Trump.

Now the question remains how these strategies, if successful, would affect US foreign policy, in particular towards Russia. In the subsequent analysis, we'll focus on the idea of polarization. **?** define polarization as the simultaneous presence of opposing or conflicting principles, tendencies or points of view. In a quantitative manner, polarization can be seen as an increase of variance of ideas and attitudes towards poltical questions. **?** argue, that this reduces the probability of group formation at the center of the political spectrum and increases the formation of groups with irreconcilable preferences. This can have peculiar effects on domestic and foreign policy. Beinart (2008) argues that polarization leads to a weakened international position of the US. First, international endeavours such as the promotion of trade agreements, UN resolutions or even military campaigns crucially depend on domestic support. Without domestic support, international allies of the US discount promises, while the US appears weak towards enemies (Beinart, 2008). Schultz (2017) supports these findings. He argues that domestic political polarization leads to three issues. First, it is more difficult to get bipartisan support for risky undertakings. Second, it gets harder to agree on lessons from failures, complicating efforts to learn. And lastly, the risk of dramatic policy swings complicates the ability to make long term commitments. The overall effect of polarization can thus be summarised as follows: A polarized society is caught up with fighting against itself. It can't reconcile large differences to find a common foreign policy strategy.

Consequently, Russia's global standing would greatly increase from a polarization of US politics. Actually, we can already see these effects in motion, especially in Syria. The US failed to gather international support for UN resolutions and seems to have no clear strategy, while Russia continues pushing its ally Assad (?). Interestingly, the other strategies described by Martin and Shapiro, such as defamation and persuasion can also be seen as tools of polarization.

Having discussed the effects of polarization, we can now try to identify Russian polarization efforts through social media.

3. Methodology

To set up our methodology, we note the close relationship between polarization and clustering techniques in the field of unsupervised learning, or community detection in the field of network science. Clustering techniques aim to detect data points that naturally belong together because they are close by some measure in some higher dimensional space, and that are far from other data points by that same measure. Similarly, in network science we try to detect communities of nodes that display high similarity in terms of their links. Nodes that form a community share many common friends, and they share very few common friends with nodes outside of their community. Essentially, polarization is the social science term to describe clustering of data points. This means that to detect polarization we can employ unsupervised learning methods to detect clusters.

In this report we aim to make use of this insight by attempting to identify clusters in the body of tweets which could provide evidence of Russian efforts to polarize the American society ahead of the 2016 presidential election. A priori, the most natural clusters we would expect to emerge are two: one cluster pulling the debate towards the extreme left, and another cluster pulling the debate towards the extreme right. However, this would not be the only type of clustering that would yield evidence for polarization. Russian polarization efforts may also have been organized along topic lines, creating two or more clusters within a chosen set of topics. Further, any effort to polarize the debate would have to be hidden among a cloud of noise in order to conceal orchestrated nature of the effort, and to appear like a body of real tweets to the American twitter users. Therefore, we do not expect the clusters to yield perfect class separation, or yield very clear separation of opinions on topics. In support of our hypothesis we would expect that, compared to a non-orchestrated discussion on Twitter, the IRA tweet corpus shows more pronounced clustering. Similarly, the level of clustering should be comparable to what are known to be very polarized debates, for example the discussion among fans ahead of a football match of two rival teams.

Since tweets are completely unstructured text documents, clustering them as is will be impossible without any labels. We therefore make use of word embeddings to represent each word in a continuous vector space. In a subsequent step, we can obtain the embedded tweets by adding each embedded words and normalizing the resulting vector by it's norm. By doing so we hope to enable clustering: Similar tweets in terms of words they use and thus topic and sentiment should be represented by similar vectors. For example, the words "good" and "bad" should be fairly dissimilar to each other, so tweets talking about "good" or "bad" politics should be differentiable inside said topic.

In order to represent the tweets in a continous subspace, we use the Word2Vec algorithm proposed by Mikolov et al. (2013). The idea behind Word2Vec is to first train a classifier on the corpus, trying to predict a given word through it's neighbors. Having trained the classiefier, one can use the parameters learned as representations of each word, as embeddings. In this setting we choose the so-called skip-gram architecture, aiming to predict the sorroundings words given the current word (Mikolov et al., 2013). In Word2Vec this prediction task is done with two-layer neural network, with an input of a one-hot encoded word-vector, a hidden layer and an output layer based on a softmax-classifier. Having trained the network, the hidden layer represents the weights each word of the input vector has for the classification task. Having trained the model, we can use the weights of the hidden layer as our embeddings. One can think of these word embedding as describing a word by the company it keeps. Words that appear more often in the same context, will have similar weights in the hidden layer, predicting to be surrounded by similar words (Mikolov et al., 2013).

To cluster the tweets we use the DBSCAN algorithm Ester et al. (1996). We choose to use DBSCAN because of two reasons: first, there is no requirement to specify the number of clusters in the data a priori which is convenient for us since we do not know how many topics we will find. Second, DBSCAN has a notion of noise. This is a particularly relevant element since we expect to find lot's of noise tweets by bots trying to pose as real users, but we are only interested in the hopefully more similar polarizing tweets.

The intuition behind DBSCAN is to find the areas of high density, separated by areas of lower density. Objects in these sparse areas - that are required to separate clusters - are usually considered to be noise and border points. Formally, the algorithm is structured around core points. A point is considered core if there are more than n_min neighbors at distance smaller than ϵ. All neighbors that are at distance smaller than ϵ from a core point are considered to be part of the same cluster as the core point. Points that are not reached by any core point are considered to be border points and are not assigned to any cluster. ϵ and n_min are tuning parameters of the algorithm Ester et al. (1996).

The algorithm allows the use of multiple distance measures. In our case, for simplicity, we picked euclidean distance. We are aware that high dimensional spaces as ours can lead to "curse of dimensionality" problems and in future research we would consider other methods suited for high dimensional spaces. Ideally, after finding the topics or clusters, we would like to find two separate groups within the cluster talking about the same topic but with completely different opinions. In other words, polarization. To measure the difference of opinions within topic/cluster, we propose to use the average cosine distance between tweets.

4. Data

To identify efforts by Russia to polarize the public opinion we make use of a dataset of nearly 3 million tweets by 3077 different English-language user accounts linked to the

Internet Research Agency and made publicly available by Twitter covering several years leading up to and after the election period.

4.1. Data preprocessing

Preprocessing of the tweet corpus is of capital importance when using natural language processing techniques where different approaches in text cleaning/filtering can lead to starkly different results during model training. In this section we are going to describe and justify the steps we followed to clean our data.

4.2. Filtering

In the data set, we find tweets in different alphabets: latin, cyrillic and arabic. The first step we take is to remove all tweets that do not use the latin alphabet. We keep English tweets only. Since the main target is to show the polarization caused by these tweets in the American society, we believe English is the most relevant language for this objective. We also drop users that don't have their user description or account language in English. Having any of the features described above would look suspicious to the "true" American users, who should trust them to be fellow, caring citizens. After removing all these tweets from our data, we keep roughly 85% of the original data set.

4.3. Cleaning

Tweets are usually dirty: no right use of grammar and punctuation, many URLs, retweets, hashtags and emojis. In order to remove noise and improve the embeddings, we proceed by removing mentions, retweets, urls, breaklines and blank spaces. A question that arises when analysing tweets is what to do with emojis. Generally speaking, emojis can be difficult to process but they carry important information about the tweet. In our case, after some preliminary results, we find that emojis don't help much on identifying polarization and, therefore, we removed them from our corpus. Another important cleaning decision regards hashtags. Hashtags provide useful information for identifying the topic and a user's opinion about that topic. Since we want to identify polarized information, we believe that these hashtags yield important information on topics, so we keep them in the corpus.

4.4. Further preprocessing

A classical approach for data cleaning in text-mining is to follow the Remove Stop Words - Stem - Lemmatize - Tokenize pipeline. However, in the case of short texts such as tweets, some recent contributions Bao et al. (2014) have argued that lemmatization and stemming lead to worse performance in terms of sentiment analysis, while feature reservation, negation transformation and repeated letters normalization improves it. Other references such as Hollink et al. (2004) support the view that lemmatizing and stemming does not improve performance of the algorithms significantly. One has to keep in mind that any morphological pre-processing of the training data reduces the amount of information that model can obtain from the corpus, which can create unnecessary noise making some sentences ambiguous. Therefore, since there is no clear evidence that we can improve performance using lemmatization and stemming, we prefer not to add this layer of complexity to the problem and stick to the cleaned tokenized raw data.

4.5. Data exploration

According to Mueller III (2019) Russia deployed a complex strategy to influence American elections through an army of troll tweeter accounts. Our hypothesis is that this was done via three main channels: politically active users, fake news accounts and bot accounts. We will call the first two channels the content providers and the third channel the content amplifier. The politically active accounts are well identified in the reports by the American intelligence agencies and also by their user descriptions. The news providers consist in a few number of users that attempted to impersonate local news agencies and spread both true and fake news. The idea behind is that, in order to get a loyal audience, they had to build up credibility providing real news for a wide range of topics. However, among these "innocent" news they also published tweets with highly polarizing socio-political content. These content provision interventions appear to then have been complemented using bot twitter users to amplify the content from content providers via retweets.

4.6. Content creators vs content amplifiers

In order to verify the underlying phenomena introduced previously we plotted the percentage of retweets per user in figure 1. We find that there is a clear separation of roles:

the amplifier is a user with a high percentage of retweets who only retweets information from other users. The content creator is a user with low percentage retweet percentage, only creating content.

In the upper right corner of figure 1, we see a high concentration of users that, despite having a large total number of tweets, hardly ever produce any content. On the other extreme we have users that produce a reasonable amount of original content tweets but hardly ever retweet. We identify an outlier that has more than 150.000 tweets, which means that over 5 years of data, this user was retweeting an average of 80 tweets per day! The plot also features some interesting 'curves' that make up perfect lines. This appears like clear evidence for bots created 'at scale' with varying retweet percentages programmed using a specified retweet/total tweet function.

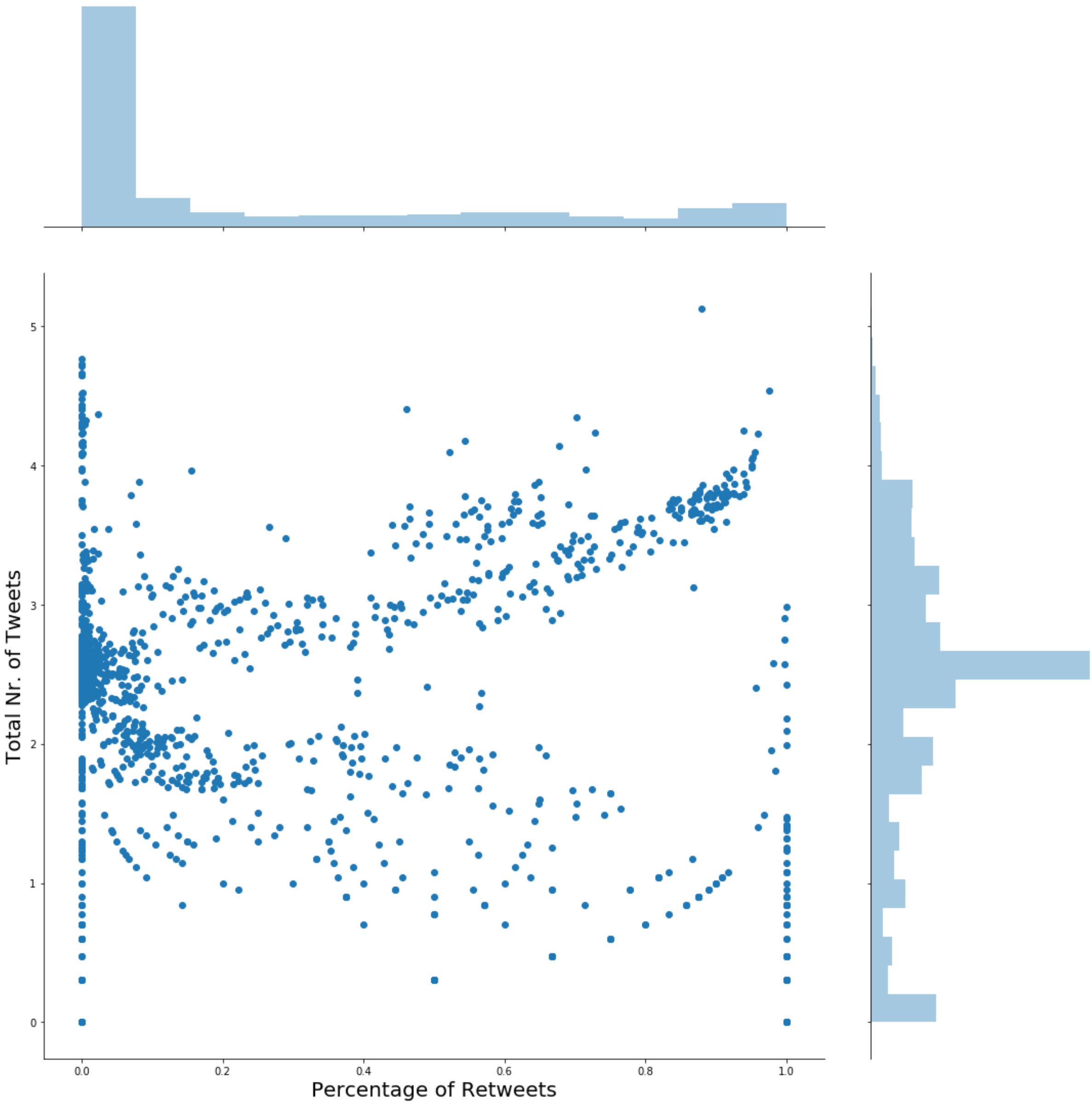
We identified the most politically active accounts and it seems that, as expected, all of them fit in the content provider category (See Tables 1 and 2). Also, we can see in table 3 and 4, that their user description gives a clear idea of their political beliefs. Regarding news providers, having an average of retweet ratio of 0.01 we can also confirm that they fit in the content provider category.

As an exploratory step, we have checked the content providers' most frequent used words and produced wordclouds for them (see Annex). In figure 4 we clearly see the preferences of right-wing partisans, mainly talking about Obama and religion related topics. In figure 5 we can clearly see the topics left-wing partisans talk about. Note, for example, a considerable amount of anti-racism vocabulary. Finally, for news providers in figure 6, we cannot identify any more a political preference nor extreme vocabulary but just a normal vocabulary a news provider would use.

4.7. Repeated tweets

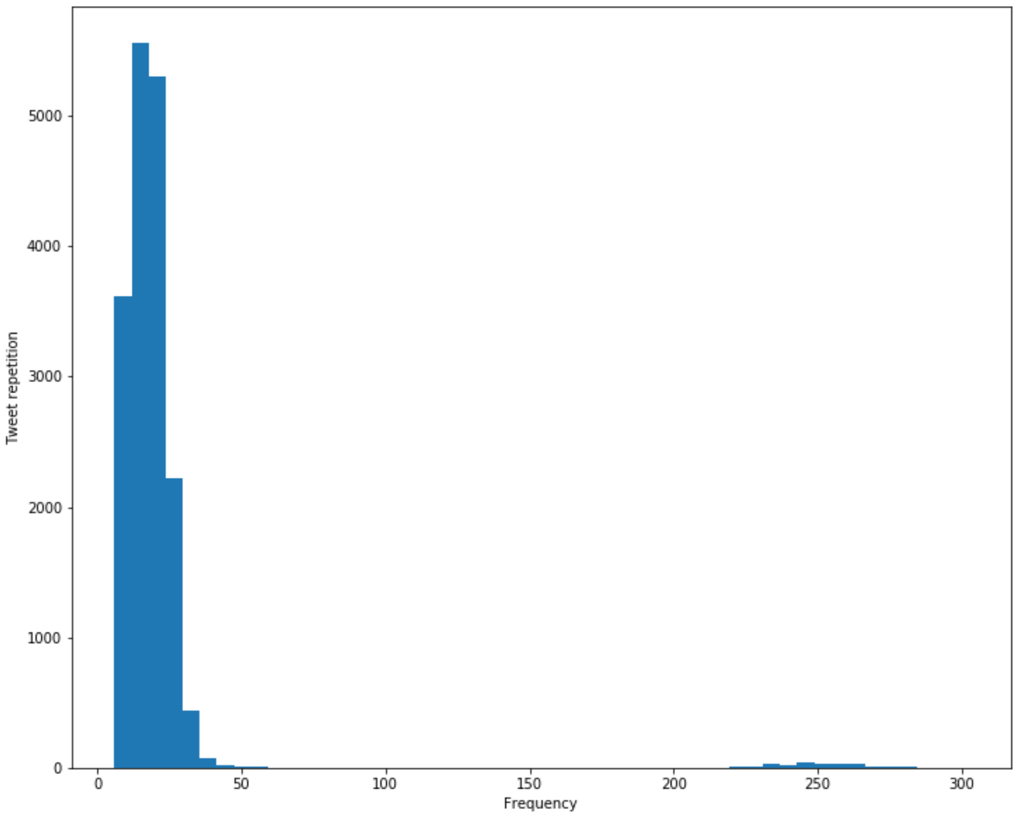
Deepening into tweet text exploration we perceived identical tweets from different users which weren't retweets. In other words, different users posted the exact same text several times. In figure 2 we plotted the distribution of how many repetitions we have from a tweet excluding retweets. We see a high concentration of repeated tweets below 50 repetitions and then a small concentration again around 250 repetitions.A first look at

Figure 1: Retweet ratios



the tweet text suggests that more frequent tweets seem to be talking about self-aid and motivation sentences whereas the less frequent tweets seem to be news repeated by the news providers. A possible explanation for this scenario is that content amplifiers or bots don't have a retweet ratio of exactly 0. Which means that, at some point, this bots have "created" some content. Therefore, they make use of a pool of motivation sentences to "create" tweets so they don't arouse suspicion. This is a particularly notable results since the repeated tweets are almost 30% of the "original" created content.

Figure 2: Repeated tweets



4.8. Hashtags

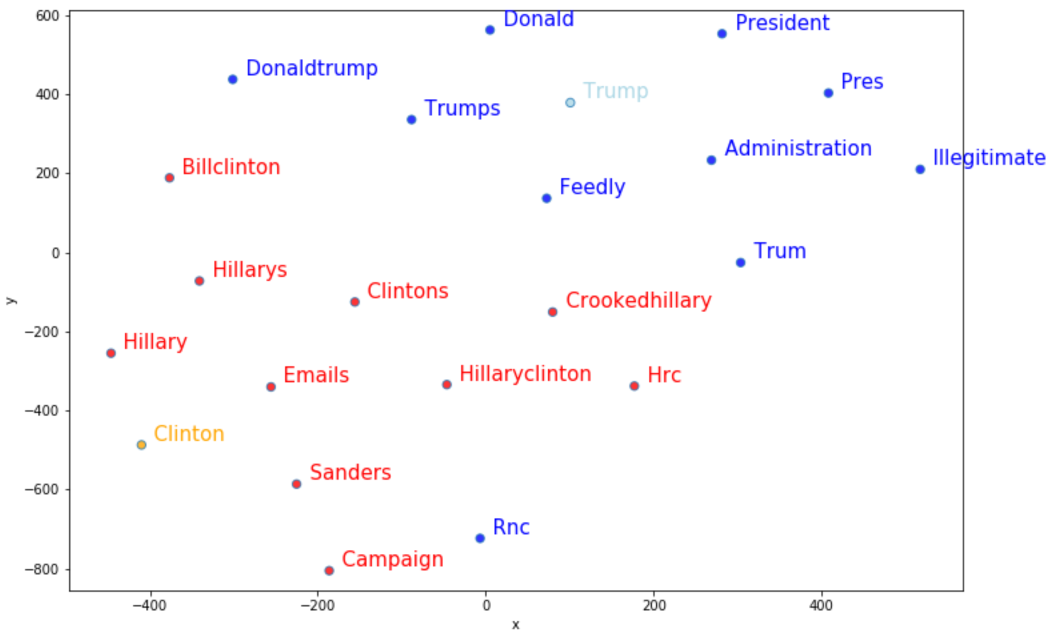
As explained previously, hashtags are relevant when analysing tweeter data since they perform as a good proxy for the topic of tweet. In figure 7 and 8 we plot most repeated hashtags among all the available data for some handpicked left and right partisans. In general, users with different political convictions don't share hashtags. This is because, usually, a hashtag has inherently a political belief behind and using it presuppose that the user is in favor of that belief. There is one particular exception with the hashtag "BlackLivesMatter" which is relevant in both groups. Apart from that, we do no see any other important information for our study.

5. Results

Having cleaned the data of non-english tweets and users, we trained the Word2Vec algorithm using Pythons Gensim library **?**. We trained the algorithm on the full corpus, using a skip-gram archtiecture with a window size of 4, embedding the tweets into a 100 dimensional subspace. Our results on individual words show promosing results towards the polarizing nature of some of these tweets. Tables 1 and 2 show the most similar words to Trump and Clinton, while 11 shows their representation in 2 dimensions using t-SNE (**?**). Overall, Clinton is connected to degoratory words such as "crookedhillary" or scandals like her e-mail server. Trump on the other hand is more similar to normal words one would expect. We see simlar results in Figures 10, 12 and 13. Each plot shows the comparison of two words in orange and light-blue and their 10 most similar words as measured by the cosine similarity of the vectors. In Figure 10 we can already make out some strinking differences between the similar vectors for Liberals and Conservatives. In a normal context, one would expect the word "Liberals" to be close "Conservatives" or other words describing the political spectrum. Here however we see, that the most similar words to liberal all seem to have a negative meaning, such as "Libtard", "Intolerant" and "Dishonest" or are connected to derogatory hashtags, like "Liberallogic". On the other hand, the similar words for Conservatives are much less extreme, highlighting the Russian bias towards Trump as a conservative candidate. Another clear indicator of the polarization efforts can be seen in figure 12, highlighting the Russian efforts to polarize the society alongside racial injustice topics. In a normal corpus, we would expect "Black" and "White" to be similar to other colors. Here we see that they are mostly social justice based, with "Black" being similar to black rights and "White" being close to notions of privilige and racism.

Having seen how the polarization efforts materialize in word similarities, we clustered the tweets. Our first effort at finding two distinct "polarization" clusters using k-means with 2 clusters failed. Without removing the news accounts and retweeters, the clustering algorithm picked up those groups as distinct clusters, showing their overall importance for the Russian efforts. Being interested in the "content-creators" described in section 4, we removed retweets and re-appearing tweets, possibly arising from a pool of tweets used by bots to appear human. Removing these users reduced the size of our dataset substantially,

Figure 3: Similarity of embedding: Clinton and Trump



leaving roughly 400.000 tweets. For efficiency reasons, we clustered a random subsample of 100.000 tweets, using DBSCAN with an ϵ-parameter of 0.6 and a minimum neighbor count of 15.

Even though only run on a fourth of the data, DBSCAN still identified meaningful clusters. Visualizing clusters through the most common words, we found that these clusters represented topics divsive topics of american politics. Figure 15 shows one of the largest clusters, which we identified as the police-brutality and race cluster. Inside of this cluster, most, if not all users identified themselves through their profile descriptions as black-rights activists, condemming police violence towards people of color. While we did not find an orthogonal cluster, pushing content like "BlueLivesMatter" the polarizing nature of these tweets is apparent. The goal of this cluster is clear: amplify disconntent and distrust with the police force in left circles, while communicating to right circles that left leaning citizens distrust the police. A similar cluster is shown in Figure 17, centered around healthcare. Here the idea is to push the extremly divsive topic of public healthcare in the US, without picking a clear side. Most of the tweets found in this cluster are coming from a single account: Politweecs. This account seems only be focused on pushing content around divsive topics, such as the healthcare bill or immigration legaslation. The drug cluster shown in Figure 18 on the other hand is clearly focused on discouraging drug liberalization, with tweets talking about the negative effects of legalization, another divisive topic. The center of these tweets is the hasthag "JunkieUS", trying to amplify the issues around drug abuse.

6. Discussion

Overall our results have shown that unsupervised learning techniques, with an emphasize on manual preprocessing and smart cleaning, can identify the Russian polarization efforts. Our analysis highlights how Russia tried to amplify divisive topics, focusing on issues which already split the american society and further pushing extreme opinions. Further, we found that Russia did not as initally expected try to polarize the disccussion inside a given topic, but rather amplify the positions they sought to be more promising or easy to influence. We saw through the word embeddings that the Russian strategy clearly favoured Trump over Clinton and trying to sow distrust between racial and party lines through the use of negative language in combination with topic-words such as "BlackLiveMatters" or "Racism".

The natural extension of our analysis would be to compare these findings and clusters to a normal twitter environment and build a measure of dissimilarity. Obviously, the political center is completly lacking from the IRA tweets, so an analysis of normal tweets could show how the IRA tweets differ in terms of polarization and embeddings, allowing an unsupervised identification of polarization in twitter discussions.

7. References

Bao, Y., Quan, C., Wang, L. and Ren, F. (2014), The role of pre-processing in twitter sentiment analysis, *in* D.-S. Huang, K.-H. Jo and L. Wang, eds, 'Intelligent Computing Methodologies', Springer International Publishing, Cham, pp. 615–624.

Beinart, P. (2008), 'When politics no longer stops at the waters edge: Partisan polarization and foreign policy', *Red and blue nation* **2**, 151–67.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X. et al. (1996), A density-based algorithm for discovering clusters in large spatial databases with noise., *in* 'Kdd', Vol. 96, pp. 226–231.

Fukuyama, F. (1989), 'The end of history?', *The national interest* (16), 3–18.

Hollink, V., Kamps, J., Monz, C. and de Rijke, M. (2004), 'Monolingual document retrieval for european languages', *Information Retrieval* **7**(1), 33–52.
URL: *https://doi.org/10.1023/B:INRT.0000009439.19151.4c*

Mikolov, T., Chen, K., Corrado, G. and Dean, J. (2013), 'Efficient estimation of word representations in vector space', *arXiv preprint arXiv:1301.3781* .

Mueller III, R. S. (2019), 'Report on the investigation into russian interference in the 2016 presidential election. volumes i & ii.(redacted version of 4/18/2019)'.

ODNI, O. o. t. D. o. N. I. (2017), 'Assessing russian activities and intentions in recent us elections', *Unclassified Version* .

Schultz, K. A. (2017), 'Perils of polarization for us foreign policy', *The Washington Quarterly* **40**(4), 7–28.

8. Annex

Table 1: 'Right' cluster retweet ratios among content providers

Twitter user name	Retweet ratio
Jenn_Abrams	0.0231
SouthLoneStar	0.0375
USA_Gunslinger	0.0770
patriototus	0.0033
redlanews	0.0038

Table 2: 'Left' cluster retweet ratios among content providers

Twitter user name	Retweet ratio
BleepThePolice	0.6768
Blk_Voice	0.0163
KaniJJackson	0.2655
LaChristie	0.0986
gloed_up	0.4606
wokeluisa	0.0822

Table 3: 'Right' cluster profile descriptions

Twitter user name	User description
USA_Gunslinger	Truth is strong, and sometime or other will prevail!
SouthLoneStar	Proud TEXAN and AMERICAN patriot #2a #prolife #Trump2016 #TrumpPence16 Fuck Islam and PC. Don't mess with Texas!
patriototus Being	patriotic means love or devotion to your homeland and readiness to defend it from any harm. United we stand, divided we fall. Conservative politics. #PJNet
redlanews	Conservative; Right and proud; Christian. Love my country and will stand against liberals and socialists.
Jenn_Abrams	Calm down, I'm not pro-Trump. I am pro-common sense. Any offers/ideas/questions? DM or email me jennabramsgmail.com (Yes, there are 3 Ns)

Table 4: 'Left' cluster profile descriptions

Twitter user name	User description
Blk_Voice	Activist. Feminist. Celebrating and highlighting Black excellence.
KaniJJackson	Follow the example set by Mrs Obama; peace, love, acceptance & vigilance #Impeach45 #Resist #GunReformNow
LaChristie	Progressive. Activist. Warrior. Inspiration. #Resistance
gloed_up	No black person is ugly #BRONZE #BlackLivesMatter #BlackToLive
BleepThePolice	For a second at least, I'm resurrecting the peace #Blacktivist #BlackLivesMatter
wokeluisa	APSA. #Blackexcellence. Political science major

Figure 4: Wordcloud for 'right' cluster

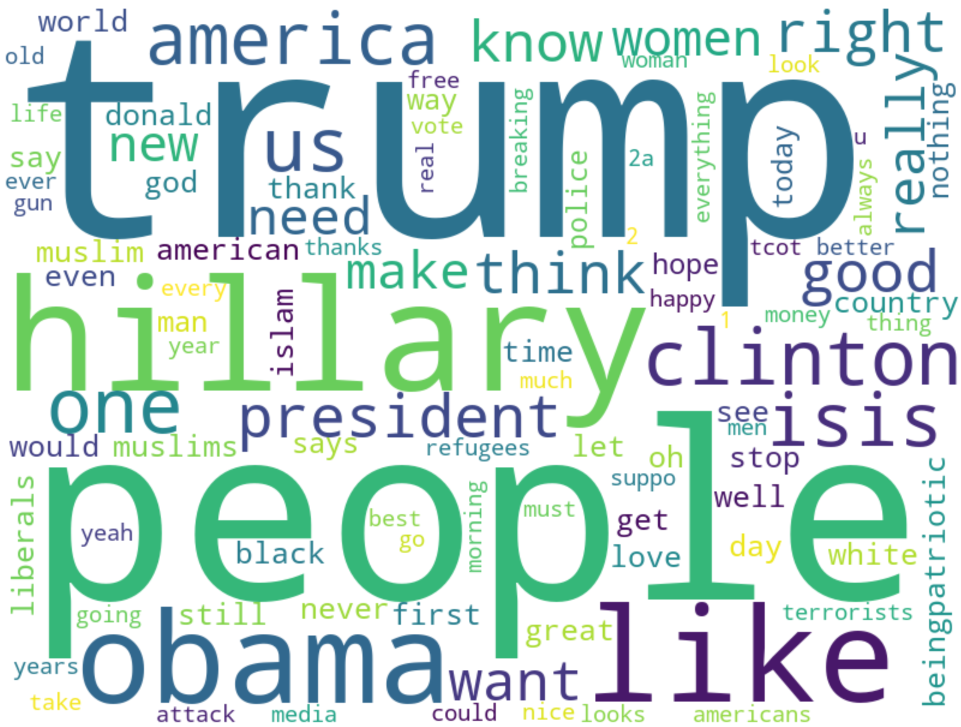


Figure 5: Wordcloud for 'left' cluster

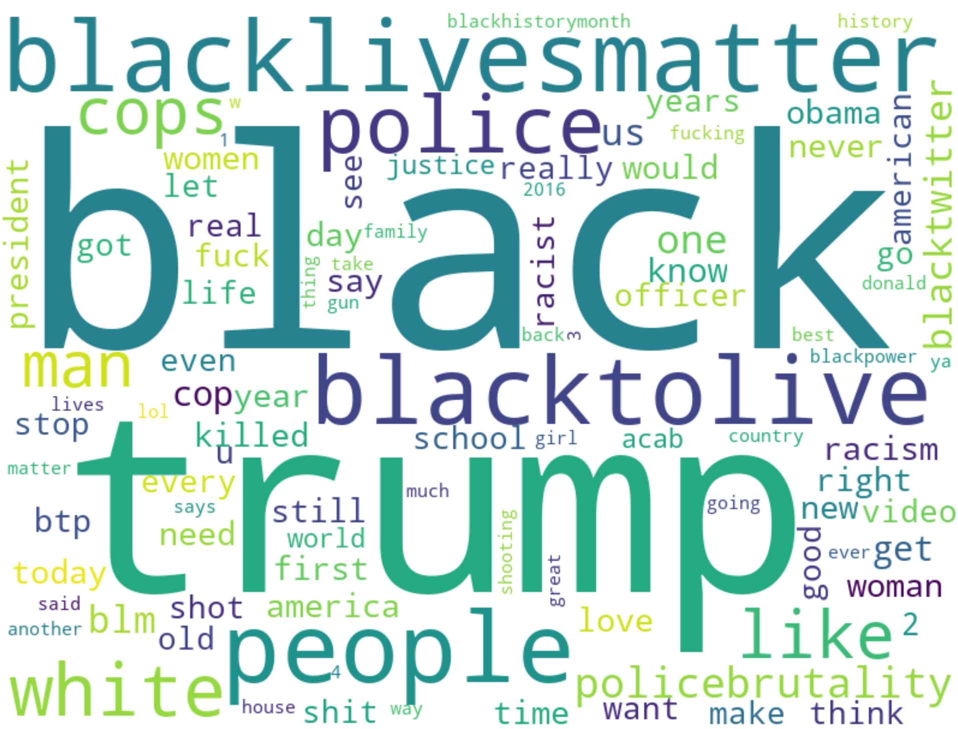


Figure 6: Wordcloud for news providers

Figure 7: Wordcloud 4

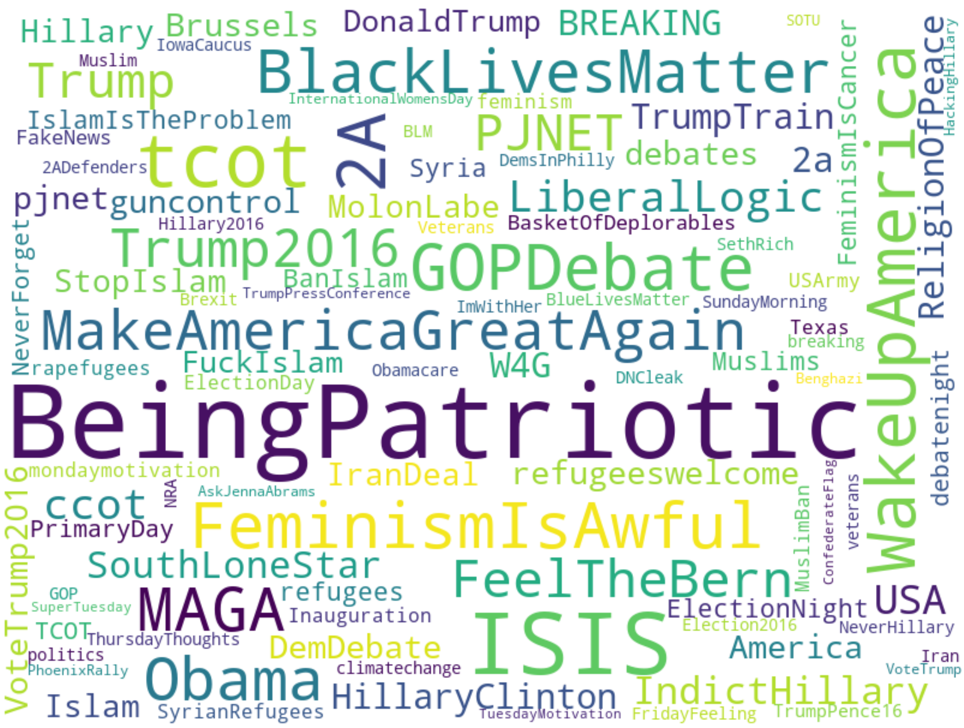


Figure 8: Wordcloud 5

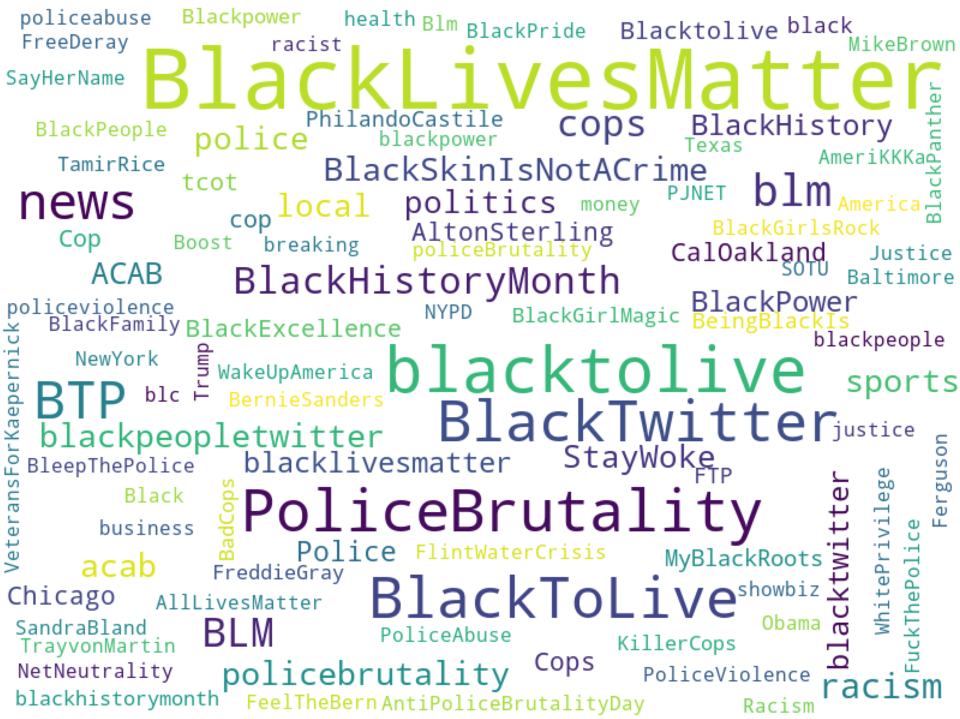


Figure 9: Wordcloud 6

Table 5: Most similar words to Trump

Skip Gram - Word	Similarity
donald	0.896358
trumps	0.791140
pres	0.770282
feedly	0.762091
illegitimate	0.755011
president	0.742422
trum	0.736769
administration	0.728688
donaldtrump	0.728006
rnc	0.725931

Table 6: Most similar words to Clinton

Skip Gram - Word	Similarity
hillary	0.935136
hillaryclinton	0.783674
clintons	0.781306
hrc	0.765259
campaign	0.741458
sanders	0.738335
emails	0.724602
crookedhillary	0.723004
hillarys	0.721376
billclinton	0.717722

Figure 10: Similarity of embedding: Liberals and Conservatives

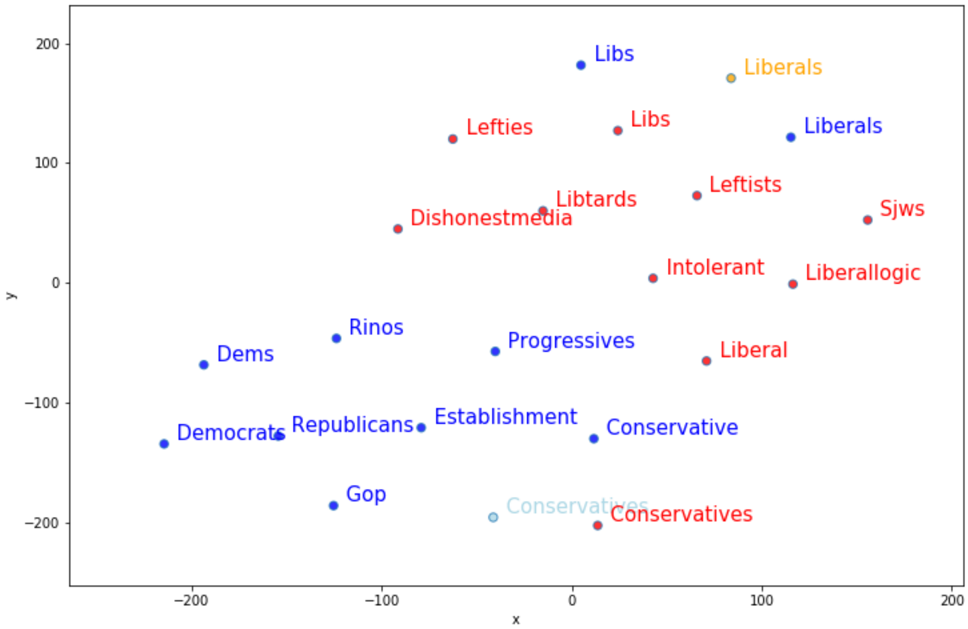


Figure 11: Similarity of embedding: Clinton and Trump

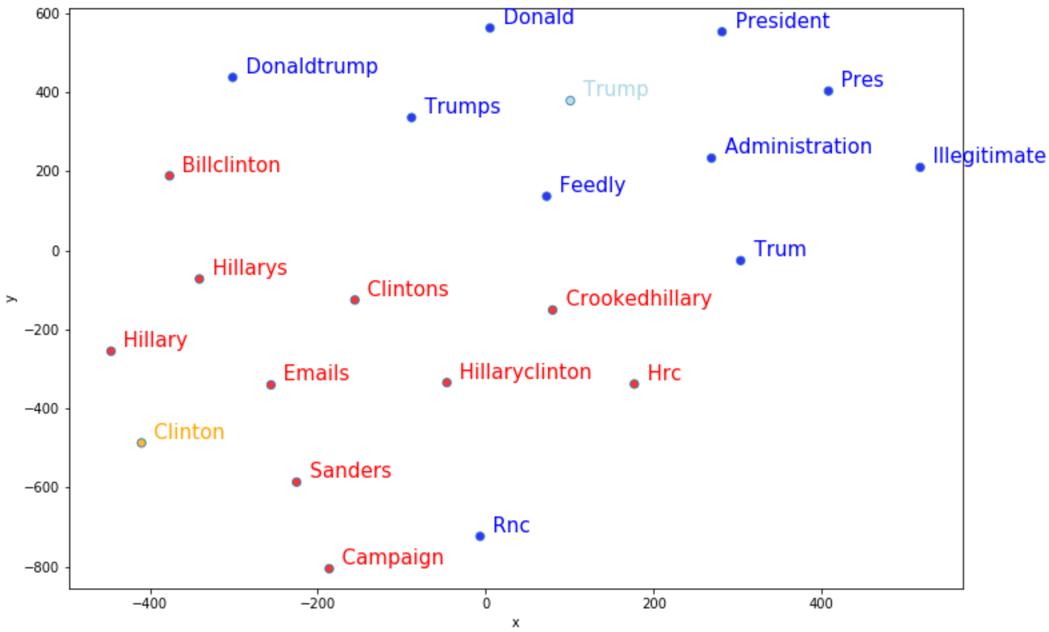


Figure 12: Similarity of embedding: Black and White

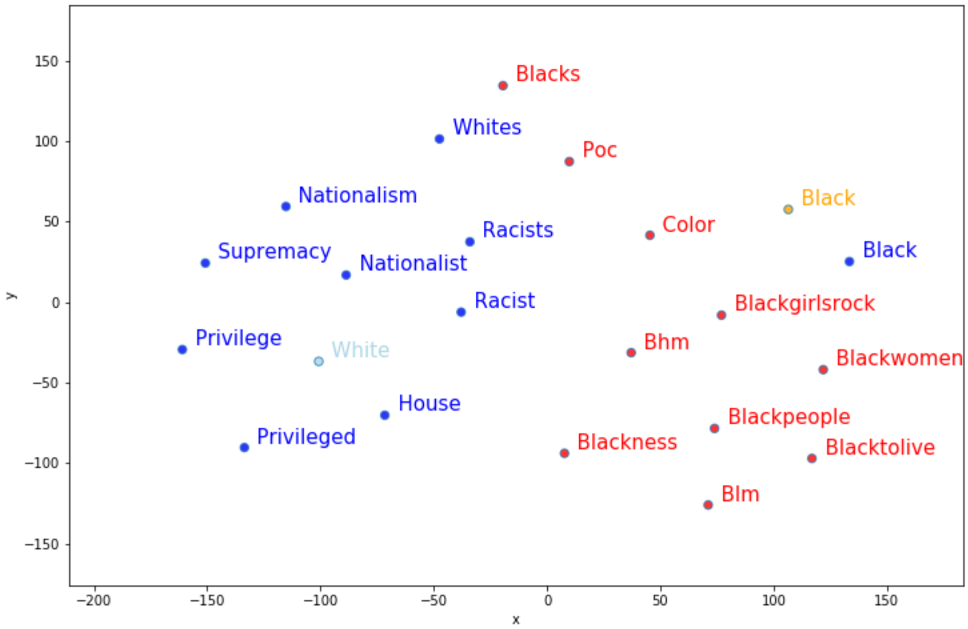


Figure 13: Similarity of embedding: Fake and True

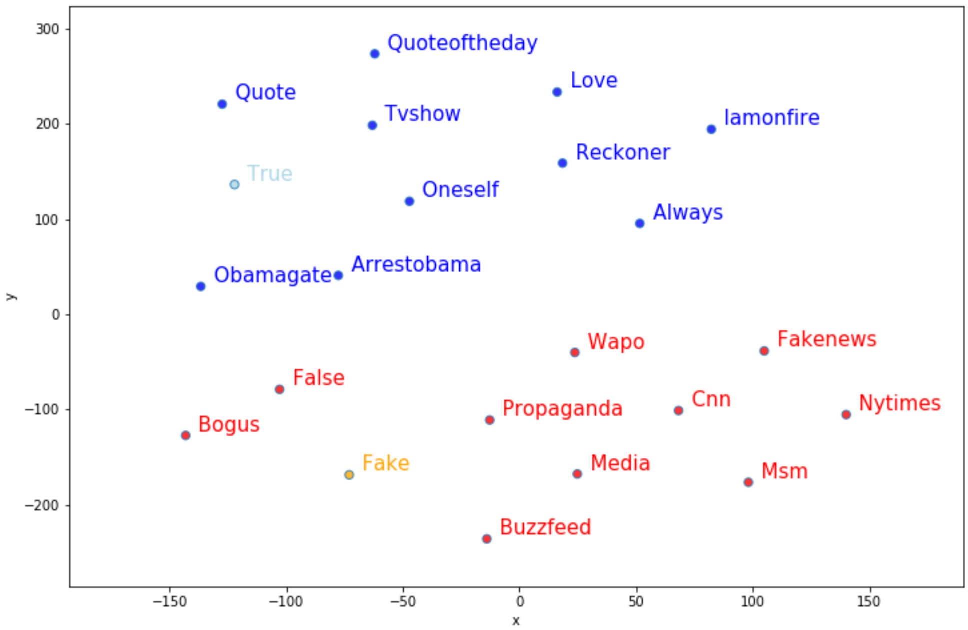


Table 7: Most similar words to Liberal

Skip Gram - Word	Similarity
liberallogic	0.760208
nutshell	0.755933
liberals	0.723288
msm	0.703458
hypocrisy	0.688949
communist	0.687983
leftists	0.671862
suppoers	0.671141
libs	0.670767
gopdebatesc	0.668273

Table 8: Most similar words to Conservative

Skip Gram - Word	Similarity
republican	0.736253
conservatives	0.693829
scprimary	0.680811
gop	0.673089
oppose	0.669254
cruz	0.656088
democratic	0.655370
democrats	0.654107
convention	0.653456
endorses	0.648152

Figure 14: Wordcloud of Noise Cluster identified by DBSCAN

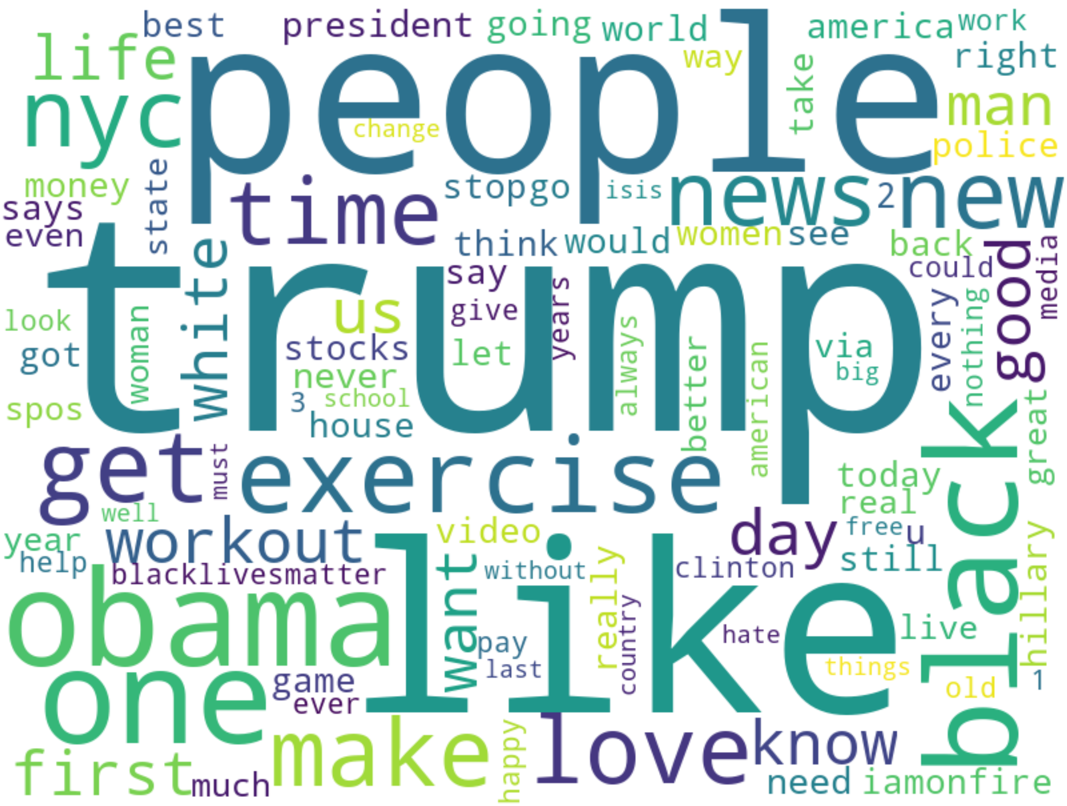


Figure 15: Wordcloud of Police Cluster identified by DBSCAN

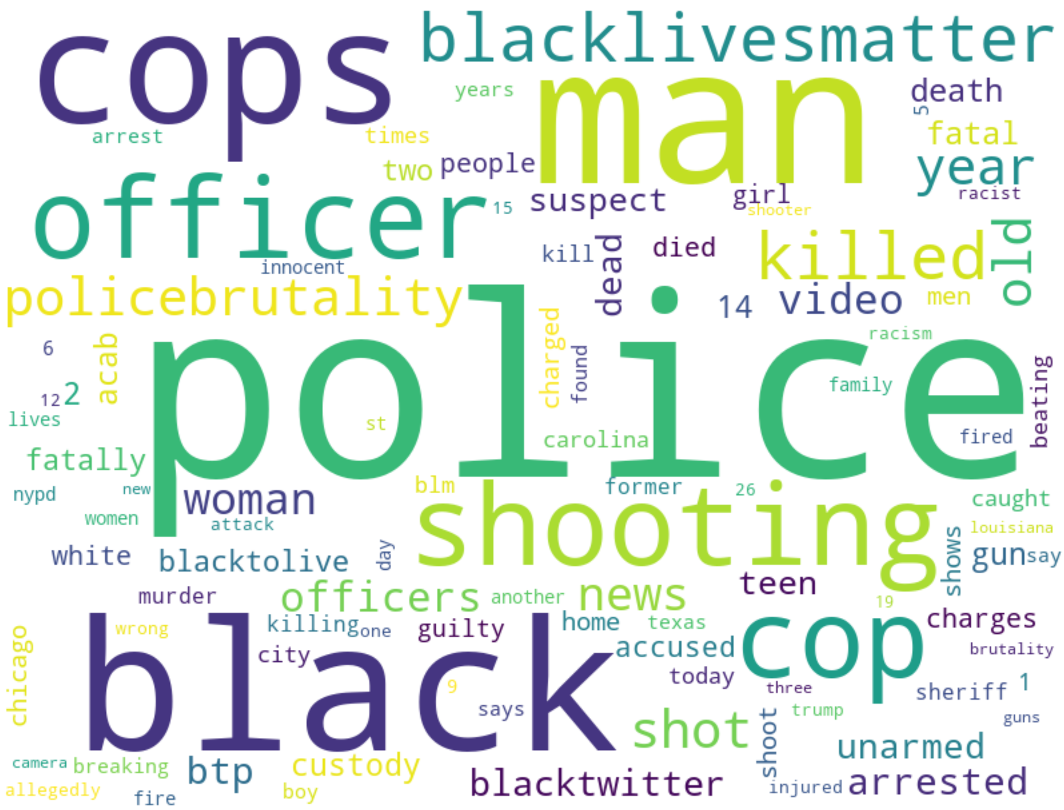


Figure 16: Wordcloud of Russia Investigation Cluster identified by DBSCAN

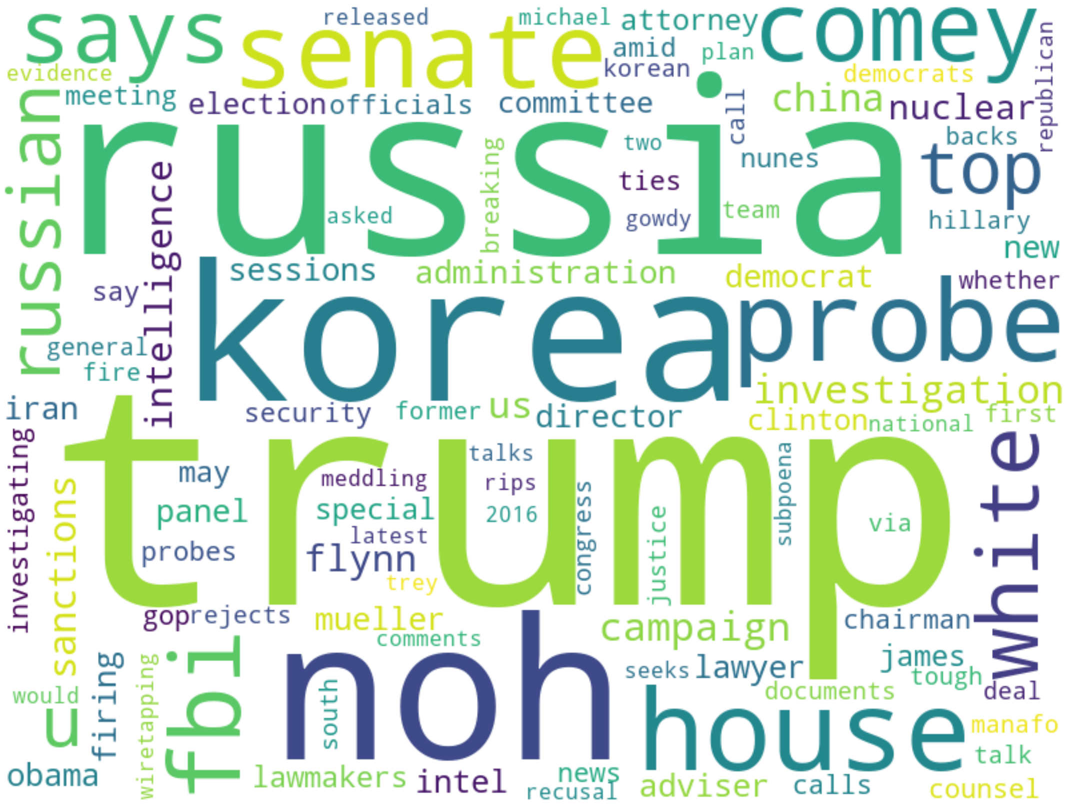


Figure 17: Wordcloud of Healthcare Cluster identified by DBSCAN

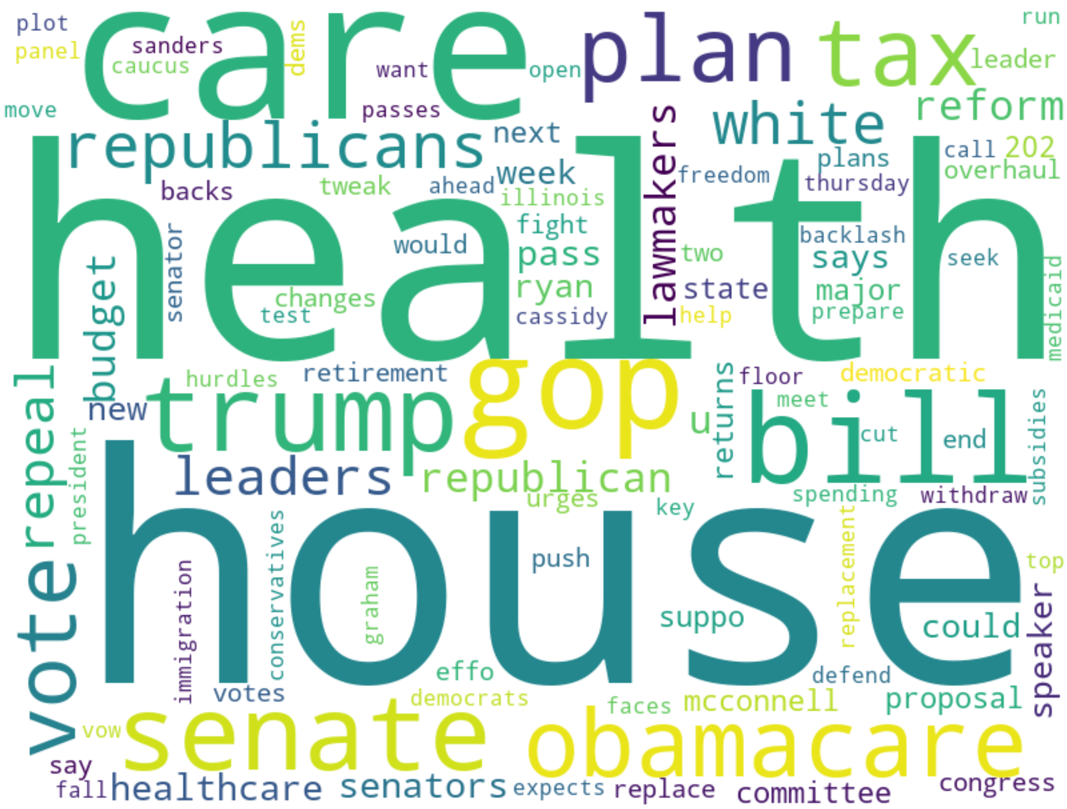


Figure 18: Wordcloud of Drug Cluser identified by DBSCAN