Capstone Project 2 - Final Report

A Few Months in the Troll Factory

An Analysis of Russian Troll Tweets in the 2016 US Election

Problem Statement

In the lead up to the 2016 US Election, and for a short time afterwards, social media users in he US were targeted by a disinformation campaign by a Russian "troll factory," designed to "sow disinformation and discord into American politics via social media."

> Earlier this year, as part of special counsel Robert Mueller's investigation, the Justice Department charged 13 Russian nationals with interfering in American electoral and political processes. The defendants worked for a well-funded "troll factory" called the Internet Research Agency, which had 400 employees, according to one Russian news report. From a bland office building in St. Petersburg, the agency ran a sophisticated and coordinated campaign to sow disinformation and discord into American politics via social media. This often involved Trump's favorite medium: Twitter. (via FiveThirtyEight.com, Why We're Sharing 3 Million Russian Troll Tweets, July 31, 2018)

I will use the Tweets to explore questions about the nature of the disinformation campaign, such as:

- Did the tweets increase in frequency or volume around the time of major events?
- Did other trolls retweet and amplify troll tweets?
- Can common topics or themes be identified?
- What were the most-used hashtags?
- Did the tweets predominantly support one candidate or political party, or seek to undermine the other?

By exploring the patterns, topics and methods of the disinformation campaign, I will seek to create insight into these efforts and understand how to recognize, identify and potentially avoid future disinformation attacks.

Client

My client for this project is the American voter, and I intend to provide analysis to aid them in discerning manufactured disinformation from "real" opinion and information.

Data

The data I will be using for this project is data that has been made available to the public by Five Thirty Eight, on their GitHub at https://github.com/fivethirtyeight/russian-troll-tweets/ .

The data was originally gathered by two professors at Clemson University; Darren Linvill and Patrick Warren, and shared with FiveThirtyEight.

> Using advanced social media tracking software, they pulled the tweets from thousands of accounts that Twitter has acknowledged as being associated with the IRA. The professors shared their data with FiveThirtyEight in the hope that other researchers, and the broader public, will explore it and share what they find. (via FiveThirtyEight.com, Why We're Sharing 3 Million Russian Troll Tweets, July 31, 2018)

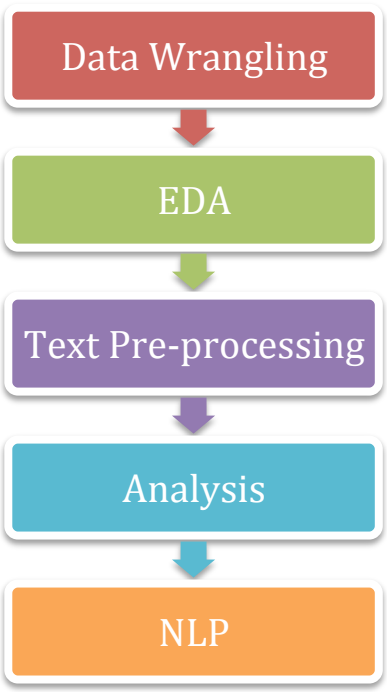
Data Dictionary

Header	Definition
`external_author_id`	An author account ID from Twitter
`author`	The handle sending the tweet
`content`	The text of the tweet
`region`	A region classification, as [determined by Social Studio](https://help.salesforce.com/articleView? id=000199367&type=1)
`language`	The language of the tweet
`publish_date`	The date and time the tweet was sent
`harvested_date`	The date and time the tweet was collected by Social Studio
`following`	The number of accounts the handle was following at the time of the tweet
`followers`	The number of followers the handle had at the time of the tweet
`updates`	The number of "update actions" on the account that authored the tweet, including tweets, retweets and likes
`post_type`	Indicates if the tweet was a retweet or a quote-tweet

`account_type`	Specific account theme, as coded by Linvill and Warren
`retweet`	A binary indicator of whether or not the tweet is a retweet
`account_category`	General account theme, as coded by Linvill and Warren
`new_june_2018`	A binary indicator of whether the handle was newly listed in June 2018

Process

The overall process for this project is listed below



Data Wrangling

The roughly 3 million-row data set results in a file size of 1.16 GB. Because this far exceeds GitHub's file size limit, the data was split into 13 separate CSV files. Code was written to iterate over each CSV file in the directory, and append each file into one CSV file. The single file was then read into a single dataframe containing all ~ 3 million rows of data.

Encoding

One challenge that was encountered during the initial Exploratory Data Analysis process was the encoding of the data set for dealing with Cyrillic characters and emojis embedded in tweet content.

When data was read in to the dataframe using Pandas read_csv() method, the data was read in as UTF-8 encoding, which then rendered the Cyrillic characters and emojis as unrecognizable text. After much research and troubleshooting, I discovered that the data had to be encoded as Unicode to be interpreted correctly. Apparently this is a common issue in NLP text pre-processing, and the Gensim NLP library provides a solution with their any2unicode class.

Data Cleaning

The data, having been prepared by the Clemson researchers, contained very few NA values. The empty fields were purposeful and did not represent "missing" values, but the valid absence of a value.

The feature 'harvested_date' was dropped, as it was unnecessary for the purpose of this analysis.

Other Potential Data Sets

If further analysis were to be undertaken, harvesting other relevant, but authentic, Twitter users who engaged in the same topics would add useful insights to the analysis. This would enable the comparison of content, topics, mentions of other users, and hashtags to the Russian Trolls to analyze the relationships between the troll users and authentic users.

Exploratory Data Analysis

Exploratory Data Analysis (EDA) was undertaken to examine the structure of the data and extract any initial findings and patterns in the data.

Initial Findings

The first area examined was around the Russian Troll authors (trolls) and quantity of tweets. The number of unique trolls was found to be 2,843, and those users accounted for 2.94 million tweets.

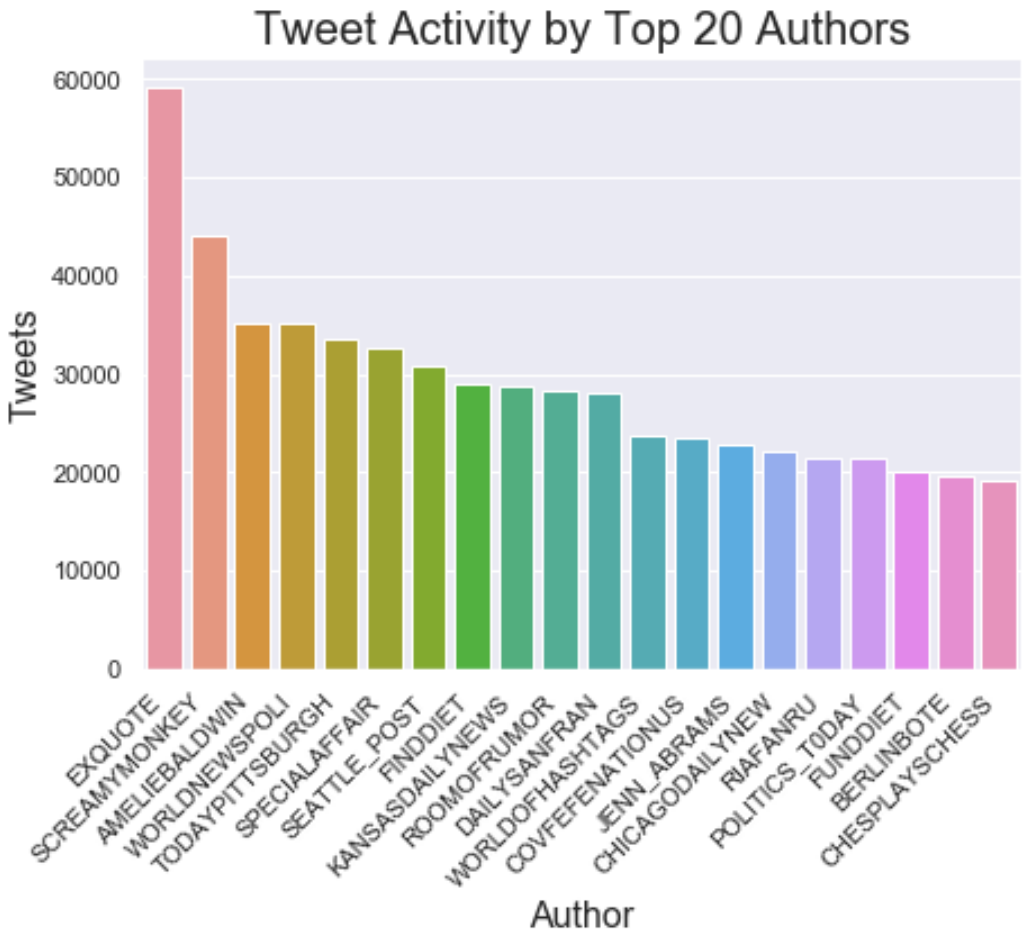
Most Active Trolls

The ten most active authors were responsible for ~350,000 tweets, accounting for only 14% of the total Troll tweets, indicating a more widespread use of trolls with tweets distributed across many users.

Top 10 Authors

Author/User	Tweet count	Followers	Following	Account Category
EXQUOTE	45,886	858	2	Commercial
SCREAMYMONKEY	44,001	13,217	14,046	NewsFeed
WORLDNEWSPOLI	41,257	1,902	4,753	RightTroll
AMELIEBALDWIN	35,261	2,473	1,929	RightTroll
TODAYPITTSBURGH	33,602	18,930	6,346	NewsFeed

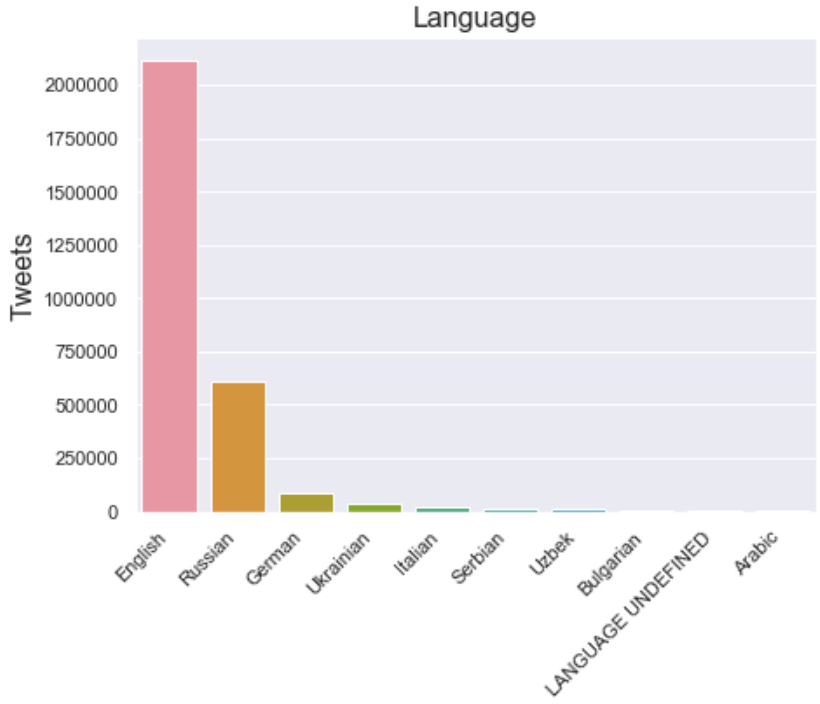
SPECIALAFFAIR	32,556	11,241	10,068	NewsFeed
SEATTLE_POST	30,793	17,188	5,252	NewsFeed
FINDDIET	29,034	400	4	Commercial
KANSASDAILYNEWS	28,806	25,368	5,171	NewsFeed
ONLINECLEVELAND	28,664	16,295	7,617	NewsFeed
Total	**349,860**	**107,872**	**55,188**	
Mean	**34,986.0**	**10,787.2**	**5,518.8**	



Language

The language that the troll tweets are written in are revealing with the majority of tweets being written in English, and Russian being the next most used language. Unfortunately, not understanding Russian, I cannot translate the Russian tweets to understand why Russian language tweets would have been used to influence an American electorate.

Tweets by Language



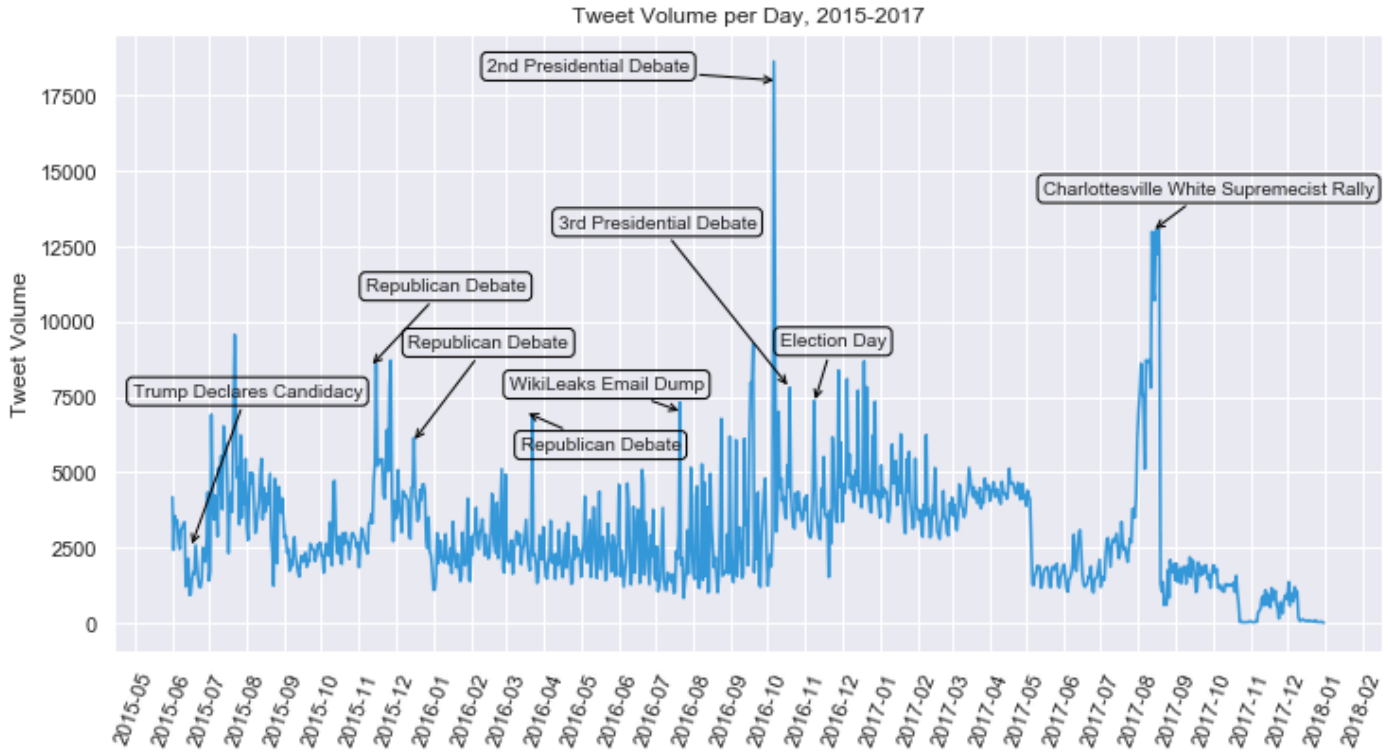
Timeline of Troll Activity

One of the primary areas of interest was the timing of the troll activity and how the activity aligned with major events in the election cycle.

While the data set provides activity from a time range from 02-02-2012 to 05-30-2018, because the primary interest is the 2016 election cycle, I chose a window of time beginning with June, 2015 and ending in January 2018, the time the troll accounts were identified as such.

In the timeline visualization below, significant events in the 2016 election were annotated to highlight the significant events on the timeline, such as party debates, presidential debates, and the Charlottesville white supremacist rally.

The timeline visualization clearly shows the spike in troll activity occurring on or in close to the day of the event, in an effort to influence or react to the event in question



Word Frequency

A basic analysis of the frequency of words occurring in the tweet content highlights the need for text pre-processing before perfoming any sort of Natual Language Processing (NLP). The most frequently occurring words are words that add no meaning to the content, and are ultimately useless to any text analysis or NLP. This analysis highlights the need for the removal of "stopwords" – the words that add no meaning. These stopwords will be removed in the text-preprocessing stage of the NLP modeling.

Natural Language Processing (NLP)

NLP is a type of data science that consists of processes for analyzing, understanding, and deriving information and meaning from the text data.

Revisiting some of the questions I seek to answer from the data, they are all the type of questions that can be answered by NLP:

- Did other trolls amplify troll tweets?
- Can common topics or themes be identified?
- What were the most-used hashtags?
- Did the tweets predominantly support one candidate or political party, or seek to undermine the other?

Text Pre-Processing

Before any form of NLP can be undertaken, text pre-processing must be undertaken. Text pre-processing consists of the following data processing techniques.

- Noise Removal
- Tokenization
- Normalization (i.e., stemming or lemmatization)

Tokenization is the process of converting a text into tokens – words or entities present in the text.

Text normalization includes:
- converting all letters to lower or upper case
- converting numbers into words or removing numbers
- removing punctuations and accent marks
- removing white spaces
- removing stop words, sparse terms, and particular words

Lemmatization
All of the processing steps listed above are common text processing and preparation steps, except for lemmatization.

The purpose of lemmatization is to reduce multiple forms of a word to a common base form. While the technique of stemming simply chops off word endings, lemmatization uses "lexical knowledge bases" to get the correct base forms of words. For example, for the words "studying" and "studies", stemming would produce "study" and "studi", respectively. Lemmatization would provide the base word (known as a lemma) "study" for both "studying" and "studies".

Hashtag Analysis
Analysis of the trolls' most-used hashtags provided limited amount of insight into the content posted by the trolls and the methods they used.

Hashtag	**Count**
News	128754
sports	48641
politics	39744
world	27558
local	25890
TopNews	15149
MAGA	14701
BlackLivesMatter	11950
health	11486
tcot*	11447

* tcot is "top conservatives on Twitter"

While there are hashtags that touch on some of the topics that were issues during the 2016 election, like “MAGA” or “BlackLivesMatter”, many of the hashtags are simply noise – like “local” and “news”.

Troll Amplification Analysis

One of the research questions I sought answers for was, are trolls amplifying other trolls with mentions or retweets of other trolls tweets. To determine this, I first looked at the top twitter user handles mentioned in the troll tweets.

User	Mentions
realDonaldTrump	13357
midnight	9094
YouTube	8713
POTUS	4880
HillaryClinton	4311
CNN	3081
FoxNews	2902
rus_improvisation	2485
TalibKweli	1852
WarfareWW	1609

I then compared the mentions in the tweet content to a list of the trolls, and found that there were 1,739 trolls who were amplifying other trolls’ messages by retweeting or mentioning them. This number was much lower than I had anticipated.

Latent Dirichlet Allocation (LDA) for Topic Modeling

There are several approaches for obtaining/modeling topics from a text, such as Term Frequency and Inverse Document Frequency (TF-IDF). Latent Dirichlet Allocation (LDA) is probably the most popular topic modeling technique, and that is the method I had chosen to use for topic modeling in this project.

LDA is used to classify text in a document and assign it to a particular topic. It builds a “topic per document” model and “words per topic” model, modeled as Dirichlet distributions.

“Each document is modeled as a multinomial distribution of topics and each topic is modeled as a multinomial distribution of words.” <Dirichlet distribution, Medium, https://towardsdatascience.com/dirichlet-distribution-a82ab942a879>

For this project, I chose to use the Gensim NLP library, which also offered other text pre-processing tools.

One main requirement for using LDA is that the data must be in a sparse vector, so we used the Bag-of-words document representation. To achieve this, each tweet's content must be cleaned of all hashtags and @ mentions, converted to a list, and then a dictionary, using Gensim's corpora library, and then finally to a bag-of-words document representation using the Gensim's gensim.corpora.Dictionary library's doc2bow() method.

The LDA model was then fit using Gensim's LdaMulticore(), to take advantage of multicore processors.

LDA Output - All Topics,

0 --- (0, '0.043*"police" + 0.039*"man" + 0.016*"woman" + 0.012*"shot" + 0.011*"ohio" + 0.011*"officer" + 0.010*"suspect" + 0.009*"school" + 0.009*"bill" + 0.009*"found"')

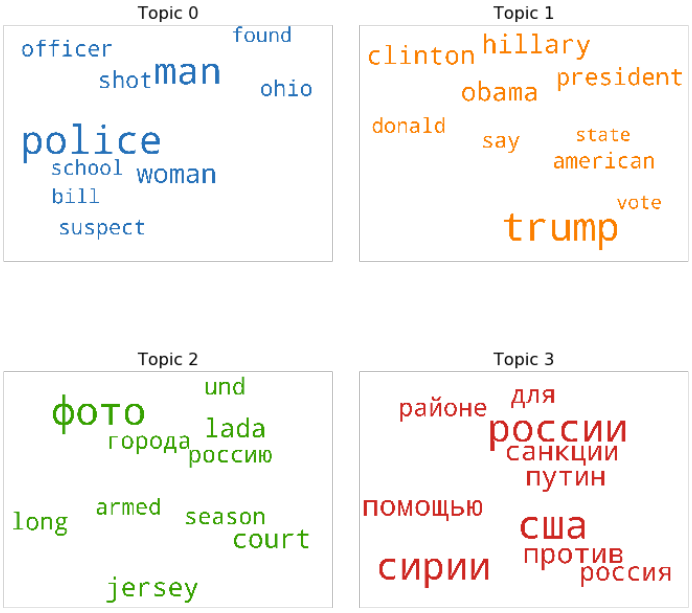
1 --- (1, '0.041*"trump" + 0.013*"obama" + 0.013*"hillary" + 0.013*"clinton" + 0.011*"president" + 0.009*"say" + 0.009*"american" + 0.008*"donald" + 0.007*"vote" + 0.007*"state"')

2 --- (2, '0.022*"фото" + 0.008*"court" + 0.008*"jersey" + 0.007*"lada" + 0.006*"long" + 0.006*"города" + 0.006*"und" + 0.006*"россию" + 0.006*"season" + 0.005*"armed"')

The output from the LDA can appear confusing at first, but is fairly straightforward. The first digit is the topic number. For example, the first topic in the sample output is Topic 0. Following that, the top 10 keywords that contribute to this topic are listed, along with the weight of keyword on the topic. So, for Topic 0 the keyword "police" has a weight of 0.043 on the topic. The weights reflect how important a keyword is to that topic.

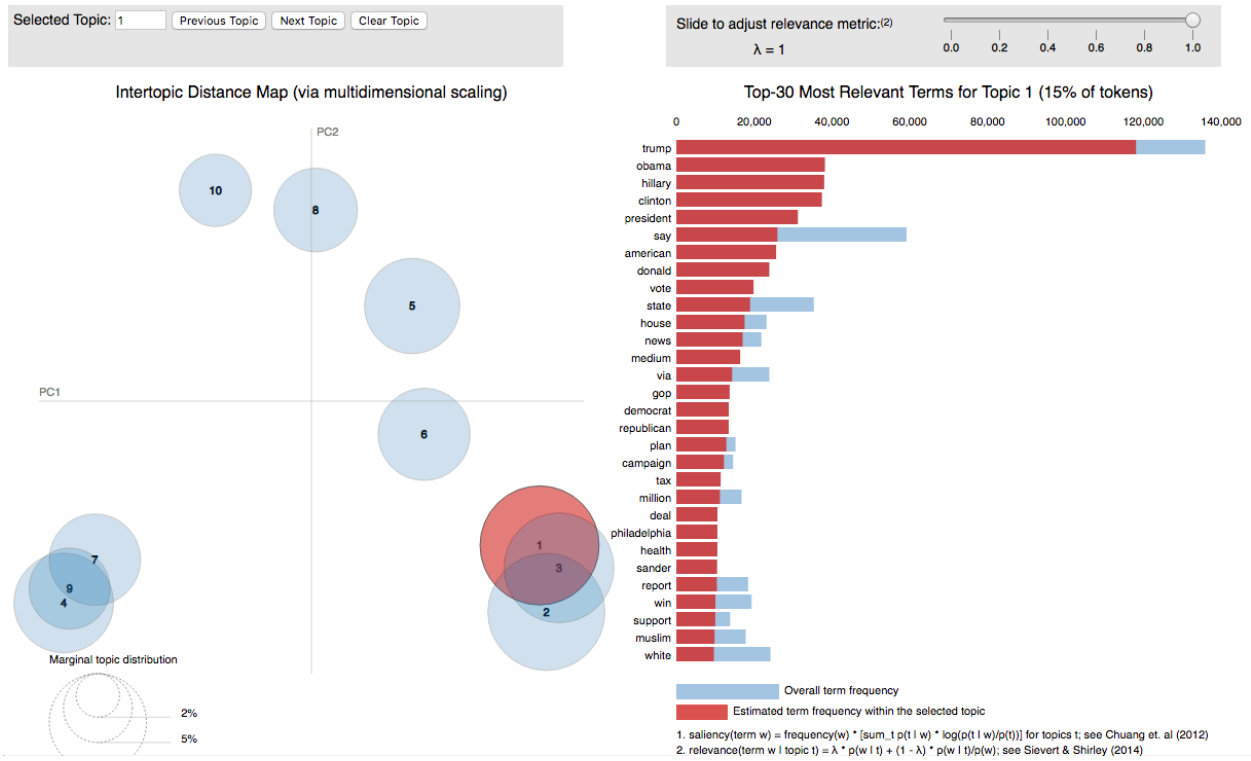
WordCloud of Topics

For visualization purposes, I also created a word cloud of the top four topics in the LDA model. The topics that the trolls are posting now become a bit more clear.



PyLDAvis

During my research, I discovered a library called PyLDAvis, which provided excellent visualization for LDA model output



On the left, the topics are plotted as circles, whose centers are defined by the computed distance between topics. The prevalence of each topic is indicated by the circle's size. On the

right, two overlaid bars showing the topic-specific frequency of each term (in red) and the corpus-wide frequency (in blue). When no topic is selected, the right panel displays the top 30 most salient terms for the dataset.

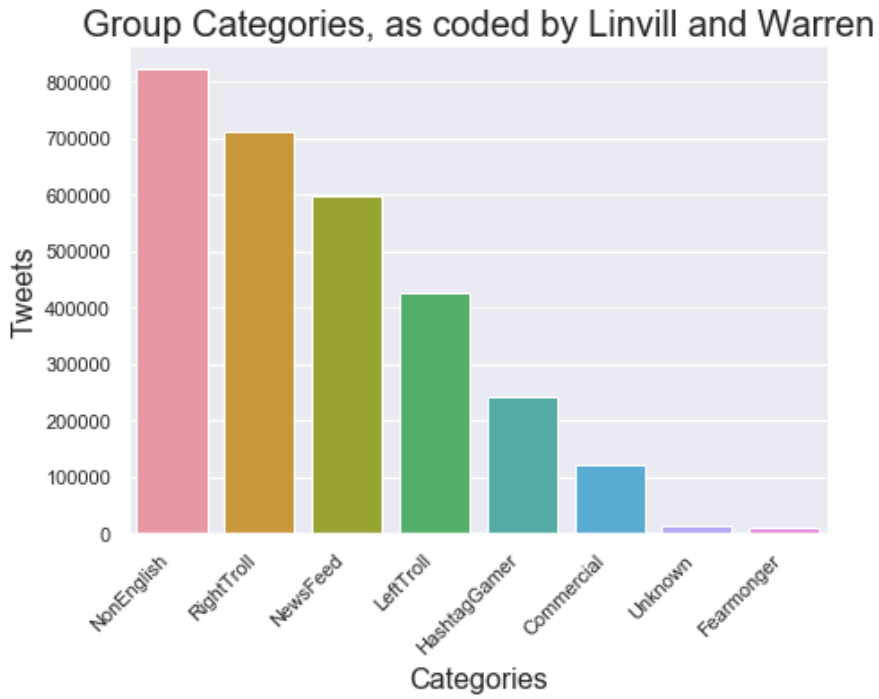
Troll Classifications

Given that the analysis of the overall tweets provided limited insight into the methods the trolls used, I turned to the five categories of "troll classes", as defined by Clemson University's Darren Linvill and Patrick Warren in their analysis of the Russian Troll Farm Twitter data, to analyze the hashtags, mentions, and topics of each of the different classes or groups.

They identified five categories of troll Twitter accounts, each with unique patterns of behaviors:

- **Right Troll**, spreading nativist and right-leaning populist messages. It supported the candidacy and Presidency of Donald Trump and denigrated the Democratic Party. It often sent divisive messages about mainstream and moderate Republicans.
- **Left Troll**, sending socially liberal messages and discussing gender, sexual, religious, and -especially- racial identity. Many tweets seemed intentionally divisive, attacking mainstream Democratic politicians, particularly Hillary Clinton, while supporting Bernie Sanders prior to the election.
- **News Feed**, overwhelmingly presenting themselves as U.S. local news aggregators, linking to legitimate regional news sources and tweeting about issues of local interest.
- **Hashtag Gamer**, dedicated almost exclusively to playing hashtag games.
- **Fearmonger**: spreading a hoax about poisoned turkeys near the 2015 Thanksgiving holiday.

(Troll Factories: The Internet Research Agency and State-Sponsored Agenda Building, Darren L. Linvill and Patrick L. Warren, July 2018, http://pwarren.people.clemson.edu/Linvill_Warren_TrollFactory.pdf)



Troll Group	**Group Tweets**
NonEnglish	820,803
RightTroll	711,666
NewsFeed	598,226
LeftTroll	427,141
HashtagGamer	241,785
Commercial	121,904
Unknown	13,539
Fearmonger	11,140

While all five groups, along with two others (Non-English and Commercial) were analyzed, I focused on Right Troll and Left Troll groups, as they best explained the strategy of the troll disinformation efforts.

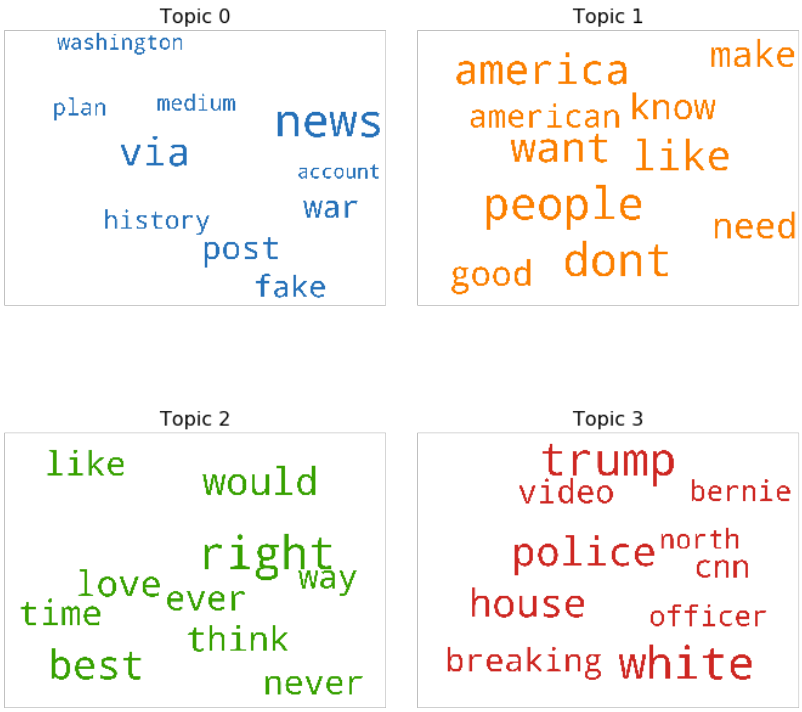
Right Troll Group

The Right Troll group appears to be strongly right-wing in their hashtag and topical themes, attacking Democrats, Barack Obama, Hillary Clinton, and those that supported them. They also attack Islam with hashtags like #IslamKills, while echoing Donald Trump's themes of #MAGA (Make America Great Again) and #2a (Second Amendment). However, the Right Troll group members also sometimes tweeted divisive messages about the Republican party as well.

Hashtag	**Count**
MAGA	14394
tcot*	10531

PJNET	10138
top	7416
news	6708
Mar	4865
topl	4399
FAKENEWS	4352
IslamKills	4209
2A+	3999
Trump	3728
WakeUpAmerica	3705
TCOT	3367
ccot	3335
GOPDebate	3210
amb	2914

* TCOT – Top Conservatives on Twitter
+ 2A – Second Amendment

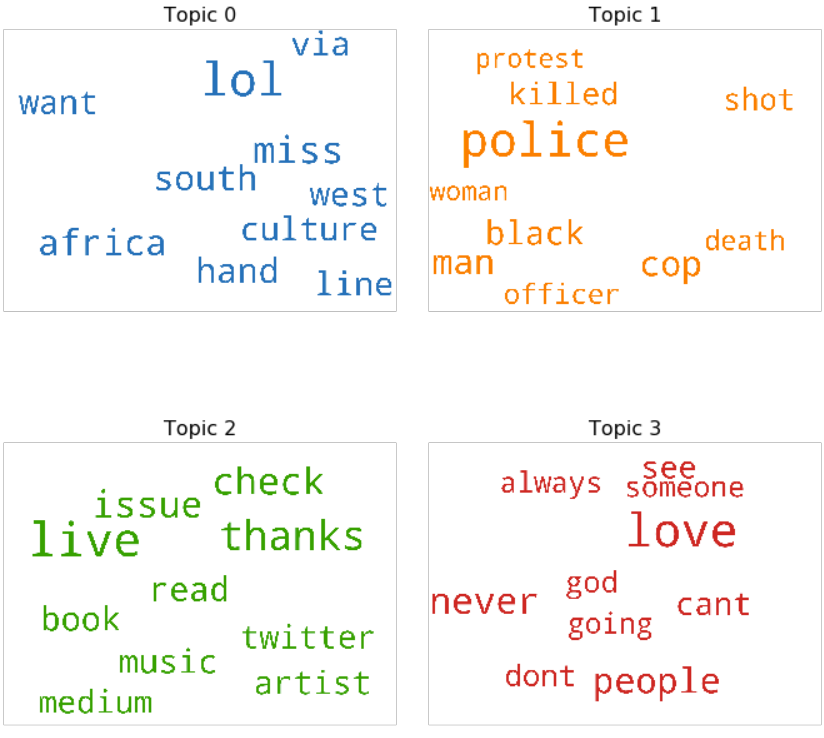


Left Troll Group

The Left Troll group played the foil to the Right Troll group, focusing on identity politics and the issues related to identity politics that were designed to provoke and inflame. The hashtags reflect the strategy of this group, with common hashtags like #BlackLivesMatter and #BlackSkinIsNotACrime. Topical themes focused on #PoliceBrutality, which was a central issues during this time period.

Much like the Right Troll group, the Left Troll group also sought to divide within the "left" by deploying divisive content pitting Bernie Sanders against Hillary Clinton,

Hashtag	Count
BlackLivesMatter	10550
NowPlaying	5241
BlackTwitter	2109
news	1779
PoliceBrutality	1615
blacklivesmatter	1595
StayWoke	1483
God	1457
BlackSkinIsNotACrime	1455
BLM	1422



Conclusions

The conclusions drawn from this analysis are less definitive than I had hoped, but several findings contradicted the assumptions held at the start of this analysis project.

I had expected a much larger "echo chamber" of trolls amplifying other trolls' messages. Analysis showed that this amplification was on a much smaller scale, with only 1,739 trolls amplifying other troll users.

The other primary finding that also contradicted my initial assumptions was the conclusion that the Russian trolls didn't appear to be supporting any one candidate, but more intent on sowing confusion, disinformation and divisiveness. As mentioned earlier, the Right Troll group made efforts to divide within the "right" of the political spectrum, and the Left Troll made efforts to divide within the "left" of the political spectrum, as evidenced by the Left Trolls pitting Bernie Sanders against Hillary Clinton.

However, it does appear that, as it became clear that Donald Trump was going to be the Republican nominee, the Russian Trolls aligned behind Trump and against Clinton. This aligns with the intelligence that has been reported, that Vladimir Putin sought to ensure that Hillary Clinton did not win the election.