The Internet, Policy & Politics Conference 2018 - Oxford Internet Institute

Protecting the Environment from Populism: Policy Implications Drawn from Sentiment Analysis of Trump Supporters' Tweets

Bushra Ferdous Khan and Abu Mohammad Shabbir Khan
North South University

Abstract

Democracy, as defined by Schumpeter, is a political system in which politicians campaign on policies that they think would best please their constituencies and earn their vote to place them in the coveted position of regulators. It is rule by the regulators to appease their voters: in democratic systems making rules or policies is almost exclusively the politicians' domain without much input from the common populace who lack the knowledge or incentives to do the mental labor necessary in this regard. Populism is a controversial flavor of democracy that is characterized by appeal to voters' emotions over reasons, and promotion of policies with ill-regard to their long-term impacts by demagogues. In this paper, we have studied the reaction of Donald J. Trump's supporters to his environmental policies, the extent to which they support him, and the themes that arise in their expressions of support by analyzing the tweets they wrote regarding those policies. Our results suggest the existence of a supporter base that shows an almost unfaltering level of support and while that supporter base is seen to converse about certain themes more frequently, there is no strong evidence to suggest that it actually cares about any theme more passionately than others.

I. The Dilemma

While the beauty of democracy lies in placing power in the hands of people to choose their leaders, it also comes with the threat of empowering a grotesque form of populism: a form of populism that caters to the present concerns of everyday men at cost of the collective's sustainable future. Democratically elected leaders are expected to cater to the needs of their base but the quickest possible way to fulfill such needs may not always lead to sustainable development outcomes. Time, however, is of the essence in electoral governance systems as democratically elected politicians are pressured to produce proof of their commitment to their base before the next round of elections. What do politicians in positions of power in the government, economic men like anyone else, who are only trying to make rational choices that best serve their interests do then? Do they sacrifice interests of future generations who they don't get to derive any benefit from to please their living-breathing constituencies who hold the power to place them in positions of power? We look into Twitter to find answer to this question.

II. The Broader Horizon: Democracy and Populism

Before we proceed to our main discussion, settling on common conceptualization of the key concepts this paper is based upon is critical.

Democracy, in this discourse, first, is accepted to be a competitive political system *"in which individuals acquire the power to decide by means of a competitive struggle for the people's vote"* as defined in 1942 by Joseph Schumpeter (Schumpeter, 2010). This particular definition of democracy, as argued in numerous literatures, stands in contrast with the classical notion which views democracy as a political apparatus for carrying out common will of the people for realizing common welfare. Schumpeter's argument, on the contrary, is that there is no common good as individuals differ on fundamental matters, and sheer application of reason does not necessarily lead to consensus about what is good for everyone (Schumpeter, 2010). Absence of common good, in effect, precludes presence of any common will of the people (Mackie, 2009; Schumpeter, 2010). Further, in his opinion, even if arrival at some agreement about what constitutes common good was feasible, electorates would still lack the knowledge or incentive of performing the mental labor[1] required for designing policies that will facilitate realization of the common welfare. Democracy, thus, he argued is much like a free market in which politicians compete for winning votes campaigning on the policy they think would appease their voters (Schumpeter, 2010).

Next, we need to operationalize the phenomenon we refer to as populism. Populism is a difficult concept to define, and there is a rich body of literature contesting its construct (e.g. Held, 1996;

---

[1] A major weakness of democracy, as argued by Schumpeter (1942) and Downs (1957), is that it does not offer sufficient incentive to voters to undertake the mental labor associated with learning about any topic in depth to vote in their own best interests. In Schumpeter's (1942) famous words: *"…the typical citizen drops down to a lower level of mental performance as soon as he enters the political field. He argues and analyzes in a way which he would readily recognize as infantile within the sphere of his real interests. He becomes a primitive again. His thinking becomes associative and affective"* (Schumpeter, 2010).

Taggart, 2000; Meny & Surel, 2002; Decker, 2003; Mudde, 2004; Karstev, 2008). Some categorically reject its acceptability (e.g. Enyedi, 2017; Wodak, 2017); some equate it with democracy (e.g. Fitzgibbon, 2017); while others provide qualified arguments in its favor (e.g. Mouffe, 2017; Stavrakakis, 2017; Rodrik, 2018). The term mostly, however, evokes negative connotations. Arguing in favor or against the different definitions of populism, or taking a position on the desirability or abhorrence towards populist occurrences is not the agenda of this paper. Rather, here we limit ourselves to only clarifying the sense the term is used to convey in this paper. Populism, in this discourse, refers to a political phenomenon in which a group of people claiming themselves to be the 'pure ones' express revolt against the oppression of 'corrupt elites' (Mudde, 2004) under the leadership of a demagogue who appeals to his electorates' emotions over reasons (Wilson, 2016). Demagogues connect to their base conflating contrasting issues to reinforce a sense of victimhood in his voters from which he promises deliverance from (Wilson, 2016). Encyclopedia Britannica's definition of populism, as quoted from Guiso, Herrera, Morelli, & Sonno (2017), summarises the key features populists are known by: *"populists claim to promote the interest of common citizens against the elites; they pander to people's fear and enthusiasm; and they promote policies without regard to the long-term consequences for the country."*

III. Narrowing the Scope Down: Contextualizing the Problem

Twitter, the micro-blogging-platform-in-chief, is widely known to be the favorite social medium of President Donald J. Trump to connect with his supporters. While much has been written in popular media about his prowess to connect to his base and attract attention through tweets, academic inquisition about the Trump-Twitter phenomenon is in limited supply. One of the two academic publications we reviewed on this topic offers an *"anthropological linguistic analysis of Twitter"* and investigates Trump's addresses to his social-media base as a speech practice (Stolee & Caton, 2018), while the other underscores how Twitter enables political dialogue that is *"simple, impulsive, and uncivil"* (Ott, 2017). It is argued that Twitter is the dwelling place of distracted minds as the platform, by its very design, is unsuitable for exchange of complex ideas, and demanding of simplistic messaging and superficial information processing (Ott, 2017; Kapko, 2016; Loh & Kanai, 2015; Carr, 2010).

In our paper, rather than studying Trump or his tweets, we study reaction of his base on Twitter to his environmental deregulations and policies. While the Trump Administration's many environmental deregulations and climate-change skepticism have been decried by almost every corner of civil societies-including by 100 members of United States Congress as recently as in last January (Congress of the United States, 2018), how are Trump supporters reacting to these?

Derision of the mainstream media or more educated cohorts of United States' voters (for a comparison of educational status of Trump supporters vs. Clinton supporters see Silver (2016)) or more traditional politicians did not stop Trump from winning the United States' Presidency. The group that discarded Trump's bid for Presidency as a bad joke is still aware of the ills of environmental mismanagement or the real threat climate change poses to United States and the world, but it is not their knowledge that mattered in 2016; what mattered was the sentiment of *"the forgotten men and women"* Trump gave a voice to and who in turn placed Trump in the Oval Office. For his base of economically disadvantaged populace, near term economic security is more important of a concern than sustainable development of the country (or of the world for that matter), and quite reasonably so. It is thus important to study this group's reaction to Trump's environmental deregulations and withdrawal of support to fight climate change to draw policy implications for future.

So in this paper we chiefly investigate two research questions:

RQ1: Are Trump voters completely supporting his environmental policies?

RQ2: How do Trump voters react to his environmental policies?

IV. Design, Data, and Dissection

IV.I Assigning Sentiments to Tweets

Since our core agenda in this paper is to draw implications for populism-proof environmental policy design, rather than creating better sentiment analysis models, we resorted to using two classifiers that have already established themselves for assigning sentiments to the tweets we collected. The first tool we used is the Valence Aware Dictionary for sEntiment Reasoning (VADER) classifier (Hutto & Gilbert, 2014) from Natural Language Toolkit (NLTK) (Bird, et al., 2009) and the second tool is the Naive Bayes (NB) classifier trained over the NLTK corpus of 2000 movie reviews (Pang & Lee, 2004) from Python library TextBlob (Loria, 2013).

The NB classifier is a popular and simple machine learning algorithm used in sentiment analysis. For computing the probability for a label given a set of features, this algorithm uses the Bayes theorem with *"the naive assumption that all features are independent"* (Bird, et al., 2009) and assigns the item being analyzed the label with the highest probability.

VADER, on the other hand, is a parsimonious rule-based model for general sentiment analysis. It contains a human validated sentiment lexicon of gold-standard quality built from drawing inspirations from well-established and extensively validated sentiment lexicons such as LIWC (Pennebaker, et al., 2007; Pennebaker, et al., 2001), ANEW (Bradley & Lang, 1999), and GI (Stone, et al., 1966) and incorporating Western-style emoticons (e.g. ":-)" that denotes a *"smiley face"*), *"sentiment-related acronyms and initialisms (e.g., LOL and WTF are both sentiment-laden initialisms), and commonly used slang with sentiment value (e.g., "nah", "meh" and "giggly")"* (Hutto & Gilbert, 2014). In addition, it combines *"these lexical features with consideration for five generalizable rules that embody grammatical and syntactical conventions that humans use when expressing or emphasizing sentiment intensity"* (Hutto & Gilbert, 2014). On a test of determining whether tweets had positive, neutral, or negative overall sentiments, VADER was shown to have significantly outperformed individual human raters as well as different NB classifiers trained on different corpora (Hutto & Gilbert, 2014).

Thus it is reasonable to presume that VADER is a better tool for tweet sentiment analysis compared to the NB classifier. However in certain cases, it gets confounded and fails to assign any sentiment to a statement. For example, "*#Trump vows to sue all female accusers as 11th woman steps forward* | *#FoxNews2016 http:// fxn.ws/2ebCGwB pic.twitter.com/7hKJaLZsIx*" and "*Large crowd outside #Trump rally in #Everett WA chanting racist-sexist-anti gay-Donald-Trump-go-away... pic.twitter.com/RqDpYWXB7U*" are samples of two negative tweets from our composed dataset where VADER failed to assign any sentiment, but the NB classifier successfully identified the negativity.

To leverage the strengths of both of these sentiment classifiers, we first checked the classification by VADER, and then by the NB classifier if VADER failed to decidedly assign any sentiment to a tweet. Specifically, if VADER assigned a polarity score smaller than or equal to -0.05 or a score greater than or equal to 0.05 to a tweet, then we regarded the tweet to have negative or positive sentiment respectively. On the other hand, if VADER assigned a polarity score between -0.05 and 0.05 to a tweet, then we utilized the NB classifier. In this scenario, we defined the difference between the positive and negative sentiment score by the NB classifier to be a tweet's polarity score. We then regarded a tweet to be negative, neutral, or positive if its polarity score was below or equal to -0.05, between -0.05 and 0.05 (exclusive), or above or equal to 0.05 respectively.

IV.II Creating List of Trump supporters

As our primary interest was to investigate the reaction of Donald Trump's base to his environmental deregulations and policies, a major task for us was to create list a of genuine Trump supports who had likely voted for him in the 2016 presidential election. Here, we put emphasis on finding out potential Trump voters as these are the actual people Trump is trying to appease.

While Trump's Twitter account has over 51 million followers (as of August 2018), a lot of those followers were simply of no interest to us as many of them are non-Americans and many of them are general Americans just following his accounts for updates. In addition, there exists the possibility that many of Trump's Twitter followers are fake accounts (Harris, et al., 2018).

So we resorted to programmatically collect tweets from 2016 with the assumption that actual Trump voters would have tweeted something regarding him in the election year. Specifically, we searched Twitter for tweets containing the hashtags #MakeAmericaGreatAgain, #MAGA, #trump2016, #donaldtrump, #trump, and #trumptrain over the timeframe between January 01, 2016 and December 31, 2016 while filtering out retweets. It should be noted that the case of the letters does not matter when searching for hashtags on Twitter.

Make America Great Again was the slogan of Trump's election campaign, and we posited that the hashtags #MakeAmericaGreatAgain and its abbreviation, #MAGA, would be popular amongst his supporters. After querying Twitter for tweets containing any or both of these hashtags, we collected 34,628 tweets from 4,991 unique user accounts. On the other hand, we found it reasonable to presume that ardent Trump supporters would also use the hashtags #trump2016, #donaldtrump, #trump, or #trumptrain in their tweets. After searching Twitter for tweets containing any combination of these hashtags, we gathered 88,783 tweets from unique 19,447 user accounts. After concatenating the two search results, we had a total of 118,308 non-duplicate tweets from 22,510 unique users. In this regard, two tweets were understood to be duplicate of each other if they had the same tweet id. If two tweets contained the same text but were tweeted by two different users, then we considered them to be non-duplicate. Additionally, if the same text was tweeted by a single user at two different times, then we still considered the tweets to be non-duplicates. The reasoning behind considering the tweets with the same text from different users or tweeted at different times from the same user non-duplicate was that these tweets were not shared by the simple use of the retweet button and likely exposed the information contained in their texts to different sets of users.

Now, as it is quite common for non-Trump supporters to also use these hashtags for criticizing Trump or making other statements, we carried out some further processing to obtain the list of genuine Trump supporters. At first, we only considered users whose tweets appeared in both of our search results with the assumption that dedicated Trump supporters would use both sets of our searched hashtags in their tweets. This gave us a list of 1,924 unique users with 63,377 tweets. Next, we assigned sentiment scores to all these tweets and calculated what percentages of all the tweets of each user were negative, neutral, and positive.

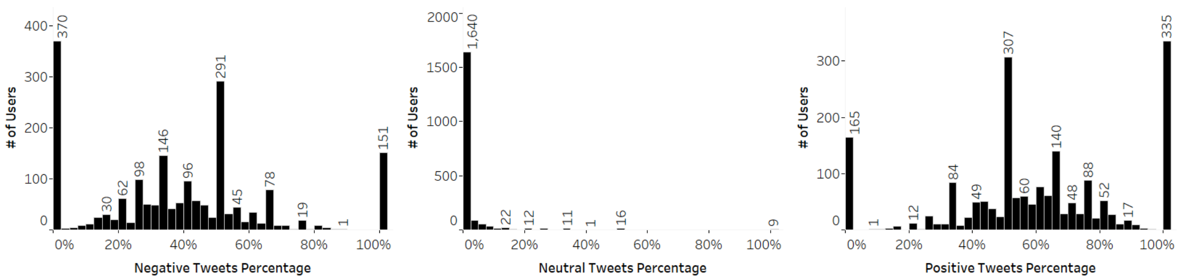

Figure 1: Distribution of users according their negative, neutral, and positive tweets percentage

Figure 1 shows the distribution of users according to what percentages of their tweets were negative, neutral, or positive. Every bin of each of the subplots has a width of 2.5%, and it includes its lower boundary but not its upper boundary in its count. Also, note that the vertical axes of the subplots are not aligned. It can be seen that a certain group of users tweeted more negative tweets while another group of users tweeted more positive tweets with our hashtags of interest. In addition, of the 1,924 users 1,640 had neutral tweets below 2.5%.

To gain a better understanding of the user distribution, we summed the neutral and positive tweet percentages of each user to get the non-negative tweet percentage of that user. Figure 2 depicts the distribution of users for their non-negative tweets percentages. As we found spikes in [50%, 52.5%) bins for both negative (left subplot of Figure 1) and non-negative tweets (Figure 2) percentages, we found it reasonable to consider users with more than or equal to 52.5% non-negative tweets with our searched

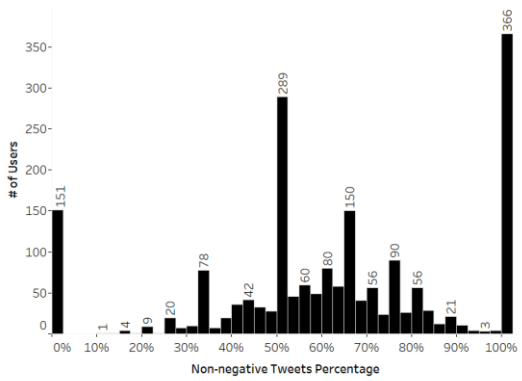

Figure 2: Distribution of users according their non-negative tweets percentage

hashtags to be actual Trump supporters based on the assumption that the true supporters would not use our hashtags of interest in a negative manner in the majority of their tweets. This consideration reduced our previous list of potential Trump supporters from 1,924 users with 63,377 tweets to 1,186 users with 41,289 tweets. As the tweets in these dataset were actually made during the 2016 election year and precede major policies taken by the Trump administration, in our further work we disregarded the content of these tweets and just utilized the list of 1,186 potential Trump supporters that contained their Twitter user handles and user names.

IV.III Creating Dataset of Tweets Related to Trump Administration's Environmental Policies

For studying the reaction of Donald Trump's supporter base to his administration's environmental deregulations and polices, we looked at tweets that were made in the vicinity of two major events, namely Trump's announcement of withdrawal from the Paris Agreement on 01 June 2017 (Milman, et al., 2017) and his signing of the executive order to approve the construction of Dakota Access and Keystone XL pipelines on 24 January 2017 (Smith & Kassam, 2017). In addition, we studied tweets that were made regarding coal and the Clean Power Plan over a course of 10 months.

As the announcement of withdrawal from the Paris Agreement and the executive order for approving Dakota Access Pipelines were discrete events, we collected the tweets regarding these events that were made over the course of eight days from the date prior to the occurrence of each of these events. Concretely, for the Paris Agreement, we collected tweets made at any time since 31 May 2017 but before 08 June 2017; and for Dakota Access and Keystone XL pipelines, we collected tweets made at any time since 23 January 2017 but before 31 January 2017. The reason for including the days previous to the dates of the events was that news about the possibility of occurrence of each of these events were circulating from the day before (Shear & Davenport, 2017; Garcia, 2017), and we wanted to capture the early reactions as well.

On the other hand, since coal and the Clean Power Plan have been discussed more or less continuously since the Trump administration came to power, we collected tweets regarding them made at any time since 09 October 2017 but before 23 August 2018. We chose 09 October 2017 as the starting date because on that day Scott Pruitt, the then administrator of the U.S. Environmental Protection Agency, announced that he will sign on a proposal to repeal the Clean Power Plan (Dennis & Eilperin, 2017), and it was one of the first official administrative actions taken against the Clean Power Plan.

It should be noted that because we were collecting tweets over a large time period for coal and the Clean Power Plan, we filtered out retweets in our search. However, as we were collecting tweets over quite a short time frame for the Paris accord and Dakota Access and Keystone XL pipelines, we searched Twitter both with and without filtering out retweets for these topics.

The specifics on what string we used to search for tweets, whether we filtered out retweets or not, what was the time range of the creation of the tweets, and how many tweets we collected from how many unique users for each search string is given in Table 1.

Topic	Search String	Retweets Filtered	Tweets Since (YYYY-MM-DD)	Tweets Before (YYYY-MM-DD)	Total Tweets	Unique Users
Paris Accord	*paris accord*	Yes	2017-05-31	2017-06-08	3530	1981
	paris accord	No	2017-05-31	2017-06-01	311	258
	paris accord	No	2017-06-01	2017-06-02	1228	876
	paris accord	No	2017-06-02	2017-06-03	1268	892
	paris accord	No	2017-06-03	2017-06-04	434	346
	paris accord	No	2017-06-04	2017-06-05	197	175
	paris accord	No	2017-06-05	2017-06-06	157	132
	paris accord	No	2017-06-06	2017-06-07	146	136
	paris accord	No	2017-06-07	2017-06-08	127	112
	paris agreement	Yes	2017-05-31	2017-06-08	4508	2218
	paris agreement	No	2017-05-31	2017-06-01	560	429
	paris agreement	No	2017-06-01	2017-06-02	1605	994
	paris agreement	No	2017-06-02	2017-06-03	1483	1000
	paris agreement	No	2017-06-03	2017-06-04	411	321
	paris agreement	No	2017-06-04	2017-06-05	199	157
	paris agreement	No	2017-06-05	2017-06-06	247	203
	paris agreement	No	2017-06-06	2017-06-07	180	155
	paris agreement	No	2017-06-07	2017-06-08	140	113
	climate	Yes	2017-05-31	2017-06-08	10794	4693
	climate	No	2017-05-31	2017-06-01	1254	874
	climate	No	2017-06-01	2017-06-02	2879	1682
	climate	No	2017-06-02	2017-06-03	3132	1912
	climate	No	2017-06-03	2017-06-04	1358	908
	climate	No	2017-06-04	2017-06-05	805	564
	climate	No	2017-06-05	2017-06-06	659	494
	climate	No	2017-06-06	2017-06-07	646	491
	climate	No	2017-06-07	2017-06-08	532	407
Dakota Access and Keystone XL Pipelines	*dakota pipeline*	Yes	2017-01-23	2017-01-31	778	530
	dakota pipeline	No	2017-01-23	2017-01-24	57	55
	dakota pipeline	No	2017-01-24	2017-01-25	403	317
	dakota pipeline	No	2017-01-25	2017-01-26	211	176
	dakota pipeline	No	2017-01-26	2017-01-27	100	92
	dakota pipeline	No	2017-01-27	2017-01-28	75	66
	dakota pipeline	No	2017-01-28	2017-01-29	59	57
	dakota pipeline	No	2017-01-29	2017-01-30	52	47
	dakota pipeline	No	2017-01-30	2017-01-31	69	61
	keystone	Yes	2017-01-23	2017-01-31	897	627
	keystone	No	2017-01-23	2017-01-24	59	57
	keystone	No	2017-01-24	2017-01-25	466	350
	keystone	No	2017-01-25	2017-01-26	233	195
	keystone	No	2017-01-26	2017-01-27	124	109
	keystone	No	2017-01-27	2017-01-28	105	96
	keystone	No	2017-01-28	2017-01-29	59	59
	keystone	No	2017-01-29	2017-01-30	58	54
	keystone	No	2017-01-30	2017-01-31	73	68
	#dapl	Yes	2017-01-23	2017-01-31	277	199
	#dapl	No	2017-01-23	2017-01-24	56	55
	#dapl	No	2017-01-24	2017-01-25	158	121
	#dapl	No	2017-01-25	2017-01-26	97	86
	#dapl	No	2017-01-26	2017-01-27	79	73
	#dapl	No	2017-01-27	2017-01-28	53	49
	#dapl	No	2017-01-28	2017-01-29	50	45
	#dapl	No	2017-01-29	2017-01-30	55	48
	#dapl	No	2017-01-30	2017-01-31	43	41

Topic	Search String	Retweets Filtered	Tweets Since (YYYY-MM-DD)	Tweets Before (YYYY-MM-DD)	Total Tweets	Unique Users
Coal and Clean Power Plan	*clean power plan*	Yes	2017-10-09	2018-08-23	909	496
	Coal	Yes	2017-10-09	2018-08-23	24153	9757

Table 1: Search details for the collection of Tweets

We thus collected 68,598 tweets. After dropping duplicate tweets in the same manner as before, we had a list of 46,023 tweets from 17,114 unique users. We then filtered the users and only took the ones whom we had previously identified as potential Trump supporters. We found that only 306 of the 17,114 users were in our potential Trump supporter list, and they had made 1,762 tweets in total.

We then collected the locations of the 306 potential Trump supporters from their Twitter profiles. As providing location information is optional in Twitter, we were not able to collect the locations of many users. Moreover, we found that many users had provided ingenuine locations like "*On your tv/radio/podcast*", "*Living rent free in your head*", "*Founder of @electricandrose*" etc. in their profiles. The valid geographical locations were also not in a uniform format as different users disclosed different level of details about their locations in varying manners while some users listed multiple locations. So we manually annotated the state of abode of each of these users when the state name was inferable from the provided location information.

While annotating the state information, we found that of the 306 user accounts, at least 7 belonged to non-Americans, 8 belonged to news and media organizations, and 1 belonged to Donald Trump himself. Moreover, 2 accounts had anti-Trump user handles and mostly posted sarcastic tweets. So we removed these 18 accounts and their tweets from our list because they were not actual Trump voters. We, thus, ended up with a list of 1,359 tweets from 288 user accounts.

Next, we assigned sentiment scores to the 1,359 tweets and found that 783 of the tweets had positive sentiments, 564 had negative sentiments and 12 had neutral sentiments. Upon manual inspection of the tweets, we found that the content of some tweets were directed towards conservatives and/or Donald Trump while some were directed towards liberals. From this, we understood that users were either supporting or opposing the Trump administration in their tweets by writing tweets of different sentiments for different audiences. We assumed that support and opposition could be modelled as binary classes S and O respectively, and the classes could be determined according to Table 2.

		Directed towards	
		Conservatives/ Donald Trump	**Liberals**
Sentiment	**Positive**	S	O
	Negative	O	S

Table 2: Coding support and opposition

Now, to figure out the classes of the tweets in our dataset, we first discarded the tweets that had neutral sentiment. This gave us 1,347 tweets from 286 users. We then provided the 1,347 tweets with their text content and sentiment as their only attributes to two of our colleagues who had good understanding of the American political system and knowledge of Trump administration's environmental policies and requested them to act as raters for independently classifying the tweets for us in accordance with the rule in Table 2. We asked them to use the sentiment assignment we provided rather than inferring sentiment on their own and to classify a tweet as I (Indeterminate) if they were not certain whether the tweet was supporting or opposing the Trump administration. The distribution of their classifications is shown in Table 3.

		Rater 1			
		S	**O**	**I**	**Row Marginals**
Rater 2	**S**	1,021	8	20	1,049
	O	5	190	17	212
	I	6	9	71	86
Column Marginals		cm^1 = 1,032	cm^2 = 207	cm^3 = 108	n = 1,347

Table 3: Class distributions among raters

To find the interrater reliability of the raters, we calculated Cohen's Kappa κ (Cohen, 1960) for Table 3 using the formula

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}$$

where $\Pr(a)$ is the actual observed agreement between the raters (0.95) and $\Pr(e)$ is the chance agreement calculated by

$$\Pr(e) = \frac{\left(\frac{cm^1 \times rm^1}{n}\right) + \left(\frac{cm^2 \times rm^2}{n}\right) + \left(\frac{cm^3 \times rm^3}{n}\right)}{n}$$

Cohen's Kappa came out to be 0.87 which represented *strong* agreement (McHugh, 2012) between the raters. We further calculated the confidence interval (CI) for κ by ($\kappa - 1.96 \times SE_\kappa$, $\kappa + 1.96 \times SE_\kappa$) at 95% confidence level where SE_κ is the standard error given by the formula

$$SE_\kappa = \sqrt{\frac{\Pr(a) \times (1 - \Pr(a))}{n \times (1 - \Pr(e))^2}}$$

We found the CI for κ to be (0.84, 0.90). The substantial inter-rater agreement led us to conjure that we could reliably utilize classifications our raters mutually agreed upon.

As we were only interested in tweets that demonstrate support for or opposition to Trump's policies by his follower base, we discarded the tweets that were not classified to be in the S or O classes by both of our raters. This led us to having a dataset of 1,211 tweets (1,021 S class and 190 O class) from 272 users. After sorting the dataset by the tweet class and user name and manually inspecting the O class tweets, it became apparent to us that the majority of the O class tweets were actually being made by a group of non-Trump supporters. For example, the tweets "*He doesn't care about the arts, healthcare, immigrants, refugees, equal pay, a living wage, the environment or the poor. #Damn #ParisAccord*" and "*Climate change IS real you horrible, HORRIBLE man! #HumanityFail #LordHelpUsAll*" were from the same user. So for each of 272 users, we calculated what percentage of her tweets in our 1,211 tweet dataset was in O class. We found that for 34 users, more than 66% of their tweets were in the O class. As a final refinement of our dataset, we decided to remove these 34 users and their tweets, and ended up with 1018 tweets (1,011 S class and 7 O class) from 238 unique users whom we could reasonably believe to be actual Trump supporters who voted for him in the 2016 presidential election. Here, we note that the other tweets of the users who made the 7 O class tweets demonstrated actual support for Trump in their other tweets.

IV.IV Statistical Analysis of Support and Opposition

The final distribution of S and O class tweets is shown in Table 4.

To determine the answer to RQ1, we note that through their tweets Trump voters either show support or do not show support to his policies. So support for Trump's policies from his voters could be said to be follow a Bernoulli distribution.

		Directed towards	
		Conservatives/ Donald Trump	**Liberals**
Sentiment	**Positive**	S = 586	O = 3
	Negative	O = 4	S = 425

Table 4: Distribution of support and opposition from Trump's supporters for his policies

From Table 4, we could see that Trump supporters have shown support to his policies 99.31% of the times. Using the binom.test() function of R (R Core Team, 2018), we calculated that even at 99% confidence level, the conservative (i.e. overestimating) CI for this support is (0.9832, 0.9980). This implies we can be 99% confident that Trump supporters would show support to his environment policies between 98.32% and 99.80% of the times.

IV.V Topic Analysis of Tweets

In addition to finding how strongly Trump's voter base supported his environmental policies, we wanted to figure out how they perceived the policies and the themes the policies pertained to. So for each category of tweets we collected, we attempted to figure out what were the major themes present in the tweets and whether certain themes were more dominant than others. So rephrase our 2nd research question as a set of research questions follows:

RQ2.1: What are the major themes present in the tweets we collected?

RQ2.2: Are certain themes more dominant than others?

RQ2.3: Do certain themes resonate more strongly amongst Trump supporters?

For extracting the themes or latent topics in our tweets dataset, we utilized the Non-negative Matrix Factorization (NMF) (Cichocki & Phan, 2009; Févotte & Idier, 2011) and Latent Dirichlet Allocation with online variational Bayes (LDA) (Hoffman, et al., 2010; Hoffman, et al., 2013) algorithms implemented in the Python library scikit-learn (Pedregosa, et al., 2011).

While NMF and LDA are both unsupervised learning algorithms, they take different approaches in figuring out the topics latent in a collection of documents. The NMF algorithm views the distribution of words across documents (tweets) as a non-negative document-word matrix V and tries to approximate V with two non-negative matrices W and H where W is the matrix of component vectors that represent the cluster of words according to topics and the H is the matrix of coefficient weights for the topics in each document (Lee & Seung, 1999). On the other hand, LDA is a generative probabilistic model that assumes that each document is a mixture of a finite number of topics, “where each topic is characterized by a distribution over words” (Blei, et al., 2003). Given a number of topics, LDA tries to determine which topic each word in each document belongs to and thus which topics are present with what probability in each document.

Before applying NMF or LDA for topic extraction, we created a bag-of-words model of our tweets since NMF and LDA do not presume that the order of words in a document (in our case tweet) matters. Moreover, as it is essential to have a clean sensible words model for NMF and LDA to work effectively, we applied some preprocessing to the tweets.

At first, we took the words that were mentions and removed the @ character from them, for example, the word *@realDonaldTrump* became *realDonalTrump*. Second, we removed the # character from all hashtags in a similar fashion, for example, *#dapl* became *dapl*. In addition, we converted the hashtags *#americafirst*, *#parisaccord,* and *#parisagreement* (after transforming all combinations of cases to all lower case) to the strings ‘america first’, ‘paris accord’, and ‘paris agreement’ respectively to match them with the scenarios in which they represented the same idea without the hashtags. We then removed all the URLs from the tweets as they generally only add noise to the data. Next, we converted the $ and % sign to the words *dollarsign* and *percentagesign* respectively as we assumed that these signs provide monetary or other significant information. We then removed two special symbols '…' and '•' from our tweets.

As common words without any significant meaning (also known as stop words) that appeared in most of our collected tweets would have hindered the search for theme specific words, we created a custom list of stop words by combining the 179 stop words list from NLTK corpus with a set of our custom picked words (showed in Table 5). We considered the lower case of each word of each tweet in our dataset and removed the word from the tweet if it was in our stop word list. Here, we note that the characters ‘*you’* and ‘*you,’* would be considered as different words because of the trailing comma in the second one. So, in our first removal of stop word the word *you* would be removed from a tweet but the word *you,* would not. Similar instances occurred with other stop words also, and we handled those instances in the end.

Custom Picked Words	Reason for Being Considered as Stop Words
http, https	In certain instances, the URLs in our tweets got divided into two portions with a space between the words http/https and the rest of the URL string. To accommodate this, we had to remove the words http/https separately after we had removed the rest of the URL strings.
can't, ca, nt	These words are not in the NLTK stop word list even though words similar to them (e.g. don't, won't) are. So we decided to consider these words as stop words too.
retweet	This word gives no information about the content of tweet but was identified as a top word in different theme categories a couple of times. So we decided to remove this word.
accord, agreement, paris, climate, dakota, keystone, xl, access, pipline, coal, dapl, keystonexl	These words were in our search strings when we collected tweets for creating our dataset. So they were in most of tweets but we did not want them to have any effect on theme identification. Besides, we had already stored which topic each tweet pertained to.
trump	The word came up in a lot of our tweets and in many of the top words list for themes. As we were analyzing tweets of Trump supporters, the mention of his last name many times does not give us any useful information and so we removed it from the tweets. However, we kept the twitter handle realDonaldTrump and the word Donald because they were not as common as the last name of Trump in our dataset.

Table 5: Custom picked stop words

After our first removal of stop words, we moved onto lemmatizing the words we had remaining in our modified tweets using the lemmatize() function of TextBlob and synsets() function of NLTK which in turn utilize WordNet's (Fellbaum, 1998) corpus. Lemmatization is the process of converting a word into a its root word, e.g., converting the words *better* and *best* to *good*, or converting the word *opened* to *open*.

Note that we had preserved the original sequence and case of the words in the tweets (see examples below). This helped us to identify which part-of-speech each word of each modified tweet belonged to when we tokenized the tweet (i.e. divided into constituent words) using TextBlob. Moreover, if a word was a noun and not in the set {*paris*, *potus*}, then we singularized (e.g. converted *liberals* to *liberal*) it before lemmatizing it.

As part of our final processing we removed the remaining punctuation marks and stop words from the tweets. We present two examples of how we created a bag-of-words from each tweet below-

Original tweet: *I thank God we finally have a President who looks after the American people FIRST! #AmericaFirst #MAGA #ParisAccord https:// twitter.com/LindaSuhler/st atus/796275041587445760 …*
Tweet after removing stop words : *thank God finally President looks American people FIRST! america first MAGA*
Tweet after lemmatization and removal of punctuation marks and remaining stop words: *thank god finally president look american person first america first maga*

Original tweet: *Thank you, President Trump! #ParisAccord #ActOnClimate #AmericaFirst https:// twitter.com/johncardillo/s tatus/870423038738747392 …*

Tweet after removing stop words : *Thank you, President Trump! ActOnClimate america first*
Tweet after lemmatization and removal of punctuation marks and remaining stop words: *thank president actonclimate america first*

After transforming each of our collected tweets into a bag-of-words, we created a list of the modified tweets. We then used that list to create different TF-IDF (term frequency-inverse document frequency) matrices (using scikit-learn's TfidfVectorizer class) for testing different NMF models and different term frequency matrices (using scikit-learn's CountVectorizer class) for testing different LDA models.

Now, scikit-learn provides two different objective functions: the Frobenius norm, and the generalized Kullback-Leibler divergence for applying NMF. We tested NMF with both of these objective functions along with LDA with features ranging from 200 to 2,500 words (2,816 being the total number of unique words we kept in the modified tweets) and topic counts ranging from 5 to 15 and found the NMF algorithm with the Frobenius norm objective function providing more coherent words in each theme the majority of times. So we further searched over various combinations of the parameters of NMF with Frobenius norm to find collection of themes that seemed most coherent to us.

However, since the choice of the best collection of themes is to an extent a subjective judgement, we only describe the parameters we used for creating the set of themes we thought to be the best representation of our dataset among all the collection of themes we looked at. For the best collection, we created a TF-IDF matrix comprised with the most frequent 2000 words of the bag-of-words tweet strings we created where each word appeared at least in 3 tweets and at most at 95% of all tweets. We did not use bigrams or higher n-grams to create our TF-IDF matrix because single words were giving us more consistent results.

Next, by setting the *n_components* (controller of the number of topics) parameter to 13, *alpha* (constant multiplier to the regularization terms) parameter to 0.1, *l1_ratio* parameter to .8 to emphasize L1 regularization over L2 regularization while using a combination of both regularization schemes, and keeping the other parameter values default, we created an NMF model using our TF-IDF matrix.

Table 6 shows the top 20 words for each theme sorted in non-ascending order of their importance weights in determining the theme. We manually assigned a name to each of the themes upon inspection of the top words for the theme using our knowledge of how the tweets in our dataset were phrased. Some themes names have an ordinal number in them because different words connotated the same major theme in different tweets.

Theme Serial	Theme Name	Top 20 Words
1	Nationalism	*america, first, put, thank, actonclimate, globalist, globalism, bye, finally, person, withdraw, john, sovereignty, winning, liberal, keep, agree, american, breaking, interest*

Theme Serial	Theme Name	Top 20 Words
2	Climate Change Skepticism	*change, manmade, real, hoax, global, stop, liberal, person, take, threat, globalist, nothing, gore, always, hilarious, leftist, terrorism, believe, hypocrite, much*
3	News 1	*order, executive, sign, advance, construction, project, reviving, forward, move, today, build, breaking, donald, day, action, create, give, baby, billion, revive*
4	Cost	*dollarsign, paid, fund, green, 1b, polluter, top, combined, billion, nation, pay, cost, would, us, global, know, toward, 100, trillion, usa*
5	American Jobs	*job, american, million, energy, independence, cost, back, kill, buy, winning, would, usa, like, worker, lower, work, create, lose, person, save*
6	News 2	*pull, keep, promise, us, right, breaking, reason, liberal, covfefe, news, melt, globalist, realdonaldtrump, go, report, snowflake, let, think, left, freak*
7	Wealth Redistribution	*wealth, world, redistribution, poor, scam, scheme, 3rd, nation, redistr, govt, funding, honest, interested, push, nationsfacilitate, give, toward, globalist, uninterested, fed*
8	Praise for Trump 1	*president, announce, dow, set, new, withdraw, record, withdrawal, thank, love, pulling, want, official, say, promise, happy, country, donald, treaty, finally*
9	Praise for Trump 2	*potus, maga, thank, realdonaldtrump, withdrawing, another, dear, globalist, agree, stand, scam, american, worker, state, say, promise, pay, make, finally, usa*
10	Anti-Liberal Sentiment 1	*deal, bad, vium, report, renegotiate, another, pulling, stand, exit, leave, iran, fraud, ye, kerry, admit, american, withdraw, alway, video, unfair*
11	Anti-Liberal Sentiment 2	*get, ready, rid, reaction, ever, rich, think, dc, leaving, back, real, see, take, country, science, build, liberal, idea, since, elon*
12	Anti-Liberal Sentiment 3	*obama, decision, legacy, react, another, maga, parisclimateaccord, kerry, expert, withdraw, one, like, know, clinton, plan, would, attack, un, make, truth*
13	Anti-Liberal Sentiment 4	*control, border, heaven, steyn, left, believes, breitbart, government, never, globalist, environment, 100, give, evening, think, wait, billion, want, parisclimatedeal, un*

Table 6: The themes discovered through the NMF model

We then calculated the associated weight for each theme for each tweet using the NMF model's transform() function on the TF-IDF matrix and normalized the weights for each theme so that that the theme scores for each tweet added up to 100. We considered themes 3 and 7 to be just news, topics 8 and 9 to be just praise for Trump, and themes 10 to 13 to be just anti-liberal sentiment for making our themes more homogenous and assigned each tweet the topic that had the highest normalized weight for that tweet. Note that, emphasizing L1 regularization during the training of our NMF model meant some words of the TF-IDF matrix had 0 weights as features. When a tweet was consisted of words receiving 0 weights, it was not possible to assign a theme to that tweet, and we assigned the theme name *Undetermined* to those tweets. We would also like to state that this process of theme assignment to tweets was done each time we tested different sets of parameters for creating our NMF model to randomly scrutinize whether the themes were actually making sense on top of having coherent words.

We can get an intuition about how Trump supporters have reacted to his environmental policies through the themes discovered in our model. Figure 4 shows the themes present in each of our tweet category and effectively answers RQ2.1.

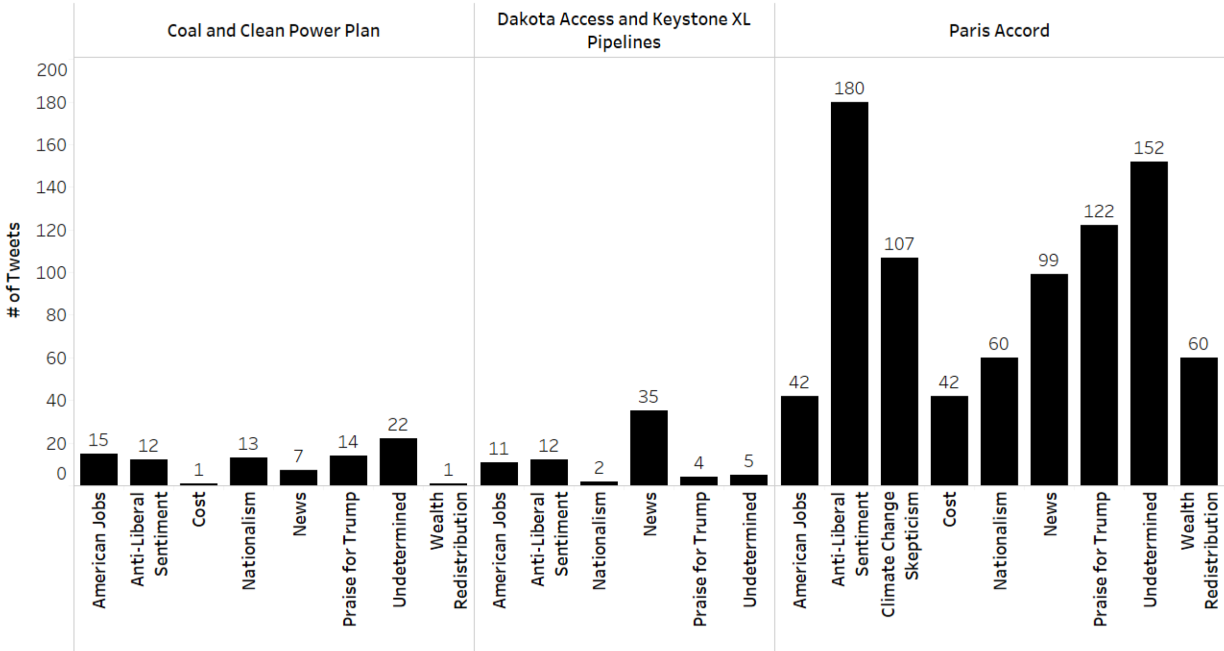

Figure 4: The distribution of tweets by themes for each major topic

Now, we differentiate between RQ2.2 and RQ2.3 by the argument that RQ2.2 asks do Trump supporters think more about certain themes compared while RQ2.3 asks whether there is any difference in how strongly they feel about different themes regardless how often they think about it. We assume that the answer to RQ2.2 could be simply be found by analyzing the frequencies of tweets of different themes while the answer to RQ2.3 could be found by analyzing the number of likes and retweets each tweet in each theme received. So we divided RQ2.3 into two parts-

RQ2.3a: Did certain themes receive more likes than other themes?

RQ2.3b: Did certain themes receive more retweets than other themes?

We present the distribution of themes along with some summary statistics for each topic of tweets we collected in Table 7. As the number of tweets per theme per topic is quite low in most cases because of smaller availability of tweets for the topics *Coal and Clean Power Plan* and *Dakota Access and Keystone XL pipelines*, we did not want to conduct any statistical test here.

Topic	Theme	# of Tweets	Total Likes	Avg. Likes	Median Likes	Total Retweets	Avg. Retweets	Median Retweets
Coal and Clean Power Plan	American Jobs	15	8,982	599	122	3,287	219	59
	Anti-Liberal Sentiment	12	4,987	416	104	1,857	155	41
	Cost	1	64	64	64	33	33	33
	Nationalism	13	1,997	154	123	919	71	37
	News	7	1,347	192	180	847	121	73
	Praise for Trump	14	4,657	333	229	2,493	178	72
	Undetermined	22	13,278	604	179	4,517	205	57
	Wealth Redistribution	1	234	234	234	52	52	52
Dakota Access and Keystone XL Pipelines	American Jobs	11	8,031	730	56	4,227	384	26
	Anti-Liberal Sentiment	12	3,369	281	109	1,497	125	50
	Nationalism	2	329	165	165	174	87	87
	News	35	11,558	330	145	4,675	134	41
	Praise for Trump	4	11,300	2,825	700	2,504	626	285
	Undetermined	5	567	113	77	307	61	32
Paris Accord	American Jobs	42	24,391	581	187	11,071	264	95
	Anti-Liberal Sentiment	180	103,708	576	138	63,931	355	85
	Climate Change Skepticism	107	55,730	521	170	33,349	312	95
	Cost	42	18,562	442	75	10,183	242	58
	Nationalism	60	37,326	622	283	23,557	393	170
	News	99	81,784	826	271	44,377	448	110
	Praise for Trump	122	59,971	492	177	29,621	243	116
	Undetermined	152	119,830	788	228	66,499	437	112
	Wealth Redistribution	60	20,573	343	116	14,194	237	91

Table 7: Summary statistics of tweets across themes for each topic

In Table 8, we present the summary statistics of our tweets datasets without considering the major topics. At this point, we find a satisfactory number of samples in each theme for conducting statistical hypothesis tests. As tweets with undetermined theme are of no use to us, we discard them in our further analysis.

Theme	# of Tweets	Total Likes	Avg. Likes	Median Likes	Total Retweets	Avg. Retweets	Median Retweets
American Jobs	68	41,404	609	132	18,585	273	90
Anti-Liberal Sentiment	204	112,064	549	132	67,285	330	82
Climate Change Skepticism	107	55,730	521	170	33,349	312	95
Cost	43	18,626	433	71	10,216	238	55
Nationalism	75	39,652	529	231	24,650	329	129
News	141	94,689	672	227	49,899	354	92
Praise for Trump	140	75,928	542	184	34,618	247	112
Undetermined	179	133,675	747	212	71,323	398	98

Theme	# of Tweets	Total Likes	Avg. Likes	Median Likes	Total Retweets	Avg. Retweets	Median Retweets
Wealth Redistribution	61	20,807	341	116	14,246	234	90

Table 8: Summary statistics of tweets in each theme

Focusing on RQ2.2, we can assume that if the certain themes were not more dominant than others, each theme would have received around 1/8th (0.125) of all tweets for which we could determine themes. So we set the following hypotheses-

Null Hypothesis (H_0): There is no significant difference in the observed proportions and expected proportions of themes.

Alternative Hypothesis (H_1): There is a significant difference in the observed proportions and expected proportions of themes.

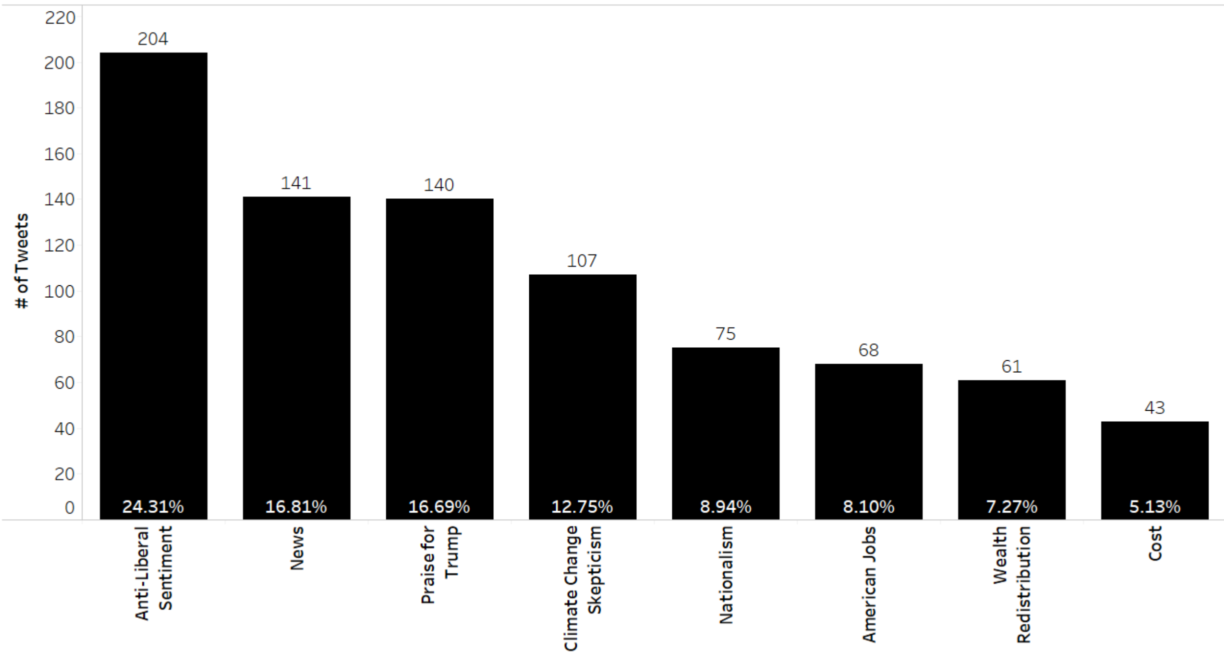

Figure 5: The distribution of tweets across themes

We present the tweets distribution across themes in non-ascending order in Figure 5. By conducting a chi-squared goodness of fit test using the chisq.test() function of R, we found the p-value to be smaller than $2.2e^{-16}$ (< 0.05). This suggests we have very strong evidence to reject these null hypothesis at 95% confidence level.

From Figure 5, we can see that Trump supporters criticize or show resentment against Trump supporters most of the times.

Next, for determining the answer to RQ2.3a, we set the following hypotheses-

Null Hypothesis (H_0): There is no difference in the # of likes the tweets in each theme received.

Alternative Hypothesis (H_1): There is a difference in the # of likes the tweets in each theme received.

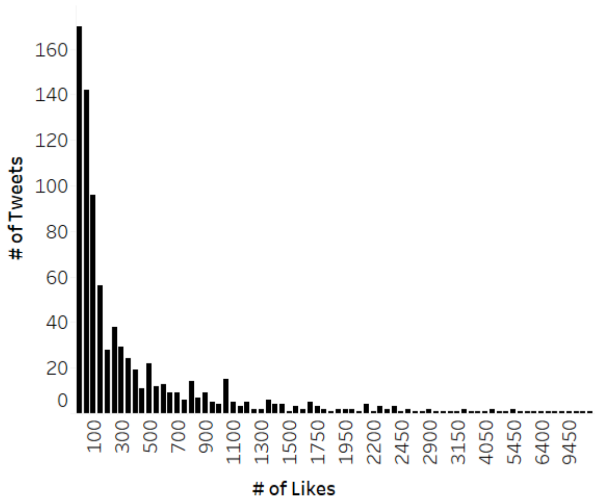

Figure 6: The distribution of tweets with different amounts of likes

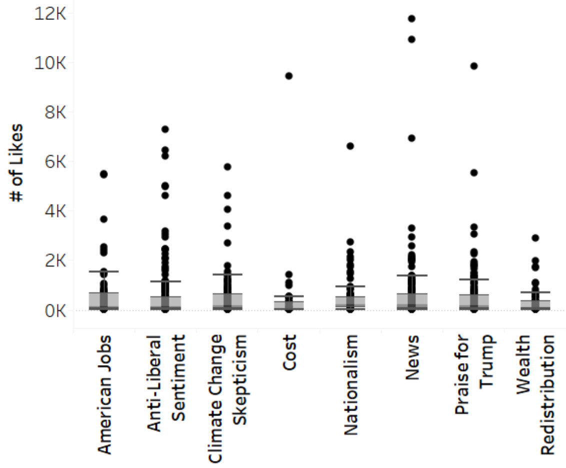

Figure 7: The distribution of likes across themes

We present the # of tweets against # of likes histogram in Figure 6 and the distribution of # likes for each theme in a box plot with whiskers extending up to 1.5 times the interquartile range from the upper and the lower quartile in Figure 7. As it is apparent that the # of likes are not normally distributed, we conduct the non-parametric Kruskal-Wallis test (Kruskal & Wallis, 1952) in R and find the p-value to be 0.0406 which means there is evidence of difference between the mean ranks of at least one pair of themes at 95% confidence level.

However, by conducting asymptotic (allowing tied ranks) pairwise theme comparisons using Wilcoxon rank sum test with the Bonferroni adjustment, we find the distribution of p-values presented in Table 9.

	American Jobs	Anti-Liberal Sentiment	Climate Change Skepticism	Cost	Nationalism	News	Praise for Trump
Anti-Liberal Sentiment	1	-	-	-	-	-	-
Climate Change Skepticism	1	1	-	-	-	-	-
Cost	1	1	0.438	-	-	-	-
Nationalism	1	1	1	0.096	-	-	-
News	1	1	1	0.098	1	-	-
Praise for Trump	1	1	1	0.142	1	1	-
Wealth Redistribution	1	1	1	0.978	1	1	1

Table 9: P-values obtained from pairwise theme comparisons using Wilcoxon rank sum test with the Bonferroni adjustment

We can see that there was not really any statistically significant variation in the # of likes during in the pairwise comparisons of themes except maybe in the comparison of *Nationalism* and *Cost;* and *News* and *Cost* at 90% confidence level. This suggests that we do not have any strong evidence that there exists any statistically significant difference between the themes in terms of # of likes.

Now, for determining the answer to RQ2.3b, we set the following hypotheses-

Null Hypothesis (H_0): There is no difference in the # of retweets the tweets in each theme received.

Alternative Hypothesis (H_1): There is a difference in the # of retweets the tweets in each theme received.

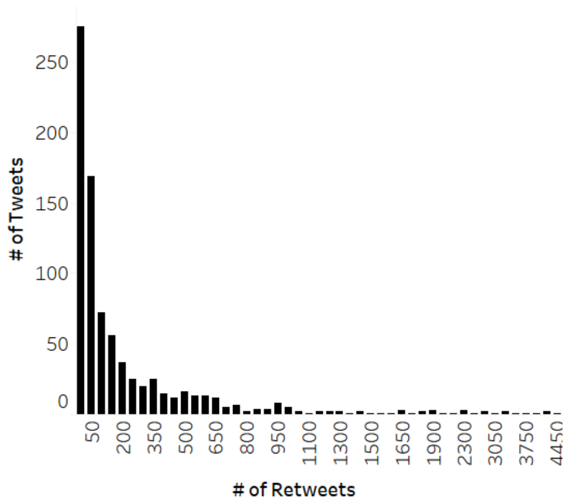

Figure 8: The distribution of tweets with different amounts of retweets

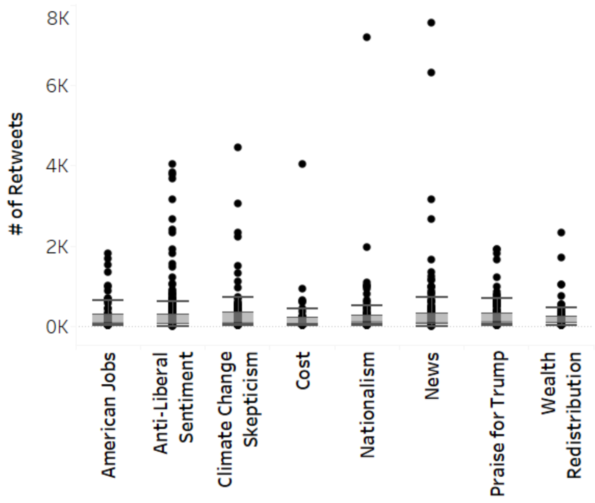

Figure 9: The distribution of retweets across themes

In a similar to what we did for RQ2.3a, we present a histogram showing the distribution tweets vs retweets in Figure 8 and boxplot showing the distribution of retweets across themes in Figure 9. We find that the # of retweets is not normally distributed too and apply the non-parametric Kruskal-Wallis test again. In this case, we find the p-value to be 0.7951 which suggests that we have no strong evidence for rejecting the null hypothesis.

So by combining the results of RQ2.3a and RQ2.3b, we cannot state any theme was more resonating or stronger reaction inducing amongst Trump supporters.

V. Conclusions

In this paper, through the analysis of the tweets of Trump's true supporters, we have found that they stay behind his environmental policies the vast majority of times. We have seen that when responding to Trump's environmental decisions, his supporters express resentment against the liberals, approval of his actions, skepticism in climate change or environmental dangers, strong sense of nationalism and concerns for their economic wellbeing. By and far, a hostile attitude towards their political opposition seems to dominate their conversations whereas how they themselves would economically benefit from the policies does not come up that often. We have also seen that despite certain themes getting mentioned more dominantly, the fervor with each theme is felt by the Trump supporter base does not show much variation. For a politician with an environment friendly outlook, the implication could be that it is not so that voters feel more strongly about certain themes regarding environmental policies, it is more so that voters have become accustomed to think about certain themes more often.

VI. Bibliography

Bird, S., Klein, E. & Loper, E., 2009. *Natural Language Processing with Python.* s.l.:O'Reilly Media Inc.

Blei, D. M., Ng, A. Y. & Jordan, M. I., 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research,* 3(Jan), pp. 993-1022.

Bradley, M. M. & Lang, P. J., 1999. *Affective Norms for English Words (ANEW): Instruction manual and affective ratings,* s.l.: NIMH Center for the Study of Emotion and Attention, University of Florida.

Cichocki, A. & Phan, A.-H., 2009. Fast Local Algorithms for Large Scale Nonnegative Matrix and Tensor Factorizations. *IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences,* 92(3), pp. 708-721.

Cohen, J., 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement,* 20(1), pp. 37-46.

Dennis, B. & Eilperin, J., 2017. EPA chief Scott Pruitt tells coal miners he will repeal power plant rule Tuesday: 'The war against coal is over'. *The Washington Post*, 09 October.

Fellbaum, C., ed., 1998. *WordNet: An Electronic Lexical Database.* Cambridge, MA: MIT Press.

Févotte, C. & Idier, J., 2011. Algorithms for Nonnegative Matrix Factorization with the β-Divergence. *Neural Computation,* 23(9), pp. 2421-2456.

Garcia, F., 2017. White House press conference: Sean Spicer suggests Donald Trump will push through Dakota Access Pipeline. *The Independent*, 23 Jaunary.

Harris, R., Dance, G. J. X. & Debelius, D., 2018. The Twitter Purge: How Many Followers Trump, Nicki Minaj and Others Lost. *The New York Times*, 13 July.

Hoffman, M. D., Blei, D. M. & Bach, F. R., 2010. *Online learning for latent dirichlet allocation.* s.l., Advances in Neural Information Processing Systems, pp. 856-864.

Hoffman, M. D., Blei, D. M., Wang, C. & Paisley, J., 2013. Stochastic Variational Inference. *The Journal of Machine Learning Research,* 14(1), pp. 1303-1347.

Hutto, C. & Gilbert, E., 2014. *VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text.* Ann Arbor, MI, s.n.

Kruskal, W. H. & Wallis, W. A., 1952. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association,* 47(260), pp. 583-621.

Lee, D. D. & Seung, H. S., 1999. Learning the parts of objects by non-negative matrix factorization. *Nature,* Volume 401, pp. 788-791.

Loria, S., 2013. *TextBlob: Simplified Text Processing.* s.l.:s.n.

McHugh, M. L., 2012. Interrater reliability: the kappa statistic. *Biochemia Medica,* 22(3), p. 276–282.

Milman, O., Smith, D. & Carrington, D., 2017. Donald Trump confirms US will quit Paris climate agreement. *The Guardian*, 1 June.

Mudde, C., 2004. The Populist Zeitgeist. *Governance and Opposition*.

Pang, B. & Lee, L., 2004. *A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts.* Barcelona, Association for Computational Linguistics.

Pedregosa, F. et al., 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research,* Volume 12, pp. 2825-2830.

Pennebaker, J. W. et al., 2007. *The development and psychometric properties of LIWC2007,* Austin, TX: LIWC net.

Pennebaker, J. W., Francis, M. & Booth, R., 2001. *Linguistic Inquiry and Word Count: LIWC 2001,* Mahwah, NJ: Erlbaum.

R Core Team, 2018. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

Schumpeter, J. A., 2010. *Capitalism, Socialism and Democracy.* s.l.:Routledge.

Shear, M. D. & Davenport, C., 2017. World Awaits Trump Decision on U.S. Future in Paris Accord. *The New York Times*, 31 May.

Smith, D. & Kassam, A., 2017. Trump orders revival of Keystone XL and Dakota Access pipelines. *The Guardian*, 24 January.

Stone, P. J., Dunphy, D. C., Smith, M. S. & Ogilvie, D. M., 1966. *General Inquirer,* Cambridge, MA: MIT Press.

Wilson, R. A., 2016. Demagogues in history: Why Trump emphasizes emotion over facts. *The Conversation*.