Analysis of the US 2020 Presidential Election Candidates' Received Attention on Twitter

Ida Johanne Austad
School of Social Sciences
University of Manchester
Manchester, United Kingdom
idajohanne.austad@postgrad.manchester.ac.uk

Hongfeng Ai
School of Social Sciences
University of Manchester
Manchester, United Kingdom
hongfeng.ai@postgrad.manchester.ac.uk

Jennifer Crawford
School of Computer Science
University of Manchester
Manchester, United Kingdom
jennifer.crawford-2@postgrad.manchester.ac.uk

Israel Ilori
School of Social Sciences
University of Manchester
Manchester, United Kingdom
israel.ilori@postgrad.manchester.ac.uk

***Abstract*—Twitter can be used to effectively determine the participating public's views on particular topics. In this instance, Democratic Presidential Candidates for the 2020 elections were considered, as well as sitting President, Donald Trump and Congresswoman Alexandria Ocasio-Cortez who both make interesting comparisons due to their known social media presence. These politician's Twitter accounts were analysed for the amount of attention that they get, and the sentiment of this attention, from the general public Twitter users. A general lexicon was used to determine positive/negative sentiments and conclusions drawn based upon perceived engagement.**

Keywords—Social Media Analytics, Sentiment Polarity Analysis, Emotion Detection, US Presidential Election.

I. INTRODUCTION

In the USA, the Candidates for the 2020 Elections are currently announcing, or have recently announced their intention to run for President. Amongst these are the four Democratic candidates Kamala Harris, Elizabeth Warren, Julian Castro and Bernie Sanders. These candidates all have a Social Media presence which can be analysed to discover how much, and what type of, attention they get. Comparing them with the Republican Candidate and current sitting President Donald Trump, and the recently elected Congresswoman Alexandria Ocasio-Cortez, is of interest as they are both known for their extensive use of Social Media to build their political identity. In this analysis Social Media Analytics is used to determine information about the attention the different politicians are getting on Twitter. The average number of responses in terms of retweets and favourites received on tweets is used as a measure of general attention. Moreover, by extracting tweets posted by the public mentioning the different candidates, Sentiment Polarity Analysis and Emotion Detection can then be undertaken, to see if the overall mentions are positive or negative or express a certain type of emotion. Comparing the results for each candidate with Trump and Ocasio-Cortez can give interesting insights into the current political climate in the USA and improved understanding of how the candidates make use of Social Media as part of building their political identity.

II. RELATED WORK

Social Media has been used in an attempt to predict several types of things, from stock market to election results, over the past few years. In particular, we are interested in analytics that cover politicians' success and Ceron, Curini and Iacus [1] show that analysing the use of social media by users can be used to give an indication of who is likely to win a particular election. They cover elections in France, Italy and USA and come to the conclusion that Sentiment Analysis can prove to be more accurate than other methods of analysis. In the USA, Twitter has been used to try to predict the 2016 presidential election by looking at sentiment analysis of tweets by users located in particular states [2]. In this case, it was found that the sentiment of tweets from users in swing states is very close to the polls run during the campaigns, although pure volume of tweets per candidate was not seen as a good indicator of success. Some will argue [3] that there are difficulties in parsing the intent behind online words, for example, taking into account sarcasm and humour. These difficulties are important to consider as, while online social media tools can be used a as gauge to help predict election outcomes, until these challenges are solved there is little difference between social media analytics and polls taken close to the election date [3]. The general consensus of the above research collectively indicates that analysing social media data is particularly useful when taking into account sentiment analysis.

III. DATA COLLECTION

Two types of Twitter data were extracted for the purpose of this analysis. The first type of Twitter Data was collected by retrieving all the tweets and retweets posted by each candidate in the timeframe of 01/01/2019 - 06/03/2019. This timeframe was chosen, as by this point, all candidates selected had announced their intention to run for President in 2020, and so replies from this point can be said to potentially be affected by this information [4]. The Twitter accounts and keywords used are presented in Table I. In total, we gathered 3,432 tweets across the 6 candidates.

The second type of data was collected using searches based on keywords present in tweet text. For each of the 6 candidates, two keywords were chosen that could identify them as being 'mentioned' in people's tweets. For all candidates these two keywords comprised firstly of either their full name or a subset of their full name, and secondly the username for the Twitter account which they use to post tweets related to their presidential campaign. In order to collect the whole tweet text, the rTweet 'search_tweets' function was used [5]. The data was limited to a maximum of 10,000 tweets per subject, from the timeframe of 24/02/2019 - 06/03/2019. Due to how the Twitter API works, only the most recent and most "relevant" tweets are returned as results, resulting in the timeline of the extracted

tweets varying depending upon how often a subject is mentioned on Twitter [6].

TABLE I. TWITTER ACCOUNTS AND KEYWORDS

Candidate	Twitter user account used in timeline search	Keywords used in Keyword Search
Kamala	KamalaHarris	"Kamala Harris+KamalaHarris"
Warren	eWarren	"Elizabeth Warren+eWarren"
Castro	JulianCastro	"Julian Castro+JulianCastro"
Sanders	BernieSanders	"Bernie Sanders+BernieSanders"
Trump	realDonaldTrump	"Donald Trump+realDonaldTrump"
Ocasio-Cortez	AOC	"Ocasio-Cortez+AOC"

IV. METHODOLOGY

To investigate how much attention the candidates were getting on Twitter the average number of favourites and retweets per tweet posted by the candidates were chosen as appropriate measures. As the Keyword search results had a large variation in time frame for the extracted tweets across the candidates, it was found that this data should not be used to measure attention (in terms of number of mentions in people`s tweets) as this would mean comparing amounts of tweets across different time spans. The average number of favourites per tweet is calculated only on the original tweets posted by the candidate, i.e. the posts with no retweeted content and retweets were the candidate has added a comment in addition to the retweeted content. Retweets without comments are not included as these cannot have favourites. For the average number of retweets, all tweets are included in the calculation as all can be retweeted. Overall, these two measures will be the indication of how much attention each candidate is getting on Twitter.

To investigate the sentiments and emotions expressed when a candidate is mentioned in a tweet, the Keyword search results are used. A large portion of the extracted tweets were retweets. Thus, the same text is represented many times in the dataset. It was decided to exclude the retweets from both the Sentiment analysis and the Emotion Detection Analysis to enable measuring the "original" sentiments and emotions expressed, rather than potentially amplified sentiments or emotions due to a viral tweet which was posted in the timeframe chosen for this analysis. Moreover, it was decided to use the Lexicon as provided in the module COMP61332 at the University of Manchester to categorize a tweet as either positive, negative or neutral. This lexicon was deemed appropriate for the purpose of this analysis as only tweets written in English were subject to analysis, and because the language used was considered to be either general everyday language or across many different domains, making it either unnecessary to use a lexicon for a specific domain or impossible to cover all domains covered across the tweets in the dataset.

Inspecting a sample of the results do however show that classification of the tweets was not always successful. Table II shows two examples of tweets where the lexicon approach to classification results in a different result compared to a manual approach. According to Ebrahimi et al. [7], Classifying tweets on dynamic events, such as discussions regarding political figures and elections, can be challenging when using an unsupervised approach such as with Lexicons. For the Emotion Detection the National Research Council Canada (NRC) [8] emotion lexicon was used which contains a list of English words and their association with eight categories of emotions: trust, anticipation, sadness, joy, anger, fear, surprise and disgust.

TABLE II. EXAMPLES OF MISCLASSIFIED TWEETS

Examples of tweet text (after cleaning)	Sentiment classification using lexicon	Sentiment classification from manual evaluation
Alexandria ocasio cortez isnt even in her s yet and shes already entangled in political corruption	Positive	Negative
Dem voter praises aoc shes a badass	Neutral	Positive

V. RESULTS

The resulting number of tweets from the timeline and keyword searches are presented in Table III. The statistics chosen to measure how much attention the different candidates are getting is presented in Fig 1. From the results one can see that the two statistics show coinciding results. Donald Trump is by far the candidate which gets the most attention and support on his tweets. The average number of favourites and retweets Trump gets is more than twice as many as Alexandria Ocasio-Cortez, which is in second place. The relative difference between the candidates are similar across both statistics, and Julian Castro gets the least attention according to both measures.

Based on the results from the Sentiment Polarity Analysis, as presented in Fig 3, one can see that there are only two candidates, Donald Trump and Elizabeth Warren, where the majority of the Twitter posts extracted are classified as positive. The opposite is true for Alexandria Ocasio-Cortez and Kamala Harris, as the results show that the majority of the extracted tweets where they are mentioned are negative. Furthermore, looking at the Emotion Detection Analysis one can get a more fine-grained understanding of what emotions are expressed in relation to the candidates, beyond whether they are positive or negative. The results, as presented in Fig 2, show that especially Bernie Sanders and Elizabeth Warren get a large proportion of tweets mentioning them, which are classified as showing trust. Donald Trump stands out in terms of how many of the tweets where he is mentioned is classified as expressing surprise. Julian Castro is the candidate which stands out the most on several measured emotions, as the tweets where he has been mentioned rank high on the positive emotions of joy, trust and anticipation. This fits well with the results of the Sentiment Analysis where Castro gets a relatively large proportion of tweets categorised as positive. Another noteworthy result is that Alexandria Ocasio-Cortez does not seem to get any larger proportion of the negative emotions compared to the other candidates although she had the largest proportion of tweets categorized as negative in the Sentiment Polarity Analysis.

TABLE III. NUMBER OF TWEETS INCLUDED IN TIMELINE AND KEYWORD ANALYSIS

Candidate	Timeline search results				Keyword search results
	Original Post	*Retweet Without Comment*	*Retweet with Comment*	*In total*	
Harris	523	19	68	610	1364
Warren	390	37	39	466	642
Castro	228	215	50	493	176
Sanders	211	73	11	295	2876
Trump	504	149	24	677	3332
Ocasio-Cortez	285	303	303	891	777

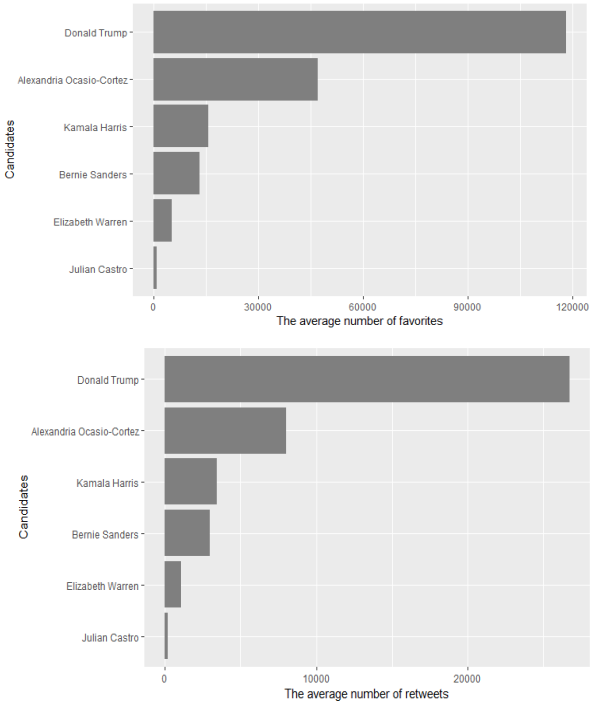

Fig 1. Average Favourites and Retweets per Candidate

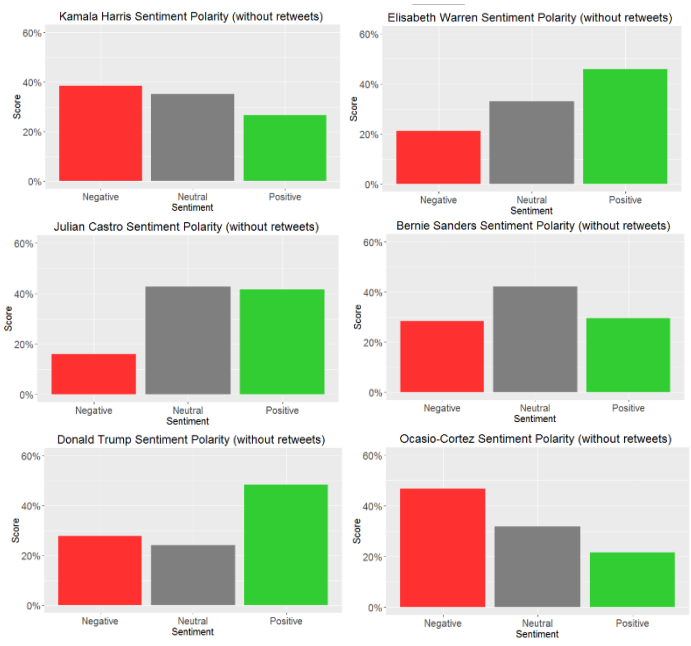

Fig 3. Sentiment Polarity Analysis results

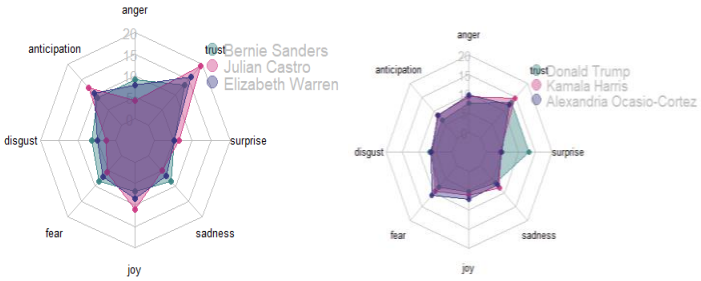

Fig 2. Emotion Detection Analysis results

I. CONCLUSION

Firstly, based on the results of this analysis one can conclude that the four presidential candidates who have announced their intention to run for president for the democrats are not getting anywhere near as much attention as Donald Trump and Alexandria Ocasio-Cortez. Donald Trump is the one garnering the most attention currently, by looking at average number of favourites and retweets. Secondly, the results from the Sentiment Polarity Analysis suggest that Donald Trump and Elizabeth Warren are the candidates were the majority of the tweets in which they are mentioned are positive, while Alexandria Ocasio-Cortez and Kamala Harris have large proportion of mentions in tweets categorized as negative. Lastly, the Emotion Detection Analysis showed that tweets mentioning Donald Trump were the ones that most often were related to Surprise, while Julian Castro got a large proportion of tweets mentioning him categorized as joy, anticipation and trust. Overall, these insights can provide us with a better understanding of the standing of the different presidential candidates for the 2020 election and can potentially help the candidates improve their Social Media strategy in the future.

REFERENCES

[1] A. Ceron, L. Curini and S. M. Iacus, Politics and Big Data, London: Routledge, 2016.

[2] B. Heredia, J. D. Prusa and T. M. Khoshgoftaar, "Social media for polling and predicting United States election outcome," Social Network Analysis and Mining, vol. 8, no. 48, 2018.

[3] M. Huberty, "Can we vote with our tweet? On the perennial difficulty of election forecasting with social media," International Journal of Forecasting, vol. 31, no. 3, pp. 992-1007, 2015.

[4] Axios, "2020 presidential election: Track which candidates are running," Axios, [Online]. Available: https://www.axios.com/2020-presidential-election-candidates-announce-running-15472039-9bf49de4-351a-46b3-bdde-b980947b21ea.html. [Accessed 4 March 2019].

[5] M. W. Kearney, "RDocumentation: search_tweets," [Online]. Available: https://www.rdocumentation.org/packages/rtweet/versions/0.6.8/topics/search_tweets. [Accessed 6 March 2019].

[6] Twitter, "Search Tweets: Frequently Asked Questions," [Online]. Available: https://developer.twitter.com/en/docs/tweets/search/FAQ. [Accessed 6 March 2019].

[7] M. Ebrahimi, A. H. Yazdavar and A. Sheth, "Challenges of Sentiment Analysis for Dynamic Events.," *IEEE Intelligent Systems*, vol. 32, no. 5, pp. 70-75, 2017.

[8] S. Mohammad and P. Turney, "NRC Word-Emotion Association Lexicon," National Research Council Canada, Ottawa, 2010.