Free Exposure for Political Candidates

Investigating Social Media Behavior of Journalists for the 2016 US Election

Jan Willem David Alderliesten (4368703)
Delft University of Technology
Mekelweg 4
2628 CD, Delft
j.w.d.alderliesten@student.tudelft.nl

ABSTRACT

This paper provides the results and analysis of a study which aimed to identify candidate exposure trends among political journalist's *Twitter* profiles during the 2016 United States election. Due to the upset outcome of the election, many have speculated about the reasons as to why the press was unable to foresee Donald Trump's victory. One of these theories revolved around additional broadcasting time being allocated to certain candidates, causing the additional exposure to act in a candidate's favor. This study aims to identify if such identical behavior existed on social media. After doing an analysis of 91919 *Tweets* of 127 journalists, it was found that a difference of 0.039 existed between the candidates, which signified a small to negligible difference, thus indicating this exposure trend did not extend to social media behavior.

Keywords

Social Web Data, Twitter, Journalism, 2016 United States Election, Trump Election, Free Exposure, Free Advertising

1. INTRODUCTION

The 2016 United States Election, in which Donald Trump was elected to the office of the presidency, was considered an election with an 'upset' outcome [1]. The election outcome was considered as such en large due to the factor that data scientists and polling agencies had been unable to predict Trump's victory[2], including a now infamous claim from the New York Times that Trump's chance of victory was only 7% [1] just before the election. Due to this inability to predict the outcome, speculation began in an attempt to find explanations for this data discrepancy.

One of the investigated issues was that journalists provided Donald Trump with more coverage than Hillary Clinton, his main rival. Before his nomination, Trump had garnered close to 2 billion (United States) dollars in free exposure [3], and estimations after his election victory place Trump close to 6 billion dollars in free exposure compared to Clinton's 2.8 billion dollars [4] in free exposure. This behavior shall be called 'free exposure' in this study, as any news about a candidate provides them with exposure to a possible voter who can become interested in them from this news. As politics is subjective, what is considered negative by the press could be considered positive by a voter, or vice-verse. Simply put, the adage 'any exposure is good exposure' is held to be true.

[1]twitter.com/nytimes/status/790228805273157632

Since this free exposure was shown to be significant within the realm of television, one might question whether journalists display the same behavior on their personal social media channels. If the same trend exists on social media, it could be that additional free exposure was provided by journalists on these channels, meaning the 6 billion figure might be larger due to the importance of social media in information sharing and the significant share that news has within Twitter's ecosystem and user focus [5]. This study aims to investigate whether the trend of free exposure for the political candidates holds true for the social media profiles of journalists covering an election.

The study found that, with a data set of 91919 *Tweets* made by 127 journalists, more exposure was given to Trump than Clinton, but in a far less extreme order of magnitude compared to the television discrepancies. It was also found that journalists from certain broadcasters were far more prone to provide additional exposure to a single candidate than the other, and that the favored candidate seems to hold per broadcaster.

2. INVESTIGATION

The investigation section outlines aspects related to the research procedure, including subject selection and data sanitation methods.

2.1 Subject Selection

To investigate the amount of exposure each political candidate obtained from political journalists, a selection of journalists had to be made and their tweets during the election period had to be analyzed. Although at first a preference was given to perform a random selection of journalists through the use of an equivalency of a Heckman estimation [6], this was also deemed impossible due to certainty that a twitter account identified was in fact a functional journalist. Hence, the selection occurred manually and non-randomly.

The selection of journalists occurred by identifying the major news broadcasting networks in the United States based on a combination of public and private broadcasting, and taking the most viewed news channels. The selection of viewers occurred by taking self-reported viewer statistics over the year 2016. The selection consisted of the networks *ABC*, *CBS*, *CNN*, *Fox News*, *MSNBC*, *NBC*, and *PBS*. After the network selection occurred, journalists were selected based on a set of requirements. These requirements consisted of being active during the candidacy announcements of the 2016 United States election up to the day after the election (12 April 2015 - 9 November 2016), having a pub-

lic record on the website of their organization of their employment, and having a label of 'political correspondent' or 'political editor.'

After this selection, the Twitter profile of a journalist was identified and utilized for the investigation. This was done through manual identification due to the existence of copycat or fake account. In the case that multiple accounts were found, the account which was labelled as being representative as their journalistic account (such as *@johndoe* versus *@johndoecnn*) was utilized.

Through the use of this subject selection methodology, a total of 127 journalists were selected representing 7 organizations providing a data set of 91919 Twitter statuses. This data set can be verified within the research repository[2].

2.2 Data Collection

To investigate the level of free exposure among journalists and compare candidate-related tweets, a method had to be developed and applied to the journalists to collect the required data.

After the journalist Twitter handles were identified and compiled into a document, a Java program was developed which employed the Twitter 1.1 API [3] to crawl all Tweet Statuses from the desired users. This crawling was done by employing the *Twitter4J* framework[4], which provides a stream accessor and API implementation for the Java language. After obtaining the Statuses, the program filters the set per use to only contain tweets created between the dates of 12 April 2015 and 9 November 2016.

The statuses are then filtered based on a selection of given keywords. The selected keywords were *Trump* and *Clinton*. These keywords were tracked on a user-by-user basis, meaning the number of Trump and Clinton tweets were shown per Twitter handle. To ensure accuracy and remove the possibility of a tweet referring to another member of the Trump or Clinton family (possible due to both families being in the public eye, including previous presidencies), additional tracker variables were created which checked the keywords of *Donald Trump* and *Hillary Clinton*.

The code utilized for the data collection can be found in the research repository[5].

2.3 Data Synthesis

To compare the exposure within the Tweet Statuses between the two candidates, a method was required which provided a mathematical overview of the occurrences of both candidates. This allows for a direct comparison of exposure to journalist's social media activity.

Given a set of journalists J containing j elements and a set of associated Twitter statuses per journalist S_j, the exposure e for a candidate c by a single journalist is defined in (1) as e_c^j. The equation sums the instances of tweets which contain 0 or 1 reference(s) to a candidate, and divides this summation by the cardinality of the status set S_j.

$$e_c^j = \frac{\sum_{s \in S_j} \min(1, c \in s)}{|S_j|} \tag{1}$$

The total exposure for a candidate c considering all journalists, given by e_c, is defined in (2) considering the set of all journalists.

$$e_c = \frac{\sum_{j \in J} e_c^j}{|J|} \tag{2}$$

Utilizing these definitions, an analysis could be done over journalists to investigate the exposure per candidate and track trends per broadcaster and candidate.

Table 1: Overview of Findings and Synthesis

Attribute	Value
Total Tweets	91919
Total Trump	17287
Total Clinton	09375
e_{Trump}	0.114
$e_{Clinton}$	0.075

3. RESULTS

The results section provides an overview of the raw findings and performs the empirical investigation as outlined in section 2.3 upon this data set. The raw results, including direct program output, can be found within the research repository[6].

3.1 Findings

A total of 127 journalists were analyzed, of which 120 had a Twitter handle which conformed to the requirements as outlined in section 2.1. Of these 120 handles, 91 were active during the election season providing a total of 91919 statuses. Of these statuses, a total of 17287 contained a reference to Donald Trump, and a total of 9375 contained a reference to Hillary Clinton.

3.2 Statistical Analysis

Based on the ratios defined in section 2.3, Donald Trump's exposure was found to be 0.114 while Hillary Clinton's exposure was .075. The information that led to these findings is summarized in Table 1.

When investigating trends per broadcaster, the averages shift slightly. Only journalists affiliated with *MSNBC* and its parent company *NBC* provide equal exposure on their social media profile at 0.133 for Trump and 0.101 for Clinton. The largest disparity can be seen with *CNN* (0.193 versus 0.088) and *Fox News* (0.145 versus 0.076), whose journalists and affiliated contributors provide exposure to Trump at almost double the rate of Clinton. An overview of the broadcaster findings is given in Table 2.

4. ANALYSIS

The analysis section aims to provide insight and highlight trends based on the results of the investigation.

4.1 Candidate Differences

When taking into account the entire data set, a difference of 0.039 exists between the exposure values of e_{Trump} and $e_{Clinton}$. The largest individual discrepancy in the data set was 0.578, and a total of 9 journalists had an exposure difference of more than 0.25. This means that those 9 journalists provided one of candidates with at least twice as many

[2]github.com/dalderliesten/JournalistExposure/tree/master/dataset
[3]developer.twitter.com
[4]twitter4j.org
[5]github.com/dalderliesten/JournalistExposure/tree/master/src
[6]github.com/dalderliesten/JournalistExposure/tree/master/results

Table 2: Overview of Broadcaster Statistics

Broadcaster	Total Tweets	Total Trump	Total Clinton	e_Trump	e_Clinton
ABC	21929	5679	3304	0.259	0.151
CBS	855	109	67	0.127	0.078
CNN	31615	6129	2801	0.193	0.088
Fox News	24959	3621	1921	0.145	0.076
(MS)NBC	9872	1317	1001	0.133	0.101
PBS	2689	432	281	0.161	0.105

tweets over the other. It is important to note that all of those journalists with a large exposure difference all provided more tweets to Trump, or formally said, favored Trump.

Looking at the average trends, the majority of journalists did provide additional social media exposure to Trump, but the majority of these journalists did not do so at a quantity above 0.05. Assuming natural differences will always occur due to differences in candidate campaign activity, this can be assumed to fall within a possible margin of error caused by this differing candidate behavior.

4.2 Overall Political Interest

Although the analyzed data set consisted of 91919 *Tweet* statuses, a total of 26662 were about the political candidates (which amounts to 29%).

Since now synthesis or data collection was done on the content of a tweet when it did not contain references to Clinton or Trump, no formal conclusion can be made regarding the nature of the other 71% of the statuses that were obtained. However, possible speculation allows the claim to be made that many political journalists either focus on less-mainstream topic when on their own personal networks, or simply focus on personal issues. To investigate this further, this study should be repeated with sentiment analysis over statuses, perhaps as suggested by Bing Liu in his 2012 paper [7]. By employing the techniques to measure sentiment, a possible discriminating factor could arise that allows non-categorized tweets from the program utilized for this research to be categorized based on personal stories, photos, or other non-election political news.

4.3 Broadcaster Trends

The most significant difference found within the data set was the difference that existed per broadcaster and the trends that came out of this of their journalists. As shown in Table 2, *CNN* and *Fox News* contain larger than average differences between the number of Clinton and Trump mentions. Compared to these networks, they very clearly provided additional exposure to Trump.

Interesting to see is that the individual accounts of their journalists match these trends. What was hypothesized to be caused by outliers was a trend that held for almost all of their associated journalist's social media accounts. It could be that social influence, the interactions of the inner circle that a person communicates with, caused all these journalists to have a closer eye on Trump than on Clinton.

Another possible explanation for this trend could be the biases that these networks are known to have. According to the crowd sourced journalism rating site *Media Bias Fact Check*, which allows anyone to submit a rating of their perceived bias of a news channel, *CNN* is rated as *left*[7] whereas *Fox News* is rated as *right*[8] on the left-right political scale. *Fox News* may therefore have a vested interest in providing Trump with positive coverage, whereas *CNN* may have a vested interest in providing negative coverage about Trump. These channel biases seem the most likely explanation for these discrepancies of the news channels mentioned when compared to the other 5.

5. DISCUSSION

The discussion section aims to address possible shortcomings with the research performed and discuss possible extensions or alternative versions of research to perform.

5.1 Limitations

The two major limitations of this research that restrict both generality and accuracy of the conclusions consist of a non-random sample size and a very focused program which only analyzed four variables due to limitations of the employed framework.

The non-random sample size introduces bias, which is unwanted and can skew the data to a certain direction. For example, if journalists who are personally anti-Trump own social media accounts more than their neutral peers (and judging by the election demographics, this can be a case), then hand-selected samples will automatically incur more anti-Trump bias, which, as aforementioned, can be beneficial for his exposure. To fix this, a methodology must be developed which allows for random selection of verified journalists to remove these biases.

Regarding framework limitations, *Twitter4J* and the *Twitter API* both limit the queries and types of queries that can be done. Originally, this research was aimed at identifying locations from where political journalists tweeted (to investigate the claim that many journalists remained in so-called liberal states). However, the geolocation tags that existed within the *Twitter API* were disabled and only returned *NULL*. Limitations related to the actual research included rate limiting and that obtaining tweets older than 5 days was a challenge and required spoofing. Limitations upon desired extensions to the research were obtaining a method of analyzing the sentiment and tags in a tweet, which was not possible at the rate limit to make it feasible. To fix this, the code could be extended to track additional variables and options in an optimal way to allow for greater conclusions regarding the journalist profiles that caused significant candidate exposure discrepancies.

5.2 Future Research

Future research which can enhance this study will involve either performing this study with a different dominating fac-

[7] mediabiasfactcheck.com/cnn/

[8] mediabiasfactcheck.com/fox-news/

tor (geolocation instead of exposure) or would aim to replicate the study while taking into account additional variables, such as tracking whether a journalist explicitly mentions they support a candidate or clearly lean to a certain political preference. These extensions can enhance the accuracy and conclusions possible from the research.

6. CONCLUSIONS

This paper summarized the procedure and findings of a study aimed at identifying the amount of social media exposure for political candidates in the 2016 election by political journalists. After a subject selection and data collection phase which gathered 91919 *Tweet* statuses from 127 journalists representing 7 news organizations. Findings suggest that, contrary to televised media, a small to almost negligible difference of 0.039 exists between the social media exposure provided between the two candidates. It was also found that only 29% of the statuses posted by political journalists during the election period focused on either candidate, further suggesting the conclusion that the free exposure is not significant enough to warrant a change in candidate outreach.

7. FURTHER READING

The further reading section aims to provide the reader with additional material and background information to allow for further research beyond the scope of this paper.

7.1 Repository Analysis

To look at the code utilized for this research, and to view the raw data set with all relevant attributes, you can read and fork the repository on GitHub[9]. In the spirit of open data and research, this repository will remain available as long as GitHub is capable of hosting it.

7.2 Twitter as a News Medium

Kwak et al. have investigated the prevalence of *Twitter* as a news medium [5]. They investigated the amount of statuses posted by users which reference to or involved news articles, and tracked their sharing, liking, and retweeting. A significant chunk of users, based on their findings, rely on *Twitter* as a serious news sharing medium, which can help explain the importance of investigating journalist behavior on *Twitter*.

7.3 Twitter Interactability

Jahng and Litau investigated the effects of journalist credibility and *Twitter* usage[8]. They found that when a journalist has a social media profile and they are willing to discuss and interact with viewers or others on those social media profiles, the credibility in the eyes of the general public increases.

8. REFERENCES

[1] B. Goldmacher, S. Schreckinger, "Trump pulls off biggest upset in u.s. history," November 2016. [Online; posted 09-November-2016].

[2] M. Braun, "How data journalism failed the american electorate in 2016," November 2016. (Accessed on 01/14/2018).

[3] N. Confessore, K. Yourish, "2 billion worth of free media for donald trump," March 2016. [Online; posted 15-March-2016].

[4] N. M. Sultan, "Election 2016: Trump's free media helped keep cost down, but fewer donors provided more of the cash," April 2017. [Online; posted 13-April-2017].

[5] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?," in *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, (New York, NY, USA), pp. 591–600, ACM, 2010.

[6] C. Winship and R. D. Mare, "Models for sample selection bias," *Annual Review of Sociology*, vol. 18, no. 1, pp. 327–350, 1992.

[7] B. Liu, *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers, 2012.

[8] M. R. Jahng and J. Littau, "Interacting is believing: Interactivity, social cue, and perceptions of journalistic credibility on twitter," *Journalism & Mass Communication Quarterly*, vol. 93, no. 1, pp. 38–58, 2016.

[9] github.com/dalderliesten/JournalistExposure