Geographic and Temporal Trends in Fake News Consumption During the 2016 US Presidential Election

Adam Fourney*, Miklos Z. Racz*, Gireeja Ranade*, Markus Mobius, Eric Horvitz
adamfo,miracz,giranade,mobius,horvitz@microsoft.com
Microsoft Research

ABSTRACT

We present an analysis of traffic to websites known for publishing fake news in the months preceding the 2016 US presidential election. The study is based on the combined instrumentation data from two popular desktop web browsers: Internet Explorer 11 and Edge. We find that social media was the primary outlet for the circulation of fake news stories and that aggregate voting patterns were strongly correlated with the average daily fraction of users visiting websites serving fake news. This correlation was observed both at the state level and at the county level, and remained stable throughout the main election season. We propose a simple model based on homophily in social networks to explain the linear association. Finally, we highlight examples of different types of fake news stories: while certain stories continue to circulate in the population, others are short-lived and die out in a few days.

CCS CONCEPTS

• **Human-centered computing → Social media**; • **Information systems → Web log analysis**;

KEYWORDS

Fake news, elections, browsing data, social media

1 INTRODUCTION

Fake news is a centuries-old problem [6] and has had a presence on the internet for as long as the medium has existed. Recently, however, social media has made it possible for an individual to rapidly share misleading information with large populations, without the overheads associated with traditional broadcast media such as newsprint or television. The potential influence of fake news spreading via social media was brought to widespread public attention following the 2016 US presidential election, and economists are already beginning to study whether fake news articles may have influenced its outcome [1]. Meanwhile, addressing fake news has become a top priority of large technology companies [3, 10], and governments worldwide have begun considering legislative action to combat its spread [2]. Together, these trends motivate a need to more deeply understand the spreading mechanisms and access patterns of fake news on the internet, and, in particular, on social media.

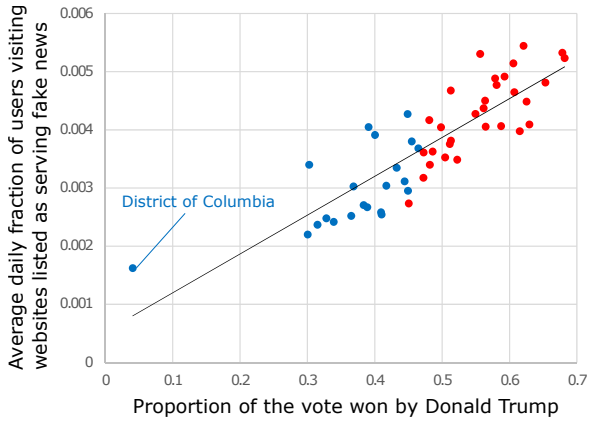


Figure 1: Correlation between voting behavior and the average daily fraction of users visiting fake news websites. Points represent states, colored blue (Democratic) or red (Republican) for the party that won the presidential race.

We report on geographic and temporal trends of the visitation of fake news websites during the 2016 US presidential election campaign. Our analyses are based on instrumentation data collected from Internet Explorer 11 and Edge, two popular desktop web browsers with hundreds of millions of users, combined. The contributions of this work are threefold: First, we confirm that social media was the primary outlet for the circulation of fake news stories (Table 1). Second, we find that the most viewed fake news stories largely exhibit one of two patterns: stories that peak and receive most of their views in 24-48 hours, and stories that persist for longer periods of time and that steadily acquire views (Fig. 3). Finally, we show that aggregate voting patterns are correlated with the average daily fraction of users visiting fake news websites, both at the state level (Fig. 1), and at the county level (Fig. 4). These correlations remained stable throughout the political campaign, and we propose a simple linear model to explain this observation.

2 DEFINING "FAKE NEWS"

There has been extensive reporting on the magnitude and nature of fake news, as well as significant debate about the definition of this term. Consistent with prior work [1], our analysis relies on lists compiled by third parties. Specifically, we leverage Wikipedia's list

* These authors contributed equally to the work, and are presented in alphabetic order.

CIKM'17 , November 6–10, 2017, Singapore, Singapore

ACM ISBN 978-1-4503-4918-5/17/11…$15.00
https://doi.org/10.1145/3132847.3133147

Domain	% of all fake news traffic	% of Referrals from Social Media
endingthefed.com	21.1%	97.6%
thepoliticalinsider.com	18.0%	80.0%
infowars.com	17.2%	10.9%
americannews.com	14.5%	98.9%
libertywritersnews.com	9.3%	96.7%

Table 1: Top five fake news domains by visitations. Together, these five domains account for 80% of the fake news visitations observed during the general election.

of fake news websites[1], together with Snopes's *Field Guide to Fake News Sites and Hoax Purveyors*[2]. Additionally, we include the web domains of the top five fake news stories, as reported by Silverman in [5]. As such, our research adopts the definitions leveraged by the maintainers of these lists. For example, Wikipedia defines fake news websites as those which "intentionally publish hoaxes and disinformation for purposes other than news satire". Notably, this definition is not limited to politics—both Wikipedia and Snopes list websites that discuss other topics, including: pseudoscience, health, and celebrity / sports gossip. We include these domains in our analysis so as to avoid editorializing, and to maintain a simple inclusion criterion.

3 DATA

We analyze 114 days of instrumentation data for Internet Explorer and Edge[3], two desktop web browsers with a combined install base of more than 10^8 machines. Our analysis begins on July 18th, 2016 (the start of the Republican national convention) and ends on November 8th, 2016 (election day). The dataset consists of a list of timestamped visits to URLs, together with anonymous user identifiers and ZIP codes. Of interest are visits to 70 fake news web domains as outlined in Section 2. Finally, we leverage Dave Leip's *Atlas of U.S. Presidential Elections*[4] for election data.

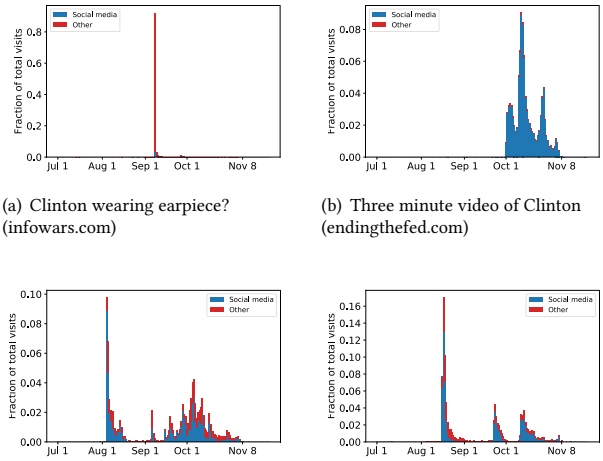
4 RESULTS

4.1 Traffic Sources and Prevalence

Consistent with past research [1], our analysis finds that social media (Facebook and Twitter) was a primary traffic source to fake news, accounting for 68% of all page visits for which traffic sources could be determined. This finding on the role of social media was especially true for four of the top five domains in our dataset (Table 1). Moreover, traffic from Facebook was orders of magnitude larger than the traffic from Twitter, with 99% of social media referrals coming from Facebook. However, the analysis also reveals that visits to fake news websites were relatively rare—on an average day during the election campaign period, only 0.34% of users visited *any* of the fake news domains that we monitored (i.e., about 1 in every 290 users).

These low visitation rates are comparable to the traffic patterns we would expect from social media advertising campaigns; though

[1]https://en.wikipedia.org/wiki/List_of_fake_news_websites, accessed on Dec. 22, 2016, and again on Jan. 25, 2017.
[2]http://www.snopes.com/2016/01/14/fake-news-sites/
[3]Browser instrumentation data is collected with user permission to support predictive services and features.
[4]http://uselectionatlas.org/



(a) Clinton wearing earpiece? (infowars.com)

(b) Three minute video of Clinton (endingthefed.com)

(c) Clinton sold weapons to ISIS (thepolicicalinsider.com)

(d) Obama bans pledge of allegiance (abcnews.com.co)

Figure 2: Histogram of visits to fake news stories from July 18 to November 8, 2016. The blue fraction of each bar represents the share of visitors referred by social media, while red represents other detectable referrers.

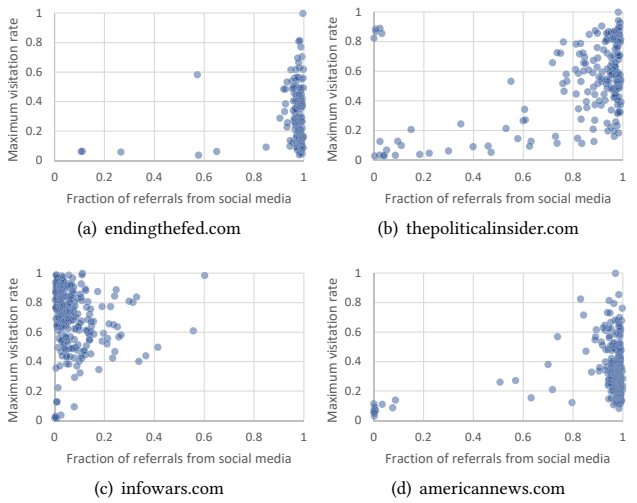
not directly comparable, it has been reported that the average click-through rate of advertisements appearing on the Facebook news feed is 0.90%.[5] If similar click-through rates apply to fake news links, then the actual daily exposure to fake news headlines in social feeds may be substantially greater than the 0.34% figure reported here.

4.2 Temporal Trends of Fake News Stories

We observed various temporal visitation patterns for high-traffic stories in our dataset. Certain stories are short-lived and get the majority of their views over a few days (e.g., Fig. 2(a)), while a second set of stories are more long-lived and receive traffic over months (e.g., Figs. 2(b)-2(d)). Figure 2 shows how the visits to four popular stories were spread over time: the first about Hillary Clinton wearing an earpiece during a forum[6], the second a viral video asking about Hillary Clintons's past[7], which endingthefed.com picked up from americannews.com, the third about Wikileaks confirming that Hillary Clinton sold weapons to ISIS[8], and the fourth about Obama banning the pledge of allegiance in schools[9]. Social media referrals are the source of a large fraction of visits for the three long-lived stories in Figs. 2(b), 2(c), and 2(d).

Figure 3 expands this analysis to the 1000 most popular stories, and four most popular websites in our dataset. We consider a story to have a high visitation rate if it gathers most of its views in the

[5]http://www.wordstream.com/blog/ws/2017/02/28/facebook-advertising-benchmarks
[6]http://www.infowars.com/was-hillary-wearing-an-earpiece-during-last-nights-presidential-forum/
[7]http://endingthefed.com/this-three-minute-video-of-hillary-just-cost-her-the-election-spread-this-now.html
[8]http://www.thepoliticalinsider.com/wikileaks-confirms-hillary-sold-weapons-isis-drops-another-bombshell-breaking-news/
[9]http://abcnews.com.co/obama-executive-order-bans-pledge-of-allegiance-in-schools/



(a) endingthefed.com (b) thepoliticalinsider.com

(c) infowars.com (d) americannews.com

Figure 3: Maximum visitation rate vs. fraction of social media referrals for the top 1000 stories published on the top four websites in the dataset.

course of a day or two, and a story to have a low visitation rate if the views are gathered slowly over a period of many days. To measure this, we define the *maximum visitation rate* of a story as the ratio of the maximum views on any day to the total views the story received. Figure 3 shows that websites like endingthefed.com and americannews.com consistently hosted longer-lived stories that were largely viewed via social media. We also found exceptions to the strong role of social media as a source of successful fake news stories. As Figure 3 shows, most of the stories hosted on infowars.com had few social media referrals and this was also the case for some stories on thepoliticalinsider.com.

Finally, although hosted on websites known to frequent in fake news, we note that articles in our dataset include a mix of opinion pieces, and biased fact-based stories that may present events out of context, in addition to articles that are entirely fabricated.

4.3 Geographic Trends of Voting Patterns

Finally, we report that the average daily fraction of users visiting fake news websites is highly correlated with geographic voting patterns at the state level (Pearson $r = 0.85$, including the District of Columbia; Figure 1) and at the level of the top 1000 FIPS counties by population (Pearson $r = 0.71$; Figure 4). States or counties experiencing more fake news visitations also tended to vote for Donald Trump. These correlations remain high throughout the election campaign, peaking in October (Table 2).

We caution readers against directly inferring any particular causal relationship between visits to fake news websites and voting patterns, since we are merely observing correlations in the data.[10] To this end, we note that geographic trends in the visitation of fake news domains during the 2016 general election campaign are also highly correlated ($r = 0.76$, $p \ll 0.0001$) with the distribution of votes won by Mitt Romney, the unsuccessful Republican candidate

[10] "*Empirically observed covariation is a necessary but not sufficient condition for causality*", [9].

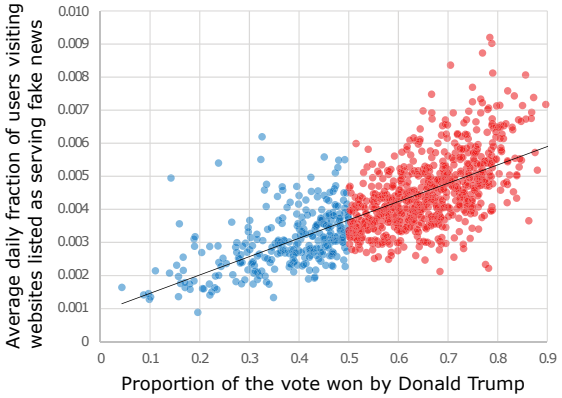


Figure 4: Correlation between average daily fraction of users visiting fake news websites and voting behavior for the top 1000 FIPS counties by population. Each point represents a county and the colors are as in Figure 1.

Date Range	Pearson r
July 18-24	0.85
August 15-21	0.84
September 19-25	0.84
October 17-23	0.88
November 2-8	0.84

Table 2: Pearson correlations between the average daily fraction of users visiting fake news domains in a given state, as measured on five distinct weeks, and the proportion of the vote finally won by Donald Trump. All values are highly statistically significant, with $p \ll 0.0001$.

who ran against president Barack Obama in the *previous* federal election of 2012. Consequently, we hypothesize that the observed correlations reflect homophily in social networks, together with the observed pro-Trump bias in fake news [1, 5]. Simply stated, we believe that an individual's political affiliation is relatively stable over time, that their neighbors in the social network will tend to have similar political beliefs, and that these connections determine the degree to which people are likely to be exposed to fake news links on social media. In the next section, we present a simple linear model based on this hypothesis.

5 MODEL

We now describe a linear model that can explain the observed correlations. Connections in online social networks capture both geographic and ideological similarities between users, and we believe this plays a major role in the observed correlation. We assume that each person has a 'type', which describes their political leaning as Democratic (D) or Republican (R). Further, we assume that every fake news article also has a type, and is either pro-Trump and anti-Clinton, which we denote by T (for Trump), or pro-Clinton and anti-Trump, which we denote by C (for Clinton).

We use *exposure* to capture the number of people who "see" a story, e.g., as a link in social media. The visitors to a story are those who *click* on it. This is a subset of those exposed (recall, our empirical data measures only this subset). The proposed model uses the key fact that homophily in social networks implies that

the probability that any person is exposed to an article depends on both their type as well as the type of the article. We model this as follows:

- A D person gets exposed to a T article with probability p_T and gets exposed to a C article with probability p_C per day.
- A R person gets exposed to a T article with probability q_T and gets exposed to a C article with probability q_C per day.

We assume the click-through rate (probability of visitation after exposure) is a constant probability b, and is independent of the type of the story and the type of the person.

Let X denote the number of C stories and Y denote the number of T stories, and note that we observe $X < Y$ (in fact, [1] reports that $X = 7.6 \times 10^6$ and $Y = 3.0 \times 10^7$, giving $X \ll Y$). Then, if a region has proportion t of type-R people, we expect that the number of clicks on articles from fake news domains per day is the population times:

$$\begin{aligned} & b \times \{t\,(q_C X + q_T Y) + (1-t)\,(p_C X + p_T Y)\} \\ & = b \times (p_C X + p_T Y) + t \times b \times \{(q_C - p_C)\,X + (q_T - p_T)\,Y\}\,. \end{aligned}$$

This is linear in t, with slope given by

$$b \times \{(q_C - p_C)\,X + (q_T - p_T)\,Y\}\,. \tag{1}$$

Now *homophily* implies that $p_C - q_C > 0$ and $p_T - q_T < 0$, that is, D people have a larger exposure to C articles than R people, and R people have a larger exposure to T articles than D people. Assuming that $q_C - p_C$ and $q_T - p_T$ have approximately the same magnitude, the fact that $X \ll Y$ implies that the slope in (1) is positive, which explains the observed correlation.

6 RELATED WORK

Our study contributes to the series of academic and journalistic works on this subject through a fine-grained geographic and temporal perspective. We discuss several representative efforts here.

Our model builds on the analysis of Silverman [5], whose data showed that the majority of the fake news stories with the most Facebook engagement favored Donald Trump. Silverman also found that in the three months preceding the election, Facebook engagement with fake news stories overtook that of stories from mainstream media outlets. A study at the MIT Media Lab [8] showed that there was very low connectivity between Trump and Clinton supporters on Twitter, which supports our model assumption of homophily in social networks.

Allcott and Gentzkow [1] use data from an online survey conducted soon after the election to estimate the impact of fake news stories. Their estimation techniques and dataset are very different from ours, and they estimate that about 1.2% of the population was exposed to the average fake news article. We note that their analysis using data from BuzzSumo aligns with our finding that fake news stories were shared on Facebook orders of magnitudes more times than on Twitter.

While our work analyzes large-scale aggregate patterns of fake news consumption, other authors have performed case studies of specific stories. For instance, a New York Times article presented the timeline of the spread of a rumor over social media after the 2016 US Presidential Election [4].

Finally, recent work has shown that a Facebook post can be classified, to a good degree, as fake or not based on the users that "like" it [7]. We hope that this growing body of work can be leveraged to raise sensitivities and frame efforts to counter the negative effects of spreading false and manipulative information.

7 DISCUSSION AND CONCLUSION

We provided an analysis of visits to fake news websites during the 2016 US presidential election campaign. We are sensitive to several limitations of our work, and to the many questions that remain unanswered.

First, our analysis is limited to considering visits in the IE and Edge browsers. It remains to be shown if similar trends occur for other browsers and in mobile scenarios—51.7% of Facebook's worldwide active monthly users access the site exclusively from mobile devices.[11]

Second, defining fake news is a complex issue, and it can be hard to verify and disambiguate fabricated stories from biased reporting. While we relied on a third party definition, we found that many of the websites in our analysis include a mix of both fabricated and non-fabricated (but possibly biased) information.

Finally, while our model can explain the observed trends, it is difficult to fit its parameters to our data — fitting requires labels for user "types" (political affiliations) and exposure rates.

ACKNOWLEDGMENTS

We thank Michael Golebiewski, Andrey Kolobov and Ryen White at Microsoft, as well as Einat Orr at SimilarWeb for helpful discussions.

REFERENCES

[1] Hunt Allcott and Matthew Gentzkow. 2017. Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives* 31, 2 (May 2017), 211–236.

[2] Anthony Faiola and Stephanie Kirchner. 2017. How do you stop fake news? In Germany, with a law. *The Washington Post* (April 2017). https://www.washingtonpost.com/world/europe/how-do-you-stop-fake-news-in-germany-with-a-law/2017/04/05/e6834ad6-1a08-11e7-bcc2-7d1a0973e7b2_story.html

[3] Justin Kosslyn and Cong Yu. 2017. Fact Check now available in Google Search and News around the world. (Apr 2017). https://blog.google/products/search/fact-check-now-available-google-search-and-news-around-world/

[4] Sapna Maheshwari. 2016. How Fake News Goes Viral: A Case Study. *The New York Times* (Nov 2016). https://www.nytimes.com/2016/11/20/business/media/how-fake-news-spreads.html?_r=2

[5] Craig Silverman. 2016. This Analysis Shows How Viral Fake Election News Stories Outperformed Real News On Facebook. *Buzzfeed* (Nov 2016). https://www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook

[6] Jacob Soll. 2016. The Long and Brutal History of Fake News. *Politico Magazine* (Dec 2016). http://www.politico.com/magazine/story/2016/12/fake-news-history-long-violent-214535

[7] Eugenio Tacchini, Gabriele Ballarin, Marco L Della Vedova, Stefano Moret, and Luca de Alfaro. 2017. Some Like it Hoax: Automated Fake News Detection in Social Networks. (2017). Preprint available at https://arxiv.org/abs/1704.07506.

[8] Alex Thompson. 2016. Parallel Narratives: Clinton and Trump supporters really don't listen to each other on Twitter. *Vice news* (Nov 2016). https://news.vice.com/story/journalists-and-trump-voters-live-in-separate-online-bubbles-mit-analysis-shows

[9] Edward Tufte. 2006. *The Cognitive Style of PowerPoint: Pitching Out Corrupts Within (2nd ed).* Graphics Press, Cheshire, Connecticut.

[10] Jen Weedon, William Nuland, and Alex Stamos. 2017. Information Operations and Facebook. (Apr 2017). https://fbnewsroomus.files.wordpress.com/2017/04/facebook-and-information-operations-v1.pdf

[11] https://venturebeat.com/2016/01/27/over-half-of-facebook-users-access-the-service-only-on-mobile/