Can we measure Internet Flooding?

Neha Lingareddy and Sarika Mahbubani

Abstract

In this paper, we explore if it is possible to measure Internet Flooding in the United States through 3 conceptual approaches. First, by performing a macro-level analysis of attention displacement regarding news events from 2013-2014; second, through a Twitter analysis of former President Trump's tweets from 2016-2020 to detect intentional tactics of flooding; and finally through an analysis of the timing of politically costly acts, namely the passage of executive orders by presidents, and how they correspond with public discourse regarding said president on Twitter.

1 Introduction

In her book 'Censored: Distraction and Diversion Inside China's Great Firewall', Margaret E. Roberts defines flooding as the "coordinated production of information by an authority with the intent of competing with or distracting from information" [9]. Flooding is a prevalent form of censorship that is cheap and accessible to authorities and groups, even in democratic countries that do not practice other forms of censorship.

Hence, to be able to understand Internet Censorship in many countries, understanding and measuring Internet flooding is a valuable perspective. However, it is unclear if flooding can even be measured. This is due to the implicit assumption of obscurity that accompanies it: namely, how would we know if an issue has been flooded out of public discourse successfully if we are never made aware of it in the first place?

Furthermore, if flooding effects can be measured, it is unclear whether we can prove intent behind such tactics. This article explores how we can begin to detect and measure flooding by discussing concepts of attention displacement, along with exploring metrics for flooding from the dominant public and political social media in the United States, Twitter.

2 Related Work

Most of the related work we were able to find were primarily from the field of political science. These papers discussed ideas related to propaganda and conventional mechanisms of internet censorship such as fear, friction or outright limits on information access.

The primary related work that we used to ask research questions about flooding was 'Censored' by Dr. Margaret Roberts [9]. In this text, flooding is introduced as a tool which displaces attention away from controversial content or sources of information. When considering whether flooding counts as a form of censorship, the detection of intent is thus important to determine if there is a deliberate suppression of access to information. Flooding can be directed on different groups: for instance, targeting the public directly through freely accessible platforms like Twitter, or targeting the media and hence indirectly reaching the public. In Roberts' view, flooding is most effective when there is high demand for low-cost information, wherein the public is not willing to spend time to understand nuanced or alternative viewpoints: actors can thus pump the information ecosystem with low-cost information to drown out sensitive topics. Flooding is not without consequences, however: it can often result in increased public distrust or the spread of misinformation online.

'Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics' by Benkler et. al [2] discusses how propaganda spreads through intentional communications by targeting specific populations. The authors believe that propaganda and misinformation are distinct and separated by intention: both work in different own realms and patterns of diffusion. The book concludes that the media environment is populated with propaganda, which represents a significant threat to democracy. We found this work to be helpful in defining intention behind flooding, and it helped us consider if our examined instances of flooding are more closely related to propaganda or misinformation.

'Fear, Friction, and Flooding: Methods of Online Information Control' by Dr. Margaret Roberts [8] discusses more theoretical concepts about flooding (the relationship between propaganda and flooding, and how the spread of information can be related to the spread of information). The paper also contains an analysis of how political campaigns in China use flooding not just through the internet but through other mediums like posters, radio, and television.

Finally, 'The politics of distraction: Evidence from presidential executive orders' by Djourelova and Durante [3], presents a historical analysis that indicates the timing of presidential executive orders, which are controversial policies due to their unilateral nature, have historically been released strategically. Namely, executive orders are signed during periods when public attention is focused on other areas, hence diluting the focus away from potential controversy. Our third approach to measuring flooding stemmed from this paper.

As far as we can tell, there were no other studies specifically studying various quantitative metrics of Internet flooding.

3 Macro Analysis of the Attention Economy

For flooding to occur successfully, a central assumption must hold true about the way the public processes information—namely, that attention is a zero sum game, wherein increased attention on a given topic detracts attention from another. To measure whether this assumption is reflected in reality, we searched for evidence that increased discussion about a given topic or event decreased discourse surrounding a secondary topic/event. To do so, we found a dataset from the University of Chile containing Tweets from 5234 news events that occurred rom August 2013 to June 2014. [10]

In order to classify the category the news event fell under, we used a dataset from the UCI Machine Learning Repository containing headlines from popular news articles along with their classification category (from 4 categories b = business, t = science and technology, e = entertainment, t = health). We then developed a logistic regression classifier from the UCI training dataset to apply classification categories to the historical Tweets dataset. Note that a politics classifier was not available from the UCI dataset: however, after a visual inspection of news stories classified by our Machine Learning model, we found that the overwhelming majority of political events were classified into the 'business' category. [5]

We then ran numerical analysis on Tweet volumes to find how the average number of tweets per news story in a given category varied across the categories. We found that the average number of tweets for entertainment topics, at over 10,000 tweets, far exceeded tweets for all others. The most commonly tweeted about event in the entertainment category was the 2014 death of Maya Angelou, with over 500,000 tweets, followed by the Superbowl and Australia's 2013 New Year celebrations.

To measure whether events in a given category tends to displace discourse surrounding an event in another category, we plotted the total number of tweets per category across time from 2013 to 2014. First, we plotted events in the entertainment category (blue, left y axis) against all events in other categories (red, right y axis). We found that for most of 2013 discourse seemed to vary with either no relation or in phase,

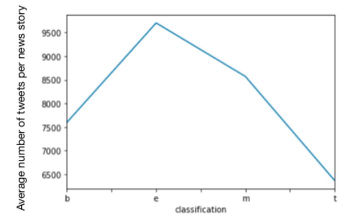

Figure 1: Mapping Discourse Quantity vs Category

whereas for 2014 there were several periods where the different categories had an inverse relationship, with tweets in a given category increasing as tweets from another category decreased.

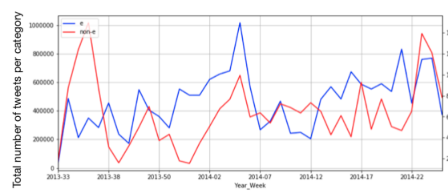

Figure 2: Entertainment vs non Entertainment Displacement

We recreated the aforementioned analysis with the categories of business events and technology events (tweets classified in the health category were too low for robust analysis). We found the same relationship with these two categories: namely that the number of tweets were in phase for some of 2013 while having an inverse relation for the rest of the time period.

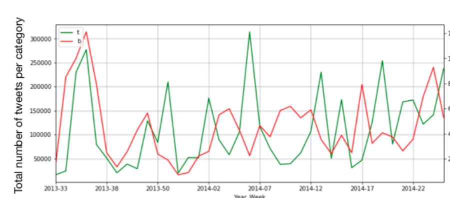

Figure 3: Business vs Technology Topic Displacement

Our findings showed there indeed are periods of time where discussion in a category is inversely related to discussion in another category. While our analysis cannot prove a causal story, such as how and why talking about the Superbowl actively made people forget about the refugee crisis, attention displacement does seem like the most intuitive explanation for the observed effects. Thus it does seem like one of the central assumptions of flooding—the zero sum nature of attention—holds weight. Thus the next question is as follows: how we can detect and measure the efficacy of intentional flooding efforts?

The code for this section is included in the cited github repository. [6]

4 Tweet analysis to detect intentional evasion tactics

A key question we had for this section was as follows: can you measure intention to distract or compete with other information?

We decided to look at social media, particularly Twitter, where information is constantly in competition based on people's attentions. In particular, we looked at President Trump's tweets during his presidency and asked - is Donald Trump using Twitter as a means of flooding? If so, what metrics best could best describe flooding, and can we differentiate between tweets that use flooding and his regular tweets?

The data collection involved first finding tweets we could use. This part involved scraping tweets from a Kaggle dataset [7], which contained all of Trump tweets as a president until June 2020. This was corroborated with another dataset of Trump's Tweets [1]. We loaded this csv into a dataframe, and took datapoints after 1st January 2016, as we hoped to study the period where Trump held a position of authority as President.

We then had to find scandals around which we could observe possible flooding. For this, we used a collated list [**mcsweeney**] of scandals during the Trump presidency (which had sources listed as NPR, the New York Times, the Washington Post). We scraped the list using Beautiful Soup, and stored the dates the scandals occurred, and their description in a dataframe.

To find data on tweets during scandals, we joined both dataframes on common dates. We then calculated the average number of the tweets, number of favorites, number of likes for all tweets, and for tweets during a scandal. We found the results in Figure 4.

In these results, we found it interesting that during periods of scandal, there was more user engagement through favorites and retweets: generally ways to indicate positive reception. We also found it interesting that on average, there were less tweets during a period of scandal than less politically contentious times. Our original assumption was that flooding would involve a high number of tweets to cause confusion and divert attention. However, these results made us think about how it could be possible to produce less content whilst still strategically engaging in flooding. If Trump posted fewer tweets that nevertheless contained emotive language unrelated to the scandal, user engagement on said distractions can still be increased to cause flooding.

calculation	tweets during scandals	all tweets
average # of tweets	13.0	14.4
average # of favorites	85420.7	69382.7
average # of retweets	19140.4	16080.6

Figure 4: Average number of tweets, favorites and retweets during a scandal and regularly

We also plotted graphs that demonstrate differences in the number of tweets/retweets/favorites per day over time (Figure 5). This visually shows that the number of tweets per day were generally high for all tweets, and that the user engagement peaks for all tweets coincide with periods of scandal, suggesting that users are engaging with Trump's tweets most during a scandal. This could be because the content that they are viewing is more engaging, familiar or aligned to their interests. It is also possible users are using Twitter more and engaging as a show of support or curiosity during a scandal.

Apart from these metrics, we also explored using wordmaps on some tweets that coincided temporally with scandals to analyze the content of said tweets. Below, there are two wordmaps that show the topics of the tweets made during a sexual assault allegation and during the Trump University scandal. As we can see the most tweeted about topics appear to be dissimilar to the topic at hand, and we wanted to see if there was a way to quantify this phenomenon.

To quantify this numerically, we found a technique called Word2Vec that embeds words into vectors and compares them through cosine similarity. We then created vectors from the text about the scandal, which was collected during parsing along with the text of the tweet. We then calculated the cosine similarity of these two texts. The average of the cosine similarity between the description of the scandal and the text of the tweets over time was -

$$0.10932879992642042$$

As expected, the cosine similarity was low between the words of the tweets and the words in the description of the scandal, which suggests that the tweets made during a time of scandal are generally unrelated to the scandal at hand.

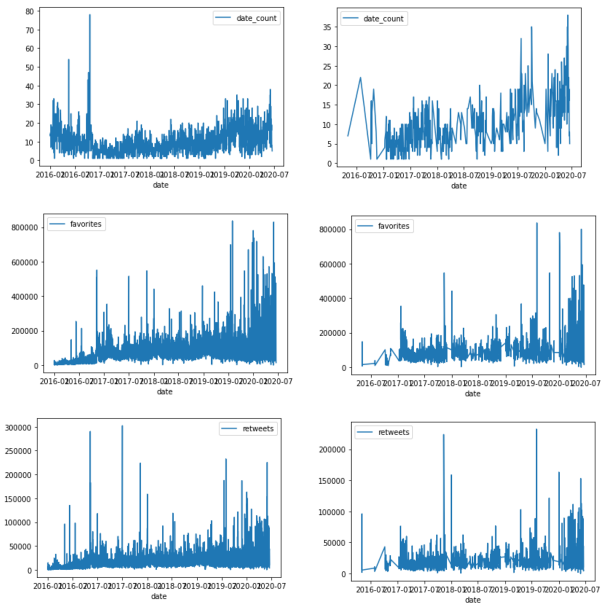

Figure 5: Graphs that show the number of tweets, favorites and retweets over time for all tweets (left) and tweets during scandals (right)

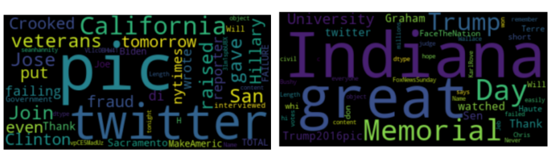

Figure 6: Wordmaps from tweets during scandals. Left shows tweets during a sexual assault allegation, and right shows tweets during the Trump University scandal

The code for this section is included in the cited github repository[4].

5 Analyzing the timing of political controversial actions and how they correspond with public discourse

Having looked at intentional efforts to distract from issues, the next question we asked was as follows: do some actors strategically time their actions to take advantage of the natural rhythms of the attention economy? To do so, we sought to replicate Djourelova and Durante's research regarding executive orders and public attention, in which they found that presidential executive orders are historically scheduled during periods where attention is diverted away from the president (as measured by analysis of primetime TV broadcasts). We decided to once again focus on Twitter and the Trump presidency for our analysis.

To measure whether public attention was focused on Trump during a given time frame, we captured his presence on Twitter's trending page in the USA from 2016-2021. Namely, we measured whether the top hashtags for a given day were related to the former president. To get a list of hashtags that related to president Trump, we used the Twitter API to write a script for finding secondary hashtags that frequently appear with trump. While this did lead to some helpful hashtags, like maga and potus, the most frequently co-occurring hashtags were politics, world and leftwing— not specific enough to relate specifically to former president Trump. We thus condensed the gathered list to specific hashtags that were directly related.

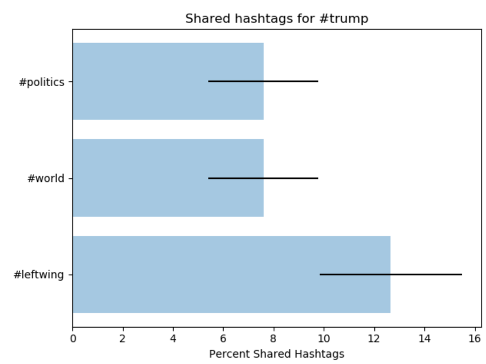

Figure 7: Shared Hashtags for Trump

To get a record of the frequency of such hashtags on the trending page, we scraped an online archive of trending topics (https://us.trend-calendar.com/) using Python's BeautifulSoup library. We then plotted the frequency of Trump's appearance on the page across time, wherein 1 indicates an appearance on the page and 0 indicates the opposite. As seen in Figure 8, Trump tended to be a trending topic very frequently during the end of his presidency, a stark contrast to the sparse beginning. This makes intuitive sense, given the history of political polarization and scandal that progressed for the duration of his presidency.

To map whether the former president's executive orders were timed during periods of sparse Trump-related discourse, we mapped every executive order passed during his presidency across time on the same map, wherein a green dot indicates the signing of a given executive order (figure 9).

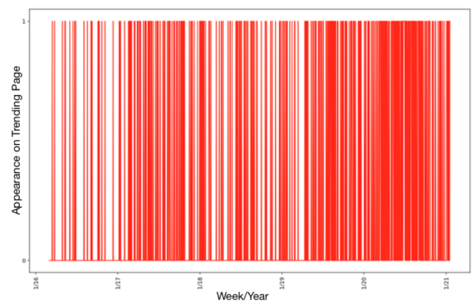

Figure 8: Frequency Map of Trump as a Trending Topic on Twitter

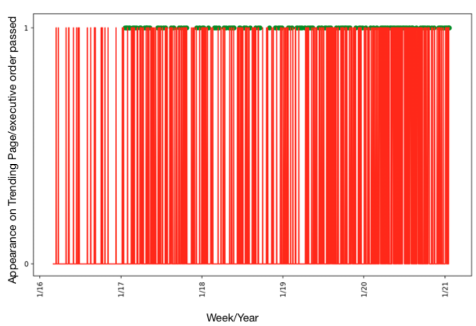

Figure 9: Mapping Trump Related Discourse against Executive Order Timing

We then zoomed in on specific time periods, as denoted by the graphs depicting executive orders in from 2017-2020, to analyze any patterns across time. We found that in the years prior to 2020, executive orders did indeed seem to occur during periods of sparse Trump-related discourse. A likely explanation for this is flooding, wherein his administration sought to take advantage of lulls in political controversy to quietly pass orders that might otherwise be latched upon during periods of controversy. This pattern, notably, was broken in 2020, as indicated in the corresponding graph. Historically, the political frenzy and polarization of 2020 drastically increased discourse related to Trump, and thus finding lulls in controversy would have been exceedingly difficult. Additionally, given that 2020 was the last year of his presidency that also occurred during a pandemic, it makes intuitive sense that previous patterns would be broken in this year.

Our findings were corroborated quantitatively: we found that the average time difference between an executive order and the *last* time Trump was trending was 4 days, and the average distance between an order and the *next* time he was trending was 3 days. Comparatively, the average difference between incidences of Trump trending was only 0.84 days. The relatively large time period between an executive order and surrounding trending incidents indicates that the administration likely waits for political lulls to reduce public attention—if collective attention is not currently on the president, then scrutiny on him is likely at a periodic low.

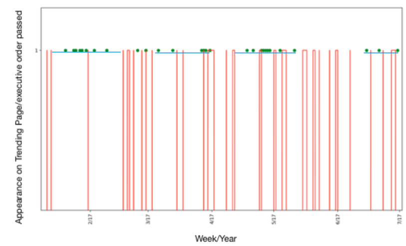

Figure 10: Zoomed in 2017

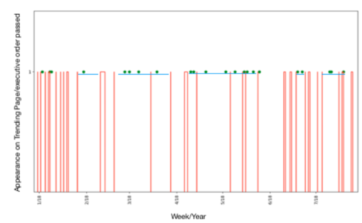

Figure 11: Zoomed in 2018

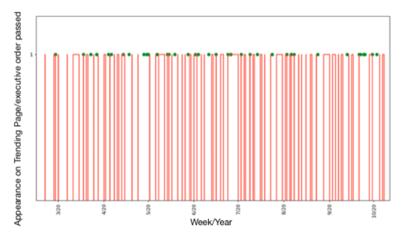

Figure 12: Zoomed in 2020

Moreover, we found that the difference between periods preceding executive orders tended to grow larger when we looked at the second, third, fourth, etc *last* incidents of trending, whilst periods *following* an executive order tended to stagnate. For instance, the time difference between an order and the second last time Trump was trending was 8 days, whereas the difference between the second next time was still 3 days. This likely occurs as a cluster of trending incidents likely represents periods of genuine controversy, whereas a one-off trending event might simply be a result of public musing. Thus looking at the last n and next n trending incidents, where n is large, is probably a more accurate measure of genuine controversy preceding and following an executive order respectively.

The fact that the average time period is relatively higher between an order and the *previous* trending incidence compared to the *consequent* incident might indicate that controversy tends to follow the executive order despite strategic timing efforts. Alternatively, it might indicate that the president felt it safe to ignite another controversy after having successfully passed an executive order quietly. Regardless, we find compelling evidence that the Trump administration did indeed use flooding efforts in the passage of executive orders.

The code for this section is included in the cited github repository. [6]

6 Conclusion

Overall, we find that our analysis includes compelling evidence of flooding. Firstly, our macro analysis of the attention economy indicates an inverse relationship between the volume of discourse across categories. While this is not enough to conclude a causal relationship, it does provide supporting evidence that attention is a zero sum game. To actually certify causality, a natural experiment needs to be set up wherein we can locate a control group. For instance, assuming political discourse occurs in every state in the US, we could measure the impact of a regionally specific natural disaster on the volume of political discourse in two states. The state not experiencing a disaster would serve as an effective control to assert causality in the state that is: if political discourse is suppressed in the latter state but not the former, that is compelling evidence that disaster-related discussion actively suppresses political discourse.

Secondly, our Twitter analysis showed that during periods of scandal, the volume of the former president's tweets lowered while the user engagement with his tweets increased. This suggests that if flooding is occurring on the platform, it aims to distract with content rather than volume. Also, after analyzing the word similarity between words relating to the scandal and tweets on the day of the scandal, we found dissimilarity between the words used. This suggests that the tweets made made during scandals could be trying to divert users' attentions. Some of the limitations of this part of the study are that we worked with a limited dataset of Donald Trump's Tweets (2016-2020) and only used some metrics to study user engagement. There are other means of measuring user engagement such as user comments, shared links, and other discussions about the tweet, and our current research does not capture that. Additionally, we did not analyze the sentiment of the tweets, so we are not sure what it is about the tweets that increases user engagement. In future work, we hope to include other measures of user engagement, and analyze language sentiment, to observe if there was any relation between these factors and flooding.

Lastly, analyzing the timing of executive orders against the incidence of trending topics indicates that the Trump administration timed their orders during lulls in discussion centered on him. While this is not enough to certify intention, deliberate flooding does seem like a likely explanation, especially in light of corroborating evidence from Djourelova et al. We could strengthen our existing analysis by mapping discourse not just as a function of time but of intensity: a higher position on the trending page indicates larger participation in discussion, presumably indicating a bigger political scandal. Furthermore, we could extend our method by analyzing the timing of other politically controversial actions, such as the appointment of controversial cabinet members or the pardoning of friends and close affiliates.

References

[1] Jason Baumgartner. "Twitter Tweets for Donald J. Trump (@realdonaldtrump)". Version V1. In: (2019). DOI: 10.7910/DVN/KJEBIL. URL: https://doi.org/10.7910/DVN/KJEBIL.

[2] Yochai Benkler, Hal Roberts, and Robert Faris. "Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics". In: (2018).

[3] Milena Djourelova and Ruben Durante. "The politics of distraction: Evidence from presidential executive orders". In: (2018).

[4] Neha Lingareddy. "GitHub repository". In: (2021). URL: https://github.com/wookiee2187/Internet-Censorship/blob/main/works.ipynb.

[5] "Machine Learning News Dataset". In: (). URL: https://archive.ics.uci.edu/ml/datasets/News+Aggregator#.

[6] Sarika Mahbubani. "GitHub repository". In: (2021). URL: https://github.com/sarikam1.

[7] Austin Reese. "Kaggle - Trump Tweets". In: (2020). [Online;] URL: https://www.kaggle.com/austinreese/trump-tweets.

[8] Margaret E. Roberts. "Fear, Friction, and Flooding: Methods of Online Information Control". In: (2014). URL: http://nrs.harvard.edu/urn-3:HUL.InstRepos:12274299.

[9] Margaret E. Roberts. "Censored: Distraction and Diversion Inside China's Great Firewall". In: (2018), pp. 80–88.

[10] "Twitter News Dataset". In: (). URL: https://users.dcc.uchile.cl/~mquezada/breakingnews/.