Analyzing Controversial Topics within Facebook

Clifford L. Short
Auburn University, cls0095@auburn.edu

Gabrielle Taylor
Auburn University, grt0007@auburn.edu

Ashish Gupta
Auburn University, azg0074@auburn.edu

Xiao Qin
Auburn University, xqin@auburn.edu

Analyzing Controversial Topics within Facebook

Emergent Research Forum (ERF)

Clifford Short
Auburn University
cls0095@auburn.edu

Xiao Qin
Auburn University
xqin@auburn.edu

Ashish Gupta
Auburn University
azg0074@auburn.edu

Gabrielle Taylor
Auburn University
grt0007@auburn.edu

Abstract

Social media plays a significant role in the dissemination of information. Now more than ever, consumers turn to social media sites (SMS) to catch up on current events and share their perspectives. While this form of communication is enjoyed by the public, it also has its drawbacks. Because many perspectives can be captured via SMS, this often leads to public discourse and in some cases, controversy. Misinformation and disinformation continue to spread throughout the internet allowing many consumers to become misinformed. This further elevates such discourse and allows for real issues to be forgotten as online debate spirals out of reality and false information gains traction. Given the issues at hand, this paper seeks to demonstrate a rudimentary measurement of curve fitting as a proof of concept for capturing controversy on Facebook using the reactions of its user base toward controversial topics.

Keyword

social media, controversy detection, Facebook, machine learning, human-computer interactions

1. Introduction

This paper addresses quantifying controversy on the social media site, Facebook, to provide possible solutions for the spread of misinformation on the site, as well as highlight possible relationships that can be made through studying the way in which the consumer base engages with controversial content. To do this, a naïve model of curve fitting is proposed to test the feasibility of such research. The following factors motivate this study to quantify controversy using Facebook interactions.

Social media plays an integral part in how people consume information and interact with each other (Levy 2021; Shearer and Mitchell 2021). With approximately 2.89 billion monthly active users during the second quarter of 2021, Facebook is the most popular social media site worldwide (Morengo et al. 2021). Controversy, and therefore controversial topics, bring various perspectives to light as the public participates in discourse as an attempt to ascertain an answer to the subject at hand. The overarching motivation for this project is an interest in helping mitigate the spread of false information and to understand the relationship between those who interact with controversial content and false information.

In this paper, we propose a naïve model using curve fitting as a baseline for the conceptualization of the patterns present in the data with a goal of creating foundational work for future studies in controversy detection concerning Facebook as a platform of interest. These models look at the ratio of "love" reactions to other reactions in a 2D plane. Definitions used for the reactions in this study are based off those provided by Merriam-Webster Dictionary. Note here that "Haha" will be perceived as "funny" for all intents of this study.

2. Related Work

Several recent studies have made attempts to capture the relationships and characteristics of controversy. These approaches to controversy detection can be summed into two categories: structural-based and content-based.

Structural-based approaches seek to detect controversy using available metadata and graphing techniques as seen in Kittur *et al.*'s work on tracking conflict and coordination on Wikipedia (Kittur et al. 2007). In another study, Choi *et al.* created graphs that tracked controversial topics and their subtopics using sentiment (Choi et al. 2010). These arrangements focus more on the metadata and clustering of related topics, hence the attention to structure.

Content-based approaches use the content of the documents to identify controversy. Content here can range from textual and sentiment analysis to argument structure. In a paper published by Garimella *et al.*, it was determined that the standard deviation of sentiment is a powerful linguistic indicator of controversy (Garimella et al. 2018). Rumshisky *et al.* demonstrated there was a positive correlation between the standard deviation of sentiment and the random walk controversy (RWC) measure (Rumshisky et al. 2017). This paper's approach would fall under structural as a determination using Facebook meta metrics is the focus.

3. Methodology

In this section, methodology will be broken down into three parts. Data collection is discussed in Section 3.1. Exploratory Analysis of the data takes place in Section 3.2. In Section 3.3, the experiment is described in further detail.

3.1 Data Collection

Facebook has recently provided researchers with a way to collect data on their platform in the form of an online dashboard called "CrowdTangle" (Boberg et al. 2020; Punziano et al. 2021). Using CrowdTangle, information on public posts such as when they were created and how consumers interact with them are available. For this paper, the focus will be on the reaction metrics, excluding likes. This comes from an assumption that the effort needed to like is minimal when compared to reacting to a post. Eleven (11) total subsets were collected with each subset examined containing roughly 100 data entries, except for a larger data set of approximately 1000 entries where experiments of time series analysis were performed. These subsets include known controversial keywords such as "abortion", "same-sex marriage", "gun control", "Israel", and a more recent topic of "COVID-19 vaccine". Patterns extracted from these subsets are then compared to each other to verify the pattern is not an anomaly and compared to 5 more subsets of non-controversial data including "tobacco", "Netflix", "birthdays", "cancer treatment", and "cats" to demonstrate patterns or the lack of those seen in the keyword data set. The larger data set is an expanded look at the "abortion" keyword to check for trends in time and if reactions can capture those trends.

3.2 Exploratory Analysis

Exploratory analysis was performed on the data and a few patterns were identified. The first of these patterns indicated a large presence of "Haha" reactions to posts that were related to the keyword "COVID-19 vaccine". This discovery is interesting as one could describe the keyword as more serious in nature considering current events. Therefore, it is believed that, along with "Angry", this sort of reaction could be thought of as a sign of resistance and perhaps signal controversy. Second, duration of availability does not seem to influence popular post consumption. Third, in data sets containing controversial keywords, there was a surprising number of posts in which reactions that show support like "Love" and "Care" are outnumbered by those that could be perceived as resistance like "Haha" and "Angry".

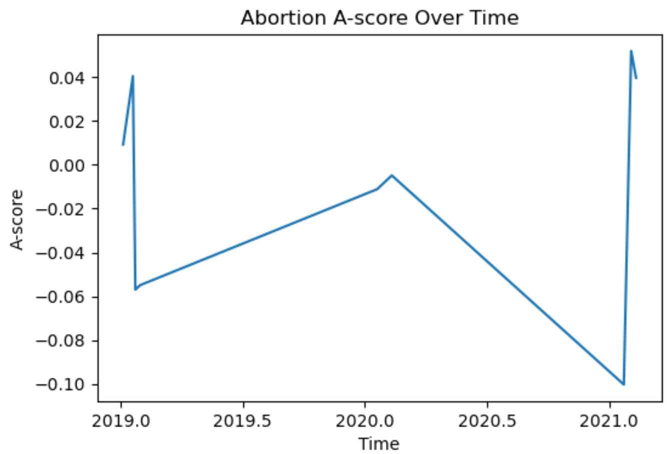


Figure 1: Time Series of Trending Controversy in Abortion

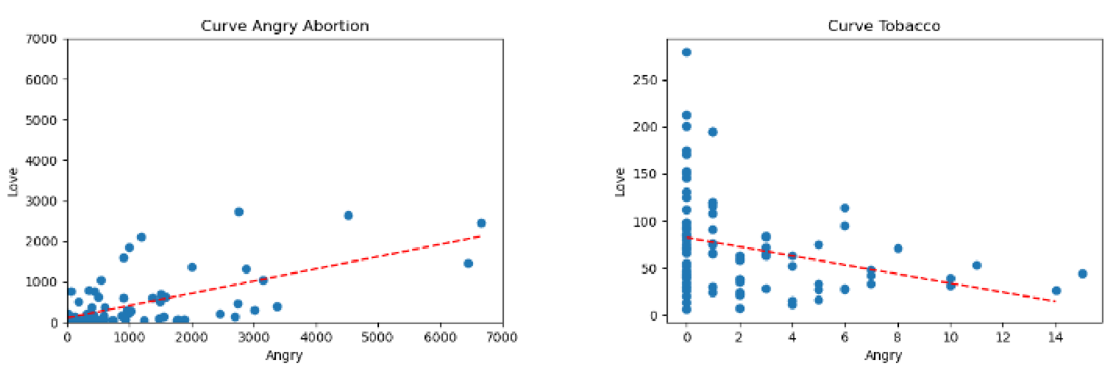
One last observation seen was that reactions were capable of tracking trends. In an experiment using the large "abortion" data set, nine equal subsets were made over a two-year period and then plotted as a time series. As demonstrated in Figure 1, abortion was covered less between 2019 and the beginning of 2021 and received less interactions as a result. In the summer of 2021, abortion is seen trending again as controversial legislature is passed by Texas lawmakers in the U.S. regarding abortion. Reactions can capture the effects of real-world events on people who are aware of them.

3.3 Quantifying Controversy

The naïve model used in this study will be curve fitting. Scatter plots are generated based on exploratory analysis that suggested there are patterns in the distribution of reactions when compared to one another. An example of these graphs can be seen in Figure 2. Please note that axes for Figure 2b are uneven. This was intentional to demonstrate what a possible non-controversial graph could look like, thus zooming in was deemed appropriate. Using this knowledge, curve fitting is adapted to capture this pattern. A simple linear regression equation

$$y_{cont} = a + b_{discourse}x, where\ b \in (-\infty, \infty)\ and\ a \in [0, \infty)$$

is used as the desired curve as those with higher degrees start to fold on themselves. Here, y_{cont} is the measurement of controversy for the topic queried on Facebook and $b_{discourse}$ is a measurement of discourse figured by the 2-dimensional plane of love and another reaction. As a is an indicator of y-intercept and has no identified relevance in this case, it is omitted from consideration. For calculation purposes, since a will always be a constant, a = 0. With the slope captured, trends can be more easily discerned given a variation in distributions from any topic.



(a) Abortion (b) Tobacco

Figure 2: Examples of Curve Fitting Results

As $b_{discourse}$ can also represent an overwhelmingly positive or negative response, this value will be normalized using the following formula:

$$c = \frac{b_{discourse} - min_b}{max_b - min_b} \; where \; c \in [0,1]$$

Before this can take place, numbers existing outside of the set [0, 1] must be handled so that the measurement is not greatly affected by the range of $b_{discourse}$. The smallest negative number is set to zero by taking the absolute value and then that value is applied to the rest of the $b_{discourse}$ scores gathered. If the number is above one, an inverse is recorded by dividing 1 by that number. After transformations have been performed, normalization takes place, and the result is compared to a threshold of 0.5.

4. Results

Using the methods prescribed in Section 3, the controversial score (c-score) for each subset was collected. Each topic's c-score and their classification based on the "Love/Angry" dimension is reported and the accuracy of the results are highlighted in Table 1. As seen by the results, all known controversial topics discussed trended controversial. The larger data set was also reported to be controversial. This demonstrates the naïve model's ability to capture this trend over a bigger collection of opinions. Regarding the "Is Correct?" column, for ground truth, a table listing controversial issues maintained on Wikipedia was used to compare the results. As this list is human moderated and consistently fact-checked, and Wikipedia has been shown to identify controversial issues, we felt this pick for ground truth was appropriate. The experiment that covered how a controversial topic trends over time revealed that current events and public awareness of the issue help to increase the controversial nature of the topic. It is also worth noting that the smaller subset of abortion covered back to summer 2021 and as such received a high c-score due to the increased interaction with the topic.

Topic	C-score	Is Controversial?	Is Correct?
Abortion	0.759588	Controversial	Yes
C-19 Vaccine	0.667919	Controversial	Yes
Same-sex Marriage	1.0	Controversial	Yes
Gun Control	0.669115	Controversial	Yes
Israel	0.627743	Controversial	Yes
Tobacco	0.337172	Non-controversial	Yes
Abortion (Large)	0.517809	Controversial	Yes
Netflix	0.514248	Controversial	No
Cancer Treatment	0.0	Non-controversial	Yes
Birthday	0.332787	Non-controversial	Yes
Cats	0.410565	Non-controversial	Yes

Table 1: Results of Curve Fitting Method

Other reaction dimensions were considered but performed significantly worse than that of the "Love/Angry" dimension. All reactions were compared to "Love" as love intuitively highlights strong support. "Care" was considered but its presence was noticeably less than "Love". When compared with "Angry", the proposed model achieves 90.9% accuracy. As assumed earlier, "Haha" reactions performed second best; however, when a secondary experiment was performed where both "Angry" and "Haha" reactions were equally considered, this resulted in a similar accuracy to that of "Haha". This attempt was labelled "Aggregate". Given that the proposed approach has shown an ability to capture controversy, its limitations will now be discussed.

These results are largely restricted by the platform in which the data was obtained. Context is an important part in measuring controversy; however, CrowdTangle does not currently allow access to comments made on these posts. Another limitation is the target audience of Facebook. Given that Facebook is maintained by U.S.-based tech company, Meta, and the data used was filtered for the English language, the target audience is largely westernized. This causes the capture of international controversies such as the status of Taiwan and the events of the South China Sea to be missed as interactions received are mostly biased toward one perspective. As discussed earlier, this model is naïve. It is not indicative of a final solution and is influenced significantly by noise. It also struggles with more broad, complex topics of controversy like Israel. For example, there are several aspects in which Israel could be considered controversial. These include the Israeli/Palestinian conflict, corruption in the government of Israel, and actions taken against the people of Gaza.

Non-controversial topics also suffer. As reported by the model, Netflix is controversial. While it is known that Netflix in general is not controversial, some posts written by mass media could have stirred up the reactions of people. One such story that occurred during the time frame captured was that Netflix was being sued by a South Korean internet service provider for the increased bandwidth traffic caused by popular South Korean drama, "Squid Game".

5. Future Work

In this paper, we presented a naïve model of curve fitting to identifying and quantifying controversy on Facebook. Future work for this includes refining the methods used to better capture controversy on Facebook. This would most likely be done by incorporating a more complex model able to capture some of the more intrinsic characteristics of controversy. Being able to apply context to this approach would probably benefit it greatly based on previous methods discussed earlier in the paper. Along with context, another possibility is assigning weights to each reaction and fine tuning a neural network to identify and measure controversy. Another consideration would be the expansion of the data scope to include more perspectives. This would diversify opinions and provide the opportunity to capture more international controversies. One last thought includes the use of sentiment analysis as another measure for controversy. Controversy is marked by the distinct divide in opinions on a topic and sentiment analysis should be able to capture this divide given context from the post. This line of work might also be useful from a sociology perspective including theories like Social Identity, conflict, and other observations of human behavior like confirmation bias, bubble filter, and echo chambers.

REFERENCES

Boberg, S., Quandt, T., Schatto-Eckrodt, T., and Frischlich, L. 2020. "Pandemic Populism: Facebook Pages of Alternative News Media and the Corona Crisis-A Computational Content Analysis." *arXiv preprint arXiv:2004:02566* (2020).

Choi, Y., Jung, Y., and Myaeng, S. 2010. "Identifying Controversial Issues and Their Sub-topics in News Articles," in *Pacific-Asia Workshop on Intelligence and Security Informatics*. Springer, pp. 140-153.

Garimella, K., Morales, G. D. F., Gionis, A., and Mathioudakis, M. 2018. "Quantifying Controversy on Social Media," in *ACM Transactions on Social Computing* 1, 1 (2018), pp. 1-27.

Kittur, A., Suh, B., Pendleton, B., and Chi, E. H. 2007. "He Says, She Says: Conflict and Coordination in Wikipedia," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp 453-462.

Marengo, D., Montag, C., Sindermann, C., Elhai, J. D., and Settanni, M. 2021. "Examining the Links Between Active Facebook Use, Received Likes, Self-esteem and Happiness: A Study Using Objective Social Media Data," in *Telematics and Informatics* 58 (2021), 101523.

Levy, R. 2021. "Social media, news consumption, and polarization: Evidence from a field experiment," *American Economic Review* 111, 3 (2021), pp. 831-870.

Punziano, G., Marrazzo, F., and Acampa, S. 2021. "An Application of Content Analysis to CrowdTangle Data: The 202 Constitutional Referendum Campaign on Facebook," in *Current Politics and Economics of Europe* 32, 4 (2021), pp. 371-397.

Shearer, E. and Mitchell, A. 2021. "News Use Across Social Media Platforms in 2020," in (2021).

Rumshisky, A., Gronas, M., Potash, P., Dubov, M., Romanov, A., Kulshreshtha, S., and Gribov, A. 2017. "Combining Network and Language Indicators for Tracking Conflict Intensity," in *International Conference on Social Informatics*, Springer, pp. 391-404.