#MeToo Twitter Data Analysis

Andres Tapia, Athalia Rochelle Handowo, Jeannine Gabe Tiemann

Advisor: Dr. Elena Zheleva

Introduction

The #MeToo movement began in 2006 with Tarana Burke to bring emphasis to sexual assault cases, and it saw a resurgence in 2017 when Alyssa Milano exposed the sexual misconduct of Harvey Weinstein. It has since then become an incredibly controversial topic, inundating Twitter's trending page with #MeToo tweets of both men and women sharing their experiences. Countless articles addressed the issue from both liberals and conservatives, and many of these articles were referenced in tweets from political figures and everyday Twitter users alike.

Our group wanted to find what topics liberals and conservatives focused on, but also we wanted to know what everyday users focused on. Did users tweet article links that aligned with their political leaning, or did they tweet from the opposing side? We analyzed the moral foundation (MF) frames[1] in tweets and news articles referenced in tweets related to the #MeToo movement. We considered these MF frames when analyzing the tweets' texts and the political leaning of the referenced news sources[2]. From this data, we can begin to draw connections between traditional news sources and unguided Twitter user opinions.

Background and Related Research

In our research, we use moral foundations defined in previous research and try to match the contents found in tweets to match these moral foundations. Moral foundations were first defined by Jonathan Haidt[1], and the list of values he initially defined has expanded into the following set of moral keys: Authority Virtue, Authority Vice, Fairness Virtue, Fairness Vice, Harm Virtue, Harm Vice, Ingroup Virtue, Ingroup Vice, Purity Virtue, Purity Vice[6]. These moral frames can be used to highlight some key differences between certain groups of people, which we specifically designate in our research as conservatives and liberals. According to Haidt, liberals

tend toward values that align with change and justice. Conservatives, on the other hands, want to preserve order and value loyalty.

Initially, research on the moral foundations utilized the Moral Foundations Questionnaire (MFQ) to test groups of people. However, this first version of the MFQ was primarily to examine any ideological differences between liberals and conservatives. Some of the questions asked when developing the next version of the MFQ were "kicking a dog in the head" that was categorized under Harm, for example. The MFQ was expanded to measure differences in "moral relevancy"[7] amongst individuals rather than comparing liberals and conservatives as groups. We can still use this to determine the moral leaning of conservatives and liberals. By using the "Big Five" personality facets (openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism[10]) and the MFQ, we can see the trends in certain moral foundations for the two groups.

In terms of obtaining data to test these moral keys, Twitter has become a hub for individuals wanting to express a variety of political and social issues. The vast array of locations that users reside in provide insight on what types of views or moral keys are prevalent in different parts of the country. Some of the analysis done on specifically liberal and conservative users included taking note of how they interacted with individuals that align with their political views and with users of the opposite viewpoint, finding the "most differentiating words" used in their tweets, and analyzing the progression of their tweets and content[9]. The results of these studies have shown conservatives to display a higher volume of identity, such as nationality, and authority, such as law or administration. Whereas liberals tended more toward emotional responses, such as feelings or tendencies toward obscene language[9].

The #MeToo movement is yet another social issue that has placed a divide between liberals and conservatives. Studies have been conducted contrasting men and women's opinions and "stereotypes" in morality alignment[11]. While some have shown morality alignments for certain religious groups like Christianity, Judaism, Islam, etc.[12] Using these studies as a basis, we analyze the morality alignment between liberal and conservative groups in the #MeToo movement discussion.

Problem Statement

In our research, we took a set of tweets that linked to articles referencing the #MeToo movement and conducted different analyses on them to determine what the conversation was. From the tweets, we wanted to compare the frequency of moral foundation (MF) keywords in news articles and the tweets referencing them. Additionally, we wanted to find any differences in MF keys between liberal and conservative new sources. By marking these differences, we could also compare the main topics being discussed by liberal sources against topics discussed by conservative sources.

Analysis

The tweets we studied were included in a dataset collected by Dr. Zheleva and the EDGES lab. To conduct our analysis, our group used Python and Jupyter to write our scripts and utilized the goose3 and pandas libraries to parse through the dataset. From the initial dataset, we first moved the news articles to a separate CSV file. With this new CSV, we were able to use the goose3 library to obtain the following article data: URL, author, article title, publication date, and article body. Using the pandas library, we could further highlight moral foundation keywords and count the frequency of these keywords. By finding these keywords, we analyzed each tweet's moral values by performing keyword matching based on the MF dictionary. Then we scraped the article for its political leaning and how far left or right it resided. Afterward, our group compared the news source's moral values to its political leaning[3].

Next, we wanted to know what major topics are being discussed. For that, we cleaned the dataset and used the Latent Dirichlet Allocation (LDA) model to decide the category that certain words mapped to. Further, we used a bag of words to track the frequency of certain words. Using LDA and a bag-of-words, we created a word cloud, and it became very easy to see the focal points of the conversations surrounding the #MeToo movement.

Results

We count the total of each news source tweet's moral values and each news source tweet's words. Then we divide each news source's moral values with each news source tweet's word count. Thus we get the frequency of each moral value of the news sources, then assemble the heat map below. We found that right leaning news sources were more likely to include HarmVice while left leaning news sources are more likely to have Morality General. Words that may have indicated HarmVice include "war," "harm," "abandon," etc. Morality General indicators may be "proper," "bad," "ethic," etc. (fig 1).
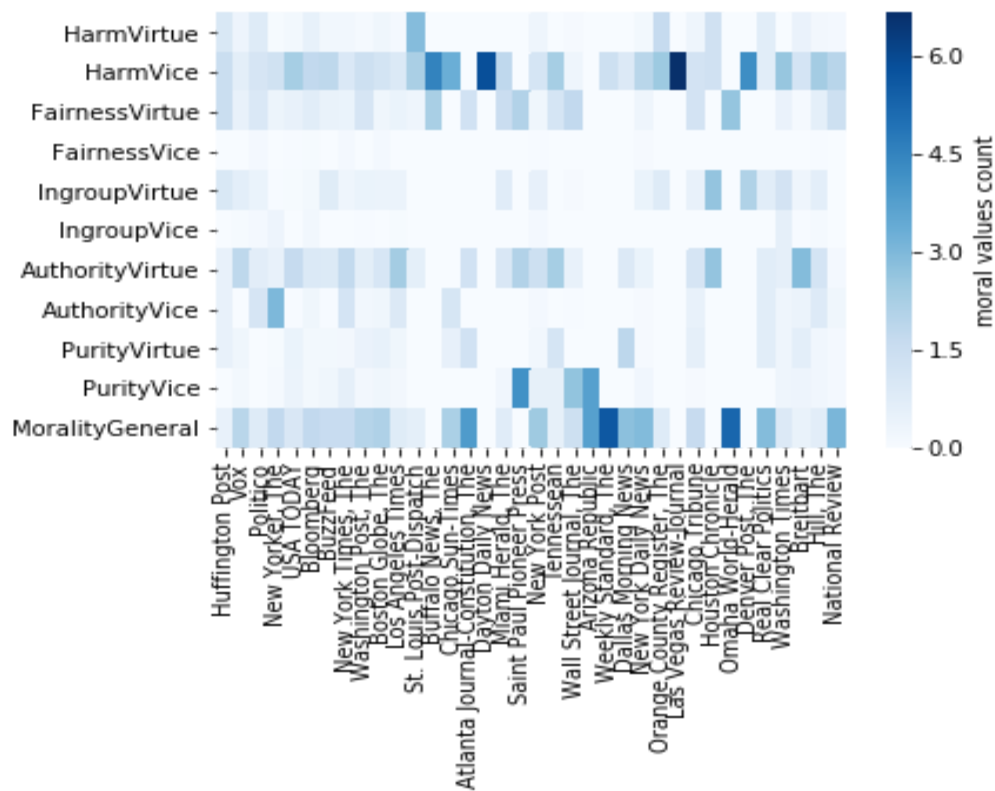

Figure 1

This word cloud was plotted using matplotlib.pyplot depicts major keywords in 3 topics relating to the tweets. The bigger the word is, the more prevalent it is in the topic. For example, Topic 3 has keywords relating to allegations and scandals. So, it can be inferred this topic is sexual harassment scandals.

Figure 2

The visualization shown in Figure 3 was created in Jupyter Notebook using the pyLDAvis and gensim libraries. Using Latent Dirichlet Allocation (LDA), the words from the #MeToo tweets were all tokenized, cleaned, and put into a LDA model. Like the word clouds (fig. 2), this model shows which words were most prevalent. However what is different from the word clouds is that the visualization shows which words were more prevalent among the whole dataset and among seven topics found and calculated using the LDA model. For example, the words "women," "abuse," and "schneiderman" are some of the most prevalent words in the dataset. It can be inferred that many of the tweets have to do with Eric Schneiderman, an attorney general who was accused of sexual abuse by four women.
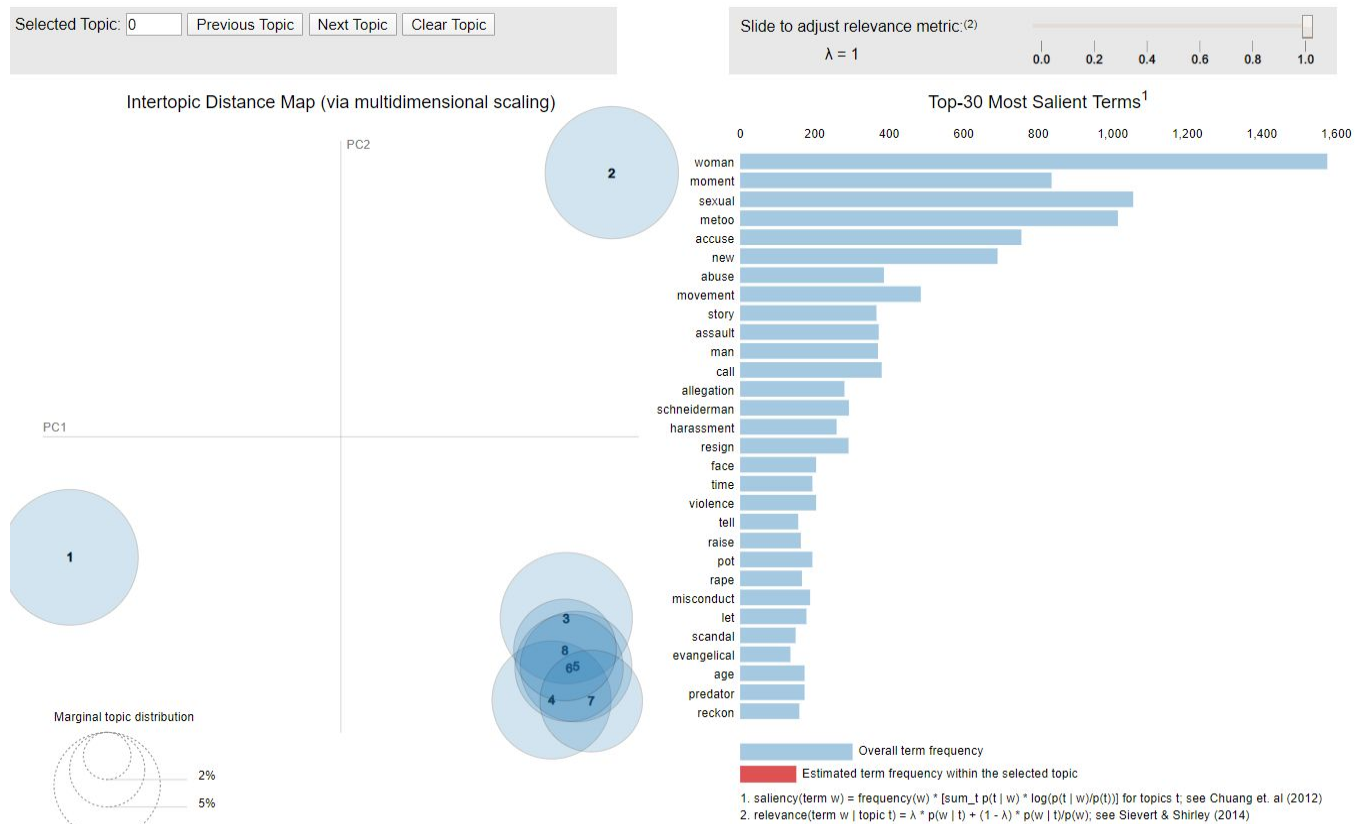
Figure 3

Conclusion

From the work already done, we can already see that right leaning and left leaning sources gravitate toward different moral keywords. We can make deductions as to whether users are tweeting these articles to confirm their beliefs or to begin a conversation with the "other side." If the work is to be continued, additional research can be done on other datasets that are based on other trends on Twitter.

Acknowledgments

References

1. Haidt, J., and Joseph, C. "Intuitive Ethics: How Innately Prepared Intuitions Generate Culturally Variable Virtues." *Daedalus*, vol. 133, no. 4, 2004, pp. 55–66., doi:10.1162/0011526042365555.

2. Quantifying moral foundations from various topics on twitter conversations. IEEE; Dec 2016.

3. Mikulincer M, Shaver PR, Cooper ML, Larsen RJ. APA handbook of personality and social psychology, volume 4: Personality processes and individual differences. US: American Psychological Association; 2015.

4. Wihbey J, Coleman TD, Joseph K, Lazer D. Exploring the ideological nature of journalists' social networks on Twitter and associations with news story content. 2017 Aug 22.

5. Clifford, S., Iyengar, V., Cabeza, R. *et al.* Moral foundations vignettes: a standardized stimulus database of scenarios based on moral foundations theory. *Behav Res* 47, 1178–1198 (2015). https://doi.org/10.3758/s13428-014-0551-2

6. Fulgoni, D., Carpenter, J., Ungar, L., Preotiuc-Pietro, D. An Empirical Exploration of Moral Foundations Theory in Partisan News Sources. http://wwbp.org/papers/mf2016lrec.pdf.

7. Graham, J., Nosek, B., Haidt, J., Iyer, R., Koleva, S., Ditto, P. Mapping the Moral Domain. J Pers Soc Psychol. 2011 Aug; 101(2): 366–385. https://doi.org/10.1037/a0021847.

8. Graham, J., Haidt, J., Nosek, B. Liberals and Conservatives Rely on Different Sets of Moral Foundations. Journal of Personality and Social Psychology, 2009, Vol. 96, No. 5, 1029 –1046. https://doi.org/10.1037/a0015141.

9. Sylwester, K., Purver, M. Twitter Language Use Reflects Psychological Differences between Democrats and Republicans. PLOS ONE. 2015 Sept. 16. https://doi.org/10.1371/journal.pone.0137422.

10. McCrae RR, John OP. An Introduction to the Five-Factor Model and Its Applications. J Pers. 1992; 60: 175–215. https://doi.org/10.1111/j.1467-6494.1992.tb00970.x.

11. Niazi, F., Inam, A., Akhtar, Z. Accuracy of consensual stereotypes in moral foundations: A gender analysis. PLOS ONE, March 5, 2020. https://doi.org/10.1371/journal.pone.0229926.

12. McPherson, R. The Influence of Spirituality, Moral Reasoning, and Personality Factors on Misogyny. The Pegasus Review: UCF Undergraduate Research Journal (URJ): Vol. 10 : Iss. 2 , Article 5. 2019 Mar. 25.