#MeToo: Polarization and Discourse in the Digital Age

Niharika Jain
niharika.jain@asu.edu

Director: Dr. Avital Simhony
Second Committee Member: Dr. Paul Lewis

Arizona State University
Barrett, the Honors College
School of Politics and Global Studies

Online platforms have become explosive tools in carrying out political discussion. The introduction of this contemporary medium inevitably impacts already-tense relations between characteristically different groups. This work discusses the divisive influences of social media on the online community, measures the impact of political polarization on the digital vernacular, and illustrates a case study with the #MeToo movement on Twitter.

Table of Contents

Author's Note.....	4
I. Online Interactions: Failures of the Fully-Connected World.....	5
♦ Filter bubbles, echo chambers, and cross-cutting interactions.....	5
♦ Group identities and polarization.....	10
♦ Attacks and outrage.....	13
♦ Discourse in social media.....	16
II. Case Study: Facilitating Uncivil Discourse.....	19
♦ Study and methods.....	19
♦ Challenges.....	22
♦ Results and discussion.....	26
▪ Angry and abusive.....	27
▪ Encouraging and empathetic.....	31
III. Final Remarks.....	34
♦ Conclusions.....	34
♦ Epilogue.....	35
IV. Reference.....	36
♦ Notes.....	36
♦ Glossary.....	37
♦ Appendix.....	38
♦ Bibliography.....	47

Author's Note

Social media is explosively popular in discussing socio-political issues. This work provides a preliminary study on how polarization occurs online. Chapter I begins by introducing limitations of the internet in maintaining a free flow of information. Not only do users seek out groups of like-minded individuals and insulate themselves from opposing views, social media platforms algorithmically curate content such that it will be in line with a user's preconceived notions of the world. The work then defines polarization and carefully discusses its most prominent causes. It then shifts focus to analyze a closely-related issue regarding political discourse: outrage, which is both a noticeable effect of and further cause of polarization. It is clearly prevalent in traditional media, but for completion, I provide a case study to measure its incidence in social media. In Chapter II, I scrutinize the language used in the #MeToo movement on Twitter and draw conclusions about the issues Twitter users focus on and how they express their views. This chapter details the method I used, the challenges I faced in designing the exploratory study, and the results I found. I benchmark patterns I find in the Twittersverse against those I find in *The Wall Street Journal*. The analysis relies upon the metric of word similarity, based on proximity of and frequency of words used together, to make distinctions about what users are most commonly saying with respect to given topics, or keywords. Chapter III closes the essay with conclusions of socio-political polarization, discourse, and outrage in social media. Finally, the essay outlines potential channels for future work.

PART I

Online Interactions: Failures of the Fully-Connected World

Filter bubbles, echo chambers, and cross-cutting interactions

Social media has become an attractive channel and one-stop shop for politics. Candidates and elected officials have direct access to voters and constituents; they can bypass the paywalls of and save time from advertising and news media by using free services to reach their targeted listeners with their messages. These social media platforms are designed to spread content – users share posts on Facebook and retweet tweets on Twitter – so these campaigns can go viral without ever having to endure the regulations and hardships (especially during election season) of paid advertising. They can also be used to generate feedback from public opinion in real-time. At the advent of any issue or controversy, users will take to their keyboards to respond. Insights from mass reactions will allow politicians to adjust their campaigns or public statements. Social media also increases the degree to which political activism can take place; voters and constituents can use it to circumvent the factors preventing large-scale organization to create more effective petitions. Ease of access is characteristic of the internet. Anyone with WiFi connectivity has a microphone: the power to communicate with strangers, augment his network and political standing, and adjust messages in response to quick feedback.

As more information flows in than any human can consume in the golden information age of the internet, there is a need to curate content to prevent information overload -- too much to choose from and too many voices to hear. In his book, *#Republic: Divided Democracy in the Age of Social Media*, Cass R. Sunstein opens with an anecdote: Nicholas Negroponte, an MIT researcher, felt the trends of the internet in 1995 presaged the idea of the “Daily Me,” where

users would only receive the information they cared about. If the user is interested in sports and poetry, but hates hearing about global politics, his Daily Me might only consist of basketball and Shakespeare, without any reference to a heightening of tensions between the United States and Iran.

From this prediction made in 1995, the internet has proved itself to largely become a Daily Me. Audiences are targeted by demographic and messages are tailored at individual levels. Relevance-based algorithms curate the content users see to ensure maximum engagement. Eli Pariser, in a 2011 TED Talk, calls this curation “filter bubbles.” These algorithms use a simple metric of number and timeliness of clicks to measure how likely one is to interact with certain information. If a user likes or comments on more of his liberal friends’ posts on Facebook than conservatives’, he will find right-leaning posts on his timeline trend toward being substituted by left-leaning ones. Google curates search results in the same way; its optimization processes oftentimes warp the significance of certain current events on the basis of likelihood of engagement with the results. Pariser illustrates an example with the Google results of the same query, “Egypt,” for two individuals. One individual’s results displayed links to news articles discussing protests; the other’s results were more centered around tourist attractions and experiences (Pariser 5:47, 2011). The speaker implies there is a degree of danger from having the events of the world “filtered” to match comfortable notions of society before it reaches audiences and without their knowledge of consent. His ideal world entails an information balance between relevant ideas similar to our already-formulated opinions and notions that will be slightly uncomfortable and challenge our worldview. Relevance-based optimization is the antithesis of what the internet actually stands for, he proposes: a free flowing of information.

Even when not hindered by social media algorithms out of users' control, they gravitate toward users who think like them. This includes those who have been raised in similar conditions, those who share the same religious values, or those who identify with the same political party. Users have a tendency toward insulating themselves from ideas contrary to their own worldviews. It's simply more comfortable to receive reinforcement from your network of friends and followers than criticism. Political experts have termed these pockets of similar ideas which have been insulated from outside opinions "echo chambers." Users will find that the information they are exposed to from others in their network are just louder echoes of their own beliefs. Echo chambers reinforce themselves. The ease and comfort of entering into echo chambers provide no attractive motivations for individuals to seek out opposition; when insulated, they can grow confident in their network and in their own validated opinions.

As one would expect, the content to which users are exposed plays a significant role in how they view the world; this seeking of aligned content leading to more insulated worldviews runs in a cycle. It also is not limited to social media platforms. In his work, Sunstein details a study made by Gregory J. Martin and Ali Yurukoglu of Stanford University exploring the effects of cable news content on voting habits. A key finding from their study showed that watching just four additional minutes per week of Fox News increased the probability of voting for the Republican candidate by 0.9 percent; on the other hand, watching MSNBC for four extra minutes weekly decreased this probability by 0.7 percent. The effects are, unsurprisingly, much greater if the viewing time is increased. (Sunstein 2017). The content that the media exposes people to is crucial to influencing their opinions!

Sunstein also discusses a disturbing finding regarding traditional news media from a 2006 study: audiences are more likely to choose their information provider based on how closely it is

known to align with their ideologies and parties. Unsurprisingly, Democrats prefer NPR and CNN, while Republicans prefer Fox News (Sunstein 61, 2017). This is true even when the news networks discuss the exact same stories! This finding indicates a perversion of the Daily Me; viewers are not basing their viewership loyalty on the topics of the content they will be exposed to, but rather on their notions of how aligned their information providers will be with respect to their own beliefs of the world. Where information comes from influences how eager people will be to listen to it.

However, one might point out a flaw in the previous finding: participants were only given five options in the study (NPR, CNN, Fox News, BBC, and “Can’t say”). It is more useful to actively track which media sources Democrats and Republicans visit. A 2016 article published by *The New York Times* details a study that does just this; the results find that a majority of people, both Democrat and Republican, consume a relatively centrist media diet: the two most common destinations were AOL.com and MSN.com. Conversely, when measuring the number of visits to media destinations instead of the number of people, the results are disproportionately skewed toward extremes: Republicans are far more likely to consume content from right-leaning sources (such as Fox News and *Breitbart*). A sizeable fraction of all Republican political news consumption was visits to these more conservative websites. The mirrored trend is found among Democrats, who largely consume content from left-leaning sources (such as *The Huffington Post* and Daily Kos). The proportion of this extreme consumption to total Democratic consumption is also sizeable, though smaller than the Republican proportion. Further, the study interestingly finds that left- and right-leaning news sources cater stories to this small group of voracious readers.

The takeaway from these two studies is that though most do not confine themselves to polarized news media, those who engage in politics often do live in media-based echo chambers of their own construction, and “exercise a disproportionate influence over our political system,” (Nyhan 2016). This trend can also be loosely mapped to social media; most users on the online platforms do not engage in political discourse, but those who do will seek out content from sources that aligns with their preconceived notions at a higher rate than those who do not. However, it is important to make the distinction between content which comes from news media and that which comes from one’s network; these studies indicate users’ self-curation with respect to the websites of news sources they frequented. It does not provide any insight on the automatic self-curation that occurs through creation of an insulated network of friends and followers, who largely share similar socio-political views.

So far, we have discussed people’s tendencies to contain themselves within closed, insulated spaces to protect themselves from information that would conflict with their preconceived notions of the world. The network of friends and followers that users on social media platforms create would indicate this, but this is not a problem exclusive to the internet; news media that has grown more aligned with specific ideologies also create a problem of exclusive viewership among members of specific parties. Finally, even if one wanted to branch out and listen to views from the other side, the algorithmic filter bubbles the internet creates to curate a “Daily Me” for each individual pose a challenge.

Does this depiction of the internet unjustly paint it as being too bleak? After all, the internet has the ability to increase access to opinions from other viewpoints; they are just a click away. Sounman Hong and Sun Hyoung Kim, in their 2016 study, “Political polarization on twitter: Implications for the use of social media in digital governments,” create a term for this

competing view to echo chambers: “cross-cutting interactions.” The hypothesis that social media outlets provide an online forum for interacting with differing political perspectives would indicate the notion that politicians who have more moderate views than their peers have a higher following, because this would pull audiences with opposing ideologies toward the center (Hong and Kim, 2016). Hong and Kim’s findings cause them to reject this hypothesis and to support the notion that social media platforms only serve to fragment the political community, not bind it (supporting the original echo chambers hypothesis).

Group identities and polarization

The structure of the internet allows users to form larger and more niche communities, which leads to a cementing of group identities. These strengthened identities polarize users on socio-political issues and create intolerance against opposing points of view. In her *Washington Post* article, “Party polarization is making us more prejudiced,” Lilliana Mason writes of members of opposing groups, “We stereotype them, we dislike them, and we evaluate them unfairly. The stronger our identities, the stronger our response to inter-group rivalry,” (Mason 2014). Mason terms the tendency to make attacks against opponents when the prestige of one’s own social group is challenged as “social polarization.” There is a human instinct to defend our social groups. Social polarization is, then, inter-group conflict birthed from in-group loyalty – the collective inter-group attacks driven by instincts for our “team” to win, or at least not to feel like losers.

There is a related phenomenon with respect to polarizing effects of groups: “group polarization,” as termed by Cass R. Sunstein, refers to the tendency of people in homogeneous groups of like-minded individuals to move toward an extreme point of view after deliberation

with and even just exposure to the other members' views (Sunstein 2017). These ideas are in the same direction the individuals' collective inclinations. Group polarization is the name for the concept that in-group membership and discussion push potentially moderate members to believe more extreme versions of the convictions they already held.

The polarizing effects of groups fuel a cyclic chain of events. Alignment among ideological, partisan, racial, and religious values in the past few decades has paved the way for groups of like-minded individuals to meet, deliberate, and push each other toward more extreme versions of their beliefs. As members of groups find more reasons to coalesce and form singular identities, there is a tendency for both heightened conflict among groups and heightened insulation within groups from challenging external views. It has already been discussed that the creation of insulated echo chambers leads to further strengthened group identities because these ideas are reinforced. It is particularly significant that inter-group conflict further strengthens group identities; this, too, inevitably fuels the cycle. Both effects – insulation and conflict – are symptomatic of a polarization problem.

Before discussing the effects of polarization, it is useful to review its causes. Polarization needs fervent group identities to survive. It occurs as a result of a limited argument pool (Sunstein 2017); when group members offer arguments in the same direction as pre-existing notions, it reinforces original convictions and pushes them away from the median position. There is an illusion of corroboration of points from new arguments, but they are unsurprisingly skewed. In an insulated, like-minded group of individuals, there are disproportionately less (if any) in favor of the opposing side. Another factor is that the volumes of all voices in the group are rarely equal. Within a group, it will be natural for individuals to adjust their own positions

toward the dominant one. People are concerned with their reputations; aligning themselves with the loudest (and often extreme) perspective provides another level of comfortable reinforcement.

The implications behind these causes are great. Exposure to other views within a group is sufficient for the occurrence of radicalization¹ toward extremes; deliberation is not necessary. Exposure to a point of view within the group will lead to more impassioned conviction in the same direction as that point of view. This is not the case for a point of view from a member external to the group. In their article, “Exposure to opposing views on social media can increase political polarization,” Christopher A. Bail et al. design an experiment that measures ideological shift in both (self-identified) Republicans and Democrats² after following an automated Twitter account which posted hourly tweets in favor of the opposing ideology; Republicans were given a monetary incentive to be exposed to a liberal Twitter bot for one month, and Democrats were exposed to a conservative Twitter bot during the same interval. At the end of the treatment, the researchers found substantial evidence to show that Republican users became more conservative than they originally were prior to exposure to challenging points of view (Bail et al., 2018). Simple exposure to an in-group argument will shift an individual’s beliefs to an extreme one in the same direction of the argument, but exposure to an out-group argument will shift the individual’s beliefs further and in the opposite direction from the argument. In other words, movement toward extremes in the direction pre-existing ideas is much easier than movement away. Regardless of which side another point of view comes from, if there is no thoughtful and open-minded deliberation, individuals exposed to it will just polarize in the directions of their own original beliefs.

Strengthened group identities and insulation from out-group arguments are not without their positive impacts. “Enclave deliberation,” another term coined by Sunstein to indicate the

form of deliberation found in insulated groups, gives a voice to historically and disproportionately underrepresented minorities. Echo chambers, characteristic of safety and reinforcement, can serve as an incubator for new ideas which would have otherwise been nonexistent or silenced among larger deliberative bodies. Today's society also places unequal importance on the views and desires of certain groups of people above others. Deliberation within mixed groups is likely to squelch the opinions and views of minorities within the group. The *#Republic* author makes a point that enclave deliberation is the most productive in this scenario when the insulation is not prolonged; it is best used as a method of nurturing the development of arguments behind new perspectives. Once these exist, enclaves have no more provisions of positive influence on the political community (Sunstein 2017).

Strengthened group identities, a shift toward extremes, willing insulation from external positions, and widening polarization naturally will have their impacts on the discourse among individuals of different backgrounds and beliefs, should they choose to have it.³ It will be crucial to understand the disparity among groups to understand the political atmosphere, especially with respect to social media.

Attacks and outrage

One of the key components of political discourse today is incivility: any indication of a lack of respect for or frustration with the views of opponents. In other words, it occurs whenever individuals are not courteous with one another (for any reason). In their paper, "From Incivility to Outrage: Political Discourse in Blogs, Talk Radio, and Cable News," Sarah Sobieraj and Jeffrey M. Berry briefly discuss the prior research on the effects of political incivility on a community: "the reigning presumption is that incivility in politics undermines faith in

government and discourages political participation” (Sobieraj and Berry 21, 2011). Since political incivility is so closely coupled with appeals to emotion and one’s sense of his own value to the community, it is (expectedly) complex. Uncivil – even outrageous – language can be productive. Negative political speech can raise individual attentions toward a particular issue and inspire greater voter turnout. Emotional speech is crucial for revolutionary movements; indeed, this country would not have seen its hallmark civil rights movement without it.

Still, it is largely considered to have negative effects: presented in particular ways, its use shifts the collective attention away from thoughtful reflection and discussion on issues. Sobieraj and Berry term the variant of incivility which has these drastic consequences as “outrage”: “a particular form of political discourse involving efforts to provoke visceral responses,” (Sobieraj and Berry 20, 2011). The researchers measure the extent of outrage in discourse in traditional media outlets: talk radio, political blogs, cable news analysis programs, and newspaper columns. While their findings are incredibly insightful, they do not correspond exactly to the platform of the internet, or specifically, social media. Cable and print news, blogs, and radio are all information providers; audiences tune in or read to remain active in their understanding of the current state of politics. These media platforms, however, are unidirectional. Lay citizens cannot use them as media through which they can transmit their own views. Traditional media only allows audiences to receive curated, expert opinions. Even political blogs, which may not necessarily need to undergo rigorous vetting to ensure analysts are well-qualified and experienced, still necessitate multiple rounds of editing and content selection. Appreciation for the effects of traditional media is fundamental to understanding the political era where everyday citizens can converse about socio-political events within an online web, but it is not sufficient to understanding it.

Outrage in political discourse is dangerous. It can engender intolerance, widen the effects of polarization, and foster distrust in the government. The researchers acknowledge its complexity and difficulty to define. Sobieraj and Berry limit it to a combination of 13 factors: “insulting language, name calling, emotional display, emotional language, verbal fighting/sparring, character assassination, misrepresentative exaggeration, mockery, conflagration, ideologically extremizing language, slippery slope, belittling, and obscene language” (Sobieraj and Berry 26, 2011). Their method segmented the content of all forms of media and analyzed number of occurrences of the aforementioned 13 factors. The overall tone of each incident they examined was negative.

Outrage saturates political commentary in traditional media. They find 89.6% of cases in their samples across all forms of media contained at least one incident of outrage. The medium which displayed the least amount of outrage was political blogs, with 82.8% of posts containing outrage in writing. The three most dominant forms of political outrage they found in their samples were mockery, misrepresented exaggeration, and insulting language (even excluding language used to name-call) (Sobieraj and Berry, 2011).

One of the least-occurring forms of political outrage was sparring; the authors attribute this surprising finding to the lack of heterogeneity in opinion found on these platforms. For example, guests on a single episode of talk radio and cable news shows rarely have radically different views. It is interesting to note this would remain true for social media. One might immediately assume there is greater chance for sparring on the internet, given ample chances for users to engage with those with opposing opinions. However, echo chambers may sufficiently insulate users to produce the same effect as traditional media; recall the cross-cutting interactions hypothesis did not find support in Hong and Kim’s study.

A final remark on the outrage discussion is that the level of outrage in media is on the rise. The authors were unable, of course, to conduct any thorough time-series studies for political blogs, cable news, and talk radio, since those only came into existence or gained popularity recently. They did, however, find that journalism in 1955 and 1975 was not impassioned in its language despite the tumultuous political atmosphere of the times. “From Incivility to Outrage” was also written in 2011; in the digital age, this is almost prehistoric. Social media has erupted as a platform and overhauled political discussion since then. Note that the revolutionary #BlackLivesMatter movement – arguably the first catalyst for political discourse on Twitter – started in 2013. Following the trends found in the study, prevalence of outrage in political discourse in all forms of media have undoubtedly risen since 2011.

Discourse in Social Media

There remains exhaustive work to be done on the discourse found in the transformative medium of social media. These online platforms can influence the political vernacular of a community. The character limit is the distinguishing feature of the tweet; at 140 characters for most of the duration of its existence and beta releases of 280 beginning a year and a half ago, Twitter is a medium for discussion on political issues with concise and easy-to-digest, albeit highly opinionated, statements from across the spectrum.

In the following chapter, I present a case study of the discourse surrounding the #MeToo movement on Twitter. It is necessary to useful to isolate a single issue to ease analysis of which aspects of the events that audiences place focus on. It also eases analysis of the language used in the Twittersverse as contrasted with reasonably unbiased language; if there is one singular issue, one can make greater assumptions about the implications behind prevalence of specific words. It

is also useful to choose a movement associated with a hashtag to reap the benefits of its use as an organizational tool; the data collection process returns significantly more relevant data since hashtags allow users to self-sort their tweets as pertaining to specific issues. It is thus an easy and relatively safe assumption to make under usage of a hashtag that all tweets collected which contain the phrase are relevant to the issue at hand.

This specific hashtag is suitable to present as a case study because individuals across the political and ideological spectrum shared (and continue to share) their thoughts through the hashtag. This is notable because it is not necessarily common for political hashtags, especially given the magnitude of the movement. Cass R. Sunstein references a 2009 study by Sarida Yardi and danah boyd which found that the hashtags users had adopted signaled their stance on a given issue (Sunstein 81, 2017). This is also seen with countless examples. The prominence of #BlackLivesMatter birthed #AllLivesMatter and the less prevalent #BlueLivesMatter. A popular hashtag among Democrats a few years ago was #ACA (for the Affordable Care Act), whereas the popular healthcare-related hashtags for Republicans were #Obamacare and #FullRepeal. In the 2016 campaign, two dominant hashtags were #CrookedHillary and #NeverTrump. While many political hashtags can contribute to polarization because users of opposing stances use them to separate their identities, #MeToo, despite its explosive scale, has maintained itself as a hashtag under which users with polar-opposite views are comfortable in publishing their thoughts.

As a final thought, it is crucial to note that this study will be inherently skewed toward the vocal users of Twitter; as discussed on page 9, most people are not politically active. Though this does provide a limitation in that it will not represent the entirety of Twitter users, it does not lessen the significance of the work: it will properly analyze discourse. Studying the language that

socio-politically active users use provides valuable insight on the impacts of social media on outrage and polarization.

PART II

Case Study: Facilitating Uncivil Discourse

Study and Methods

This work analyzes the differences in language of text written on the same topic over two media, Twitter and *The Wall Street Journal* (WSJ). The hypothesis is that users on Twitter employ extreme language. This includes words with abusive or strong emotional connotations, characteristic of ad hominem and character assassination attacks. The language of WSJ articles is intended to serve as a comparison group by being an impartial setting. This publication was chosen over others due to its reputation across all members of the political spectrum. According to a Pew Research Study in October 2014, “Political Polarization & Media Habits,” it is the only publication more trusted than distrusted across audiences within each of the five categories of the political spectrum: consistently liberal, mostly liberal, mixed, mostly conservative, and consistently conservative (Mitchell et al.). Other high-ranking publications include the Economist, BBC, and Google News.

For data collection, I made queries on the ASU library’s repository of *The Wall Street Journal* for all news articles⁴ including the term “#MeToo.” 521 articles were scraped, amounting to a vocabulary of 16,601 unique words (excluding those in a list of 153 stop words: words too common to contribute any meaningful information). I similarly made queries on Twitter via its premium search API on the developer platform. This work processes a collection of 83,412 tweets, written in English, which related to the #MeToo hashtag, ranging from the time of the first allegation against Harvey Weinstein in October 2017 to early March of this year. This corpus included a vocabulary of 34,055 unique words. All punctuation marks (including hyphens

and apostrophes) were removed and all letters were converted to lowercase as a preprocessing step ensuring uniqueness in the vocabulary (e.g. to prevent “assault” and “assault.” from being considered separate words).

The method I used to perform language analysis maps all words in the vocabulary of the two corpora to a lower-dimensional space, where distances between words correspond to their degrees of similarity. This process involves using a family of algorithms called “Word2Vec” which uses a two-layer artificial neural network to learn an encoding for each word in the vocabulary based on its surrounding words in all its occurrences (Mikolov 1, 2013). Each encoding is a 100-dimensional vector. I use the Gensim library implementation of Word2Vec for my study (Řehůřek 2010). Words that are more similar are closer together in their vector space. The distance metric used is cosine similarity, which measures the angle between the two vectors; the smaller the angle, the more similar the words. In this application, “similarity” will refer to any way two words can be used within a given context; these similar words have been used together in the same sentence. They fall under the same thematic categories: an example of a similar pair can be found in the original word embedding created by the first Word2Vec paper: the words “king” and “queen” display a close relationship. They are not quite synonyms, but they can often be used interchangeably in the same position of a sentence, which retains its thematic meaning (such as “The king ruled over the kingdom.” and “The queen ruled over the kingdom.”). It is also critical to note that word similarity allows for “rotations” in the semantic meaning behind the word; this metric of word similarity would place antonyms as being very similar, since they fall within the same thematic categories. Thus, words marked as similar are not necessarily similar in their underlying meanings, but rather in their underlying contexts.

The key analysis in this work will discuss how these close relationships differ across the two corpora, *The Wall Street Journal* and Twitter. The most similar words will be found for the same keywords across both platforms. Since *The Wall Street Journal* was chosen as a reasonably unbiased source, it will be considered under the assumption of having little to no extremism in its use of language. The sentiment for a given keyword's similar words over political tweets will be weighed and considered against the keyword's similar words over the WSJ articles. Some might argue the technique of analyzing similar word pairs does not grant enough evidence to draw conclusions about the socio-political vernacular and atmosphere pertaining to a given movement, since the single most similar word to a keyword does not need to have been used extremely often with the keyword; it only needs to have been used with the keyword more often than other words have.⁵ Comparing distances among word pairs' similarity will allow readers to understand the relative frequencies of word pairs to provide the evidence necessary to make inferences. I consider the relation between the cosine similarities of notable word pairs on page 32 at the end of *Results and Discussion* and in Figures 1-4 in the *Appendix*.

Of course, there are some disparities in the data which should be considered before making claims. *The Wall Street Journal* publishes formal writing. Each article is around 500 words, written by a professional in the field who must submit it for several rounds of editing. On the other hand, Twitter is used by the general public to articulate immediate thoughts and feelings. Tweets are constrained to 280 characters. Language of this corpus will inherently use a different, more casual, style of English. However, the meaningful results I find do not display a disparity in the formality of tone, but rather a disparity in the classes of words used – both among other tweets and articles in *The Wall Street Journal* – when discussing political and social issues over a social media platform.

There are also limitations in using Word2Vec and word similarity in my analysis. The most obvious fault in the study is that I am restricted to analyzing single words, and I cannot extend my analysis to phrases. Secondly, Word2Vec creates a unique embedding for each word, irrespective of semantic meaning. The embedding for the word “bank” would incorporate both the definitions, “financial institution” and “slope bordering a stream.” This certainly would pose a problem in understanding the potential contexts of ambiguous words. Fortunately, my results are not too ambiguous. Also, given the specificity of the corpus (tweets and articles related to the #MeToo movement), it is easy to rule out less-probable meanings.

Challenges

The data collection process was difficult and time-consuming. I chose to use the ASU library’s resources to extract WSJ articles. The program I used to scrape these articles used software designed for automated testing. Selenium, a Python library, allows a developer to automate the process of opening an internet browser and navigate to and within webpages (@SeleniumHQ, 2019). My code logged into the ASU Library’s archive portal and made a Basic Search query, “#MeToo.” It filtered the results to exclude editorials and scanned through all six pages of results to collect and the individual links to each article. I then used another Python library, Newspaper3k, to download just the text of the article from each link; this would exclude titles, author names, image captions, etc. (Ou-Yang, 2019). Individually downloading all 521 articles took several tries, and several hours for each try. The Newspaper3k library has a short default timeout value for all downloads; after 12 seconds, if the article has not successfully downloaded, the code will raise an error and terminate. The download speed slowed considerably after the first 50 articles were scraped. Understanding this behavior required poring

through documentation of the library. An unexpected challenge of using an ASU resource was the requirement to be connected to the internet through an ASU network during the download process. All employees of ASU must register for two-factor authentication; employees outside of ASU's network range must perform some small action on a separate device than the one attempting to access an ASU resource to verify they are the ones making the access request. This effectively prohibits automated software outside the ASU network to access an ASU resource. The implications for my research, I originally believed, were that I must plan to have my computer run through the hours-long download process while I was on campus. After several failed attempts, I was made aware of a mobility client to allow my computer to create a secure and verified connection to the ASU network, even while out of the physical range of WiFi. Upon this realization and increased flexibility in my data collection process, I successfully scraped all 521 articles.

To scrape tweets, I was constrained to using the Twitter Developer Platform. This platform is designed with enterprises in mind; it aids organizations in analyzing user activity to create effective ad campaigns and make advantageous business decisions. There is also a substantial research community which uses the Twitter API endpoints. I created a Twitter Developer account and completed an application for approval detailing the data I would collect and the aims of my study. In under a week, Twitter Developer approved my account to start using its endpoints. Separately, I applied for funding through the Honors College. Considering Twitter Developer largely serves enterprises, the prices of its services are substantial. I purchased a Premium tier subscription to the Twitter Search: Full Archive API. This gave me full access to search within all tweets published on the platform, dating from 2006. Twitter Developer generates access tokens for each approved account. Developers use these tokens to authenticate a

request for a bearer token; this bearer token is verified to access all Twitter API endpoints. This two-step authentication process was challenging to learn, especially with sparse and decentralized Twitter Developer documentation on the matter. After having obtained my authentication tokens, I was able to make the request to the service.

Twitter Developer recommends using Twurl to make all requests. This is a Twitter wrapper for cURL (Client for URLs), a command-line tool to receive or send files using URL syntax. I instead opted to use cURL to avoid further installation and setup. I learned the correct syntax to make requests to the Twitter API endpoint using both cURL, and requests, a Python library that created a programmer interface for this service. After I fully understood how to authenticate myself and access the Twitter API endpoint, I was left to make the search queries and collect the results. Considering the previous Twitter Developer hassles, this was remarkably straightforward, since Twitter Developer provides extensive documentation for using its endpoints (“Premium Search”). A slightly confusing feature of the API is that results were split up in “pages,” so each search query need to input a token indicating the next page (excepting, of course, the first). Twitter API endpoints are also rate-limited. Developers are only permitted to make 15 requests every 15 minutes. I wrote code that automatically froze the program for 15 minutes for every 15 requests it made. This naturally slowed the data collection process.

There were also several non-technical challenges in using the Twitter Developer API. The Search: Full Archive API returns a list of tweets in reverse-chronological order. It was necessary for the case study to collect ample tweets from over a year ago, since the #MeToo movement started in October 2017; the results were undoubtedly skewed after having only collected tweets from the past three months. I then collected more tweets, inputting “fromDate” and “toDate” parameters as October 5, 2017, and October 5, 2018, respectively, in my query to

alleviate this issue. After undergoing this second collection process, the final sample of tweets I analyzed can be considered to cover a similar period as the WSJ articles.

The original case study design entailed an exploration of how the language of liberal and conservative users differs. In relevant literature, researchers analyze followers of prominent elected officials, who self-identify with a specific party. They assume ideological/partisan alignment, or that followers of Republicans are conservatives and followers of Democrats are liberal. However, my case study was not simply analyzing tweets of followers, but the language of tweets specifically related to the #MeToo movement. This would have involved creating a search request for each unique follower. This would have proved to be unconceivably costly to collect a number of tweets appropriate for Word2Vec; my \$249 subscription tier caps the number of requests to 250.

Another interesting path I was unable to explore was performing sentiment analysis on the corpus of tweets and calculating the average degrees of negativity, positivity, and neutrality. These average numbers would have given insight into the degrees of outrage and polarization, as well as the encouragement and empathetic support symptomatic of convergence upon mutual understanding. One of the most popular Python libraries to perform this task is VADER by the Natural Language Toolkit (@nltk, 2019). Unfortunately, it did not prove very accurate for the corpus of tweets, incorrectly marking the largest majority of them as “neutral,” despite indications amply evident to humans of sharp negative sentiment. To illustrate, the tweet containing the phrase, “I literally want to see R Kelly Suffer here on earth first before he goes to Hell” returned a neutral score. If time had permitted, it might have been conducive to this research to allow human subjects on a data collection service (such as Amazon’s Mechanical

Turk) to hand-label the collection of tweets. This method requires approval from a review board to ensure all research practices are ethical.

Lastly, I faced challenges unrelated to the case study at all. #MeToo is a sensitive issue; it has unsurprisingly created a collective sentiment of anger and fear. During a short interval where I left my computer unlocked and unattended at my place of work, an anonymous individual typed a message onto an open webpage: “Manipulating the #MeToo movement to your advantage is disgusting. I have been summoned.” This aggressive invasion of personal space and threatening message to a student conducting a thorough study and analysis of the collective culture of a movement is disappointing. It is a real-world example of how a sufficiently polarized political movement can create an atmosphere of hostility and resentment.

Results and Discussion

There are two opposing ideas in the #MeToo debate: either victims cannot be taken seriously by only their word (following our system of criminal justice) or the public has a moral obligation to believe the good intentions of victims without reservation. The vocal majority on Twitter, by this study, seems to lean further toward the former. Still, it is necessary to analyze the language of those with the other stance.

Upon inspection, language in Twitter regarding the “#MeToo” news and movement seems to take one of two main forms – angry and abusive, or encouraging and empathetic. Users on both sides of the debate are angry and abusive, but it is only for those who lean toward believing the good intentions of victims that there are significant results to show encouraging and empathetic sentiment. Not shockingly, any sharp contrast in sentiment or motivation is largely absent among *The Wall Street Journal*’s list of similar words.

The most similar word to “metoo” from *The Wall Street Journal* is “sexualliberation.” From Twitter, this word is “horrifically.” Neither most similar word is found in the other corpus. The WSJ articles evidently treat the movement as an opportunity to discuss the effects it has had on sexuality. This isn’t to say the sentiment of these articles will be positive (even though the term “sexual liberation” exudes a positive connotation), but it is clear that the focus of the #MeToo Journal articles is not on the emotional and mental state of individual victims in any stage after the crime is committed, but rather on the collective consequences on populations and cultures – perhaps that the culture of harassment and assault is born from pursuing a culture of sexual liberation, or that the culture of sexual liberation will be restrained by the fear of harassment accusations. There is a similar discussion over the moral foundations of the #MeToo movement on Twitter; the most similar word to “culture” on Twitter is “purity,” so there is evidence Twitter discusses the influence of the movement from and impacts on culture in addition to discussing individual experiences; it provides a holistic analysis of individuals and the society they live in. The most similar word to “culture” on the other corpus is “conversations.” *The Wall Street Journal*, it seems, aims to emphasize cultural impacts and norms in the light of the #MeToo movement, and has much less of an emphasis on individual experiences (whether it be of the victim or the alleged) than Twitter does.

Angry and Abusive:

This sharp disparity in relevant connotation continues further for several keywords. A subset of the closest words to “allegation” or “allegations” used in *The Wall Street Journal* are “complaints,” “misconduct” and “(non)/consensual,” while in Twitter it is “credible” and “unsubstantiated.” The evidence suggests that these unbiased news articles do not frequently

discuss the degree of credibility of allegations made against individuals for sexual violence, but that they instead choose to focus on the content of these statements (indicating that they are complaints that certain activities were nonconsensual and amounted to misconduct). On Twitter, however, the credibility debate is unmistakably common. There is discussion as to when suspicion should be raised regarding a victim's statements.

Ideas in this vein continue when viewing the three most similar words to the keyword, “guilty” on Twitter: “innocent,” “proven,” and “unless.” It should be obvious in which context these words are frequently found together: innocent unless proven guilty. This indicates users on Twitter very frequently cite – for allegations of sexual misconduct – the tradition of presumption of innocence under due process for criminal cases. In summary, due to discussions on the guilt of alleged sexual predators, conversation in the Twitter-sphere has shifted away from the content of the allegations, going instead toward whether the uninvolved populace ought to believe the words of an individual (the victim) or reserve their judgments until the crime can be unquestionably shown to have occurred. This finding is again supported in discovering that the closest word to “evidence” used in Twitter is “without.” Twitter users speaking under the hashtag #MeToo evidently are largely uncomfortable with accusing individuals of severe misconduct without investigations or tangible evidence to conclude they committed the alleged crime.

It might be interesting to delve into the possible motivations behind these reservations. Since it is difficult to conclude much about polarizing language from analyzing only one word at a time, this work will follow a snowballing technique to understand possible contexts the tweets and articles were written in. For example, if the closest word to a keyword, “piano” is “violin,” then “violin” may also be analyzed to discover more insight about the original keyword. The

goal of this method is to mimic how subjects and frequently used words in a conversation can evolve. Of course, there are limitations in using such an approach; a discussion can shift its topic and language used entirely. The similar words found for snowballed keywords should be analyzed in their degree of similarity to the original keyword. If the cosine similarities between them and the original keywords are too large, one cannot conclude they provide any additional information.

An interesting similar word to “allegations” on Twitter was “unsubstantiated.” Using “unsubstantiated” as a snowballed keyword, one finds the most similar word on Twitter is “weaponizing.” A set of the most similar words to a popular synonym, “uncorroborated,” brings “commonplace” and “malicious.” Similar words to these keywords for the comparison-group corpus include “debate,” and “decades-old.” A similar word in The Wall Street Journal to “allegation” was “consensual.” Using “consent” as a snowballed keyword, a similar keyword on Twitter is “coercion.” A general trend can be drawn from the Twitter findings that those who discuss lack of evidence to allegations also discuss possible ill-intentions of those making the claims. They are either perpetrating or responding to fears that unsubstantiated claims can be used as weapons against the alleged predator. The form of outrage in “From Incivility to Outrage” closely indicated by this finding is “character assassination,” (Sobieraj and Berry, 2011) These arguments focus heavily on ad hominem attacks against an individual, rather than against the contextual substance of the individual’s statements. This is further supported by the earlier finding that Twitter focuses more on the credibility of a victim when discussing allegations than on the actual content of the allegations. There is discussion on how much the media, the audience, or even the victim may have sensationalized a #MeToo story about coercion. Overall, doubt against victims’ stories is a common theme on Twitter.

Why? It is well-known the consequences for #MeToo allegations have been fierce. In October 2018, one year after the first allegation against Harvey Weinstein was released, *The New York Times* reported that 201 prominent men lost jobs or major roles after sexual assault allegations went public (Carlsen, 2018). It is a disturbing prospect, then, that someone who may not have committed a crime of misconduct be fired from a position simply because someone else said they were guilty. This is likely how the aforementioned Twitter users view the circumstances of allegations; indeed, the most similar words to “ruin” in Twitter are “reputation,” and a familiar word, “weaponized.” The second-most similar word to “career” is “ended.” The WSJ similar words for the same keywords are “press” and “clients.” Press perhaps is a synonym for the media. “Clients” is a word often used in the same context as “career,” invariant of the #MeToo context. It is clear that a discussion over the safety of reputations and careers under the light of sexual allegations is prominent on Twitter, but largely nonexistent in unbiased and nonpartisan writing.

Observation of the opposing side of the debate unearths another prominent component of the dialog on Twitter: an argument over which, if any, implicit advantages and disadvantages men or women have. In the context of #MeToo, the most similar word to “male” on Twitter is “privilege.” A similar word to “entitlement” (a synonym, whose third-most similar word is “privilege”) is “dominated.” The #MeToo movement undoubtedly revealed a power clash between the two (main) genders, a topic which was discussed to great length among users on the social media platform. The users who discuss this issue might be aiming to shed light on the disparity between male and female experiences in society. Users may be claiming that those who identify as male often have the advantage of certain privileges and entitlements which females may not enjoy (with a sharply negative connotation as indicated by the suggestion that it allows

for a creation of domination in one gender and further maintaining a power disparity), or that they may be using the words “privilege” and “entitlement” in tweets to rebut this idea. This debate uses sharp words which correspond closely to negative words.

Encouraging and Empathetic:

Civil discourse among people who disagree necessitates the creation of mutual trust and willingness to listen actively before deliberation occurs. This task is so difficult, it often requires facilitators trained in conflict resolution. Certain communication channels do not afford this creation; Twitter is often a volatile platform since it was designed to share short, unfiltered content quickly. It is unsurprising, then, that outrage manifested as angry and abusive language would be fueled and perpetuated on the website more so than encouraging and empathetic language.

There is, however, a shred of positivity in the violently negative discussion on the propagated violence. The most similar word to “victim” on Twitter is “survivor.” Surviving the event connotes a positive, empowering experience. Conventional usage of the term indicates that one is a victim of sexual violence when the emotional wounds are still open and fresh. One who has not yet battled through the harrowing hours of therapy, self-blaming, fear from trusting others and herself, fear from speaking out, or gaslighting. One who has not overcome her trials. Victimhood is a powerful word which recalls the nature of the crime; survival is a powerful word which recalls the strength of healing. The most similar word to “victim” on *The Wall Street Journal* is “hugging.” The context behind this word is ambiguous; perhaps it exudes a positive connotation, too, but it is not a carefully-chosen word to inspire a hopeful experience from a brutally negative act. It could also imply again a focus on the collective community – hugging is,

of course, an act of comfort, support, and sympathy from another person – instead of on the experiences of the individual who endured the violence. Collective support is vital to the movement, but the most similar WSJ word does not indicate a stance of solidarity and togetherness. There is another context under which “hugging” could be used. These articles were collected before public outcry over Joe Biden’s behavior with women, but it is possible the word is used to discuss a shift in the way individuals interact in the workplace; since the explosion of victims coming forward with sexual harassment allegations, individuals act with caution and limit their participation in this behavior, even if it is innocuous.

Twitter users who focus on the individuality of the victim focus on validating their experiences and reflections. It was earlier mentioned that the most similar word to “metoo” on Twitter is “horrifically.” This similarity indicates that there is continuous affirmation among Twitter users of the unimaginable negativity surrounding the event. Its aftermath grants the victim a sense that she is understood, almost as if she does not have to suffer her trials alone. Despite the negative connotation of the word, it can actually be viewed as a positive tweet exhibiting empathy and emotional support.

As mentioned on page 20, each pair of words can be measured in their degrees of similarity by using a distance metric which computes the angle between their two vectors (where, the smaller the angle, the closer the words are in contextual meaning). The similarity scores for all notable pairs that were discussed here are portrayed in Figures 1 and 2. Note that though the similarity scores for the WSJ corpus are much higher than those for the Twitter corpus, this does not mean that words used in the WSJ are *more* similar than those used on Twitter. This disparity in similarity scores is due to the disparate sizes of corpora used; recall that the Twitter corpus’s vocabulary size is over double that of the WSJ corpus. While pairs of words

cannot be compared in their degrees of similarity across corpora, it may be useful to compare the similarity of words within a corpus. Figures 3 and 4 show word pairs with their similarity scaled down such that the least similar word pair in that corpus has a score of 0, and the most similar has a score of 1. This allows a user to infer how frequently words were used together in the corpus. Notice that “horribly,” despite having been the single most-used word with “metoo,” has a significantly lower similarity score than the rest of the word pairs. This can be explained because the word, “metoo” is used in all tweets I investigated, and “horribly” was not used in nearly as many tweets. However, it is evident that a vast number of tweets which included the word “guilty” included “innocent,” as well.

These similarity scores can be used to assess how sharply the Twitter users focused on certain topics. Twitter users did not discuss the implications of the #MeToo movement on victims (such as their horrific experiences or their empowered status as survivors) as fervently as male privilege, reputations, and the role of presumption of innocence.

PART III

Final Remarks

Conclusions

Analyzed against the non-polarizing articles from *The Wall Street Journal*, the language on Twitter follows certain trends which may shed light on its correlation with polarization on the social media platform. It is important to note that contributors on Twitter use more words not found in the WSJ articles than vice-versa. These words include horrifically, sensationalizing, weaponizing, maletelling, liarsguilty, accusationsdemocrats, roil, and witchhunt (note that compound words may have been hashtags or had punctuation marks such as hyphens as separators which were removed in data preprocessing). All words have significantly negative connotations symptomatic of outrage. A lesser-seen but meaningful pattern was usage of emotional words. It is not surprising that language relating to emotional experiences would be omitted from a formal news publication, as it may reduce the formality of the content published.⁷ But it is also appealing to consider that it may be omitted to prevent compromising the impartiality of the content. Perhaps Twitter's usage of more personal or language responding to heart-wrenching stories may create a shift in opinions further from center toward both ends of the spectrum.⁸ Lastly, Twitter does not stray from discussing the #MeToo movement from the lens of individuals – either those who have faced the violent crimes or those who are alleged of having perpetrated them. *The Wall Street Journal* largely ignores individual experiences and instead chooses to discuss the violence and consequent #MeToo movement in the light of its impact on the community (or perhaps reflexively, discusses how the culture of the community corresponds to the crimes committed and the transformation of the movement).

Epilogue

This case study did not analyze the discourse within and among individual groups (such as liberals and conservatives or men and women), but rather the language used in reference to the collective movement. While group polarization is certainly occurring on Twitter, as represented by Hong, Kim, and Sunstein, the study aims to focus on the manifestations of political outrage for a specific movement on Twitter. A further study can improve upon this work to analyze how unique dialog differs among groups. To illustrate, it might be appealing to examine how liberal language compares to conservative language for a given issue on Twitter. Sobieraj and Berry found that the usage of outrage language in all forms of traditional media they investigated – talk radio, political blogs, cable news analysis programs, and newspaper columns – is stronger among Conservatives than among Liberals (30). There is ample reason to believe this would be no different for another form of media, but it will require ample exploration to confirm that this hypothesis holds, and to what degree.

Secondly, with recent advances to the field of natural language processing and natural language inference, performing sentiment analysis on the entirety of tweets for a movement, or for tweets from individuals with specific group identities, may prove immensely valuable. Understanding the average degree of negativity, positivity, and neutrality in the collective Twittersverse will reveal the extent of polarization, empathy, and impartiality of the views held by the general populace pertaining to critical socio-political movements. While I was not able to perform this study in limited time with an existing Python library, there is room for future work in this area, which would require hand-labeling tweet sentiment. A further analysis into the in-group and collective vernacular and sentiments will provide rich insights into how polarization occurs on social media, and how it affects socio-political discourse.

Reference

Notes

1. Sunstein clarifies his usage of this term: “You can become radicalized in the sense that you come to believe, firmly, a position that is within the political mainstream” (75, 2017). In other words, the radicalization comes from the firmness of belief.
 2. There is an implicit assumption in this work of ideological/partisan alignment.
 3. Fervent commitment often invites outrage, but it is important not to conflate extremism or polarization in views with incivility. For example, supporters of state socialism and supporters of laissez-faire capitalism are extreme views at polar ends of the economic spectrum, but this fact alone does not imply there is incivility between the two groups.
 4. This excludes editorials, which usually lean further from center and are more opinionated than the articles of a publication.
 5. Thank you to Dr. Lewis for raising this concern and possible counter-argument with me!
 6. While not within the scope of this essay, it is intriguing to understand how the movement and the language represented by it have shaped and even limited each other. There is a targeted message of empowerment in the #MeToo Movement, most clearly seen with the collective replacement of the word, “victim” with “survivor.” Blog writer Kate Simon (2018) offers an article scrutinizing the societal repercussions of this explosively popular yet carefully chosen replacement: it shifts the focus away from the crime and criminal and offers an alternative narrative of a blameless act; to illustrate, society does not call one a survivor of a robbery, but it calls one a survivor of a natural disaster. Despite her deep reservations with the term, she feels compelled to call herself a survivor in writing simply because the term has become characteristic of the #MeToo movement, and she wants to see it succeed. Her final hope for the movement is for it to allow more expressivity. “We need to get to a place where language can be more nuanced, more telling, more personal. We need to be able to speak our truths however we want to articulate them, to fully own our stories, and not just donate them to the cause.”
- Simon, Katie. “I Was Raped. Call Me a Victim, Not a 'Survivor.'” *The Lily*, 23 Apr. 2018, Web.
7. Interestingly, this would not apply for television news channels, which tend to emphasize emotions.
 8. This is manifest in a movement that spun off from #MeToo: #BelieveWomen. This permanent trust in victims in their allegations was a measure of counter-anger against those who argue vehemently for presumption of innocence of the alleged.

Glossary

cosine similarity

distance metric which measures the angle between any two given word vectors; the smaller the angle, the more similar the words

group polarization

the tendency of people in homogeneous groups of like-minded individuals to move toward an extreme point of view after deliberation with or even exposure to the other members' views

incivility

any indication of a lack of respect for or frustration with the views of opponents

outrage

a particular form of political discourse involving efforts to provoke visceral responses

similarity

the degree to which two words have been used together within a given context; similar words have been used frequently together and fall within the same thematic categories

social polarization

the tendency to make attacks against opponents when the prestige of one's own social group is challenged

Appendix

Degrees of Similarity on Twitter

Keyword	Similar Word	Cosine Similarity
metoo	horrifically	0.5237208
culture	purity	0.7111308
allegation	credible	0.77100766
allegations	unsubstantiated	0.67431444
guilty	innocent	0.8950322
guilty	proven	0.8297467
guilty	unless	0.7498757
evidence	without	0.76594007
unsubstantiated	‘weaponizing’	0.7594503
uncorroborated	commonplace	0.7586543
uncorroborated	malicious	0.7378206
consent	coercion	0.6924712
ruin	reputation	0.83047396
ruin	weaponized	0.64953774
career	ended	0.69298136
male	privilege	0.7435811
entitlement	dominated	0.7210374
victim	survivor	0.67541957
female	maletelling	0.6146249
malicious	liarsguilty	0.82200164
malicious	accusationsdemocrats	0.7786152
unproven	roil	0.7162616
destroy	witchhunt	0.6697526

Figure 1: similarity scores for prominent word pairs on Twitter. The possible range is [-1,1], where -1 would indicate an extremely dissimilar word, and 1 would indicate an extremely similar word. Similarity scores represent a proportion between frequency the “similar word” used with the keyword and the total usage frequency of the keyword. It is unsurprising, then that despite “horrifically” being the single most used word with “#MeToo,” it has a substantially lower similarity score than other word pairs, since all 83,412 tweets contained the word “metoo.”

Degrees of Similarity on *The Wall Street Journal*

Keyword	Similar Word	Cosine Similarity
metoo	sexualliberation	0.9433428
culture	conversations	0.99949706
allegation	complaints	0.9961927
allegation	consensual	0.9957324
allegations	misconduct	0.9867864
allegations	nonconsensual	0.98159915
guilty	statements	0.9989534
uncorroborated	debate	0.9948064
uncorroborated	decadesold	0.994082
ruin	press	0.99505115
career	clients	0.99973357
victim	hugging	0.9990991

Figure 2: similarity scores for prominent word pairs on *The Wall Street Journal*. The possible range is the same as in Figure 1. Since the size of the vocabulary is substantially smaller in the WSJ corpus than in the Twitter corpus, similarity between words is much closer.

Degrees of Similarity on Twitter, Normalized

Keyword	Similar Word	Adjusted Similarity Score
metoo	horrifically	0
female	maletelling	0.24481904
ruin	weaponized	0.33884481
destroy	witchhunt	0.39328661
allegations	unsubstantiated	0.40557236
victim	survivor	0.40854865
consent	coercion	0.45447137
career	ended	0.45584531
culture	purity	0.50472461
unproven	roil	0.51854266
entitlement	dominated	0.53140464
uncorroborated	malicious	0.57660443
male	privilege	0.59211837
guilty	unless	0.60907072
uncorroborated	commonplace	0.63271287
unsubstantiated	weaponizing'	0.63485662
evidence	without	0.65233459
allegation	credible	0.66598241
malicious	accusationsdemocrats	0.68647071
malicious	liarsguilty	0.80331722
guilty	proven	0.82417588
ruin	reputation	0.82613451
guilty	innocent	1

Figure 3: Values from Figure 1 have been normalized, so the keyword-similar word pair in the subset with the least similarity now has a score of 0, and the pair in the subset with the highest has a score of 1. This allows a user to clearly identify similarities in the table relative to one another.

Degrees of Similarity on *The Wall Street Journal*, Normalized

Keyword	Similar Word	Adjusted Similarity Score
metoo	sexualliberation	0
allegations	nonconsensual	0.6784151
allegations	misconduct	0.77040267
uncorroborated	decadesold	0.89977846
uncorroborated	debate	0.91262453
ruin	press	0.91696478
allegation	consensual	0.92904566
allegation	complaints	0.93720834
guilty	statements	0.98616493
victim	hugging	0.98874869
culture	conversations	0.99580587
career	clients	1

Figure 4: Values from Figure 2 have been normalized as done between Figures 1 and 3.

Two-Dimensional Mapping of Twitter Embedding for Selected Words

Figure 5: The Gensim implementation of Word2Vec represents each word (by default) as a 100-dimensional vector. To display the results, one can apply Principal Component Analysis to retain only the most expressive features in each vector and map them to two dimensions. Words that are plotted closer are more similar. However, it is important to note that, while PCA calculates the two most expressive features, 198 features have been lost. For some close cases, it may shift which words appear closer together. For example, “weaponizing” was not the most similar word in 100 dimensions to “metoo.”

Bibliography

- Bail, Christopher A., et al. "Exposure to Opposing Views on Social Media Can Increase Political Polarization." *Proceedings of the National Academy of Sciences*, vol. 115, no. 37, 2018, pp. 9216–9221.
- Carlsen, Audrey, et al. "#MeToo Brought Down 201 Powerful Men. Nearly Half of Their Replacements Are Women." *The New York Times*, The New York Times, 23 Oct. 2018.
- Hong, Souman, and Sun Hyoung Kim. "Political Polarization on Twitter: Implications for the Use of Social Media in Digital Governments." *Government Information Quarterly*, vol. 33, no. 4, 2016, pp. 777–782.
- Mason, Lilliana. "Party Polarization Is Making Us More Prejudiced." *The Washington Post*, WP Company, 28 Jan. 2014.
- Mikolov, Tomas; Sutskever, Ilya; Chen, Kai; Corrado, Greg S.; Dean, Jeff (2013). *Distributed representations of words and phrases and their compositionality*. Advances in Neural Information Processing Systems.
- Mitchell, Amy, et al. "Political Polarization & Media Habits." *Pew Research Center's Journalism Project*, 26 Apr. 2018.
- Nyhan, Brendan. "Relatively Few Americans Live in Partisan Media Bubble, but They're Influential." *The New York Times*, 7 Sept. 2016.
- @nltk. "Natural Language Toolkit." *GitHub*. 2019. <https://github.com/nltk/nltk>
- @SeleniumHQ. "selenium." *GitHub*. 2019. <https://github.com/SeleniumHQ/selenium>*
- Pariser, Eli. "Beware online 'filter bubbles.'" *TED*. Mar. 2011
- "Premium Search APIs - Twitter Developers." *Twitter*.

* There is no agreed method on citing Python libraries. Here, I have cited the page to the GitHub repositories for the package code when citing preferences were not provided.

Ou-Yang, Lucas (@codelucas). “Newspaper3k: Article scraping and curation” *GitHub*. 2019.

<https://github.com/codelucas/newspaper>*

Řehůřek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora.

LREC.*

Sobieraj, Sarah, and Jeffrey M. Berry. “From Incivility to Outrage: Political Discourse in Blogs, Talk Radio, and Cable News.” *Political Communication*, vol. 28, no. 1, 2011, pp. 19–41.

Sunstein, Cass R. *#Republic: Divided Democracy in the Age of Social Media*. Princeton UP – M.U.A, 2017.