

Uncovering Latent Topics in Text: Using Topic Models to Identify Discussion Themes in
the Brett Kavanaugh Controversy

Simon Rodier

A Thesis
in
The Department
of
Education

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Arts (Educational Technology) at
Concordia University
Montréal, Québec, Canada

February 2021

© Simon Rodier, 2021

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: **Simon Rodier**

Entitled: **Uncovering Latent Topics in Text: Using Topic Models to Identify Discussion Themes in the Brett Kavanaugh Controversy**

and submitted in partial fulfillment of the requirements for the degree of

Master of Arts (Educational Technology)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

_____ Examiner
Dr. Julie Corrigan

_____ Supervisor
Dr. Vivek Venkatesh

_____ Supervisor
Dr. Richard Schmid

Approved by _____
Dr. Saul Carliner Graduate Program Director

_____ Dean of Faculty
Dr. Pascal Sicotte

Date _____

Abstract

Uncovering Latent Topics in Text: Using Topic Models to Identify Discussion Themes in the Brett Kavanaugh Controversy

Simon Rodier

Social media texts are abundant and generated at a rapid pace; often, they discuss contentious issues in ways that can be extreme, with people defending counter-intuitive points of view. This project develops an understanding of how individuals discuss contentious issues in online fora. It examines gender as a contentious issue, and also examines the controversy surrounding Brett Kavanaugh's nomination to the U.S. Supreme Court amidst sexual assault allegations in 2018 as a gendered controversy. Social media posts about the controversy were collected from *reddit* and analyzed with a framework using topic models, which help uncover latent topics in text. Significant discussion throughout the corpus centered on four major themes: the search for and evaluation of evidence; the importance of a Supreme Court nomination and of an investigation into the allegations; sexual assault and gendered perceptions and expectations; and finally, a discussion of the judiciary hearing into the allegations. Discussion of the results suggests that notions of power and loss aversion may be at play throughout the corpus. Additional discussion reflects on the use of topic models as a qualitative research tool, suggesting that they can be an effective way of exploring broad themes within larger corpora.

Acknowledgements

While this thesis bears my name, it was also shaped by a number of individuals whose contributions helped me complete and refine the work.

I would like to thank my supervisors: Dr. Vivek Venkatesh, for listening to my ideas and giving me every opportunity to explore them and put them into practice; and Dr. Richard Schmid, for critically engaging with my proposed methods and challenging me to think about them carefully throughout the process. I would also like to thank Dr. Julie Corrigan for providing helpful critique of this work.

Finally, a special thanks to my colleagues, my friends and my family for their support throughout this endeavour. In particular — Mom, Dad — I'm forever grateful, this never would have happened without you.

Contents

List of Tables	vii
Content Warning	1
Introduction	1
Research Problem	1
Background	2
We’re Not Quite Rational	2
Identity is Situational	7
Gender as a Contentious Issue	9
Modes of Inquiry	12
Qualitative	12
Computational	13
Topic Models	17
The Brett Kavanaugh Controversy	19
Addressing Subjectivity	20
Research Questions	21
Methodology	21
Setting and Data Collection	21
Overview of Analytic Methods	24
Credibility	26
Analysis	27
Selecting a Number of Topics for Analysis	27
Summary of Topic Modeling Output	29
Topic 1	32
Topic 2	33

Topic 3	35
Topic 4	38
Topic 5	40
Topic 6	41
Topic 7	42
Topic 8	45
Topic 9	46
Topic 10	48
Topic 11	49
Topic 12	51
Topic 13	53
Topic 14	54
Topic 15	55
Topic 16	58
Topic 17	60
Topic 18	61
Topic 19	64
Topic 20	65
Interpretation of Topic Modeling Output	66
View From Men’s Interests Subreddits	67
Thematic Groupings	69
Thematic Grouping 1: Evidence & the Search for Truth	70
Thematic Grouping 2: Importance of a Nomination to the Supreme Court	71
Thematic Grouping 3: Sexual Assault and Gendered Norms	72
Thematic Grouping 4: The Hearing	74
Summary of Themes	75
Discussion	76

Interpreting Findings	76
Topic Modeling as a Qualitative Research Tool	78
Practical Applications	82
Conclusion	83
References	84
Appendix	89
Appendix I	89
Appendix II	89
Appendix III	89
Appendix IV	89

List of Tables

1	Ten most representative words for each of the twenty topics.	31
2	Brief summary of topics and their observed grouping type.	32
3	Discussion about the Brett Kavanaugh controversy from late 2018 from the menslib and mensrights subreddits.	68
4	Topic groupings by topic ID with a brief description of their themes.	70

Uncovering Latent Topics in Text: Using Topic Models to Identify Discussion Themes in
the Brett Kavanaugh Controversy

Content Warning

This project examines online discussion of gender-based issues and the allegations of sexual assault about Brett Kavanaugh in his nomination to the Supreme Court in 2018. This document thus contains some discussion of sexual assault, rape, sexual violence, and sexualized hateful speech.

Introduction

Research Problem

This thesis was initially motivated by a desire to examine the ways that people discuss contentious issues through social media. Anecdotally, discussions about contentious issues - for instance, gun control in the United States, feminism and our changing conceptions of gender, or immigration - in online spheres seem to include more and more extreme rhetoric and disagreement. Individuals seem to take perspectives in stark contrast to one another, and arguments seem to quickly polarize individuals into diametrically opposed camps. While the phenomenon caught my interest, the number of conversations evolving across the internet makes it difficult to find ways to characterize the discussions surrounding these issues. Still anecdotally, these conversations and arguments often seem repetitive: each side of an argument puts forth a perspective, points are endlessly repeated, while each side tries to break apart the arguments coming from the opposing side. Very little listening and negotiation seems to happen within these discussions, however. This project seeks to make a modest contribution to help interested stakeholders systematically get a sense of the conversations happening around contentious issues, and to give them tools to identify the major issues on all sides of a debate. Better understanding the conversations around these contentious issues is a first step to crafting better and more relevant interventions.

Background

In this introductory section, I review research that examines the ways that individuals argue contentious issues online. In particular, I examine research that deals with opinion formation, its links to identity, and how these are manifested and exhibited. The research suggests an informal framework for conceptualizing these behaviours: people aren't purely rational, they want to fit in to groups, and have contextually-triggered identities. These three factors can easily lead individuals to take perspectives that can be not-quite-rational, or even fully counter to other established beliefs the individual might hold.

We're Not Quite Rational

To start, it is important to establish that not only are conversations anecdotally becoming more polarized and contentious in online spaces, but that this is an actual and documented phenomenon. A classical view of the human as a thinking, rational being would see an individual start with an idea, and then every time the individual is confronted with a new piece of information about that idea, the individual evaluates that new piece of information independently of their currently held beliefs and updates their opinions as a result of this evaluation (Taber & Lodge, 2006). Acting on this model, individuals would rationally update their beliefs, and, as they are exposed to new information, would change their opinion to reflect the evidence available to them. However, Taber and Lodge (2006) show that when dealing with contentious political issues, individuals have a tendency not to reason in such a rational manner. Taber and Lodge let participants explore arguments by a variety of parties about two contentious issues (gun control and affirmative action). They find a series of concerning effects. Notably, individuals who already felt strongly about an issue evaluated arguments supporting their position more favorably than arguments opposing their prior attitude. They also find that individuals spent more energy counter-arguing opposing arguments compared to arguments supporting their own side, showing a greater willingness to be critical of opposing viewpoints than their own. To add to the tally, Taber and Lodge find evidence for a *confirmation bias*, finding that individuals were

more likely to seek out arguments confirming their point of view (e.g., for an individual against gun control, they were more likely to seek out arguments from the National Rifle Association than from a Democrat politician). Furthermore, they find evidence that even when subjected to an equal number of opposing and supporting arguments, individuals' attitudes were likely to solidify and polarize. The authors refer to this group of tendencies as motivated bias, a tendency to evaluate new information to fit in with one's existing world-view. They also find evidence for two modulating effects on this type of bias: pre-existing opinions and knowledge are both associated with a tendency to engage in motivated reasoning: the stronger either of these are, the greater the tendency of participants to engage in the motivated reasoning described above.

These findings are not only relegated to the issues of affirmative action and gun control: Redlawsk (2002) performed a similar study, this time studying the information-seeking and decision-making patterns of individuals in a simulated election environment, with mock candidates representing real political parties towards which participants already had a pre-established affinity. Redlawsk's findings support those by Taber and Lodge (2006), particularly when it comes to the time individuals spend counter-arguing opposing view points, and time spent seeking out supportive arguments. Redlawsk also finds support for the notion that individuals will strengthen their support for their preferred candidate in the face of information against the candidate. Redlawsk further finds support for the idea that if an individual develops positive affect towards a candidate early in the election process, then even as information against this candidate appears, the individual's positive affect for the candidate will rise – *even if* information about a candidate that better matches the individual's opinions arises. Redlawsk's findings suggest that people do not update their beliefs based on the logical quality of evidence presented to them, but on their affective reactions to these new pieces of evidence. All is not lost however; Redlawsk does reveal that some of these biases are attenuated when participants were told they would be tested on their recall of arguments for and against various candidates. While this suggests that humans *can* in fact reason logically, it also suggests that this behaviour needs to be specifically

triggered; that by default, we engage in simplistic reasoning that relies on mental heuristics and affect.

These effects are more pervasive than many might anticipate. While it may be easy to grant their existence in the values-based realm of politics and moral debates, one might expect the effects to be lesser in fact-based discussions around scientific issues. However, Hart and Nisbet (2012) find that similar motivated reasoning occurs when individuals contemplate the effects of climate change. Hart and Nisbet find that the way one describes the victims of climate change is influential in determining participants' support for various government policy interventions to combat climate change. The authors find that when the victims are presented as having many superficial differences from participants (different country of origin, language, etc.), then participants that originally were against government policy interventions towards climate change are even less likely to support these interventions, even if presented with factual information that – *logically* – should have swayed their opinion.

All of these findings describing human biases and ways of acting irrationally are congruent with a growing body of research in psychology and economics. Starting in the 1970s, psychologists have explored the tendency of humans to act in ways that are generally independent of logic and rationality, unless a specific effort is made to behave rationally (Kahneman, 2011).

The above findings also fit in coherently with the *elaboration likelihood model* of persuasion (ELM) developed by Petty and Cacioppo (1986), which describes the ways in which individuals may be persuaded to take on new attitudes (more precisely, Petty and Cacioppo operationalize a change in attitude as the individual's ability to elaborate arguments in favour of a given position). Their model is built upon a substantial body of literature, and contains seven postulates, which are summarized below:

1. Individuals are motivated to hold correct attitudes.
2. Despite their motivation to hold correct attitudes, individual and situational factors

can influence individuals' willingness or ability to evaluate messages.

3. The amount and extent of attitude change displayed after a persuasion attempt can be affected by a series of variables. These include convincing, persuasive arguments (the so-called "central" road to attitude change), peripheral cues (for instance, perceived speaker credibility), which are most influential when the ability and/or motivation to critically evaluate arguments is low, and factors that affect elaboration in an objective or biased manner (see items 4 and 6, respectively).
4. An individual's ability (or motivation) to evaluate a message objectively can be altered by increasing or reducing message scrutiny. Distraction, repetition, personal relevance, personal responsibility, and an individual's general disposition towards cognitive evaluation are all factors that may affect objective elaboration.
5. In an act of persuasion, as the recipient's motivation and/or ability to process arguments is decreased, the influence of peripheral cues towards persuading the recipient increases. Similarly, as the motivation/ability to process arguments increases, the importance of peripheral cues decreases. These peripheral cues include cues about the source (credibility, reliability) and cues about the message (e.g.: the number of arguments included can be salient when personal relevance is low, however the strength of arguments is more salient when personal relevance is high).
6. An individual's level of motivation and/or ability to process a message can be affected in a relatively biased way (i.e.: motivation/ability can be affected by something other than message content). Significantly, prior knowledge about the issue and forewarnings (either about the message topic or the fact that a persuasion attempt is forthcoming) can have this effect.
7. Changes in attitude that are the result of a (mostly) "central road" persuasion attempt (i.e.: convincing, issue-related arguments) last longer, are more effective predictors of behaviour, and are more resistant to change in the future than those made from

peripheral cues.

The model above is a useful way to conceptualize attitude change: when individuals purposefully focus on the merits of arguments, they may be better equipped to recall and agree with those arguments at a later time. If, however, an individual is not actively tending to the quality of arguments, then a host of other factors may be influential, such as the perceived credibility of a speaker, the sheer number of times an argument has been made (as opposed to its quality), or the individual's initial disposition toward the argument being made (Petty & Cacioppo, 1986).

To add to this, Walton and Banaji (2004) find that we can be persuaded to think of others differently simply by altering the grammatical structure of a sentence referring to them. The authors explore whether the types of words and phrases that we use to describe others and ourselves have an impact on the types of perceptions that we form of the subject of the sentence. Specifically, they examine the effect of using a noun or a verb phrase on perceptions of personal preferences. For example, how does an outside observer perceive John's enjoyment of chocolate if it is framed in these two ways: "John is a chocolate eater" (noun phrase), versus, "John eats lots of chocolate" (verb phrase)?

Walton and Banaji (2004) uncover two key findings: first, using noun phrases to describe others leads to a stronger and more stable perception of the attitude of the subject of a sentence (i.e.: typical study participants will identify John's enjoyment of chocolate as stronger and more stable when John is described with the noun phrase). Furthermore, when describing themselves with a noun phrase (as opposed to a verb phrase), participants were also more likely to view their own attitudes as stronger and more stable. The authors note that this second effect isn't as strong as the first, but that it is still statistically significant.

This study is particularly interesting because it suggests that word choices do matter when an individual is tasked with processing a message. It also seems to be consistent with other studies that view identity itself as a situational construct.

Identity is Situational

Thus far, attitudes have been described as largely affective and not-necessarily-rational, while acknowledging that we humans, holders of those attitudes, *do* have the capacity to analyze arguments critically, if we would only consciously take the time to do so. In other cases, we are easily convinced by our own affect, or by situational cues. In this section, I shift the focus to identity, the backdrop on which our attitudes are held and formed. The literature shows that identity may be interpreted as a performed, situated phenomenon.

Anecdotally, the popular view of identity is something often colloquially referred to as the *true self*. This expression underscores the idea that there is a singular, underlying essence to personality. The evidence, however, paints a very different picture. Notably, Bucholtz and Hall (2005) performs an extensive review of evidence from the fields of sociology, cultural studies and linguistics to offer five principles describing identity. Together, these five principles illustrate how identity is situated in a specific time, place and context, and is not a fixed, underlying feature of individual personalities. The principles are as follows:

1. Emergence. Identity is the emergent product - and not the underlying source - of linguistic and other semiotic practices.
2. Positionality. Identities are positioned at various levels - from macro-level demographic categories (e.g. gender) to local, ethnographically-specific stances and roles (e.g. an individual in a position of power in one scenario may behave differently than another scenario where they do not have this power).
3. Indexicality. Language can mark the speaker's, or others' identity in a variety of ways: either by using overt labels, using implicit language to mark an individual's belonging (or not) to a group, and/or by using linguistic structures typical of a group.
4. Relationality. Identities are intersubjectively constructed in interaction by a speaker and others, through overlapping, complementary relations. These include: similarity/difference, genuineness/artifice, and authority/delegitimacy.

5. Partialness. Any construction of identity may (or may not) be consciously or deliberately constructed, and is in part a result of interactional negotiation, in part an outcome of perceptions, and in part an effect of larger ideologies. Identity may therefore be constantly shifting, even as an interaction unfolds.

Bucholtz and Hall (2005) thus characterize identity as a performed product, one that identifies oneself and others as part of or apart from groups, and that is deeply entwined with the context in which the performance occurs. These conclusions lead to the interpretation of an individual's identity as a shifting phenomenon that can hold radically different properties in different contexts, depending on the factors surrounding it.

Oyserman (2009) takes this notion one step further. Importantly, her findings are consistent with those of Bucholtz and Hall (2005), supporting the idea that identity is susceptible to contextual cues, but Oyserman pushes this further, finding that individuals are likely to make different choices in different contexts, largely dependent on which identity is most salient at the time. Oyserman offers a few key findings that are particularly relevant for our purposes, linking an individual's contextually-cued identity and their behaviour. In order to draw this link, Oyserman first defines two types of 'readiness' for behaviour. The first is action-readiness: the way that we are primed to act in a given context. The second is procedural-readiness: the way that we are primed to think (or process information) in a given context. The author's findings in this domain are consistent with other results seen so far. First, individuals are generally motivated to act in ways that are congruent with their identity. Second, the options an individual has before them for action and procedural readiness are triggered by identity and as such, are context dependent. Finally, action readiness does not necessarily prime an individual to act in their own best interest. For instance, a student-athlete who has resolved to get better grades when their student identity is active may make decisions that forego academics when their athlete identity is active.

The findings presented by Oyserman (2009) provide an insightful glimpse into some of the reasons individuals might make certain irrational, self-contradictory decisions. If the behaviour in question can be interpreted as a manifestation of action- or procedural-

readiness, then we might say that the individual was behaving in contextually-relevant, identity-congruent ways. These findings suggest, in particular, that an individual can be primed to process information or to act in seemingly contradictory ways, so long as they can be primed to take on the appropriate identity in a given context. This conceptualization of identity will serve as an important backdrop to the ways that individuals talk about contentious issues in online spaces.

Gender as a Contentious Issue

To some readers, gender might seem to be a completely benign topic - one that is not ripe for controversy. To others, it might be self-evident that this is a hot-button issue. This section makes that controversy explicit by exploring some of the ways that gender is discussed, experienced, and navigated in digital spaces.

In a pair of articles, Jane (2014a, 2014b) describes in depth the vitriol that women receive online, opting to call this type of discourse *e-bile*, and hopes that by giving it serious attention, we can begin to address the issue. As for examples of what this e-bile might consist of, both of Jane's (2014a) articles are titled with tame exemplars of the type of e-bile she is describing: "Back to the kitchen, cunt" (Jane, 2014a) and "Your a ugly, whorish, slut" (Jane, 2014b). More typical, vile examples include comments like this one, discussing an online video of a cheerleader falling from the top of a human pyramid: "She gave great blowjobs before her fall, now imagine the pleasure she will bring with out her front teeth" (Jane, 2014a), or this one, commenting on a researcher's "rapeability" after she publicly disclosed having been sexually assaulted at knifepoint: "She's so fugly, I wouldn't even bother raping her from behind with a box cutter" (Jane, 2014a). These examples seem extreme, and they should be understood as such. They are not, however, isolated incidents. Jane's (2014a) work cites countless examples in countless fora, clearly identifying this as an ongoing phenomenon. Jane characterizes this type of speech as an invective verbal assault that is often sexually explicit and gendered. Jane (2014a) points out that women are the overwhelming majority of the recipients of this type of discourse, but that men sometimes

receive it, too. However, in their case, the sexualized invective is often aimed at a female loved one; for instance their wives, mothers and/or daughters. Turning this invective into a real catch-22, Jane describes the ultimate irony of this discourse: it stipulates that women exist for nothing more than the sexual gratification of men, but then also looks down upon women in disdain for the exact same reason.

Jane (2014b) defines this e-bile as “any text or speech act which relies on technology for communication and/or publication, and is perceived by a sender, receiver or outside observer as involving hostility” (Jane, 2014b). Her definition may seem overly broad, but it is done so purposefully, largely in response to what she characterizes as the ‘second wave’ of research on flaming (another type of online invective), which defined flames as text acts which were of hostile intent by the sender, and recognized as such by a receiver and an observer. This trifecta of offence was incredibly difficult to achieve, and thus would leave a large number of offensive speech acts unexamined. Furthermore, Jane recognizes that some of this vitriolic speech doesn’t have hateful or hurtful intent, but might be better classified as “boredom” or “gaming” speech: speech acts that involve a bit of one-upmanship among its speakers, seeing who can outdo their peers in shock value and offensiveness. However, Jane (2014b) makes the case that this doesn’t make the speech any less problematic, particularly because disguising this cruelty as humour has the additional effect of de-legitimizing the suffering experienced by its recipients. Ultimately, this type of discourse leads to reduced inclusivity in digital spaces, as women withdraw from them because the price of participating is simply too high (Filipovic, 2007; Jane, 2014b).

Hurtful gendered speech doesn’t end with e-bile, however. Shepherd, Harvey, Jordan, Srauy, and Miltner (2015) focus on the constellation of hurtful behaviour contained within the GamerGate movement, wherein GamerGaters perpetrated hateful acts against women, all while claiming that their true intent was a moral crusade for “ethics in video game journalism”. Shepherd et al. report that while the talking points of the movement focused on this question of ethics, the actions of the group largely targeted (through e-bile, threats, doxxing, etc.) any person who dared question gender roles in video games, or who dared

promote the inclusion of typically under-represented individuals in game culture. The authors posit that GamerGaters may have viewed questions of inclusion in video game culture as a zero-sum game: one where the inclusion of others within the culture would reduce their own involvement, as opposed to viewing it as a growing culture within which more and more people could find something to enjoy. This perception of a zero-sum game is conducive to hateful behaviour, as loss-averse individuals take a stand to protect themselves from any perceived losses.

This shouldn't seem out of place, though. Individuals within the GamerGate movement largely vocalized that theirs was a moral struggle. This situates well within the bounds of the research previously examined in this document: individuals strive to hold correct attitudes, yet can take on other and sometimes contradictory perspectives through a series of biases. In this case, GamerGaters also potentially may have felt threatened that they were at risk of losing something dear to them – and in a community of like-minded peers where their behaviour may have been normalized – the conditions were ripe for hurtful behaviours to rise to the surface.

While gender may be a contentious issue online, digital spaces do exist where women do have the right and ability to speak and be heard, without threat of retaliation or unprovoked attack. Newsom and Lengel (2004) report that these spaces do exist, but that they are characterized by significant limitations; they call these spaces of *contained empowerment*. These spaces are forums, blogs and other sites, intentionally created to foster dialogue with individuals on equal footing, challenging traditional power hierarchies. While Newsom and Lengel recognize that these digital spaces give new power to women's voices, they argue that these spaces can only challenge hegemonic structures indirectly. Indeed, the power these sites offer is bounded in time and space to the specific online location where it occurs. The authors posit that these sites still offer hope, though. By giving women these networks within which to share with one another and to create and discover new meanings, change might 'spill' outside the boundaries of the hegemony and ultimately help to indirectly change existing power structures.

Modes of Inquiry

The previous section examined the ways in which humans engage in conversations about contentious issues in digital spaces. This section examines some of the ways these conversations have been analysed previously, both qualitatively and computationally, before delving into a discussion of topic modeling, a specific mode of computational inquiry applicable to this domain.

Qualitative

Sobieraj (2018) specifically looked at how women's voices, and women's very presence in digital public spaces are impacted by digital sexism. Her findings are consistent with much of the work examined in the previous section: Sobieraj finds that aggressors intimidate, shame and otherwise discredit women's voices, and undermine women's contributions to the conversation by using their femininity and femaleness against them, using gendered insults that draw attention to the female body and sexuality as a constant reminder that gender is at the heart of the attack. Sobieraj's (2018) methodology consists of qualitative interviews with 38 women who were victims of such attacks online, as well as informal conversations with internet safety experts and content moderators. While such research is valuable and helps us gain insight into the phenomenon, it is limited in the size of trends that can be analysed, given the costly and time-consuming nature of the in-depth interviews.

Switching focus slightly, but remaining within the realm of gender, Sutherland, Forbes, Hodgson, and McLaren (2014) explore the ways that conceptions of pregnant women's bodies are formed, negotiated and performed through language using discursive analysis. This takes the form of the authors reading the text and studying it for "conventional discourse analytic concepts, such as variation in language, design or structure of responses, their rhetorical organization, [and] accountability" (Sutherland et al., 2014, p. 106). Their intent with this methodology was informed by social justice concerns, and the authors hoped to uncover the processes which lead to injustice by focusing on how inequalities are evoked, produced and managed in a given moment and interaction. Discourse analysis, in this

regard, proves useful. Sutherland et al. are able to uncover how participants construct their posts, and for what purpose. While this is highly useful information, it is also highly resource-intensive, and the authors acknowledge that such a methodology is only feasible with a restrained data set.

Computational

Blevins et al. (2016) analyse tweets by gang-involved youth, attempting to automatically classify whether the tweets in question expressed the sentiments of loss and/or aggression, with the goal of using these classifications to signal community stakeholders to intervene in moments of loss in order to prevent reactionary aggression. The authors use a number of helpful techniques: such as a customized automated part-of-speech tagger to handle the non-standard English found in these tweets, and they include emotion scores of words used in the tweets by using the Dictionary of Affective Language. The authors also translate slang and emoji into standard English using a standardized phrasebook in order to use them in the classification task. This study is particularly interesting in its ways of coping with informal English, but does have one weakness: it is a supervised-learning method, meaning that the classifier can only be built using a large number of previously manually-annotated tweets, from which the statistical classification model is built.

Continuing on the gang-related theme, Blandfort et al. (2018) paired social work researchers and computer scientists together to analyze public tweets. They built an automatic classifier that uses the text of a tweet, and any image included within, to predict whether the tweet expresses any of three psychosocial concepts: aggression, loss and substance use. The authors posit that these three psychosocial concepts are important as they are potential avenues towards violence. The classifiers built by Blandfort et al. are trained on features such as unigrams and bigrams (continuous sequences of one and two words, respectively), unigrams and bigrams tagged with their corresponding parts of speech, and the minimum and maximum pleasantness, activation and imagery scores of the text of each tweet, as determined by lookups in the Dictionary of Affective Language. The authors find

that coupled with automatically extracted visual cues from images in tweets, these features are useful in detecting aggression and substance use in tweets. They also note, however, that the process of manual annotation is critically important to the task, and therefore, that we are far from a completely automated process.

Chang et al. (2018) also aimed to build automatic classifiers to detect loss and aggression in gang-related tweets, working from the assumption that events of loss (i.e.: losing a loved one) in a gang-related context could result in retaliatory aggression. The authors also tried to somewhat sidestep the costly problem of needing manually-annotated data by creating corpus-specific word embeddings (numeric representations of words, which define the relationships among words in a corpus in a high-dimensional space). Also, given the highly-specialized type of language in these tweets, the authors used a semi-automated method to generate a list of loss and aggression related words. They asked manual annotators to generate a short list (29 and 40 for aggression and loss, respectively), and then used the SENTPROP algorithm to generate a probability of association for each of the words in the corpus to the concepts of loss and aggression. The authors also augmented the content of each tweet with contextual information, including the results of the analysis on the author’s previous tweets, and tweets from other users that included the author. Their best experimental results show an F1 score (combination of precision and recall) of 42 for aggression and 75.9 for loss, showing that while the method is useful to a certain extent, there is still plenty of room for human intervention.

Shifting away from the gang-related studies, Saleem, Dillon, Benesch, and Ruths (2017) identify hateful speech in public fora in an automated fashion with as little human intervention as possible. Their study aims to identify hateful speech in online social spaces while also bypassing the need for an extensive manual annotation process (as required by many machine learning algorithms, including those discussed above). The authors point out that keyword-based approaches are insufficient to identifying hateful speech as there is an overlap in vocabulary use between *hate* groups and *support* groups (i.e.: keyword-based methods can identify speech relevant to the target group, but not necessarily hateful speech

towards that group). The authors test a new method, targeting three groups particularly vulnerable to hateful speech online: black, plus-size and female.

The central question motivating Saleem et al.'s (2017) study is: given posts from self-identified hateful subreddits and support group subreddits for each of the three groups, can an automatic classifier distinguish examples of hateful speech from examples of non-hateful speech? In this task, the training data is never explicitly annotated: the authors simply make the assumption that *all* posts from the self-identified hate group are hateful, and that *all* posts from the support group are supportive.

Saleem et al. (2017) find that classifiers trained in this way are more effective at identifying hateful speech than traditional keyword-based approaches. For each domain (black, plus-size, female), the community-language based classifiers outperformed comparable keyword-based classifiers. Switching to a community-language based model leads to a 10-20% increase in accuracy. This method is of particular interest because of its success in using few manually-generated resources, yet still outputting good, usable results.

Looking at argumentative discourse more generally, Danescu-Niculescu-Mizil, Lee, Pang, and Kleinberg (2012) look specifically at power relationships. They study Wikipedia Talk pages and Supreme Court discussions between lawyers and judges, and identify patterns of speech that identified power relationships among the speakers. They also investigate how discourse changes as individuals change in power level. In analysing these texts, the authors consider a wide variety of linguistic features, as follows: linguistic style markers (marked by the use of function words with little semantic meaning – articles, auxiliary verbs, conjunctions, high-frequency adverbs, impersonal pronouns, personal pronouns, prepositions and quantifiers), coordination measures (how much an individual adopts the previous speakers' use of markers in an immediate reply, as compared to the baseline of their typical word usage when replying to the same speaker), and coordination toward a group (generalizing the previous, except counting all the replies to the members of the group).

Danescu-Niculescu-Mizil et al. (2012) find that users generally coordinate their behaviour more with higher-powered individuals than lower-powered individuals. While the

authors hypothesized that high-powered individuals would not coordinate with lower-powered ones, the Wikipedia data shows otherwise. They explain this by hypothesizing and showing that high-powered individuals (in this case, admins) had to earn that rank, and that over time, they displayed that as a group, they were more coordinating than other individuals (in contrast, individuals who wanted admin status but did not get it did not show the same level of baseline coordinating behavior). When controlling for personality, they find evidence that lends support to their hypothesis. They see the same behaviour in judge/lawyer interactions, where the lawyer is usually dependent on the judge for a favourable ruling. Lawyers coordinated more with unfavourable judges, and unfavourable judges coordinated less with lawyers.

It is interesting to observe that the coordination described in the previous study (Danescu-Niculescu-Mizil et al., 2012) is largely unconscious: it is reasonable to assume individuals are not counting their use of auxiliary verbs to match another individual's, for instance. The effect is thus largely sub-conscious and possibly inadvertent.

Continuing this move towards more abstract studies of debates, Somasundaran and Wiebe (2010) aimed to automatically recognize the stance of a particular text in ideological online debates. To recognize a stance, the authors hypothesize that we must understand the sentiment of a text toward its target (e.g.: negative sentiment towards health insurance companies can be interpreted as a pro-healthcare stance). They create an argument lexicon from a dataset, which is annotated with arguing “trigger expressions” (expressions that introduce an argument). Intuitively, the authors hope that identifying sentiment and expressions that introduce arguments within a text will lead to a better automatic classification of the stance of the text. The authors build Support Vector Machine (SVM) classifiers on four topics: gun rights, gay rights, abortion and creationism. As a baseline comparison, the authors use an SVM trained with a feature set of unigrams – this classifier performs well, even outperforming a classifier trained on sentiment data alone. However, combining Argument and Sentiment lexicons leads to a better classifier. Additionally, given the descriptiveness of their feature set, their classifier might have more explanatory power.

All of the above studies present interesting ways of looking at debate, stance, and controversy in online settings. However, except the work by Saleem et al. (2017), every study looked at so far relied on the *supervised* machine learning paradigm. Supervised machine learning methods allow a computer to learn a mathematical model based on a number of previously manually-annotated examples. While these methods may be effective, they are also resource intensive, as they require experts to manually go through a large number of examples before a model can be built. In this project, I instead made use of *unsupervised* machine learning methods. These do not require previously annotated examples to produce results, although the types of questions they can answer are somewhat different. Nonetheless, these methods can still be leveraged in order to automatically glean insights on large quantities of text with little manual effort.

Topic Models

The unsupervised method I used in this project is known as topic modeling. A topic model is a statistical model that describes abstract concepts (“topics”) that occur in a corpus of documents (Blei & Lafferty, 2009). Topic modeling is the process of uncovering these abstract topics. This process takes as input a collection of documents and generates a list of probabilities that each document was generated from each of a list of potential topics uncovered in the corpus. Furthermore, the *topics* predicted by topic modeling are each described by the most common words found in that topic in the corpus.

The Latent Dirichlet Allocation (LDA) algorithm by Blei, Ng, and Jordan (2003) is one of the standard ways of automatically finding topic models for large corpora. Fundamentally, the algorithm assumes that a corpus is generated by a set of k topics, where each of the topics is associated with some set of words. The algorithm then assumes that each document in the corpus is a mixture of these k topics, and thus has some likelihood of containing words from each of the topics which generated it. We can illustrate this with a simple example: if one of the categories in a corpus is *dessert*, the words associated with the category might be *cake*, *bake*, *pastry* and *fruit*. Another category, *pets*, could be associated

with the words *dog*, *cat*, *play* and *companion*. A document might then be generated from one or both of these categories using the words that define the category. If we see more words like *bake* and *cake* in the text, we might assign a greater probability to the document of being generated by the *dessert* category. If the document contains few words from the *pet* category, we might assign a low probability that the document was generated primarily by the *pet* category.

With these assumptions, Blei et al.'s (2003) algorithm not only generates a list of probabilities that each document is generated from each topic, but also determines the list of words associated with each of the k topics, which can be of help to a human analyst to determine the semantic significance of each of these topics. In this project, I worked specifically with the implementation of the LDA algorithm provided by the Mallet software package in Java (McCallum, 2002).

While the above assumptions may be gross simplifications of the processes by which documents in a corpus are *actually* generated, other applications of topic modeling provide convincing evidence that these simplifying assumptions can still lead to usable results.

Wallace (2012) uses an LDA-based approach to parse out the major narrative threads of David Foster Wallace's novel, *Infinite Jest*. Paul and Dredze (2012) finds that applying topic modeling to posts on a drug discussion forum could help to automatically infer information about the drug discussed in a post, the associated delivery method (e.g. ingestion, vaccine) and some aspect of drug use (i.e. health, culture or effects). Gerber (2014) uses topic models, augmented with geolocation data of tweets from Chicago and historical crime maps from the city's available data, to assist in predicting future crimes. Althoff, Danescu-Niculescu-Mizil, and Jurafsky (2014) uses topic modeling on posts from the subreddit *Random Acts of Pizza* to determine that the narrative behind a request for a gift of free pizza was a significant predictor of the success of that request. Of particular interest to this project, Baumer, Mimno, Guha, Quan, and Gay (2017) finds that topic models can converge with grounded theory qualitative analyses, although the models may require some human interpretation, contextualization and combination to match the themes uncovered

by grounded theory analyses.

The Brett Kavanaugh Controversy

Previous sections of this thesis examined a framework in which to understand how individuals argue and discuss contentious issues online, established gender as one of those contentious issues, and outlined a variety of ways that contentious issues have previously been studied. This project aims to use computational methods to uncover themes present throughout the discussion of a contentious issue. To do so, I have specifically narrowed in on one recent issue from the news that generated a significant amount of controversy online. This section broadly outlines the issue of Brett Kavanaugh's nomination to the United States Supreme Court in 2018, and the controversy surrounding that nomination amid allegations of sexual assault in Kavanaugh's past.

For context, Roberts (2018) describes the timeline of events as follows: in July 2018, U.S. President Donald Trump nominated Brett Kavanaugh to replace an outgoing Justice. Days prior to the nomination (according to testimony), Christine Blasey Ford contacted her congressional representative and the Washington Post's tip line (the latter, anonymously) to inform them that a potential candidate to the Supreme Court had sexually assaulted her when the pair were teenagers. In late July, Ford sent a letter to Diane Feinstein (a Democrat on the Senate Judiciary Committee) about the assault, requesting confidentiality. Kavanaugh's confirmation hearings began in early September 2018, when reports began to be made about the letter sent to Feinstein. Feinstein stated she had submitted the issue to investigative authorities, and did not share the accuser's identity given the request for confidentiality. On September 14, Brett Kavanaugh denied any wrongdoing, before Ford publicly made her claims in a Washington Post article on September 16 (citing that several journalists had learned of her identity as the reason why she chose to go public with the story). The Senate Judiciary Committee's nomination vote was thus delayed so that the committee could hear from both Kavanaugh and Ford. In the meantime, other women came forward, accusing Kavanaugh of behaviours ranging from inappropriate to abusive. In late

September, both Kavanaugh and Ford testified individually at the Senate panel's hearing, which was viewed live by a wide audience. A deluge of social media posts came forward both about this specific issue, but also launching forth discussion of sexual assault more largely, with a large number of (mostly) women describing their own stories of sexual assault with the hashtags #WhyIDidntReport and #BelieveSurvivors. Many women protested the nomination, which triggered a number of protestor arrests. Less than a day after the hearing, the Senate committee ultimately decided to back Kavanaugh as a nominee to the Supreme Court, after asking the Trump administration to launch an FBI investigation into the allegations. Trump agreed and said the investigation should take less than a week. The investigation finally could not corroborate the allegations. The conclusion was controversial, however. Democrats called the investigation incomplete, as Ford claimed that neither she nor other witnesses she listed had been interviewed by the FBI. Republicans were satisfied by the investigation, and questioned why Democrats and Feinstein particularly had not taken the matter to the FBI earlier. Kavanaugh was ultimately appointed to the Supreme Court in early October 2018.

The controversy around Brett Kavanaugh's nomination in the midst of these accusations was often gendered. It included discussions of gender roles portrayed by the concerned parties in their testimonies regarding the events and the different perceived affordances given to men, women, and the alleged survivors and perpetrators of sexual assault (Crary, 2018; Malone, 2018; Rhor, 2018; Vesoulis, 2018). It thus appears to fit as a controversial issue within the broader theme of gender, which as described previously, is a significantly contentious issue in online spaces.

Addressing Subjectivity

Researcher subjectivity is inescapable. Indeed, as previously argued, individuals are constantly subjected to competing influencing factors, and subject to a range of identities which may shape their actions and ways of interpreting information. I am no exception. I herein aim to provide some account of my own subjectivities and biases, insofar as I am

conscious of them and of the fact that they have influenced my performance as researcher. I am, by profession, a software developer with a background in computer science. It is my inclination to use computational tools to assist humans in solving problems. This certainly played a part in my choice of methodologies to consider in tackling this problem. Furthermore, I recognize my own inclinations towards discussions in social media spaces: I value listening and comprehension in interpersonal exchanges, and appreciate the notion that one should understand another’s perspective before launching into debate. Personally observing the divisiveness, hostility and entrenchedness of a number of social media discussions propelled me to consider such a research topic, and necessarily affected many of the decisions I made throughout this work. These factors are not exhaustive descriptors of my subjectivity, but it is my hope that they address some of the major ways in which my subjective experience guided this project.

Research Questions

The earlier description of the Kavanaugh controversy leads to the first research question that this project dealt with: “What are the different topics or themes that are involved in discussion of the Brett Kavanaugh controversy in online spaces?”

Complementarily, the second research question was concerned with *how* to respond to the first: “Given the large quantities of social media texts available to us, how can these topics or themes be discovered using computational tools?”

Methodology

Setting and Data Collection

This research project’s primary goal was to delve into the online discussion surrounding the Brett Kavanaugh controversy in 2018, which garnered substantial attention on social media (Cummings, 2018; Gelman & Wilson, 2018). This project focused specifically on reactions posted to the social media site *reddit*¹, as its structure is conducive to the

¹<https://www.reddit.com>

proposed analytic method. reddit is a popular social media website; as of January 2020, it ranked sixth (Amazon.com, 2020) among the top sites viewed in the United States, above Wikipedia and Ebay, and behind Google, YouTube, Amazon, Facebook and Yahoo. reddit is divided into a number of distinct social media communities called *subreddits*, which vary wildly in topic of interest, but also in their code of conduct, set of administrators, and so on. For instance, the subreddit *Random Acts of Pizza*² is devoted to individuals requesting and offering pizza across the world; while the subreddit *Change My View*³ is dedicated to individuals presenting a stance they hold, and asking other users of the subreddit to argue against it and to change the original poster's stance. More directly useful to this project, several subreddits self-identify specifically on lines of how they view gender. This particular property is interesting in order to examine the way people with different views on gender talk about a given gender-charged issue. In particular, in this project, I gathered discussions from a number of subreddits. First, I considered the Politics⁴ subreddit, dedicated to discussion of U.S. political issues, with no specific stance on gender. The Politics subreddit has been established since 2007 and has over 5 million users. I also included posts from the subreddit TwoXChromosomes⁵ (established 2009, over 12 million members), a subreddit specifically dedicated to considering issues from women's perspectives. Additional posts were collected from subreddits dedicated to men's perspectives, notably, MensRights⁶ and MensLib⁷. These however have much smaller user bases (245,000 and 105,000, respectively), and are perceived to represent different viewpoints. MensRights is perceived by some to be misogynistic (Caffier, 2017), while many on the MensRights subreddit view the MensLib subreddit to be misandristic and rife with toxic femininity⁸. While the size of the men's

²https://www.reddit.com/r/Random_Acts_Of_Pizza/

³<https://www.reddit.com/r/changemyview/>

⁴<https://www.reddit.com/r/politics/>

⁵<https://www.reddit.com/r/TwoXChromosomes/>

⁶<https://www.reddit.com/r/MensRights/>

⁷<https://www.reddit.com/r/MensLib/>

⁸Some of the top results when searching for the keyword "menslib" on the MensRights subreddit are the following threads: "I've just lost my final shred of respect for r/menslib. They have officially become a group of man/self haters." https://www.reddit.com/r/MensRights/comments/cnf3jm/ive_just_lost_my_final_shred_of_respect_for/, "Toxic Femininity on r/MensLib" https://www.reddit.com/r/MensRights/comments/eia7v6/toxic_femininity_on_rmenslib/, "r/MensLib is a misandrist joke" https://www.reddit.com/r/MensRights/comments/dirtzm/rmenslib_is_a_misandrist_joke/, and "An-

interests subreddits are not sufficient for the computational methods proposed in the next section, they may provide some clues for interpreting its results, and thus were included in the data collection phase.

In general, reddit threads follow a common structure. One main post shares an idea - be it a blurb of text, an image, a link to a video or article, etc. Redditors can then respond to the main post and to one another, creating a branching tree-like discussion about the main post. reddit also allows users to upvote and downvote each post, which influences the visibility to the post, as well as giving other users some idea of the public reaction to said post. reddit's application programming interface (API) permits us to extract these conversation trees for any public thread, which is particularly useful for the purpose of this study.

The data was collected by searching for different combinations of the keywords "Kavanaugh," "Brett Kavanaugh," "Blasey Ford" and "Christine Blasey Ford", selected from late 2018, when the issue was front and center in the news. Among the results, the aim was to select a small number of large discussions, and as much as possible, about similar facets of the issue at hand - in this case, Brett Kavanaugh's nomination while being accused of sexual assault, and the U.S. Senate Judiciary Committee's hearings on the issue. For data collection from the Politics and TwoXChromosomes subreddits, special attention was made to selecting threads which discussed the same news articles, where possible. A total of over 3,000 posts was collected from each of the two subreddits, which formed the basis of the dataset for exploration with the topic modeling approach.

For further gender-related context to the Kavanaugh nomination, threads were gathered from the MensRights and MensLib subreddits from late 2018, where the thread title contained one of the keywords mentioned above, and where the thread contained at least 35 comments, in order to ensure that the threads being considered had gained some traction in their respective communities.⁹

other example of why Menslib is for spineless pussies. They are too afraid to talk about one of the most important men's issues" https://www.reddit.com/r/MensRights/comments/cz5e1e/another_example_of_why_menslib_is_for_spineless/

⁹A full list of reddit threads selected in the data collection is available as Appendix I - this includes

All selected posts were ultimately put into a database containing the post contents and some of the metadata about the post including reddit's unique identifier for the post, the post it was written in reply to, and its score (a total of the number of upvotes and downvotes it received). The next section examines the analytic method applied to the posts collected from the Politics and TwoXChromosomes subreddits.

Overview of Analytic Methods

This project had two main goals: one, to identify narratives used to discuss the Brett Kavanaugh controversy on social media, and two, to examine more automated methods for gaining these insights. To these ends, I used *topic modeling* to automatically uncover topics found in the social media posts. In Modes of Inquiry (p. 12), I described the topic modeling process as one that accepts a corpus of documents (essentially each a series of words) and describes a set of k topics used within the corpus by the words that most frequently appear in each topic. As described, this method has been tried and tested in a variety of contexts. This sections examines how I employed the method to address the project's goals, largely built upon previous work I completed with Project Someone¹⁰.

The method largely consists of the following steps, which I describe in greater depth below:

1. Preprocessing: standardize the text into machine-actionable format.
2. Generate sets of k topics using the LDA algorithm, for values of k in the set $\{[2 - 15], 20, 25, 30, 40, 50\}$
3. Select one of the sets of k topics with sufficient interpretive salience.
4. Analyze the selected set of topics for themes.

The preprocessing phase involves transforming the raw text collected from Reddit into a standardized format that can be read and worked on by a computer. The first step

threads from the Politics, TwoXChromosomes, MensRights and MensLib subreddits.

¹⁰See Rodier (2019), discussed further in Practical Applications (p. 82)

is to *tokenize* the text. That is, the raw text is transformed from a string of words into a list of *tokens*, where each token is an individual word from the text. Each token is then transformed to lowercase (to avoid considering different capitalizations of a word as different words, e.g. “Action” and “action” will be equivalent). From this list of tokens is removed any token that is not comprised solely of letters, or that is present in a list of *stop words*, words that are common in a language and would likely appear in texts regardless of their content (e.g.: common words like *the* and *a* which would yield no interesting results for our analysis).¹¹

The LDA topic modeling algorithm requires the user to specify the number of topics present in the text. This is unfortunately counter to my aims: the goal being rather to uncover an unknown number of latent topics within the text. Therefore, I performed a series of analyses, running the topic modeling algorithm for differing numbers of topics, and then examining the results for each. I ran the LDA topic modeling algorithm on the corpus a total of 19 times, once for each number of topics ranging from 2 to 15, and 20, 25, 30, 40, and 50.

These 19 candidate runs’ outputs were then analysed further. For each run, I examined the twenty words best representing each proposed topic, as well as the twenty documents (thread posts) deemed most representative of each topic. Admittedly, twenty is a fairly arbitrary number, but enough to usually allow a reader to see the most important words and documents in a topic. If twenty were to be insufficient, the remainder of the data remained accessible in a database if it was deemed necessary to drill further down into the topic.

This process produced a set of 19 results packages, each containing a set of topics (each of which is in turn characterized by a set of representative words and posts). From these, I formed an initial interpretation of the topics by reading through the individual results packages, and selected the set of topics with the most interpretive salience. Rodier

¹¹I used a list of stopwords found at <https://gist.github.com/sebleier/554280> by user *rishad2m8*, for its thorough list of common words and inclusion of contractions like *it’s* which would otherwise dominate the topic groupings.

(2019) and Baumer et al. (2017) both find that there is significant overlap between these topics and those a human interpreter might find.

This set of topics provides high-level summaries of some of the most frequent co-occurrences of words throughout the corpus, which should enable meaningful qualitative interpretation: indeed, the algorithm is primarily concerned with mapping common utterances together, thus grounding all further analysis directly in the words of the source texts. These topical groupings will subsequently be used to inform my interpretation of discussion themes happening throughout the corpus.

Credibility

Care was taken in the execution of this project to try to remain credible. In a project such as this one, where there is a large number of internet users (often anonymous) contributing to the gathered corpus, it is practically impossible to ask each redditor to participate in member checking. A compromise, however, is to remain as close as possible to the actual utterances as they appear. In this manner, the topic modeling approach is particularly useful: it is grounded in the actual words of users, and uses the collocation of words in their texts to categorize them based on word-usage frequency patterns within the corpus. This allows the grouping of documents together based on the actual language within, and not on the basis of interpretation. Throughout my analysis, and in particular, in the step of interpreting the output of the LDA algorithm, I attempted to focus my interpretation using the most representative words for each topic, and how they appeared in individual redditors' posts.

While the findings and interpretation of those topics are my own, additional data sources were consulted throughout this document to add understanding from different perspectives. In particular, data was collected from some of the men's interests subreddits which both challenged and confirmed some of the findings (as discussed further in this thesis). A set of news articles was also read - not only about the Kavanaugh controversy, but also about its fallout (as cited in *The Brett Kavanaugh Controversy* (p. 19)). Adding views

from these two different types of sources with wildly different viewpoints assisted in ensuring that my understanding was broader and included additional perspectives with which to view the data.

Furthermore, I have made all of the materials that informed my interpretations available through this document's appendix. Not only the raw data has been made available, but the various presentations of the raw data that may have influenced my interpretation along the way. For instance, I have included the specific ordering of topics and posts in the order I read them, and the metadata that I viewed at each step of the analytic process. Where relevant and not explicitly stated in this document, I have also included in the appendix notes I made along the way as a means for the reader to retrace my reasoning.

Analysis

Selecting a Number of Topics for Analysis

For the dataset collected from the TwoXChromosomes and Politics subreddits, the LDA algorithm was run a total of 19 times, considering for each run a number of topics k among the values of 2 through 15, and 20, 25, 30, 40 and 50. For each of these runs, I then read through the top words and documents for each topic. The run generating the most coherent topics - or more specifically, those with the greatest subjective interpretive salience - was selected for further analysis.

The literature unfortunately does not offer a definitive method for choosing k , and this likely remains an open research problem which might be tackled from both computer and social science domains. Some prior studies essentially use either a static number of topics (Althoff et al., 2014; Wallace, 2012), while others pick one value among several that they experiment with (Baumer et al., 2017; Gerber, 2014), without diving deeply into how a selection was made.

In order to select a number of topics, I read through the 20 most representative posts generated by each "topic" in a given run of the LDA algorithm, as well as the 20 most representative words for that topic. None of the reddit-based metadata for each post was

viewed during this process to avoid introducing further subjective bias. The only metadata included in this phase were the top 20 words' representativeness of the topic, and each post's score as likely as being generated from a given topic. This helped to clarify how likely the algorithm considered a given word or post to be for a given topic. With this information, I wrote a brief summary of each topic, and indicated if it was *noisy* or not (e.g.: if the topic was coherent and cohesive, or if a narrative was spotty, or interspersed with posts that didn't fit the narrative). The results packages for each run of the algorithm, as well as the notes I took are available as an appendix to this document, and it is my hope that they shine a light on some of the subjectivity required to use this computational method. Indeed, while the aim is to use computational and statistical methods to automate a part of the work of qualitative discovery, it is important to note that a computational or statistical approach does not rid the process of subjective interpretation.¹²

The chief source of subjectivity I found to be introduced was unsurprisingly in the reader's (in this case, my own) interpretation of posts. Further, the *order* of reading through the runs' results also played a part. I read through the results going from $k = 2$, upwards towards $k = 50$. This means that the first results packages I read through contained very broad topics, and as the value of k increased, the results packages contained more and more fine-grained topics. The process gave the impression that some of the topics were indeed "crystallizing" as I moved towards an optimal number of topics. For instance, one broad topic that existed for lower values of k combined discussions of the FBI investigation, the burden of proof for determining guilt, and Brett Kavanaugh's character. As k increased, much more clearly-defined topics emerged for each of these themes. Similarly, for most of the result packages, discussions of the presumption of innocence and the burden of proof required to assign guilt were bundled together. Once the number of topics hit 50, however, they suddenly appeared to be their own contained topics. While this crystallization did in fact offer some insight, it is reasonable to question if I would have reached the same (or

¹²Appendix II presents the raw results from each of the 19 runs of the LDA algorithm (i.e. the "topics" found in each run, represented by their top words and posts). Appendix III includes the notes I made regarding each topic in each of the results package. These went on to form the basis of the rationale for the selection of a results package to work with.

different) insights if I viewed the results packages for varied values of k in a different order.

Another potential source of subjective bias can come from the context in which reddit posts were initially viewed. For instance, upon an initial reading of the results package for a low value of k , I might have come across a post categorized in a topic seemingly about “bias in selecting a Supreme Court Justice”. In a later results package for a different number of topics, the same post might have been included in a topic about the desired character of a Supreme Court Justice, or in Brett Kavanaugh’s emotional displays during his hearing. It is reasonable to question what subjective effect this order of viewing might have on interpreting a post’s theme or topic.

Summary of Topic Modeling Output

This section examines the output of the topic modeling algorithm with the number of topics set to 20, as this results package displayed the best combination of consistency and coherence within topics. In this portion of the analysis, the results package for analysis also includes metadata for each post within a topic: specifically, the subreddit and thread in which the post was made in order to get a sense of the difference in discourse between the threads and subreddits.¹³

I will spend a substantial number of pages here diving into each topic; later in this document, I will recombine some of these topics and interpret what they can tell us in order to answer the first research question, which seeks to identify the different narratives used to discuss this contentious issue online.

Table 1 lists the twenty generated topics with their ten most representative words. This provides an adequate launching pad to jump into a discussion of each topic. A brief summary of each topic is presented in table 2, although these will be explored in more depth in the coming pages. A note that this table also describes a topic “type”. While the topic modeling algorithm, by definition, sorts documents (which are, in this case, reddit posts) into groups based on shared tokens (i.e. words) within the documents, the nature of the

¹³Appendix IV presents this specific results package with the additional metadata, which was used to inform the analysis presented in this section.

similarity between documents in a topic is not always the same. Often, documents were grouped thematically (for instance, discussion of polygraphs). Other times however, they were linked by type of dialogue (or discourse), and in one instance, by typographic style. Two topics were too noisy to truly decipher, and are labeled as such.

Topic 1 it's don't i'm that's you're didn't people he's doesn't can't	Topic 2 years kavanaugh ford ago happened party remember therapist people time	Topic 3 word ideas medical found amount mental hysteria lazy general ban	Topic 4 court supreme kavanaugh judge job justice highest scotus lifetime appointment	Topic 5 shit fucking people fuck yeah time post trump thing bullshit
Topic 6 sexual assault victim victims rape women woman abuse forward movement	Topic 7 feinstein letter ford people public leaked story office trump job	Topic 8 people person make good time things point thing feel agree	Topic 9 constitution law conservatives states legal rights abortion history government american	Topic 10 polygraph polygraphs test lie lying questions truth science tests results
Topic 11 democrats republicans trump party vote political republican time allegation dems	Topic 12 women men sexual white woman assault kavanaugh man believed sexually	Topic 13 questions hearing question kavanaugh answer lied anger asked job response	Topic 14 drinking beer school drunk high yale kavanaugh triangle game alcohol	Topic 15 money reddit white race gender work give response speech bloated
Topic 16 guilty innocent job proven evidence trial proof interview prove innocence	Topic 17 accused life rape man innocent angry accusations falsely rapist ruined	Topic 18 investigation fbi investigate kavanaugh confirmation process truth democrats committee senate	Topic 19 evidence people testimony credible case ford kavanaugh fact reason public	Topic 20 hysterical kavanaugh man called white women black woman emotional men

Table 1

Ten most representative words for each of the twenty topics.

Topic ID	Description	Type
1	Alternate apostrophe	Typographic
2	Recollection of alleged assault	Thematic
3	Hyperlinks	Noisy
4	Nomination to Supreme Court	Thematic
5	Swearing	Discursive
6	Sexual assault; perception of accused and accusers	Thematic
7	Going public with Dr. Ford's story	Thematic
8	(Dis)agreement, opinion & argumentation	Discursive
9	Constitution, U.S. history, laws	Thematic
10	Polygraphs	Thematic
11	Partisanship of Supreme Court nomination	Thematic
12	Sexual assault, gender & race	Thematic
13	Perception & interpretation of behaviour in hearing	Thematic
14	Brett Kavanaugh: lying during hearing?	Thematic
15	Extreme viewpoints?	Noisy
16	Burden of proof; presumption of innocence	Thematic
17	Accusations of rape & their impact	Thematic
18	Investigation into allegations	Thematic
19	Quality of evidence (particularly testimony)	Thematic
20	Gendered & racialized perception of emotional displays	Thematic

Table 2

Brief summary of topics and their observed grouping type.

Topic 1

The first topic is strange insofar as it is difficult to tease a distinct narrative from it. Its list of representative words are all short words, and nine of the ten noticeably contain an apostrophe. Furthermore, the majority of these words are included in the list of stopwords - words so common that they should be ignored by the topic modeling algorithm (see Overview of Analytic Methods (p. 24)). While their inclusion was initially a point of confusion, it turns out that while the majority of apostrophes are written using the key next to the ENTER button on a standard QWERTY keyboard, some typists will use a single quote using a different character. The difference can be seen by comparing the former and

latter here: ’ vs ’. This first topic is thus a collection of short posts that include words with this alternate apostrophe. As an illustration, here are some of the most representative posts in this topic:

1. Stupidest statement I’ve ever heard. Gross exaggeration of maybe a few crazies. ¹⁴
2. It’s slut shaming someone... people do see each other’s yearbooks. Even if they don’t, it’s wrong.
3. Just so you’re aware, she isn’t claiming she was raped

The presence of oft-used words like “I’ve”, “it’s”, “don’t” and “you’re” with the alternate apostrophe creates a series of posts that are linked through their usage of these alternate forms of our stop words. While this grouping is not fundamentally informative about the themes within the corpus, it does illustrate the algorithm’s functioning, and is therefore informative nonetheless.

Topic 2

The second topic provides a clear narrative thread running through its posts. Notably, the discussion here centres around people’s memories of the events surrounding a past assault. Some of the topic’s most representative words immediately point us in the direction of memory and the events of the allegation: “years”, “ago”, “party”, “remember”, and “happened”. The ways in which these words appear in some of the topic’s most representative documents is examined below.

The most representative post in this topic includes one reddit user contradicting another user’s interpretation of some of the allegations:

- 1.> Mark Judge’s ex gf says Mark Judge admitted to gang raping a drunk girl.¹⁵

Not what she said.

> Razor says that Judge “told her ashamedly” about an incident in which he and other boys took turns having sex with a drunk woman, per the New Yorker. He seemed to think it was

¹⁴The bold number at the beginning of any quote refers to its position among the most representative posts for a given topic, 1 being the most representative post of the topic as output by the LDA algorithm.

¹⁵Note: the > character in a reddit post signals that the line is a quote; either of another reddit user or from outside material.

consensual, she added, and he didn't name anyone else who had been involved. There is no indication that Kavanaugh was one of the boys, the New Yorker notes.

Also.

> In a statement, Judge's attorney said that he "categorically denies" Rasor's account.

The second most representative post also uses a quote to contradict another user, this time using an individual's recollection of rape to support the allegations against Brett Kavanaugh:

2. From Julie Swetnick's statement:

>In approximately 1982, I became the victim of one of these "gang" or "train" rapes where Mark Judge and Brett Kavanaugh were present. Shortly after the incident, I shared what had transpired with at least two other people. During the incident, I was incapacitated without my consent and unable to fight off the boys raping me. I believe I was drugged using Quaaludes or something similar placed in what I was drinking.

So actually, yes, Brett did rape someone.

Both previous examples quote an individual's memory of prior events to illustrate their point. That tendency continues in the next example, which attempts to lay out the likelihood of Brett Kavanaugh being an assailant of Dr. Ford based on her recollections. An excerpt, for illustrative purposes:

7 (quoted by 3).[...]

Do the facts make Brett Kavanaugh a likely assailant of Dr. Ford around 36 years ago?

Dr. Ford does not recall the year, the month or day the alleged incident took place.

She does not recall whose house or street that the assault took place

She does not recall how she got there, or how she got home

She does recall having just one beer at the party

She recalls four other people at the party: Kavanaugh, Mark Judge, Patrick PJ Smith, Leland Keyser (whom she claims is her lifelong friend). Kavanaugh/Smith/Judge/Keyser have all issued statements *under penalty of perjury*, that they do not recall attending any such party, or any assault taking place.

Dr. Ford stated that she revealed details of her assault to her therapist in May 2012. The therapist's notes told her there were "four boys" in the bedroom, not two as she has stated.

The notes do not name Kavanaugh. The notes also indicate Ford said she was in her late teens when she was assaulted. But Ford now says she may have been only 15.

[...]

We see in this topic an importance granted to the memories shared by individuals involved with the allegations. The topic also includes some descriptions or questioning of what Dr. Ford recalled in a session with her therapist; for example:

11. Those therapist notes have been proven inaccurate. There are more than a few holes in the polygraph as evidence claim. As far as she told many people about this incident, if I recall only her husband has stated that she named Kavanaugh prior to her letter to the senator.

and

12. >the only thing i was implying is that dr ford had told people many years ago that this happened, that it was brett kavanaugh.

Who did she tell it was specifically Brett Kavanaugh. She didn't even tell her therapist that.

Recollection of the events and what Dr. Ford told her therapist are consistent themes throughout this topic; as is the practice of reddit users interpreting and using individuals' recollections to build their own interpretation of the events in question.

It should be noted finally that posts in this topic emanated from both the TwoXChromosomes and Politics subreddits.

Topic 3

Most of the topics presented with the number of topics set to 20 have some interpretative salience. Topic 3 is one of the noisiest, however, with no clear single narrative thread keeping the various posts in order.

The top four posts in this topic are form-type posts: two of them are the generic introductory post in each thread on the Politics subreddit:

1. As a reminder, this subreddit [is for civil discussion.] (/r/politics/wiki/index#wiki_be_civil)

16

¹⁶The format [text](web/address) represents hyperlinks on reddit. The text within square brackets is what is visible to the reader, and the brackets contain the URL to which the link points.

In general, be courteous to others. Attack ideas, not users. Personal insults, shill or troll accusations, hate speech, **any** advocating or wishing death/physical harm, and other rule violations can result in a permanent ban.

If you see comments in violation of our rules, please report them.

I am a bot, and this action was performed automatically. Please [contact the moderators of this subreddit](/message/compose/?to=/r/politics) if you have any questions or concerns.

The other two are the messages generated on the Politics subreddit after a comment has been deleted for not meeting community standards:

2. Hi 'Nazisarebadpeople'. Thank you for participating in /r/Politics. However, [your comment](<https://www.reddit.com/r/politics/comments/9k5arb/-/e6wqa7p/>) has been removed for the following reason(s):

* Your comment does not meet our [comment civility rules] (https://www.reddit.com/r/politics/wiki/index#wiki_be_civil). Please be civil.

If you have any questions about this removal, please feel free to [message the moderators.] ([https://www.reddit.com/message/compose?to=/r/politics&subject=Question regarding the removal of this comment by /u/Nazisarebadpeople&message=I have a question regarding the removal of this \[comment.\]](https://www.reddit.com/message/compose?to=/r/politics&subject=Question%20regarding%20the%20removal%20of%20this%20comment%20by%20/u/Nazisarebadpeople&message=I%20have%20a%20question%20regarding%20the%20removal%20of%20this%20comment.) (<https://www.reddit.com/r/politics/comments/9k5arb/-/e6wqa7p/?context=10000>))

Other posts among the most representative also include hyperlinks - one of the possible similarities that has grouped these particular posts together, for instance:

6. Here is what I found by searching for 5 minutes. Google is a great thing.

<https://www.youtube.com/watch?v=9hsPpghm3U4>

<https://slate.com/news-and-politics/2018/09/brett-kavanaugh-lies-senate-testimony-supreme-court.html>

<https://www.motherjones.com/politics/2018/09/five-times-brett-kavanaugh-appears-to-have-lied-to-congress-while-under-oath/>

https://www.washingtonpost.com/politics/2018/09/17/did-brett-kavanaugh-give-false-testimony-under-oath/?utm_term=.5ffa77c14ea9

<http://www.chicagotribune.com/news/opinion/zorn/ct-perspec-zorn-kavanaugh-ford-renalte-crying-20180930-story.html>

<https://www.theatlantic.com/politics/archive/2018/07/how-kavanaughs-last-confirmation-hearing-could-haunt-him/565304/>

<https://www.wsj.com/livecoverage/kavanaugh/card/1538060501>

With this in mind, one might reasonably infer that this topic is perhaps generated not by a common narrative or discussion of the issue, but rather the form of the posts. I therefore group this topic with topic 1 insofar as both topics find some feature of the posts that are common, without giving us greater narrative insight, as topic 2 offered.

As the results package shows as well¹⁷, the top words scores are very low for this topic compared to the others. In addition, the scores for most representative posts drop off precipitously after the first four posts (from scores above 0.91 to below 0.72, where 1 is the maximum score). Looking down the list of representative posts, the posts eventually become so noisy that it is difficult to see any relationship to the previous ones. For instance:

12. Liberals tend to be atheistic, they tend to be more open to changing their worldview, they (sometimes) exhibit signs of mental illness. They also tend to have a higher iq than your average Republican, but it was within the study's margin of error barely. People who have a higher iq are significantly more likely to exhibit signs of mental illness than your average (in terms of iq) person. Take Einstein for example he displayed symptoms of being manic depressive. I happen to be a high functioning autistic with a savant ability in memory (I can almost instantly recall anything I've read). I happen to be left leaning because I believe we should all work together toward making the most amount of people happy that we can.

and

14. When I read this, my eyes misfired and saw "A high boner" at the end.

And you know what? I nodded in sage agreement. It is indeed a high boner.

As mentioned, the scores for these posts (around 0.6 for both) indicate that they are not particularly representative of the topic, but are rather included here to show their divergence from the highly representative posts as posts become noisier and noisier.

¹⁷See Appendix IV, topic 3.

Topic 4

This next topic offers a clearer narrative thread than the last. Some of the most representative words already begin painting a picture: “court”, “supreme”, “judge”, “scotus”, “lifetime” and “appointment”. When taken in conjunction with the posts taken from this topic, a thread of discussion about the process of nominating Brett Kavanaugh (or attempting to block his nomination) to the American Supreme Court becomes apparent.

The first two posts are near copies of each other, and are a call to action to oppose Kavanaugh’s nomination. Here is one:

1. If you are pro-choice, now is the time to reach out to your Senators. In my opinion, your best options is [5 Calls](<https://5calls.org/>) (site takes a minute to load). They give you the relevant numbers based on your location and a script to use. You can adjust the script for yourself to make it more natural, but at least your have a guideline. The site takes a minute to load, so don’t navigate away right away if it doesn’t seem to be working. Calls during east coast business hours generally lead to a conversation with a staff assistant, calls outside of east coast business hours usually lead to voicemail.

Right now, there are multiple scripts for Kavanaugh: [DEMAND JUDICIARY COMMITTEE THOROUGHLY REVIEW ALL KAVANAUGH DOCUMENTS](<https://5calls.org/issue/judiciary-committee-review-kavanaugh-documents>) /// [BLOCK BRETT KAVANAUGH TO PROTECT VOTING RIGHTS](<https://5calls.org/issue/brett-kavanaugh-voting-rights>) /// [OPPOSE SUPREME COURT NOMINEE BRETT KAVANAUGH FOR HIS HOSTILITY TO REPRODUCTIVE RIGHTS](<https://5calls.org/issue/oppose-kavanaugh-reproductive-rights>) /// [PROTECT THE SPECIAL COUNSEL: VOTE NO ON KAVANAUGH](<https://5calls.org/issue/special-counsel-brett-kavanaugh>) (links take a minute to load). You can cut and paste the scripts into word to make your own script.

If you can’t handle the phone, text RESIST to 50409 and Resistbot will help you send a fax. Apparently they’re also hand delivering letters on the Hill right now.

Other posts deal more specifically with the significance of Kavanaugh’s nomination. The following posts discuss the relevance of a hearing into Brett Kavanaugh’s conduct; largely by appealing to the importance of the Supreme Court and the standard to which a potential nominee should be held.

3. His potential employer hasn't accused him of anything. His potential employer is the President of the United States who has been actively promoting his confirmation.

Ten days for a nomination hearing is nothing. One nominee had to endure two years of debate. This guy is going to be employed for life. It's worth finding out how he behaves, drunk or sober.

According to federal statute, the Court normally consists of the Chief Justice of the United States and eight associate justices who are nominated by the President and confirmed by the Senate. Once appointed, justices have lifetime tenure unless they resign, retire, or are removed from office.

5. I think you undersold it by saying "interview for the highest court"

It's one of 9 lifetime appointments that can check both Congress and the POTUS. "Interview for one of the most powerful jobs in the world" would be more appropriate.

8. It's over a SC seat though. This is about whether he gets to be one on nine people deciding law for all of us or if he's merely going to be a federal judge. Denying him a seat is no more an injustice than denying anyone else a seat. There's no standard of evidence needed to not seat him, a senator could vote against him because they didn't like his tie. We don't need to prove he did it, we don't even need to show it's likely it happened. The Supreme Court should be above suspicion.

While a majority of posts in this topic are about the importance of a position on the Supreme Court and the standard to which a nominee should be held, some posts, like the ones below, make a comparison to a standard job hiring process:

4. >It's not a trial. but if you were interviewing for, say, your chief accountant for the firm - and one candidate has unsubstantiated mutterings of financial misbehaviour following him around, and the others are squeaky clean - why would you pick the one who is short of perfect?

Because the one with unsubstantiated mutterings is your preferred candidate.

If the financial misconduct was true, there would be evidence of it.

Taken together though, this is a topic which relatively clearly has a central focus: nomination to the Supreme Court, and the process that these redditors think need be followed to proceed with Brett Kavanaugh's nomination.

Topic 5

This topic is another that suffers from a certain amount of noise. The existence of these topics shouldn't be that surprising: social media often contains short posts, and these might be short on distinctive words that get them grouped into a significant topic. This topic groups a number of short posts that are often informal and vulgar, as evidenced by some of the topic's top words: "shit", "fucking", "fuck" and "bullshit".

While the subject of many of the posts in this topic appear to be distinct, the posts are short and non-specific enough that the vulgar language rises in relative importance. Let us look at some examples, where I've italicized the vulgar or insulting language characteristic of posts in the topic:

1. "Where is his GoFundMe?"

On the Go Fund Me web site?

You did start one for him, didn't you?

Anyone can start one, if you think he needs one, and he hasn't got one, why haven't you done anything about it?

As for *twonks* threatening people, that is never acceptable, but I'm pretty sure security comes with his job.

2. This fear of flying rebuttal may be the dumbest *shit* I've seen yet. Thousands of people have a fear of flying and do it all the time. Is that really all it took to convince you?

3. America is NOT turning into the middle east. *Jesus christ*, get a grip on reality.

Edit: I think it's nice to edit posts for clarification but people who post insane *shit* then delete it because they're going to reddit karma hell (and seriously who cares about that even?) Are so annoying. You might post crazy *shit* but atleast have some conviction. That I can respect.

4. Lol. At the time of watching, my co-workers and I kinda felt bad for him assuming that his dad (and some of his friends) had died.

But nope.

He's just an *asshole*.

These examples are fairly representative of the topic. An observation: only two of the ten most representative posts in this topic came from the TwoXChromosomes subreddit;

the other eight came from Politics. Looking at the top 20, the Politics subreddit supplies a whole 15 posts. This might point to different styles of discourse emanating from each subreddit, which may be indicative of the prevalence of the type of discourse in the threads collected from each of the subreddits.

Topic 6

Another clear narrative thread emerges in topic six. Namely, this topic concerns sexual assault and the attitudes and perceptions of the accused and of alleged victims. The illustrative words for this topic clearly suggest this interpretation, as well: “sexual”, “assault”, “victims”, “women”, “rape” and “abuse”, among others, are clear indicators of the topic of discussion. However, coupling those words with some of the most illustrative posts for the topic offers clear insight into a discussion about how alleged perpetrators and victims of sexual assault and rape are perceived. Notably, some redditors on the Politics subreddit critiqued statements by Kelly Anne Conway (a counsellor to President Trump) about women’s responsibility vis-à-vis assault:

1. If Kelly Ann said women are responsible for their own conduct and what happens to them then, then by her own words she’s not a victim because she is responsible for her alleged assault on her.

7. wow. kellyanne conway was just on cnn a few minutes ago and said women like dr ford are responsible for their own conduct and what happens to them - !! - then went on to admit she’s been the victim of sexual abuse too (while conveniently omitting what *she* did in her particular case, and whether *she* reported the incident to the appropriate authorities, as she continues to criticize dr ford for not doing).

more victim-shaming from a republican woman who claims to be a victim herself.

Meanwhile, some on the TwoXChromosomes sub also expressed disagreement with placing the responsibility on women’s shoulders:

5. This is how my mom thinks. Just because she grew up in a world of rampant sexual assault and harassment that the women "complaining" about it today ought to simply "toughen up." It's sad that the world broke her (and many, many other women).

There is also a thread of discussion, notably in TwoXChromosomes, about dismissals of inappropriate behaviour from young men under the reasoning that “boys will be boys”. The following posts illustrate that discussion:

3. Their age does not excuse the behavior.

Being hormonal does not excuse the behavior.

Being an "awkward teenager" does not excuse the behavior.

High School aged boys are not idiots and should know that holding people down and covering their mouths in an attempt to have sex with someone is wrong on all levels. It's rape, or in this case attempted rape. You can "explore you sexuality" with consent, and this encounter was clearly not consensual.

4. So like, wtf? How can any woman be OK with that? If all boys act that way in HS you should be screaming to separate boys from girls and regulate their behavior. You should be working to protect your daughters and fellow women.

But they know that not every boy acts that way. It's a shitty boys who don't need to fear repercussions or consequences.saying “what boy hasn't done that in high school” is just a way for them to excuse the men and boys they want to protect while abandoning the people they have no use for.

The previous examples serve as illustrations to make clear that this topic is fairly coherent as a broad discussion of the attitudes surrounding sexual assault and the perception of both victims and the accused.

Topic 7

The previous topic was a fairly broad one which included separate yet thematically-linked conversations. This topic finds a much narrower theme of conversation: chiefly about the decision to tell Dr. Ford's story. The topic contains plenty of discussion about this decision, ranging from how Dr. Ford decided to come forward with the story, to how the decision was made to go public with the accusations. As the top words for this topic suggest, a “letter” that Dr. Ford wrote to Dianne “Feinstein” is one top narrative, as is how the “story” “leaked” to the “public”.

Some of the most representative posts for the topic discuss Dr. Ford's indecision with going public, before the letter she wrote to Feinstein was leaked:

1. > ...she grappled with concerns about what going public would mean for her and her family — and what she said was her duty as a citizen to tell the story...

> ...By late August, Ford had decided not to come forward, calculating that doing so would upend her life and probably would not affect Kavanaugh's confirmation...

She was thinking about going public, later decided it probably wasn't worth it, then somebody leaked the story and everything else came with it.

3. *She could of kept it confidential and still brought it to the SJC to be investigated. She could of sent a letter to the FBI and kept it confidential.*

Facts are she didn't, she didn't even ask him in the 1v1 interview they had.

There are 2 people who could of leaked the letter blowing up her confidentiality. They are Feinstein, and Ford herself.

In fact there is an [investigation](<https://www.politico.com/story/2018/09/30/cotton-feinstein-ford-leaked-letter-854019>) on Feinstein's office to determine if she leaked it.

Edit: Also that article cites a Bar investigation on Ford's lawyers. They were recommended by Feinstein.

Some redditors show that they are skeptical about the timing of the story going public:

4. So she comes forward with this in July. We go through the entire nomination hearings, where thousands of documents are presented and questioned, and it's not mentioned or brought up once? It's not even hinted at until after the hearings. Then it's bright forward when it's clear he's going to be confirmed.

That's not fishy. Ok. Keep those blinders on.

Several posts across both subreddits discuss who knew of the story and who might have leaked it. The first quote below is a redditor clarifying who received the letter from Ford, while the subsequent ones discuss the story going public.

6. Feinstein received the original letter, not Pelosi.

7. The senator could have sent it to the FBI and not released to the public.

Her staffers leaked it regardless at the most political opportune time. Smells fishy.

8. If you watched the hearing you'd know that Feinstein respected Dr. Ford's wish for privacy, confident her office didn't leak the info. It likely came from a friend of Dr. Ford (she admitted that she told multiple friends if contacting Feinstein's office), at which point, Dr. Ford & Feinstein went public.

10. She reported this months ago, and it's taken that long for the allegations to become public. She had actually given up, and the allegations ended up leaking and forcing her to come public:

>She contacted The Post through a tip line in early July, when it had become clear that Kavanaugh was on the shortlist of possible nominees to replace retiring justice Anthony M. Kennedy but before Trump announced his name publicly. A registered Democrat who has made small contributions to political organizations, she contacted her congresswoman, Democrat Anna G. Eshoo, around the same time. In late July, she sent a letter via Eshoo's office to Sen. Dianne Feinstein of California, the ranking Democrat on the Judiciary Committee.

>In the letter, which was read to The Post, Ford described the incident and said she expected her story to be kept confidential. She signed the letter as Christine Blasey, the name she uses professionally.

>Though Ford had contacted The Post, for weeks, she declined to speak on the record as she grappled with concerns about what going public would mean for her and her family — and what she said was her duty as a citizen to tell the story.

>She engaged Debra Katz, a Washington lawyer known for her work on sexual harassment cases. On the advice of Katz, who believed Ford would be attacked as a liar if she came forward, Ford took a polygraph test administered by a former FBI agent in early August. The results, which Katz provided to The Post, concluded that Ford was being truthful when she said a statement summarizing her allegations was accurate.

>By late August, Ford had decided not to come forward, calculating that doing so would upend her life and probably would not affect Kavanaugh's confirmation. "Why suffer through the annihilation if it's not going to matter?" she said.

>Her story leaked anyway.

These posts altogether paint a fairly coherent thematic topic: one that discusses the narrative surrounding Dr. Ford's story going public; from Dr. Ford's own hesitations to possible sources that may have leaked it.

Topic 8

The eighth topic is another, like topic 5 with its abundance of cursing, that is notable because it brings together a type of dialogue rather than discussions about a narrative present in the Brett Kavanaugh nomination controversy. Its top words (including “people”, “person”, “make” and “good”) don’t immediately paint a clear picture. However, in reading through the posts, a common thread emerges: these are posts of agreement or disagreement, of opinions and argumentation. These range from civil and genuine-seeming attempts at argumentation and (dis)agreement all the way to sarcastic and hostile.

This first post explicitly states disagreement with another redditor, while attempting to remain friendly with them:

1. [...] Plus, I disagree with your main point that the public is smarter than the individual.

I think it is the exact opposite currently.

[...] That is why all faith has been lost completely in the system. I am glad that people like you exist and are optimistic about it. But to me, it is broken.

The second offers clarification that they might already be in agreement with a previous poster:

2. ? What. Yeah. I was trying to point out that it’s a good thing. Sorry if that wasn’t clear.

I’m about as liberal as they come, but try to ride the middle line in debates so I can at least understand the other side’s points. I’m fairly entertained I got mistaken for a conservative.

Other posts appear more argumentative and sarcastic. While sarcasm can be hard to detect through a text-only medium, this redditor states they appreciate being accepted, when the rest of their statement indicates their position wasn’t accepted at all.

5. So now I am making claim and mine isn’t believed but hers is? Excellent logic. I appreciate the acceptance

Or the following post, which suggests its sarcastic intents by implying its target is a child (emphasis added for illustrative purposes):

6. What a thoughtful and well reasoned reply. You really showed me.

Only problem is that there were actually pro Clinton sentiments in there, but hey, reading is hard and who really has the time, amirite?

Sorry about that. *I’ll write it in crayon and use smaller words and pictures next time.*

In all, this topic is interesting for the insight it offers into discourse. Based on the occurrence of language indicating (dis)agreement, it offers a look at specific instances of agreement and disagreement throughout the text as well as an idea of the spirit of these across the corpus.

Topic 9

This topic is another that has a thematic link between the individual posts. That theme is already present in some of its key words, such as “constitution”, “law”, “states” and “history”. This topic is one that largely discusses the American constitution and the legal history of America. However, while it contains many comments about law throughout the topic, these seem diverse in nature.

For instance, there is some discussion about the constitution’s role in the American slave trade:

1. You know that the slavery in the Constitution was a huge compromise that allowed this country to form right? The founds absolutely did not want slavery and it was necessary to get enough votes to declare independence from England

As well as this post, written in response to the previous one:

4. >You know that the slavery in the Constitution was a huge compromise that allowed this country to form right?

Yes.

> The founds absolutely did not want slavery and it was necessary to get enough votes to declare independence from England

So you’re say that the founding fathers, many of whom owned slaves, who *compromised* on the issue, "absolutely did not want slavery?" That’s a revision of history I haven’t heard before.

And this excerpt from a lengthy post which describes the American founding fathers’ takes on slavery:

6. This is not revision at all my friend. Here is Abraham Lincoln on the founding fathers and their take on slavery:

The argument of "Necessity" was the only argument they ever admitted in favor of slavery; and so far, and so far only as it carried them, did they ever go. They found the institution existing

among us, which they could not help; and they cast blame upon the British King for having permitted its introduction. BEFORE the constitution, they prohibited its introduction [...]

The discussion isn't limited to American slavery, however. Discussions also focus on the law, the constitution, and what role those have in a society. The following posts illustrate the tension in viewing the law as an infallible ideal, versus conceptualizing it as a tool that serves society.

8. I have never really understood their ilk. Yes, I consider the constitution a sacred document, but not an infallible one; Constitutionists seem to see it as actually being /divine/, unquestionable and untouchable. Culture, science, and time must stagnate in service of leaving it untouched. But the Founding Fathers were not perfect, they were as human as anyone, and it took a few iterations to get the documents we have. They screwed up more than a few times in the drafting period.

9. The law is a tool. It does not exist on its own, it exists to serve a purpose. An enlightened civilized society would be able to bend the law when it no longer served that purpose. Unfortunately, conservatives, especially American conservatives, see law as an end in itself.

Still other discussions speak to specific aspects of the legal system, such as gun ownership and the second amendment (first post below), or the behaviour a Supreme Court judge is expected to display (second):

7. Democrat judges voted to ban sawed of shotguns for not being useful in a militia. This passed.

Democrats judges voted guns like the AR-15 were exempt from civilian ownership since they were intended for use in a militia, and since we dont have a militia, we dont need AR-15s. This did not pass.

But point being, these two beliefs cannot coexist. The 2nd amendment is clearly defined in it's own words and any interpretation by a judge is a bastardization to suit themselves. Problems like this should not arise.

10. >Article III

>Section 1.

>The judicial power of the United States, shall be vested in one Supreme Court, and in such inferior courts as the Congress may from time to time ordain and establish. **The judges,

both of the supreme and inferior courts, shall hold their offices during good behaviour**, and shall, at stated times, receive for their services, a compensation, which shall not be diminished during their continuance in office. [...]

This topic does highlight one weakness of topic modeling as an approach to exploring a corpus: while it is interesting that this topic presents and gathers all of the posts that deal with the themes of the American Constitution and American legal history, it does not automatically piece the discussions together within the context of the Kavanaugh controversy. Indeed, the approach is good at indicating that this theme is present in the discourse, but it requires further digging to contextualize the discussion. On the other hand, the very knowledge that so much discussion involves the constitution and the American legal system, on top of the discussion having shifted far from the specifics of the Kavanaugh controversy, is interesting in itself.

Topic 10

Topic 10 is another clear, thematic topic. However, in contrast to the previous topic, this one is more restrained in its range. It contains discussion about polygraphs and their efficiency and accuracy at distinguishing truths from lies. The tendency towards this topic is visible through some of its most representative words: “polygraph”, “test”, “lie”, “truth”, and “science”.

There is some discussion about how polygraph tests are used for security clearance:

1. TS SAR SCI almost always requires poly, TS SCI w/ poly is also an established level of clearance for working within the IC. Non IC work, like DOD - 24th AF and TS SAR SCI work usually doesn't require a poly, just the chance you may have to provide one.

Also, Polys come in two flavors, full scope, or CI. One of them is “any question is allowable” the other is 6 questions which surmount to “are you a spy”

20. Have had two types of security clearances for projects, did not need a polygraph.

As well as some talk about how polygraphs are used:

3. They don't "go off", there isn't like a buzzer that says "HA THAT DUDE IS LYING". It is a guy trained to read polygraph charts looking at graphs showing your various autonomic

functions that you can't see.

A larger part of the discussion, however, is centered around the reliability of the polygraph:

2. Actually it could prove they might have high blood pressure..... I mean that's what the polygraph was invented as a way to test for hypertension not a lie. how it became used as a lie detector is pretty much thanks to Hollywood.

5. The 2 Question Polygraph that literally asked her if she was lying or nah? That one? That's good enough to destroy a man's life? You do know polygraph tests don't actually determine if you're telling the truth or not, right? They read your physiological responses, such as pulse, heart rate, whatever. There Is No evidence that proves that any one physiological response can show deception. This is why polygraph tests are NOT admissible in court. They are not accurate and they mean less than nothing.

6.SWIM was trained to misinterpret the questions to trigger a "truthful" response.

Sometimes this can be bending the english language to answer a different question that the Test Administrator actually asked, and sometimes its by providing false positives on his read questions like your name, age, etc.

Also many prescription drugs throw off heart-rate, blood pressure and respirations. Allowing you to just straight up lie your ass off.

Theres a reason Lie Detectors arent admissible in court.

This topic thus suggests that there is a significant discussion about polygraphs happening within the corpus. Furthermore, it is interesting to observe that 19 of the 20 most representative posts for this topic come from the TwoXChromosomes subreddit, which points to that topic being of greater interest or more salient to that community (at least in the sample of discussion selected for this project).

Topic 11

Topic 11 centres around a discussion of partisanship in Supreme Court nominations, discussing the motivations of Democrats and Republicans to have ideologically-friendly justices in the Supreme Court. While the topic's top words don't offer direct clues to

this interpretation, we do see words that suggest partisan political themes (“democrats”, “republicans”, “party”, “political”).

This first redditor describes their understanding of this partisanship:

1. They will, and they have. I don't think that's disputed. Both sides want to confirm Justice's that lean more towards their values.

This can't really be disputed when party lines say they will vote no on any justice a President picks before the nominations are announced. Democrats did that with Neil and Brett (before they were announced) and Republicans did that (in essence) by saying we won't hold a vote (effectively voting no while holding majority) for Merrick (before Obama announced him). So I don't think it is a "should we pack the courts", it is a "when we pack the courts".

At least, that's my take on it.

This second post specifically uses the assumption that the political party in control of the American Senate controls the nomination of a Supreme Court justice:

2. [...] It wasn't their seat. It required a majority Senate vote and Republicans controlled the Senate. Sure, some may have votes across the aisle but we don't know that and it seems in this political environment it is unlikely. Regardless, you can't say Democrats, by right, should have had that seat when the opposition controlled the majority required to confirm.

The next post also speaks to the notion that parties strategically try to nominate Supreme Court justices based on their level of control of the Senate and timing of the nomination:

3. You have to look at the situation. Merrick was a moderate left acceptable over some further left potentials. However, the GOP realized they didn't need to accept moderate left because they could delay until a moderate right was nominated.

Democrats are doing the same thing, if midterms were over they would accept a moderate right nominee as the best of the "worst" (meaning any justice that's right leaning). Now they see a chance to delay until midterms and the potential opportunity to force a moderate left nominee.

I dunno man, I see this as two methods to accomplish the same goal. It's about stifling the other side.

Others even try to find ways to eliminate the partiality:

10. It's virtually impossible for political partisanship to not exist in the SCOTUS. Nominees are selected by a president that is affiliated with a political party and confirmed by a Congress that is affiliated with political parties. Anyone not partisan enough will never get through this process. At the very least they will need to "lean" a certain way.

Perhaps a random name of a Federal judge should be chosen by lottery and then vetted by Congress?

This topic is fairly clear in its theme. Even some of the noisier posts that begin to stray from the topic still show elements of discussion about how parties act in the nomination process, as exemplified by the two following posts. The first alludes to the Republican party (GOP, the "Grand Old Party")) being willing partners in letting the people's representation express itself freely, while the second mocks the idea of an impartial Republican party.

14. There is no deadline. If the voters choose different representation during this process and that representation won't confirm him, then I believe the GOP is on record saying we should wait for that vote.

15. Not the Republicans!! They would never delay an appointment. They certainly wouldn't delay it indefinitely!

Topic 11 is thus of note because it is another topic that is underpinned by a common theme in discussions, and that theme is fairly narrow and consistent throughout its individual posts.

Topic 12

This topic is one that specifically enters the domain of gender, which was one of this project's main concerns. Specifically, this topic exhibits a discussion and debate of the privileges afforded to different intersectionalities of gender and race, particularly as it pertains to sexual assault. This appears in some of the topic's words: "women", "men", "sexual", "assault", "white", and "believed". Perhaps unsurprisingly, much of the discussion (15 of the top 20 posts) in this topic are pulled from the TwoXChromosomes subreddit, which self-identifies as being intended for women's perspectives.

One of the contentious issues at play in this topic is the perception that white men are not held accountable for their actions, or conversely, that white men are particularly vulnerable to accusations.

These redditors specifically discuss white men’s perceived accountability:

2. Privileged white men are rarely held accountable.... except men receive 60% higher sentences than women [for the same crimes](https://huffpost.com/us/entry/us_1874742). How often do you see a male teacher get off for having sex with an underage student? How often does it happen with women? Not to mention polygraphs are worthless and she freaking volunteered for it.

The previous poster makes reference to the severity of punishment men receive for crimes compared to women, implying that men are more likely to be harshly punished, particularly in the case of “having sex with an underage” person. In contrast, this following post quotes and responds to another redditor’s post about men’s accountability:

5. > privileged white men are rarely held accountable for their actions

I’m sorry, this misandry/racism, which is against rule #1.¹⁸ I can understand if you believe this to be true, but you can’t just throw that out and expect it to be okay.

This redditor also pushes back against the notion that men are “rarely held accountable”, instead focusing on women’s lack of accountability:

18. ‘Rarely held accountable’ I believe you are thinking of the cis white female. Who can rape a child or as the news calls it "Sex romp" and get like 2 years probation. You can’t call white men privileged when they can be raped and their rapist not get prison time.

There is also backlash against the idea that women are not typically believed when it comes to sexual assault:

10. That is stupid. Woman are not believed? All around the world, woman can just come ans point on any guy saying..he raped me, and that guy is done! Lately, in USA, they can just come and say"he touched my knee 35 yrs ago"

¹⁸Rule number 1 of TwoXChromosomes, the subreddit from which this post is taken, reads: “No hatred, bigotry, assholery, misogyny, misandry, transphobia, homophobia, racism or otherwise disrespectful commentary. Please follow reddiquette.”

Many of the other posts in the topic are shorter and make reference to this broad topic without being so easily interpretable. For instance, the first post below alludes to a woman's credibility, while the second directly references white privileged men:

15. People on r/news are already assuming she's lying and talking about how she was drinking and such...

14. I'm confused, are you talking about the 'white privileged men' part?

The following post is included in this topic, even though it is outside of the bounds of the discussion about white men. Here, discussion ventures to the case of Clarence Thomas (who was in the process of being nominated to the Supreme Court in 1991), accused of sexual harassment by Anita Hill, a black woman.

4. Ding ding ding. If the Thomas complainant had been a white woman, he would never have been confirmed. I remember thinking that in 1991, as well.

This topic is somewhat sprawling in its area of focus, but clearly brings together posts that discuss issues of gender, race and privilege, and how these are related to sexual assault.

Topic 13

Topic 13 deals with a fairly constrained topic: behaviours displayed during the hearing, and the perceptions of those behaviours. Again, some of the topic's key words illustrate this: "questions", "hearing", "kavanaugh", "answer", "lied" and "anger". Many of the posts in this topic specifically deal with Brett Kavanaugh's comportment during the hearing.

The following posts all discuss Kavanaugh's behaviour in refusing to answer questions during the hearing, as well as what was perceived as a lack of grace under pressure.

1. To be on the SCOTUS, you're supposed to be rational & refusing to answer questions, trying to lecture people, acting like going to a good college means he couldn't have done it, is crazy. He refused to answer about half the questions, even asked a Senator one of the questions that he refused to answer. I don't know any job that would be okay with it. And he showed he's completely partisan and conspiratorial with his little rant.

3. It isn't a binary response. He could express his anger in a measured manner. He could have expressed his empathy for Dr. Ford and still questioned her charges. He could have NOT

LIED. He could have asked for an investigation and claimed it is revenge by the Clintons. Seriously? The Clintons again.

6. We're not discussing whether or not he has the right to be pissed off here. That remains to be seen. We're discussing Kavanaugh's inability to conduct himself with grace under the slightest bit of pressure.

Seriously, how did Hillary stand up to decades of this, but Kavanaugh can't make it through a single week of scrutiny without proving he's bitch who can't handle the pressure of the national spotlight?

11. Also consider how Kavanaugh was responding to questions the first 30 minutes. It was a disaster.

This topic is interesting insofar as it singles out Brett Kavanaugh's behaviour during the hearings, but also interesting insofar as the discussion here is not explicitly gendered. Gendered and racialized interpretations of the behaviour also emerged in a later topic - however, it is of note that these two sets of documents were grouped separately.

Topic 14

Building on the previous topic, this topic also addresses the Kavanaugh hearings. More specifically, it discusses whether Kavanaugh was truthful about a relatively narrow subset of topics. Some of the top words in this topic are quite descriptive: "drinking", "beer", "yale", "drunk", "triangle", and looking slightly outside of the top 10: "lying" and "lies". Indeed, many of the posts centre around Brett Kavanaugh's testimony as to the definition of the "Devil's Triangle" (which he defined as a drinking game), the work he put in to be admitted to Yale, his father's calendar and "boofing". The dominant theme in this topic questions the veracity of Kavanaugh's claims from the hearing, as evidenced by the posts below:

1. Brett Kavanaugh said the "Devil's Triangle" was a drinking game and acted like his dad was dead the way he talked about how he keeps a calendar just like him.

Kavanaugh's father is alive.

3. >Whatever your opinion on of Kavanaugh, a little intellectual honesty would be nice.

How about being intellectually honest about Brett Kavanaugh's obvious lies about:

- "The Devil's Triangle" being a drinking game played w/ quarters
- That he "busted his butt" to get into Yale b/c he didn't "know anyone" w/out mentioning that his grandpa was a Yale alum.

10. Most folks who watched his performance know what I'm referring to.

- "The Devil's Triangle" is a drinking game played w/ quarters.
- He "busted his butt" to get into Yale and "didn't know anyone" w/out mentioning that his Grandpa was an alum.
- "Boofing" is farting.
- In response to a question about whether he has ever been black-out drunk he said "he went to Yale."

12. "Didn't know anyone at yale" grandpa alum

"Devil's triangle is a drinking game" yeah ok.

"Boofing is farting" even what damn near anyone from the seventies and eighties knows notwithstanding, it doesn't even work in the context of the sentence it was used in, "have you boofed yet?" Asking someone if they have ever farted yet? Are you kidding me? And these are just the weak little lies that shouldn't even mean anything. What else?

This topic is rather consistent throughout, regularly addressing the same issues of Kavanaugh's testimony and calling his entire testimony into question based on what many considered to be easily disprovable statements.

Topic 15

Topic 15 is one of the noisier ones where a consistent theme or use of language does not evidently shine through. The scores for representative words and posts are low across the board, which provides further evidence for its lack of cohesion. At most, the topic does seem to group together posts that talk about extreme viewpoints, and communities that house those extreme viewpoints on the web. For instance, the following posts discuss extreme viewpoints and censorship thereof:

17. Obviously you ignore the right extremists.

They're always going to exist unfortunately. Going to take a few generations to weed out.

Doesn't justify silencing the entire human race in the process. That's an OVERREACTION. And if you want to go down the route of hate speech inciting violence, which I'd agree with, this is the downfall of global connectivity and clashing of cultures. Our toughest battle is to come together over the coming years, diffuse as many wars as possible, and love each other. Sounds cheesy but I feel it's our best bet. Don't drive more separation through group identity and compelled speech.

7. Silence the entire human race??? A bunch of subs spreading both "ironic" and unironic white supremacism got a warning sign telling you that there are edgelords inside before you enter! Even if they did get banned, calling that "silencing the entire human race" is a massive bloody overstatement.

Like it or not, shunning people who call for genocide or reinstatement of apartheid *works*. Given how many supremacist movements work using slimy and manipulative tactics, actively stopping them from indoctrinating others is absolutely morally and ethically justifiable. Calls for violence, genocide, or attempts at dehumanisation of ethnic, racial, or religious groups need to be combatted. Divisions aren't overcome by saying that the KKK are just fellow citizens and need to be heard out and respected, because they will not only radicalise the group they consider their own base, they'll send their victims towards radicals as well. Identity exists, and always will exist. People will identify with their race, their gender, their sexual orientation, their favourite brands, and their preferred pizza toppings, what needs to be prevented is supremacism and the sort of relentlessly mean spirited "ironic" edgy behaviour that the KKK themselves espouse as a great tool for indoctrination and normalisation of their views.

4. The subs weren't even banned though, all you get is a warning about their contents! These are communities who regularly demonstrate their "humour" to be blatant racism, sexism, or some form of calling for the deaths of others with no punchline to go with it, and they're not even being censored for it!

Again, driving stuff "underground" doesn't work the way you think it does. Look at the relationship between Reddit and Voat, and what happened with 4chan and /pol/. The parts of Reddit that did get censored are now there, festering in a pool of completely unabashed white supremacy and neonaziism, but it's tiny and doesn't have a fraction of the reach Reddit has. /pol/ started out as a containment board for political "discussion" on 4chan and ended up becoming the largest board on the site, massively influencing the hateful, reactionary, and edgily inflammatory politics that is now constantly present on almost every board. There's actually a study about this that I'll reply to you with when I find it, but right now I'll just

say that containment and just-leaving-them-alone doesn't work like you seem to think it does, and actively fosters indoctrination and radicalisation of the people on the platform they share. Keep in mind, these are not just subs which complain about taxes, or economic policy, or immigration policy, they're actively engaged in ("ironically") promoting violence against religious and ethnic groups. The line seems to be "no encouraging or glorifying violence or dehumanisation, especially against demographic groups", which is why /r/fullcommunism got slapped with a quarantine as well. If simply quarantining them is "the pendulum swinging too far the other way" to you, then I doubt we can ever agree.

Some of the posts above do make mention of race and gender, which are among the topic's most representative words. There are also some much shorter posts that make mention of the concepts. For instance, this post makes mention of "white"-ness, even if the post itself is not specifically about race:

6. > A bloated white carcass starts scream-crying during his job interview

Oh, that is beauty! This is why I come to Reddit!

And this post about Serena Williams also mentions race and gender, although it is unclear that the discussion is linked to extremism and censorship, as in the first examples.

9. > But let's not defend Serena. She violated the code several times over, then tried to hide behind race and gender cards. Such acts do a disservice to the movement and give feminism a bad name.

You're not wrong. However...

[John McEnroe defends Serena Williams: 'I have said far worse'](https://www.tennisworldusa.org/tennis/news/Serena_Williams/60311/john-mcenroe-defends-serena-williams-i-have-said-far-worse/)

While race and gender seem to be constant through the previous examples, there is also a substantial amount of noise in this topic. The following posts appear to have no solid ties to the ones we discussed previously.

11. Shoulda coulda woulda stfu

12, 13, 14 (all identical). https://www.reddit.com/r/politics/comments/9k5arb/comment/e6womgl?st=JMP37BP8&sh=2738d19d

15. On Monday it will be released that she had ECT, electroconvulsive therapy.

[...]

If you look at a high resolution image of her, you can see the scars between her eyebrows and above the eyebrows where the electrodes scarred her skin. Instead of the more temporal placement of the electrodes, in her case they were place above her eyebrows.

The above gives a sampling of the kind of diversity of posts seen in this topic. Even though there is some evidence of the algorithm placing posts in this topic that share common words, this particular topic does not present as cohesive a thematic or syntactic underpinning as many of the others explored thus far.

Topic 16

This topic presents a coherent narrative theme: it discusses the burden of proof required to render a judgment of guilt, and the presumption of innocence until that point. There is evidence of this clearly in the topic's top words: "guilty", "innocent", "proven", "evidence" and so on. Examining some of the topic's most representative posts reveals this narrative thread which is on display consistently throughout the topic.

The first, most representative post is mostly centred around a description of the burden of proof and how one of America's foremost civil liberties is the presumption of innocence until the required burden of proof has been met.

1. Even IF polygraphs were reliable, in America defendants are innocent until proven guilty beyond a reasonable doubt in a court of law, one of our most important civil liberties. Thus the burden of proof always rests on the plaintiff, not the defendant. Unless she can prove his guilt, he has no need to prove his innocence. If there are no other witnesses or material evidence, then there is more than a reasonable doubt about her claim.

This second post frames the discussion in light of differences between criminal and civil proceedings, describing the different assumptions in required proof in each venue:

2. [...] This isn't really accurate the way you frame it. In a criminal trial, the *government* has to prove beyond a reasonable doubt that the defendant is guilty in order to secure a conviction. But in a civil lawsuit, the accuser (the plaintiff) only has to prove that it's "more likely than not" that a harm has occurred (a "preponderance of the evidence").

This is why O.J. Simpson was acquitted (found not guilty) in his criminal trial but found liable in his civil trial for wrongful death. The state couldn't prove its case beyond a reasonable doubt but the victims' families proved their case by a preponderance of the evidence.

This debate over the level of proof needed to find someone guilty continues through the next few posts in the topic. The first two repeat the argument that the burden of proof falls with the accuser. The third suggests that the Kavanaugh hearings are not criminal or civil proceedings at all, implying that the case does not need to be held to the same standards.

3. I believe her but The burden of proof is on the plaintiff, not the defendant. In our court system, it is the duty of the plaintiff to prove that the accused did it, not the other way around.

4. Because it isnt his duty to prove his innocence, it's the prosecutors duty to prove, beyond a shadow of a doubt, that he's guilty. Just because the allegation is rape doesnt change how our courts function and in this country until you are proven or plead guilty you are presumed innocent in the eyes of the court.

5. Have you noticed that this isn't a criminal proceeding under which one must rule beyond a reasonable doubt?

Have you similarly noticed that this isn't even a civil proceeding under which "a preponderance of the evidence" guides the ruling?

This is a job interview. Kavanaugh is not under criminal proceedings.

The entire thread of argumentation here seems to be underpinned by a fear that baseless accusations could potentially have negative effects on an individual. This next post illustrates the fear, using an example of baseless accusation to respond to the notion that the burden of proof might be lowered in the context of a "job interview".

9. So if I accuse you of embezzlement in your job, but there is no evidence that you did besides my word, you should still be denied a promotion on the grounds that you could have embezzled?

Others try to establish some nuance between innocence and guilt, highlighting that the presumption of innocence does not exclude the possibly of guilt:

12. [...] I pointed out that he wasn't found guilty or innocent because no investigation has taken place.

The distinction being that although we should presume someone innocent until proven guilty, we can't disqualify the possibility of guilt without due process.

The debate finds no resolution within the topic, but the issue generates significant discussion within the corpus. Many are clearly concerned that we should only punish those individuals whose crimes are documented by sufficient evidence. Others are weary of granting an important position to someone with significant allegations made against them.

Debate notwithstanding, this topic presents a clear and coherent narrative thread within the discussion around Brett Kavanaugh's nomination to the Supreme Court.

Topic 17

This topic deals specifically with the theme of accusations of rape and the impact of such accusations. The topic majoritarially deals with the notion that an innocent person might be accused of the crime, and what the implications are if this is the case. This is reflected by some of the most representative words for the topic: "accused", "rape", "life", "ruined", and "innocent". Some of the most illustrative posts illustrate that the topic deals largely with Brett Kavanaugh's reactions to what some redditors perceive as "baseless" accusations of a crime that Kavanaugh "didn't do". While there is some implicit and explicit discussion of women's role in this in some of the later posts, a common theme of concern over the impact on a man's career and livelihood is noteworthy throughout the collection of posts below.

1. Imagine being accused of something you didn't do, something horrific. Whether or not you believe Ford, its foolish to assume that this isn't appropriate behavior for a man who was accused of rape.
2. Maybe he's gettin' angry because his career is being destroyed on baseless accusations, just sayin' #WalkAway
8. Of course he's angry, they accused him of organizing gang rape with no evidence. I'd like to see how everyone else would react if accused of the same thing

Other redditors are not willing to proclaim his innocence, but explain Kavanaugh's angered reactions by calling on the severity of the accusations.

4. He's being accused of not only attempted rape by her, but having multiple gang rape train drug parties. I'm not saying he's innocent because I don't know if he is, but if I was accused of that I would be pissed. His anger could absolutely be justified.

19. Men have had lives ruined from false allegations. If he is innocent then I don't blame him for being upset.

This discussion also starts to take on an implicit gendered tone against women, as one redditor describes the fallout of the accusations as something significant that he will have to deal with, while overlooking any costs and harms a woman may face for going public with the allegations:

6. well that...and the death threats he and his family are also receiving plus the absolute trashing of his reputation and end of his reputation and career, whilst she gets go walk right back into hers

But gendered concerns aren't just subtle and implicit, they are also on full display in comments like this one, which imply that the accuser's gender gives her a free "pass".

5. Well when you're a lying fraud trying to ruin a man's life and career just to win a political maneuver, you don't get to be angry when everybody sees right through it and denys you your pussy pass.

The present topic thus adds to a collection of narrative topics; this one concerned with accusations of rape and their fallout, largely focusing on how a man might react to such allegations and the perceived negative impacts they might inflict on his career and livelihood.

Topic 18

Topic 18 is another cohesive thematic topic. It revolves around the theme of an investigation into the Kavanaugh allegations. This is clearly illustrated by some of the topic's key words: "investigation", "fbi", "investigate" and "kavanaugh". Some of the most

representative posts in this topic suggest the discussion on investigations is largely centered around the role of political partisanship in said investigation, although there is also some discussion on the role an investigation should play in the allegations.

To begin, there is some discussion into the role of partisanship in the investigation. This first post states that Democrats aren't asking for a criminal investigation, but rather for a deepened background check.

1. No, the request for an FBI investigation was and is requested because they do federal background checks, for federal offices.

No Democrat senator was asking for the FBI to do a criminal investigation. They were asking that the background check be re-opened since new information has come to light... [...]

The next post states that the Democrats may gain an advantage because of the time it would take to perform an investigation.

3. The deadline for him to be confirmed is soon, and an investigation is going to obviously last longer than that, allowing for the Democrats to gain an advantage from the Midterms.

The following suggests that Republicans want a quick resolution in order to secure a confirmation of Kavanaugh's appointment to the Supreme Court.

11. That's why there should have been an investigation rather than trying to hamfist his confirmation. The Republicans are the ones to blame. They gave an ultimatum to have her testify right away. They were banking on her being unprepared.

Another point of discussion is whom exactly is calling for an investigation, and why. The following posts discuss the issue.

4. Both sides? One of those sides has repeatedly asked for a legitimate investigation and the other has tried to sweep this under the rug.

7. > One of those sides has repeatedly asked for a legitimate investigation and the other has tried to sweep this under the rug.

They weren't asking for an investigation when they got the information about the rape, they didn't ask for an investigation when they talked with Kavanaugh other times. They had more than ample enough time for an investigation. They're only asking for an investigation *now.* The Republicans are (rightfully) saying this is political opportunism to extend the nomination proceedings.

9. The Democrats didn't call for the hearing. Republicans did.

Ford came forward prior to Kavanaugh's nomination when she heard he was on the list to her congresswoman. Ford wanted her name to be private, so Democrats sent her story to the FBI to be investigated. Republicans said an FBI investigation was not needed and we ended up with the bullshit hearing on Thursday where they refused to talk to her.

Now, because they don't have enough votes to approve him, they are allowing an FBI investigation... But it must be *one week or less*.

So that's why. There's no statute of limitations on rape in Maryland but they're trying to deny her a real investigation and trial. Idk why that's allowed.

Another related thread of discussion within the topic looks at the role of an investigation into the allegations. This post describes whose duty it is to perform an investigation:

2. It's not the Senate's job to perform a full investigation of the crime. Nor is it the FBI's. It is the responsibility of the state or local authorities in Maryland after a complaint was filed. But since a complaint was never filed...yes, this was all political theater by all sides.

Relatedly, this next post discusses the redditor's personal opinion on the credibility of the allegations, but also chimes in on jurisdictional responsibilities, which is also mentioned in a few posts within the topic.

14. [...] There should absolutely be further investigation. All allegations should be examined. Sexual assault is no joke. Neither is false accusation of such crimes.

I find, in this situation, the accusers are not credible specifically because of the FBI scrutiny he has already undergone and the lack of corroboration to any of the claims by any of the purported witnesses.

Additionally, the accusations are not within federal jurisdiction, they are state crimes. They should be investigated by the Maryland State Bureau of Investigation. The FBI has already failed if the accusations are true...six times.

Through the above posts, there is a clear thread of discussion about the role of an investigation in the Brett Kavanaugh nomination controversy; and thus another topic that informs us about thematic underpinnings throughout the corpus.

Topic 19

This topic is centred around discussion of the quality of evidence present in the case, and notably, on the value of testimony as evidence. There are hints of the topic's theme through some of its most representative words, such as "evidence", "testimony" and "credible".

First, the topic groups together some discussion about what evidence there is in the case. The following example has one redditor directly asking another to list the evidence:

3. [...] Can you tell me what evidence there actually is that would stick on a actual court of law? If not, then how is it a credible claim? If it was credible t would stand on any court of law. If no such claim can be made that would pass a court of law, why are we destroying a man's career over accusations? This is not how the world should work. [...]

There are a significant number of posts that dig into testimony as the basis of evidence. Some, like the redditor below, supports testimony as credible, suggesting that there is no personal motive for the accuser to lie.

5. [...] It's a cold stone fact that people don't just lie under oath for no reason, backed up by evidence. It's a cold stone fact that there's no plausible motive for her to lie. You need to stop confusing your opinion with fact. You don't like the evidence which says it's unlikely she's lying, so you try to act like it doesn't exist. [...]

However, a larger number of posts in this topic question the value of testimony at all, including some examples of this line of discussion below.

7. [...] Witness testimony or the accusers testimony on its own is usually worthless without any corroborating evidence.

Ie if you accuse someone of a crime, that accusation means jack unless you can also prove they were in the area, you have physical evidence of the crime in relation to the accused, etc.

Its not a issue of strength unless multiple people witnessed the crime, but even then if someone has a alibi with multiple other witnesses it's still useless.

So basically, yes, witness testimony is general useless in the absence of other evidence.

8. I know that it's happened in the past and while I'm not disputing Dr. Ford there's no evidence to corroborate her story. And while I sympathize with her, it's a dangerous precedent to take people at their word and only their word in court.

There is also some frustration against the idea that testimony isn't credible evidence, particularly when irrefutable evidence might be impossible to provide:

4. So in other words, rape basically isn't a crime because it's impossible to provide irrefutable evidence of it in nearly all circumstances.

In all, this topic provides another cohesive thematic grouping in the corpus, focusing on evidence and the quality and credibility of testimony as evidence.

Topic 20

The final topic presents perceptions of emotional displays, particularly along gendered and racialized lines. There is here a considerable amount of discussion on how certain behaviours are perceived differently depending on the individual performing them. This is reflected in some words describing emotional states in the topic's key words: "hysterical", "emotional", and just outside the top ten, "angry". There are also words indicating race and gender, such as "man", "white", "women", and "black". Throughout the topic, there is a constant interplay between emotional states and the concepts of race and gender. For instance, some redditors discuss the difference in perception afforded to members of different racialized and gendered groups:

1. > Women: hysterical
 - > Black Men: overly aggressive
 - > Black Women: uppity, loud
 - > White males: great testimony with passion
- edit: adjective

More often, though, discussion is centred around gender and how individuals might interpret similar behaviours from individuals with different genders:

4. Dominant males would describe Kavanaugh's erratic behavior as a sign of strength and confidence. If the same behavior were to be exhibited by Dr. Christine Blasey Ford she would be deemed hysterical and a drama queen, even crazy.

Even the privilege to act like an asshole on national TV is gendered.

8. Again it is a double standard. Hysterical man - strong and defending his honor. Hysterical women - weak and probably trying get sympathy fro something she did to herself.

That is a double standard and it just sucks to be called on it.

There is also some evidence of individuals defending Kavanaugh's behaviour during the hearing, citing that had he behaved differently, that also would have been criticized:

15. Kavanaugh would have been called "cold" and "definitely guilty" if he hadn't shown any emotion.

And here, an example of a post that challenges the supposed double standard by invoking the accusations of "hysterical" behaviour :

16. Sure but Kavanaugh WAS called hysterical, so what's the double standard?

As a side note, while this topic is relatively free from noise, it does contain one post that doesn't exactly fit the theme - yet it uses some of the same language - particularly with regards to race and gender:

13. The year is 1950. That's the year the conservatives perpetually live in. Blacks, women, homosexuals, progressives, all oppressed.

Overall though, these posts illustrate the final theme that the topic modeling algorithm presents: perception of emotional behaviour as viewed through the lens of gender and race.

Interpretation of Topic Modeling Output

Gathering 20 individual topics may be interesting in its own right; however, the goal remains to answer one of the key research questions of this thesis. Namely, what kinds of narratives are used to discuss the Brett Kavanaugh controversy online?

This section takes a step back from discussing the specifics of each of the 20 found topics and recontextualizes them within the corpus as a whole, looking to identify broad trends and narratives within them that can help us answer the above question. To do so, I took the full list of topics as presented in table 2 and looked specifically at the thematic

topics. I used these to identify if there are particularly interesting groupings in the text and to examine these groupings in some depth to gain greater insight into the corpus' themes.

Before diving into that analysis however, it is relevant to observe a little bit of additional context. The following section presents a brief overview of how the controversy was discussed in some of the men's interests subreddits.

View From Men's Interests Subreddits

While a deep dive into how this controversy was discussed in the men's interests subreddits is outside of the scope of this thesis, it may be relevant, for context, to set the stage here with the types of discussion happening in the Men's Lib and Men's Rights subreddits. Table 3 displays the titles of reddit threads from each of the two men's subreddits from late 2018 that dealt with the Brett Kavanaugh controversy and had a minimum of 35 posts at the time of data collection. While this cutoff is somewhat arbitrary, the goal was to only include threads that received a fair amount of traction and discussion within their respective communities. In any case, these headings should only be viewed as a very general overview of the *kind* of discussion happening in each of the subreddits, as this may inform interpretation of some of the previously-described topics.

	Subreddit	Title	Num. of Posts
1	menslib	Given the Kavanaugh hearing yesterday, can we talk about what lessons we should be taking away from this for ourselves and our children?	372
2	menslib	1,600 men voice support for Christine Blasey Ford in New York Times ad US news	88
3	menslib	Amid Kavanaugh Allegations, Rethinking The Common Refrain 'Boys Will Be Boys'.	51
4	menslib	In light of the recent Kavanaugh/Supreme Court Ford events — how do you feel about your own behavior towards women as a teenager?	40
5	menslib	Family Conversations about Abuse, Post-Kavanaugh	39
6	mensrights	The Kavanaugh Smears Broke My Decades-Long Support For Feminism	377
7	mensrights	In delay tactic, Kavanaugh accuser Christine Blasey Ford says she won't testify about her accusations until a thorough FBI investigation is conducted first. Can't commit perjury if you don't testify.	228
8	mensrights	Should the Women who Made Knowingly False Statement in the Kavanaugh Hearing Be Criminally Prosecuted?	194

9	mensrights	I don't see anyone saying it, so I'm going to. I don't give a damn of your political beliefs, but the Kavanaugh fiasco should be concerning to everyone here.	109
10	mensrights	Poll finds 40% of Democrats now say Kavanaugh confirmation process a "national disgrace"	88
11	mensrights	As a male high school student, all this stuff with Kavanaugh and the boy where girls banded together to falsely accuse him scares the hell out of me. What can I do to prevent stuff like this happening to me?	81
12	mensrights	Woman Who Claimed Justice Kavanaugh Raped Her Now Admits They've Never Even Met. She's Been Referred to DOJ/FBI for Investigation and Could Soon Be in Serious Legal Trouble	74
13	mensrights	Kavanaugh Accuser CONFESSED To Making False Accusation	65
14	mensrights	Revsisting the false allegations made against me 16 years later	46
15	mensrights	Female Democrat Senator: Kavanaugh Not Entitled To Presumption Of Innocence Due To His Ideological Views	44
16	mensrights	As usual, woman who accused Kavanaugh of raping her admits they never even met	43
17	mensrights	Not one single mainstream outlet is talking about the fact that two of the three women accusing Kavanaugh of sexual	42
18	mensrights	Jordan Peterson: Kavanaugh Should Step Down from Supreme Court to 'Have His Name Cleared'	39
19	mensrights	Got banned from r/socialism for saying MAYBE kavanaugh isnt a rapist. "Note: Male"	37
20	mensrights	"Witches planning a 'hexing' of Kavanaugh, with proceeds from the event going to feminist groups." I've been saying for ages there's a strong link between feminists & 'Witches' (or nutcases). Here's an example.	36

Table 3

Discussion about the Brett Kavanaugh controversy from late 2018 from the menslib and mensrights subreddits.

Of the five threads collected from the menslib subreddit, there is one about men voicing support for a woman speaking out about sexual assault allegations (thread number 2), and four that suggest some level of reflection on individual males' roles in conversations about abuse and assault in light of the Kavanaugh controversy (threads 1, 3, 4 and 5).

Conversely, of the 15 threads in mensrights, there are nine threads that deal with the topic of false accusations of sexual assault and/or rape (threads 7, 8, 9, 11, 12, 13, 14, 16 and 17). Three of the remaining six threads deal with political ideology and bias towards accusations of sexual assault and rape (threads 10, 15 and 19). Of the remaining three, two decry feminism (threads 6 and 20), and one calls on Brett Kavanaugh to step down from

the Supreme Court in order to clear his name (thread 18).

An interesting dichotomy between the two subreddits is that the subset of threads from menslib seems introspective and reflective, and suggest an attempt at growth in light of the Kavanaugh controversy. For instance, thread 1 asks “what lessons should we be taking away from this for ourselves and our children?”, while thread 3 suggests “rethinking the common refrain ‘Boys will be boys.’” In contrast, the threads from mensrights seem outward looking, examining false accusations *against men*, also how political ideology may target certain men or devalue their perspectives, and how feminism fails men. There is a narrative about false accusations in many of the headlines, for instance thread 7 says of Brett Kavanaugh’s accuser, Christine Blasey Ford, that she “can’t commit perjury if [she doesn’t] testify”. There are individuals recounting their own stories, such as in thread 14: “Revisiting the false allegations made against me”, or thread 11, which describes one male’s fears of having false allegations thrown their way: “all this stuff with Kavanaugh where girls banded together to falsely accuse him scares the hell out of me”.

While this analysis is too simplistic to generalize, there is a difference in tone between threads from the same time period in the two different subreddits. In particular, threads that gained traction in the menslib subreddit suggest recontextualizing men’s roles in light of what’s been learned from the controversy, while threads that gained traction in the mensrights subreddit centre against feminism and focus on the injustice of false claims made against men.

Thematic Groupings

We now return to the topics generated from the posts in the TwoXChromosomes and Politics subreddits. Looking through the topics, we can find thematic groupings among them - groups of topics that share similar themes or discussion points. In this section, we will highlight and discuss some of those groupings, and use the groupings as a launchpad to further examine some of the broad conversations happening within the corpus. An overview of these topic groupings can be seen in table 4. We will dive deeper into them in the text

Topic Group	Topic IDs	Description
1	2, 10, 16, 19	Evidence
2	4, 7, 11, 18	Importance of Supreme Court nomination & investigation
3	6, 12, 17	Sexual assault & gender
4	13, 14, 20	The Hearing

Table 4

Topic groupings by topic ID with a brief description of their themes.

below.

Thematic Grouping 1: Evidence & the Search for Truth

Taken together, the first thematic grouping can be read as a moral quest for truth. There are attempts throughout to reconstruct the *truth* of the allegations, and attempts to judge whether or not *guilt* can be inferred from the known evidence. Indeed, Topic 2 (p. 33) sees a deconstruction and examination of individuals' memories, Topic 10 (p. 48) debates the reliability of polygraphic evidence, Topic 19 (p. 64) questions the quality of testimony as evidence, and Topic 16 (p. 58) argues over the burden of proof required to infer guilt.

There is thus in this thematic grouping a tendency towards debating and interpreting the facts as they are known, particularly when it comes to reconstructing the events of the alleged assault, or in debating the quality of the amassed evidence. Topic 19, in debating the quality of evidence - and chiefly, that of testimony as evidence - almost subverts this grouping by demonstrating its futility. In this case, the alleged crime is in the past and hard, incontrovertible proof to incriminate or exonerate the accused will almost certainly never be known. Given this, it is hard to imagine either side of the debate acquiescing. Topic 16 tries to deal with this by debating the threshold of evidence needed to infer guilt, or to act upon the accusations lobbed at Kavanaugh.

The findings for this particular thematic grouping are not explicitly gendered: in general, there are not calls to evidence based on gender characteristics. However, all four topics contain debates about evidence, and it is interesting to see the dichotomy between the

arguments attacking and defending Kavanaugh. Against Kavanaugh, we have the word, recollection and testimony of women. Defending Kavanaugh, we have competing testimonies, but also - a questioning of the validity of testimony at all. We also see a debate over the level of proof required to judge him (whether it be for an individual to personally consider him guilty, for him to be found criminally guilty, or for him to be found guilty enough to not be considered for a Supreme Court nomination). Truth is thus negotiated in two manners in this corpus. First, there is a debate over what can actually be considered fact. Second, there is a debate over how those facts can be used and interpreted.

This negotiation of truth points to an underlying tension which might in fact be gendered. This tension is largely unresolved in the text, but on the one hand, there is a contingent that seek justice for wrongful behaviour or believe that any level of doubt is enough to eliminate Kavanaugh from contention for what is perceived to be one of the most powerful positions in the United States, if not the world. On the other, there is a resistance to this idea based on the notion that one should be proven guilty beyond a shadow of a doubt if one is to suffer any ill consequences for an accusation levelled against them.

What should be noted, however, is that the narrative here is clearly focused on a story where Brett Kavanaugh's nomination to the Supreme Court is the central figure. Evidence is discussed and debated to determine the truth of the allegations, and the framework for deciding guilt is about whether or not he should be a Supreme Court justice nominee. In terms of consequences, topics 16 and 19 extrapolate what Kavanaugh's guilt or innocence might mean for him personally and career-wise, and also for the United States as a whole, given its impact on his nomination to the Supreme Court. Perhaps conspicuous by its absence, one can note that *not* discussed in this topic is what truth would mean for the accuser and alleged victim, Christine Blasey Ford.

Thematic Grouping 2: Importance of a Nomination to the Supreme Court

The second grouping deals with the Supreme Court and generally discusses the importance of a nomination to the Supreme Court. It groups three topics: Topic 4 (p. 38) deals

with a nomination to the Supreme Court, Topic 11 (p. 49) deals with political partisanship in Supreme Court nominations, and Topic 18 (p. 61) discusses the investigation into the allegations (particularly in the context of Brett Kavanaugh's Supreme Court nomination), while Topic 7 (p. 42) deals with how the allegations by Dr. Ford became public.

This thematic grouping is largely concerned with the power granted to a Supreme Court justice, and the partisanship involved in nominating a new justice to the Court. Topic 4 is largely concerned with establishing the power of the position and the process to attain it. Topic 11 delves into how and why partisanship influences the nomination process. Topic 7 looks at how Dr. Ford's story went public, and in part, debates whether the allegations are a strategic ploy by Democrats. Topic 18 looks at the investigation into the allegations and examines how the same partisanship that affects Supreme Court nominations might impact the investigative process.

In this thematic grouping, we see a conception of the American political apparatus as highly partisan: a position of power - namely, a seat on the American Supreme Court - is decided by partisan politicians. Obtaining this power is thus the result of a calculated game of chess. There is concern herein about which sides are being fair, and which are simply trying to flex their political muscle to produce a desired outcome. In topic 11, these nominations are partisan in the best of times. In topics 7 and 18, the production of sexual assault allegations, and investigation into them are debated as having been produced by a partisan process, and wielded strategically by partisan interests.

Thematic Grouping 3: Sexual Assault and Gendered Norms

The third grouping broadly encompasses discussions and interpretations of gendered norms as it pertains to sexual assault. It contains Topic 6 (p. 41) dealing with perceptions of those accused of sexual assault and their accusers, Topic 12 (p. 51), which includes discussions of sexual assault through the lens of gender and race, and Topic 17 (p. 60) which deals with accusations of rape & their impact.

This thematic grouping is high in its inclusion of explicitly gender-oriented discussion.

Topic 6 includes a discussion of *responsibility* for sexual assault and to what extent a woman is responsible for an assault committed against her, and also some discussion of the “boys will be boys” narrative, exploring the idea that men may simply be more prone to certain behaviours. Topic 12 is largely concerned with white men and whether or not they are too often held unaccountable for sexual assault, or prime targets for false accusations. It also debates whether white men are punished more severely than others for a comparable crime, and if white women are able to act with impunity and make false accusations. In topic 17, concern is shown for the negative impact that rape allegations can have on a man’s career - though interestingly enough, there does not seem to be a significant complementary discussion about the impact that *making* the accusation, or being victim of the crime, may have on a woman’s life. This topic is not brought to the forefront in any significant way by the topic modeling algorithm. It also was not evident in looking at the headings of discussions in either of the men’s interests subreddits.

Central to this grouping is the implicit notion that women are the primary targets of sexual assault. There is some pushback to this in topic 6, with some stating that women aren’t the only victims, but it exists as a tacit assumption in topic 12, where white men are seen either as aggressors who often get away with crimes, or as targets for false accusations and severe punishments. In contrast, women - when not viewed explicitly as victims - are portrayed as capable of acting with impunity or able to make false accusations against men (which, as we might recall, was one of the central themes found in the headings of the discussions in the mensrights subreddit, p. 67).

Topic 17, which deals with the negative repercussions that rape allegations can have on a man’s career, makes explicit some of the consequences of guilt that were discussed in the first thematic grouping: guilt in this context can have an impact on a man’s life (and perhaps deservedly so, depending on the redditor), and may be a determinant in allowing an individual a seat on the Supreme Court. However we again note the absence of discussion of consequences for women who may have been assaulted, or for Dr. Ford if her story is true. This perspective appears to not appear systematically enough in the corpus (if at all)

to bubble to the surface of any of the topics presented by the LDA algorithm.

Interestingly, the sources that seem to most consider this perspective are some of the posts in the menslib subreddit, which questioned the role of men and pondered what men could learn from this controversy moving forward. Still - those conversations superficially appear to still be about men's perspectives. The corpus appears to contain little explicit discussion of outcomes for Christine Blasey Ford, or for women more generally.

Thematic Grouping 4: The Hearing

Finally, this is a grouping of topics that deals with the hearing into the Kavanaugh accusations. Specifically, it includes Topic 13 (p. 53) on the perception and interpretation of behaviours during the hearing, Topic 14 (p. 54) about whether Brett Kavanaugh was lying during the hearing, and Topic 20 (p. 65), which includes discussions of gendered and racialized interpretations of emotional behaviour (largely displayed during the hearing).

Topic 13 is particularly concerned with what is perceived as a lack of composure on Brett Kavanaugh's part during the hearing. His behaviour is largely described as lacking characteristics that many expect from a justice of the Supreme Court: such as being "rational", "measured", and having the ability to "conduct himself with grace" under pressure.

Topic 20 further examines what is perceived as Kavanaugh's unacceptable behaviour during the hearing, but this time paints that picture in gendered terms, pointing to the notion that had a non-white, non-male person behaved similarly in the same context, their behaviour would be interpreted as "hysterical" (if a woman) or "overly aggressive" if Black.

Topic 14 debates what some identified as possibly being lies told by Kavanaugh during the hearing. Oft-cited were "Devil's Triangle", "boofing", and his connections at Yale. Many posts in the topic outright challenge Kavanaugh's telling of the truth, although a small number of posts also defend him in this context.

As a whole, this thematic grouping is clearly concerned with Brett Kavanaugh's comportment during the hearing, and paints complementary interpretations of his behaviour. We see an examination of how that behaviour compares to the "ideal" behaviour of a

Supreme Court justice: both by directly pointing out the perceived flaws in his behaviour in topic 13, and the possible lies told in topic 20. We also see a discussion of how said behaviours are often interpreted depending upon the actors: topic 14 essentially puts forth the debate that individuals other than white men might not be granted some of the same affordances as Kavanaugh might be.

Summary of Themes

Stepping back from the individual fine-grained topics and viewing them as part of larger groupings helps contextualize them, as the thematic groupings show that they often form complementary aspects of larger discussions. While I have chosen one particular manner of grouping the topics above, it should be noted that there is significant overlap within these groupings. The goal in this exercise was not to find definitive, clean-cut topics, but rather to get a broad overview of the kind of discussion happening within the corpus. These groupings have thus suggested one way to answer the question: “How are people discussing the Brett Kavanaugh controversy within this particular threads on reddit?”

The four areas of discussion offer a better sense of the stakes of the discussion. There is significant discussion about the role of evidence in establishing truth and the veracity of the allegations. It is interesting to note that this topic seems complementary to the dominant theme of the falseness of the allegations in the mensrights subreddit, as explored in the View From Men’s Interests Subreddits (p. 67) section. The importance of the Supreme Court nomination is another important discussion in its own right, as is the partisanship that surrounds it. Given the nature of the allegations, discussion about sexual assault also plays an important role in the corpus. Finally, interpretation and discussion of the hearing itself also proved salient among redditors posting in the selected threads.

I also described some themes that seemed to be *absent* from the topic modeling output. Notably, I discussed the absence of significant discussion about consideration for what making accusations means for the accuser, as well as the absence of discussion over what sexual assault means for the victims of the crime. While the omission of these topics

from the topic modeling output is noteworthy, one must be careful in interpreting what that entails: it *does not* mean that there was *no* discussion of that idea throughout the text (that can only be confirmed by a complete read-through of the raw data). It does however indicate that *if* there was such a discussion in the corpus, language was consistently dissimilar enough within those posts to evade being categorized together. One way to verify the omission was to look through my notes in Appendix III for selecting a number of topics to work with. No other quantity of topics significantly raised these issues, either. In the topic modeling output starting with the number of topics set to 40, I did observe *some* discussion of victims, but even then, the topic was a noisy one that didn't lead to a single, coherent interpretation.

Discussion

In this thesis project, I have sought to perform two parallel tasks: first, to analyze the discourse around a contentious issue; and second, to present a methodology for extracting insights from large corpora. In order to achieve both goals, I used a methodology centred around topic modeling to examine the discourse around the nomination of Brett Kavanaugh to the United States Supreme Court in 2018 in a number of threads on the social media site *reddit*. In this section I reflect on both sets of findings: first, I consider the thematic findings from the previous section in the context of the body of literature about argumentation examined earlier in this document. Secondly, I reflect on the use of the methodology as a research tool, examining strengths and weaknesses of the method. Finally, I describe the development of this project and its grounding in practical applications.

Interpreting Findings

In the introduction to this document, I examined the forces that shape argumentation, and used this as a framework to understand how opinions are formulated and expressed in online fora. It has been my assumption that individuals are capable of rational thought, have a desire to be correct, but are subject to a host of factors which may motivate bias

in their reasoning. These are forces that I have assumed we are all subjected to, to various extents.

The particular issue at study in this thesis - that is, the controversy around the nomination of Brett Kavanaugh to the United States Supreme Court - is particularly contentious because it exists at the intersection of gendered and politicized controversies, two types of controversies that, as we saw in the introduction, are subject to forces that bias our reasoning. These forces are likely to induce motivated reasoning on everyone (this author included) at some points in time, and are thus just as much at play for Democrats as they are for Republicans; for feminists and men's rights activists.

With this in mind, it is important to note that in this thesis, I can make no assumption on how specific forces influenced any one individual's line of argumentation, or contribution to the overall discussion. Indeed, the form of analysis presented in this work can only concern itself with the collections of words used by redditors in their posts, and, as no background information is known about these individuals, no inference can be made about their disposition towards any specific type of motivated reasoning. An interesting avenue for future research may be to investigate the specific ways that motivated reasoning occurs throughout these discussions. For instance, a question that seems particularly interesting to me in light of this study's findings is the following: in the debates over facts and what can be considered evidence (Thematic Grouping 1: Evidence & the Search for Truth (p. 70)), how do individuals from different political backgrounds argue? Is testimony regarded as valid evidence by individuals identifying with one political party more so than those of another? Furthermore, would individuals' opinions of the validity of testimony hold in a similar controversy if the political roles were reversed? Another interesting avenue for future work may be to consider the multi-dimensional nature of this controversy. As this controversy is both politicized and gendered, it would be interesting to consider the ways that multiple identities (in this case, gender and political) interact in the production of argumentation.

In the introduction, GamerGate was discussed as a gendered controversy where Gamer-

Gaters may have feared losing cultural space in a zero-sum game (Shepherd et al., 2015). Given the explicit zero-sum nature of power in a Supreme Court nomination (i.e. a Republican judge necessarily takes a potential seat away from a Democrat judge), it is fair to consider the similarities between these two controversies, and the ways in which loss aversion may have influenced people's opinions.

Indeed, one of the underlying thematic elements that became apparent to me after stepping back from the four major themes in the corpus is the focus on *power*:

Politically, power is discussed in terms of who has it and who wants it, and coincides largely with the second thematic grouping (Thematic Grouping 2: Importance of a Nomination to the Supreme Court (p. 71)). Power was also discussed with respect to gender: the third thematic grouping presented discussions of men having power over women with respect to performing acts of sexual assault, but also of women potentially having power over men in terms of accusing them of assault. Finally, in discussing the facts of the case and the evidence within, the first thematic grouping examines and tries to establish who has the power, and under what circumstances, to render a verdict and assign consequences to the accused.

There seems to be an implicit assumption throughout the corpus that power is zero-sum - that one person's gain of power is another's loss. It is fair to wonder if this could in fact be one of the ways that motivated reasoning is triggered throughout the discussion, on top of the other factors examined in the introduction. Indeed, loss aversion is a documented bias exhibited by humans, wherein we go to great lengths to avoid suffering perceived losses (Kahneman & Tversky, 1979; Rabin & Thaler, 2001). This provides yet another avenue for future research: to what extent do motivations to obtain and maintain power shape discussion and argumentation?

Topic Modeling as a Qualitative Research Tool

One of the two central aspects of this project has been the use of computational tools to uncover narratives in social media text. I used the LDA topic modeling algorithm

to uncover fine-grained, latent topics within a relatively large corpus, and then grouped those topics thematically to get a broad sense of the discussions happening throughout. In this section, I examine some of the strengths and weaknesses of the approach discovered along the way, as well as other considerations to be taken into account when applying this methodology and interpreting its results.

One obvious strength of the topic modeling approach is that it allows one to take a large corpus and boil it down to a set of latent topics. This divides the texts into meaningful subsets, where each subset is centred around a particular group of words, and thus, likely focus on a similar subject or employ similar discursive patterns. The ability to perform this initial triage on arbitrarily large sets of data can be immediately useful in identifying some of the most salient threads of discussion within a sizeable corpus.

Another clear strength of the topic modeling approach is the ability to ground research in speakers' own words. The large-scale division of the corpus into topics is conducted automatically based on patterns of co-occurrence of tokens in text, which mitigates one possible source of interpretive bias by human analysts. The work of interpretation remains a human task, but the initial division is produced systematically, and the list of representative keywords for a topic provide a grounding for interpretation in the actual words of the individuals speaking them.

Topic models are a rather blunt instrument, however. While they may give a good sense of the narrative threads happening throughout the corpus, the basic building block of each topic is a set of tokens, and an initial understanding of each topic uses that set of tokens and a set of representative posts. Both of these convey a substantial amount of information, yet omit much nuance. Without further digging, one loses out on the structure of the conversation, and does not immediately see the order of discussion, nor what points are made in response to which posts.

Because the basic building block of each topic is a set of tokens, some ambiguity is also introduced if a token has multiple definitions within the same corpus. While the discussion was not significant enough to form a topic in this project with the number of topics set to 20,

in results packages with the number of topics set to 40 or 50, some topics which ostensibly appeared to be about Hillary Clinton included the token “bill”, in reference to her husband Bill Clinton. Some unrelated posts were also included about comedian Bill Maher - largely because those posts shared the token “bill” in common with the posts about the Clintons.

For the purposes of this project, a list of stopwords was used to disqualify certain words from inclusion in topic models; the idea being that these words (like *the*, *a*, *I’m*, and *their*, for instance) are too common to reliably be indicators of one latent topic or another. Other approaches to using a list may be valid, however. For instance, the NLTK library (Bird, Klein, & Loper, 2009) in Python allows one to dynamically generate lists of stopwords by only considering words that appear in a certain proportion of documents, thereby omitting words that appear in too few or too many documents to be considered informative. Experimentation with different lists and methods of eliminating stopwords may help fine-tune results.

It should be noted that all of the topics generated by the algorithm include what might be called *noise*, or words or documents that don’t seem to fit the underlying theme of a topic, even if it shares some language with others in the topic. Some amount of subjectivity is required to discern when to dismiss a document as “noisy” within a topic (for instance, if its language is similar but it is clearly discussing a different idea than other documents in the topic), or to use the inclusion of such noisy posts to wonder if one’s interpretation of a topic is indeed “on the mark”.

While one of the selling points of this method is the ability to analyze large corpora, it should be noted that selecting the number of topics is subjective and somewhat labour-intensive. The process was (subjectively) helpful in narrowing in on what some of the key topics at play in the corpus were, but could be difficult to implement for much larger corpora with larger numbers of topics. While some computational evaluation measures like *complexity* attempt to evaluate the coherence of topics generated by the algorithm, these measures do not always align with human interpretation. Future work may want to address this particular concern. However, it should be said that while choosing the number of topics

to use throughout the analysis was time-consuming, analysis of a specific results package was much more straightforward.

To continue on the idea of selecting a number of topics, one interesting observation is how specific and narrow some of the topics became as the number of topics was increased (to, for instance, 40 or 50). At that threshold, topics began to have very small numbers of highly-relevant posts, but these were often much more narrowly focused. Conversely, running the algorithm with fewer topics led to much broader, and sometimes hard-to-distinguish topics (insofar as if *too many* topics are bundled together, it was difficult to see any theme binding them together). This project struck a compromise by selecting the number of topics which provided the most straightforward interpretation. However, future work may attempt to divide a corpus into more, narrower topics, and then reconstitute them into broader themes (as was done here), or divide a corpus into fewer topics, and then some subset of those topics through the algorithm again to achieve further division.

Finally, while this method gives an overview of *what* themes of conversation are happening within the corpus, the algorithm itself does nothing to inform of the stances people have or quality of discussion around those themes, nor does it create in the reader the required background knowledge of a situation to properly interpret its output. So while the algorithm's output clearly demonstrates some conversation about evidence and the burden of proof, it does not indicate what exactly people are saying about it, which remains unknown unless one reads the source material. To make that inference requires a human reader, or possibly, additional computational methods from sentiment analysis or other fields. This project has aimed to keep a human analyst still very much involved in the process, however, as computational tools are not always best at understanding the context in which discourse happens. This may also be the subject of future work, as the larger and more far-reaching a corpus is in subject matter, the more we might need to seek out computational methods to help us make sense of context.

Practical Applications

The bulk of the methodology in this project was first developed in work done with Project Someone¹⁹. I collaborated on the development of the Words in Context project²⁰, where I developed a database and a web interface to examine the output of the team's analysis. The work largely sought to extract insight from social media texts to examine hate discourse online, in order to inform stakeholders of patterns and trends in discourse around online hate, particularly in relation to tensions in both Lebanon and Canada.

In parallel to the development of the database and web interface, I developed an initial version of this methodology using a subset of the raw data used to inform the team's analysis. This first attempt (Rodier, 2019) illustrated the power of the methodology to find relevant narratives within a corpus, and inspired me to pursue it further for this thesis. This work adapts the original write-up and adds depth to the methodology to gain even more valuable high-level insights from the raw text, while applying it to a different context.

This type of work is important in promoting understanding among interested stakeholders; whether they be local governments, community groups, researchers, or others. Understanding the narratives and discussion themes present on social media is not always a straightforward task: anecdotally, we may all individually have our experiences navigating social media sites, but the volume of information is often too great to parse, and thus, computational methods can be of valuable assistance in helping us extract trends from texts. Be it for policy-making or crafting educational interventions, it is important for stakeholders to possess a critical understanding of the individuals and audiences with which they mean to interact. For interested stakeholders, methodologies such as the one I have presented in this thesis are vital in efforts to manage and gain insight from text-based corpora.

¹⁹<https://projectsomeone.ca>

²⁰<https://projectsomeone.ca/wordsincontext>

Conclusion

This project has sought to answer two specific questions. One was aimed particularly at uncovering themes in social media discussions revolving around the Brett Kavanaugh controversy in 2018, and the second, aimed at using computational tools to assist in the discovery of those themes.

I began by examining the literature about how people interact online, how individuals form opinions and debate in online spaces, and how contentious issues are navigated in these spaces. I examined how gender is one such contentious issue before selecting the Brett Kavanaugh controversy as a specific case to examine.

Approximately 6,500 posts were extracted from the social media site *reddit* that discussed the Brett Kavanaugh controversy. These posts were analyzed using a framework centered around the Latent Dirichlet Allocation (LDA) algorithm which generates *topic models*: clusters of documents within the corpus that use similar language. These clusters of documents are characterized by their most representative words as well as the most representative documents within them. These clusters were then used to find common topics of discussion within the corpus, as well as thematic tendencies across topics.

The analytic method uncovered four broad themes throughout the discussion of the Brett Kavanaugh controversy. These include the search for and evaluation of evidence; the importance of a Supreme Court nomination and related investigation into the Kavanaugh controversy; sexual assault and gendered perceptions and expectations; and finally, a discussion of the events of the judiciary hearing into the Kavanaugh assault allegations. These findings, generated in part algorithmically, and in part by human analysis and interpretation, are grounded in the specific language of the documents in the corpus.

I also took the opportunity to reflect on the use of topic models as a qualitative research tool. While their capacity to deal with large corpora is a feature, certain considerations remain, and future work may be required to perfect their usage in this domain. Their use, however, may very well aid in the analysis of larger corpora, where more labour-intensive qualitative analyses may be impossible.

References

- Althoff, T., Danescu-Niculescu-Mizil, C., & Jurafsky, D. (2014, May). How to Ask for a Favor: A Case Study on the Success of Altruistic Requests. *arXiv:1405.3282 [physics]*. Retrieved 2018-10-21, from <http://arxiv.org/abs/1405.3282> (arXiv: 1405.3282)
- Amazon.com. (2020). *Top sites in United States*. Retrieved from <https://www.alexa.com/topsites/countries/US> (Accessed 2020-01-04)
- Baumer, E. P., Mimno, D., Guha, S., Quan, E., & Gay, G. K. (2017). Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? *Journal of the Association for Information Science and Technology*, 68(6), 1397–1410.
- Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with python* (1st ed.). O'Reilly Media, Inc.
- Blandfort, P., Patton, D., Frey, W. R., Karaman, S., Bhargava, S., Lee, F.-T., ... Chang, S.-F. (2018, July). Multimodal Social Media Analysis for Gang Violence Prevention. *arXiv:1807.08465 [cs, stat]*. Retrieved 2018-10-22, from <http://arxiv.org/abs/1807.08465> (arXiv: 1807.08465)
- Blei, D. M., & Lafferty, J. D. (2009). Topic models. In *Text mining* (pp. 101–124). Chapman and Hall/CRC.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993–1022.
- Blevins, T., Kwiatkowski, R., Macbeth, J., McKeown, K., Patton, D., & Rambow, O. (2016). Automatically processing tweets from gang-involved youth: towards detecting loss and aggression. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers* (pp. 2196–2206).
- Bucholtz, M., & Hall, K. (2005, October). Identity and interaction: a sociocultural linguistic approach. *Discourse Studies*, 7(4-5), 585–614. Retrieved 2018-10-22, from <http://journals.sagepub.com/doi/abs/10.1177/1461445605054407> doi: 10.1177/1461445605054407
- Caffier, J. (2017, 09). Here are reddit's whiniest, most low-key toxic subreddits.

- Vice*. Retrieved from https://www.vice.com/en_us/article/8xxymb/here-are-reddits-whiniest-most-low-key-toxic-subreddits (Accessed 2020-01-04)
- Chang, S., Zhong, R., Adams, E., Lee, F.-T., Varia, S., Patton, D., ... McKeown, K. (2018, September). Detecting Gang-Involved Escalation on Social Media Using Context. *arXiv:1809.03632 [cs]*. Retrieved 2018-10-22, from <http://arxiv.org/abs/1809.03632> (arXiv: 1809.03632)
- Crary, D. (2018, 09). Kavanaugh-Ford hearing: A dramatic lesson on gender roles. *Associated Press News*. Retrieved from <https://apnews.com/c3bd7b16ffdd4320a781d2edd5f52dea/Kavanaugh-Ford-hearing:-A-dramatic-lesson-on-gender-roles> (Accessed 2019-12-15)
- Cummings, W. (2018, 10). Kavanaugh supporters celebrate, spark social media outrage with #beers4brett hashtag. *USA Today*. Retrieved from <https://www.usatoday.com/story/news/politics/onpolitics/2018/10/08/kavanaugh-supporters-beers-4-brett/1563416002/> (Accessed 2019-12-17)
- Danescu-Niculescu-Mizil, C., Lee, L., Pang, B., & Kleinberg, J. (2012). Echoes of Power: Language Effects and Power Differences in Social Interaction. In *Proceedings of the 21st International Conference on World Wide Web* (pp. 699–708). New York, NY, USA: ACM. Retrieved 2018-10-22, from <http://doi.acm.org/10.1145/2187836.2187931> doi: 10.1145/2187836.2187931
- Filipovic, J. (2007). Blogging While Female: How Internet Misogyny Parallels Real-World Harassment. *Yale Journal of Law and Feminism*, 19, 295. Retrieved from <https://heinonline.org/HOL/Page?handle=hein.journals/yjfem19&id=299&div=&collection=>
- Gelman, J., & Wilson, S. (2018, 11). Will Kavanaugh's confirmation affect the midterms? Here's what showed up on social media. *The Washinton Post*. Retrieved from <https://www.washingtonpost.com/news/monkey-cage/wp/2018/11/05/will-kavanaughs-confirmation-affect-the-midterms-heres-what-showed-up-in-social-media/> (Accessed 2019-12-17)

- Gerber, M. S. (2014, May). Predicting crime using Twitter and kernel density estimation. *Decision Support Systems*, *61*, 115–125. Retrieved 2018-10-27, from <http://www.sciencedirect.com/science/article/pii/S0167923614000268> doi: 10.1016/j.dss.2014.02.003
- Hart, P. S., & Nisbet, E. C. (2012, December). Boomerang Effects in Science Communication: How Motivated Reasoning and Identity Cues Amplify Opinion Polarization About Climate Mitigation Policies. *Communication Research*, *39*(6), 701–723. Retrieved 2018-11-18, from <http://journals.sagepub.com/doi/10.1177/0093650211416646> doi: 10.1177/0093650211416646
- Jane, E. A. (2014a, July). ‘Back to the kitchen, cunt’: speaking the unspeakable about online misogyny. *Continuum*, *28*(4), 558–570. Retrieved 2018-11-14, from <https://doi.org/10.1080/10304312.2014.924479> doi: 10.1080/10304312.2014.924479
- Jane, E. A. (2014b, July). “Your a Ugly, Whorish, Slut”. *Feminist Media Studies*, *14*(4), 531–546. Retrieved 2018-11-14, from <https://doi.org/10.1080/14680777.2012.741073> doi: 10.1080/14680777.2012.741073
- Kahneman, D. (2011). *Thinking, fast and slow*. Canada: Doubleday Canada.
- Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–291. Retrieved 2020-12-26, from <http://www.jstor.org/stable/1914185> (Publisher: [Wiley, Econometric Society]) doi: 10.2307/1914185
- Malone, C. (2018, 09). The Ford-Kavanaugh hearing was a clash of gender politics. *FiveThirtyEight*. Retrieved from <https://fivethirtyeight.com/features/the-ford-kavanaugh-hearing-was-a-clash-of-gender-politics/> (Accessed 2019-12-15)
- McCallum, A. K. (2002). *Mallet: A machine learning for language toolkit*. Retrieved from <http://mallet.cs.umass.edu>
- Newsom, V., & Lengel, L. (2004). The Culture of Computing: Gender Online as Contained Empowerment. *Conference Papers – International Communication Association*, *1*. Retrieved 2018-11-12, from <http://0-search.ebscohost.com/mercury.concordia>

.ca/login.aspx?direct=true&db=ufh&AN=45283735&site=eds-live

- Oyserman, D. (2009). Identity-based motivation: Implications for action-readiness, procedural-readiness, and consumer behavior. *Journal of Consumer Psychology*, *19*(3), 250–260.
- Paul, M. J., & Dredze, M. (2012). Experimenting with Drugs (and Topic Models): Multi-Dimensional Exploration of Recreational Drug Discussions. In *AAAI Fall Symposium: Information Retrieval and Knowledge Discovery in Biomedical Text*.
- Petty, R. E., & Cacioppo, J. T. (1986). The Elaboration Likelihood Model of Persuasion. In R. E. Petty & J. T. Cacioppo (Eds.), *Communication and Persuasion: Central and Peripheral Routes to Attitude Change* (pp. 1–24). New York, NY: Springer New York. Retrieved 2018-10-22, from https://doi.org/10.1007/978-1-4612-4964-1_1 doi: 10.1007/978-1-4612-4964-1_1
- Rabin, M., & Thaler, R. H. (2001). Anomalies: risk aversion. *Journal of Economic perspectives*, *15*(1), 219–232.
- Redlawsk, D. P. (2002). Hot cognition or cool consideration? Testing the effects of motivated reasoning on political decision making. *The Journal of Politics*, *64*(4), 1021–1044.
- Rhor, M. (2018, 09). Brett Kavanaugh supreme court confirmation hearing shows divide between men, women. *USA Today*. Retrieved from <https://www.usatoday.com/story/news/2018/09/28/brett-kavanaugh-christine-blasey-ford-assault-claim-gender-divide/1459557002/> (Accessed 2019-12-15)
- Rodier, S. (2019). *Making sense of text with topic models: Using computational methods to reveal latent document groupings in a corpus*. Retrieved from <https://projectsomeone.ca/wp-content/uploads/2019/11/Making-sense-of-text-with-topic-models.pdf> (Accessed 2020-01-04)
- Saleem, H. M., Dillon, K. P., Benesch, S., & Ruths, D. (2017). A web of hate: Tackling hateful speech in online social spaces. *arXiv preprint arXiv:1709.10159*.
- Shepherd, T., Harvey, A., Jordan, T., Srauy, S., & Miltner, K. (2015, July). Histories of Hating. *Social Media + Society*, *1*(2), 2056305115603997. Retrieved 2018-11-12, from

<https://doi.org/10.1177/2056305115603997> doi: 10.1177/2056305115603997

Sobieraj, S. (2018, November). Bitch, slut, skank, cunt: patterned resistance to women's visibility in digital publics. *Information, Communication & Society*, 21(11), 1700–1714. Retrieved 2018-11-09, from <https://doi.org/10.1080/1369118X.2017.1348535> doi: 10.1080/1369118X.2017.1348535

Somasundaran, S., & Wiebe, J. (2010). Recognizing stances in ideological on-line debates. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text* (pp. 116–124). Association for Computational Linguistics.

Sutherland, O., Forbes, L., Hodgson, B., & McLaren, K. (2014, November). Digital actualizations of gender and embodiment: Microanalysis of online pregnancy discourse. *Women's Studies International Forum*, 47, 102–114. Retrieved 2018-11-09, from <http://www.sciencedirect.com/science/article/pii/S027753951400082X> doi: 10.1016/j.wsif.2014.05.005

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3), 755–769.

Vesoulis, A. (2018, 09). The Kavanaugh-Ford hearings showed how men and women talk differently. *Time*. Retrieved from <https://time.com/5409469/christine-blasey-ford-brett-kavanaugh-gender/> (Accessed 2019-12-15)

Wallace, B. C. (2012). Multiple Narrative Disentanglement: Unraveling Infinite Jest. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 1–10). Stroudsburg, PA, USA: Association for Computational Linguistics. Retrieved 2018-10-23, from <http://dl.acm.org/citation.cfm?id=2382029.2382031>

Walton, G. M., & Banaji, M. R. (2004). Being what you say: The effect of essentialist linguistic labels on preferences. *Social Cognition*, 22(2), 193–213.

Appendix

Due to space concerns, the appendices are hosted online at the links below. They can be accessed as a group at <https://github.com/mopigeons/ma-thesis-supplementary-materials>, however direct links for each set of documents are presented below.

Appendix I

This spreadsheet contains a list of all of the reddit threads selected for analysis.

https://github.com/mopigeons/ma-thesis-supplementary-materials/raw/main/appendix01_selectedThreads.xlsx

Appendix II

This zip file contains the output from the 20 runs of the topic modeling algorithm, for numbers of topic in the set [2-15], 20, 25, 30, 40, 50. This is the data that informed my selection of a number of topics for the analysis portion of this project.

https://github.com/mopigeons/ma-thesis-supplementary-materials/raw/main/appendix02_dataForTopicSelection.zip

Appendix III

This spreadsheet contains my notes for each of the results packages and topics contained in Appendix II, and were the basis of my selection of the number of topics for the analysis portion of this project.

https://github.com/mopigeons/ma-thesis-supplementary-materials/raw/main/appendix03_topicSelectionNotes.xlsx

Appendix IV

This text file contains the output of the topic modeling algorithm where the number of topics is set to 20. It is largely similar to the corresponding results package from Appendix II, but contains additional metadata.

[https://github.com/mopigeons/ma-thesis-supplementary-materials/blob/main/
appendix04_postsWithMetadataForAnalysis.txt](https://github.com/mopigeons/ma-thesis-supplementary-materials/blob/main/appendix04_postsWithMetadataForAnalysis.txt)