Media and Communications

Media@LSE MSc Dissertation Series

Editors: Bart Cammaerts and Lisa Derand

Mapping Networks of Moral Language on U.S. Presidential Primary Campaigns, 2016-2020

Kobi Hackenburg

Published by Media@LSE, London School of Economics and Political Science ("LSE"), Houghton Street, London WC2A 2AE. The LSE is a School of the University of London. It is a Charity and is incorporated in England as a company limited by guarantee under the Companies Act (Reg number 70527).

ABSTRACT

During U.S. presidential elections, candidates engage in a connected discourse, contemporaneously offering competing visions and assessments of America. Research reveals that the moral language used to articulate these messages can significantly influence the way individuals form and act on their judgments about political issues and candidates; however, the moral language actually used by campaigning Democratic and Republican candidates has gone largely unmeasured. Using a Twitter dataset of every tweet (N = 139,412) published by 39 U.S. presidential candidates during the 2016 and 2020 primary elections, moral language is first extracted and then used to construct network models illustrating its role in connecting or differentiating both candidates and parties. The results find that Democratic and Republican candidates appeal to voters along differing moral dimensions, with Republican candidates emphasizing in-group loyalty and respect for social hierarchies, and Democratic candidates emphasizing careful and fair treatment of individuals. The results also illustrate how shared patterns of moral expression between competing candidates form intra-party norms which define the moral-rhetorical connectedness of a given primary, with 2016 Republican candidates espousing less unified moral rhetoric than their Democratic counterparts. Finally, the results find that candidates can insulate themselves within — or isolate themselves from — their party and competition through their use of moral language, making empirical and visible the extent to which political outsiders like Donald Trump rhetorically distanced themselves from the political establishment. This study finds that unique methods of text network analysis can be effective in studies of politics, and on campaigns in particular, and addresses the research gap that exists regarding the way candidates use moral language and how these patterns of expression establish rhetorical networks of partisan division — and unity.

Mapping Networks of Moral Language on U.S. Presidential Primary Campaigns, 2016-2020

Kobi Hackenburg

INTRODUCTION

During a 2016 rally in Iowa, Democratic presidential nominee Hillary Clinton stood behind a dais and stated a fact so obvious it seemed absurd: "*Words matter*" (Keneally, 2016). Clinton was referring to recent comments made by her general election rival, Donald Trump, who had excoriated her as a "traitor", "criminal", and "devil", culminating in the ominous statement that maybe gun rights activists should "do something about her" (Corasaniti & Haberman, 2016). Hillary's statement — words matter — became a solemn refrain as the Republican party embraced a vitriolic outsider, but its assured delivery belied another, more desperate fact: that the wildly enthusiastic supporters who flocked to Trump's rallies in the wake of his comments had not done so because of a legislative policy dispute, or out of rational self-interest; in fact, many supporters were animated by those words, motivated by an emotional desire to confront what they felt was an existential threat to their values and way of life (Khazan, 2018). Clinton therefore left her dais having expressed one truth and acknowledged two others: first, that words do indeed matter. Second, that partisan divisions in the U.S. are not purely political, resting instead on fundamental disagreements over deeply held values and moral convictions. Finally, that these words are especially meaningful — and these moral divisions especially clear — during contentious presidential elections.

Liberal and conservative voters in the U.S. tend to diverge in the ways they make moral judgments (Haidt & Graham, 2007; Graham *et al*., 2009). In fact, research has repeatedly shown that moral intuitions and political behavior are closely linked (Janoff-Bulman, 2009; Morgan *et al*., 2010) and often determine the way individuals approach political choices. Exposure to familiar moral language, for example, can have significant persuasive effects (Feinberg & Willer, 2015), greatly influencing the degree to which a liberal or conservative individual is likely to endorse a political issue (Feinberg & Willer, 2019). Conversely, exposure to unfamiliar or opposing moral language can entrench individuals in their existing views and exacerbate political polarization (Tetlock *et al*., 2000). Research even suggests that specific use of moral language can effectively induce or reduce support for political candidates, bearing important implications for electoral outcomes (Voelkel & Feinberg, 2018; Voelkel & Willer, 2019). All of this convincingly suggests that moral language influences the way Americans think about, understand, and act on their political judgments, especially in electoral contexts where the importance of such judgments is heightened (Hart, 2009). This begs the question: what sorts of moral language are political candidates using on their campaigns? How do patterns of moral expression delineate parties and candidates, and when are those boundaries crossed?

Surprisingly, in spite of these questions, the moral language used by campaigning politicians in the U.S. has gone unmeasured. This study begins to fill this research gap by mapping this previously uncharted domain, revealing not just the moral language used by candidates, but how this moral language connects candidates and parties to form a moral-political landscape. To this end, this research utilizes a comprehensive dataset of candidate Twitter discourse to examine the moral language used by 39 candidates during the 2016 and 2020 primaries in the U.S., paying particular attention to the ways moral language has connected and distinguished individual candidates and parties. This study will achieve this first through an extraction of moral language using automated text analysis by word counting, or "dictionary analysis", and finally through the construction of network models illustrating both inter- and intra-party trends.

This dissertation will contain five additional sections. First, a theoretical chapter will present a literature review of relevant concepts including rhetorical positioning on political campaigns, morality and political attitudes, moral foundations theory, moral language and political persuasion, and Twitter's role as a distributor of political rhetoric. This will be followed by a presentation of the conceptual framework used in this study — emphasizing the networked nature of campaign rhetoric — and a statement of the three research questions which will guide this work. The next chapter will rationalize the research strategy developed for this study, discuss the selection, tuning, and validation of a moral foundations dictionary, and outline the construction of text networks based on the extracted moral language. The following section will present the results as they relate to each research question proposed. Next, a discussion will suggest implications of the findings, address limitations, and advance questions for further study. Finally, the conclusion will summarize the results of the present research.

THEORETICAL CHAPTER

This chapter is organized as follows. First, a literature review will discuss and position the concepts that structure and motivate this research. The second section will include a statement of the conceptual framework used for this dissertation. Finally, the third section will include a statement of the three main research questions that guide this work.

Literature Review

The following section will identify and synthesize the concepts relating to political behavior and human moral reasoning which underpin this study. In spite of the myriad implications of their intersection, an investigation of existing research on these topics — including rhetorical positioning on political campaigns, morality and political attitudes, moral foundations theory, moral language and political persuasion, and the ascent of Twitter as a tool for the distribution of campaign discourse — reveals an absence of research examining the way moral language is used by campaigning politicians. The present research will be contextualized and positioned as a means to address this gap.

Rhetorical Positioning on Political Campaigns

Campaigns are, in their simplest form, an attempt to persuade a public audience. Arguments for the use of rhetoric in political audience communications can be traced to Aristotle, who proposed that when aiming to establish what is just and true in the face of a public jury or assembly, rhetorical devices are essential (Christof, 2010). To Aristotle, the ability to affect the decisions of a public audience necessarily depended on rhetorical choices, not only the dissemination of facts. Aristotle's conclusions paved the way for more recent scholarship examining the role played by rhetoric in contemporary political campaigning (for example: Håkansson, 1997; Strömbäck & Kiousis, 2014; Feld *et al.*, 2014).

Most relevant to the present study is research on rhetorical positioning in political marketing, which examines how candidates and issues are rhetorically "framed" through the use of specific language (Smith, 2005). Diverging from traditional positioning in political marketing — which often emphasizes candidate positioning based on policy positions (Baines *et al*., 1999) — *rhetorical* positioning on campaigns emphasizes the process of using specific language to select "those associations to be built upon and emphasised, and those associations to be removed or de-emphasised" (Aaker & Shansby, 1982, p. 56). Such research has found that the language used by candidates to argue for or against policies has a significant impact on the groups of people who support both the candidate and the policy, and consequently bears significantly on the outcomes of that election or legislative initiative (Smith, 2005). Illustrating the power of rhetorical positioning in politics, Lakoff (2002) points to the conservative description of the U.S. gun control debate — as an issue of individual liberty, rather than of public health and safety — and notes how this has effectively prevented any meaningful legislation on gun-related issues, even in the face of exponential increases in gun-related deaths.

Research by Jerit (2004) and Smith (2005) has suggested that rhetorical positioning on modern political campaigns is most often achieved through the use of emotional language. Their research aligns with Esser & Stromback (2014), who argue that emotional rhetoric is increasingly incentivized as a result of the colonization of the political sphere by media logics favoring emotional discourse that elicits wider, more prolonged attention from viewers. Further research by Brady *et al*. (2017) found that when moral-emotional language is present in political messages, the messages diffuse more rapidly through digital social networks, and receive more engagement. However, in spite of such research explicitly indicating the rising role of moral-emotional rhetoric in shaping political campaign discourse, no research has comprehensively examined or measured the use of such language by campaigning politicians along distinct moral or emotional dimensions. As this study will argue in the discussion, to address these current trends in rhetorical positioning on campaigns, further study is needed — not just of what politicians say, but of the moral-emotional language they invoke as they say it.

Morality and Political Attitudes

Political psychology has also emphasized the role of moral intuitions in the construction of political attitudes (Janoff-Bulman, 2009; Morgan, Skitka, & Wisneski, 2010), offering an explanation for why moral-emotional rhetoric might be especially effective in eliciting political reactions. In a significant contribution to the understanding these attitudes, Marietta (2008) found that when engaged in political reasoning, individuals tend to view politics either through a consequentialist or absolutist lens. Whereas consequentialist political reasoning involves cost-benefit analysis, compromise, and nuanced decision-making, absolutist reasoning involves "unwavering stances and the rejection of any form of compromise" (Feinberg & Willer, 2019, p. 2). This has proven to be critical in understanding the impact of moral attitudes on political reasoning: Tetlock *et al*. (2000) found that emphasizing moral sacredness of issues consistently engenders more extreme absolutist reasoning, and that these patterns of absolutist thinking lead to polarization, inducing more extreme political stances from individuals. This builds on research finding that individuals treat morally sacred issues differently from other issues, and tend to reject rational reasoning (such as cost-benefit calculations) in defense of their moral principles (Skitka & Mullen, 2002), and that judgments about moral priorities often rely upon emotional intuitions (Haidt, 2012). Similarly, Feinberg *et al*. (2012) conclude that when confronted by a moral issue, often the "gut reaction" determines whether the behavior is deemed moral or immoral, and this judgment forms the basis for future decisions.

An irrational, emotional relationship with moralized political issues takes on added significance when it is applied to partisan ideology: liberal and conservative individuals often hold divergent moral intuitions from one another, and issues found to be perfectly acceptable by liberals may be morally repugnant to conservatives, and vice versa (Haidt, 2012). For example, conservatives report stronger moral judgements regarding "sexual purity" (Haidt & Hersh, 2001), while liberals report stronger moral intuitions about protecting the environment (Feinberg & Willer, 2013), which often leads to divergent emotional responses to, for example, same-sex marriage and climate change. This evidence suggests that partisan polarization around issues is not purely a matter of policy disagreement and differing cost-benefit analysis, but is instead related to the irrational role played by individual moral constructions in the formation of political opinions. Existing research, however, fails to examine the manner in which political actors engage with these individual moral constructions through their rhetoric. The present research thus begins to address this gap by assessing the manner in which campaigning politicians of different parties diverge in their moralized rhetoric.

Moral Foundations Theory

Subsequent sections of this literature review will discuss the relationship between political behavior and moral language; overwhelmingly, this research has been operationalized using moral foundations theory, proposed by Haidt and Joseph (2004). The most significant contribution of moral foundations theory is its definition of specific moral dimensions, allowing for categorical measurement and comparison of moral inclinations. While previous studies of moral reasoning tended to emphasize only moral dimensions of "harm" and "fairness" — therefore leaving out much of what many individuals explicitly include in their moral reasoning — Haidt and Joseph (2004) found that human moral reasoning tends to take place along five moral dimensions, or "foundations": care/harm, fairness/reciprocity, ingroup/loyalty, authority/respect, and purity/sanctity. Their framework has proven able to illustrate how cultures and individuals place varying degrees of weight on specific moral foundations (Graham *et al*., 2013). Although some critics have argued that each "foundation" can ultimately still be distilled into a "care/harm" binary (Schein & Gray, 2017), repeated studies, as well as continual updates to the theory (Haidt & Graham, 2007; Haidt, 2012) have shown a "foundations" approach to be effective and useful for comparing psychological frameworks for moral reasoning (Graham *et al*., 2013). It has since become the most effective social psychological theory for illustrating the ways in which humans form moral judgments (Graham *et al*., 2011).

Although moral foundations theory was created to explore variations in moral reasoning across different cultures, subsequent scholarship has found it to be useful in studies of political ideology (Graham *et al*., 2009). Research has repeatedly found that liberal individuals are more responsive to moral foundations of care and fairness, framing their support for policies based on notions of compassion, nurturance and social equality (Haidt & Graham, 2007; Graham *et al.*, 2009). Conversely, the same research found that conservative individuals are more responsive to moral foundations of loyalty, authority, and sanctity, framing their support for policies based on notions of tradition, patriotism, and religious purity. The present research uses moral foundations theory as a framework for measuring moral campaign language, and as such it will be further discussed in section **2.2.**

Moral Language and Political Persuasion

Moral foundations theory has aided in the study of moral expression: research has found that the rhetorical strategies employed by both liberal and conservative individuals often reflect their differing moral convictions (Higgins & Lakoff, 1998), with liberals constructing their arguments using care and fairness language, and conservatives framing their arguments using sanctity, authority, and loyalty language (Feinberg & Willer, 2015). While potentially useful for motivating a target individual who shares the moral and political views of the individual, this framing is likely to be ineffective at persuading a citizen with differing views. In fact, exposure to arguments framed using unfamiliar or opposing moral frameworks results in "increased commitment to one's existing stance and greater animosity towards those on the other side" (Feinberg & Willer, 2019, p. 2), increasing levels of political polarization.

Conversely, Feinberg and Willer (2019) found that just as unfamiliar moral language can be polarizing, familiar moral language can be persuasive: their study on "moral re-framing"[1] illustrated that by arguing in favor of a partisan policy priority or political agenda using moral language most commonly endorsed by a political rival, political communicators may be able to increase bipartisan support without changing their policy positions. Voelkel and Willer (2019) argue that familiar moral language is persuasive because it gives skeptical audiences a chance to re-envision a policy based on moral reasoning frameworks to which they already subscribe, thereby transforming "positions that would otherwise seem morally wrong … into something morally acceptable or even desirable" (Feinberg & Willer, 2019, p. 2).

1 Moral re-framing is defined as a rhetorical persuasion technique in which "a position an individual would not normally support is framed in a way that it is consistent with that individual's moral values" (Feinberg & Willer, 2019: 2)

“Morally re-framed” rhetoric has been shown to persuade liberal and conservative voters in the U.S. to support positions associated with the opposing party. One study found, for example, that conservatives were more likely to express support for pro-environment legislation when confronted with a purity-based language arguing for climate protection—emphasizing how “dirty”, “disgusting”, and “impure” environmental degradation is — than they were when given a standard “liberal” argument emphasizing the danger and harm which can be caused by environmental destruction (Feinberg & Willer, 2013). Feinberg and Willer (2015) also showed similar results when examining liberal support for military spending: Democratic voters were more likely to support increased military spending when the role of the military in overcoming inequality was emphasized, whereas they offered less support when confronted with arguments emphasizing national loyalty and respect for authority. Finally, Bloemraad *et al*. (2016) found that conservatives were more likely to support pro-immigration measures when confronted by arguments which were centered on appeals to family unity and moral foundations of loyalty.

Specific use of moral language might also induce support for political candidates: a study by Voelkel and Feinberg (2018) offered groups of American participants two reasons why they should not support Donald Trump in his re-election bid: the first, emphasizing the moral foundation of fairness, was that Trump discriminates against minorities and stokes prejudice. The second, emphasizing the moral foundation of loyalty, was that Trump was disloyal to his country by dodging the draft during the Vietnam war. The study found that conservative voters were less likely to support Trump after being exposed to the second argument — which emphasized a typically conservative moral foundation of loyalty — while the first argument had no effect. Likewise, Voelkel and Willer (2019) found that a hypothetical progressive candidate who discussed liberal policies using conservative moral foundations was supported by much higher numbers of conservative survey respondents. However, while the persuasive, polarizing, and consequential effects of moral language have been identified in a lab setting — using hypothetical political arguments and hypothetical political candidates — these theories remain unexamined in actually existing campaigns. Feinberg and Willer (2015) found that when asked to write an argument that the opposing party would find persuasive, individuals engaged in small amounts of moral-reframing behavior instinctively, but it is unclear whether politicians are aware of the moral language they use at all, or make attempts to measure and analyze their own patterns of moral expression.

It is also worth noting that during a primary election, incentives for bipartisan appeal are paired with incentives for partisan appeal, as candidates most actively solicit votes from those in their own party. Still, there are a number of reasons why primary candidates might value bipartisan appeal, such as to distinguish themselves from their competitors or to appear "electable". A supplemental analysis in this study examines the shift in moral language used by primary victors between the primary and general election stages. It finds evidence that the post-primary rhetorical moderation hypothesis (Acree *et al*., 2018) exists — albeit mildly — along a moral dimension, with primary winners increasing their use of moral foundations associated with the opposing party as they orient themselves toward the general election; this study can be found in **Appendix I**.

Twitter and Political Campaigns in the U.S.

Of increasing interest in studies of modern campaigning is the role played by Twitter in the publication and distribution of candidate rhetoric (Conway *et al*., 2015). Of course, Twitter is not the only platform used on recent political campaigns: research has noted the degree to which platforms like Facebook retain prominence for their fundraising utility (Auter & Fine, 2017), and research has also shown that digital content strategies across numerous platforms have become an integral part of the contemporary political campaigns around the world (Dimitrova & Matthes, 2018; see also Lilleker *et al*., 2011; Koc-Michalska *et al*., 2016). Recent research convincingly suggests that campaigning on social media writ large improves electoral outcomes for individual candidates (Bright *et al*., 2019).

Still, recent studies have found that Twitter plays a uniquely large role in the distribution of candidate speech (Bode & Dalrymple, 2014; Jungherr, 2015). The prominence of Twitter as a tool for propagating political discourse during campaigns has been traced to its effectiveness as a broadcasting tool (Vergeer *et al*., 2011; LaMarre & Suzuki-Lambrecht, 2013; Kruikemeier, 2014) and its efficacy in influencing the frames used by journalists covering elections (Kreiss, 2014; Conway *et al.*, 2015). This effectiveness has resulted in Twitter playing a high-profile role in the distribution of candidate rhetoric during recent elections in the U.S. (Conway *et al.*, 2015). While research has addressed use of Twitter by presidential candidates as early as 2012 (LaMarre & Suzuki-Lambrecht, 2013; Kreiss, 2014), scholars have paid particular attention to its exponential rise in popularity since the prolific use of the platform by Donald Trump during the 2016 election (Stolee & Caton, 2018; Pain & Masullo Chen, 2019).

Research finds that recent U.S. political candidates from both parties use Twitter to vociferously defend and discuss substantive policy positions, fundraise, and organize their supporters[2] (Enli, 2017); popular debate has even evolved to ask whether Twitter has too much influence on election outcomes (Suciu, 2020). However, despite its newly prominent role in political campaigning, there is a paucity of research which takes advantage of the massive quantity of political campaign rhetoric made available by candidate Twitter use. Instead, existing research tends to restrict analysis to a very small number of candidates (Conway *et al.*, 2013) or a very short span of time, resulting in samples containing only fractions of the campaign discourse that actually exists (Adams & McCorkindale, 2013). The present research differentiates itself from much of the research on political rhetoric on Twitter by measuring the moral language used by 39 presidential candidates through the collection of a complete data set of all tweets published throughout each primary campaign.

Positioning the Present Research

Situating the contribution of the present research is difficult, as it intersects with many adjacent fields but separates itself from standard contributions to each in important ways. For example, it diverges from much research on political positioning by emphasizing *moral* positioning, rather than positioning through policy stances (Baines *et al.*, 1999). It diverges from political marketing management research by ignoring the extent to which campaigns were aware of the moral language they were using and the degree to which strategies were intentionally constructed using traditional marketing management frameworks (Henneberg,

2 A complete supplemental text network analysis of the various topics discussed on Twitter by candidates in the 2016 Republican and 2020 Democratic primaries can be found in Appendix II, displaying precisely what discursive categories were present and how they were variously connected to one another.

2009). It is also a non-traditional contribution in the field of rhetoric, as the issue frames constructed by the moral language are not examined, nor is context analyzed (Condor *et al.*, 2013). Moreover, while leveraging political psychological processes, this study does not attempt to draw conclusions about the impacts of moral language use on the outcomes of the elections examined in the data. Finally, it also diverges from rational choice models of political behavior (Petracca, 1991) as it instead understands voters from a behavioral-psychological perspective: as irrational, emotional, and influenced by imperfect cognition. And yet, even as this research diverges from traditional contributions in these fields, the linguistic trends measured and analyzed in this work hold clear implications for each. These implications will be outlined in the discussion.

In summary, this research aims to combine tools and findings from a multiplicity of fields, making empirical — and visible — for perhaps the first time, the linguistic connectivities between speakers and groups of speakers, as well as their spatialized position within a selected body of discourse. Because of its focus on the moral language used by political candidates, this research primarily relates to rhetorical positioning on campaigns. However, it can secondarily contribute to other fields by offering methodological approaches for the analysis of text which might augment existing research practices.

Statement of Conceptual Framework

This research further diverges from traditional approaches to the study of rhetoric and positioning in political marketing by making explicit a phenomenon usually left inferred: namely, that rhetorical choices made by candidates during campaigns can be interpreted as network inputs (even the very nature of the word "positioning" implies a spatial relationship between candidates). It is the nature of competitive primary elections that contrast between candidates is emphasized; candidates compete contemporaneously and engage in a continuous and connected discourse. In this way, it is not what a candidate says that matters, it is what a candidate says *given what the other candidates are saying.* To truly understand the role played by moral rhetoric in a primary campaign, one must therefore understand how use of moral language connects or distinguishes individual candidates within a party.

This research finds concert with scholars such as Hart (2009) and Rule *et al*. (2018), who argue that by choosing to use some words and not to use others, speakers create a sociolinguistic map that can be reconstructed and analyzed. Viewing the rhetorical dimension of a primary election as a networked structure helps make visible the ways in which rhetoric acts both as a differentiator and as a binding agent, offering insights into party ideology, candidate positioning, and political outcomes. This study will create those maps as networks of digital political speech, illustrating how moral words chosen by individual candidates connect to form party trends and ideological norms, isolating some candidates and clustering others in the process.

The framework provided by moral foundations theory (Haidt & Joseph, 2004) offers a well-established means through which moral expression can be organized and understood. As such, moral foundations theory is the operative concept that will be used to understand human moral reasoning in this study. Moral foundations theory contains five moral foundations. In this study, they will be referred to as care, fairness, authority, loyalty, and sanctity. Each contains a positive and a negative valence. **Figure 1** provides a visualization of the complete framework.

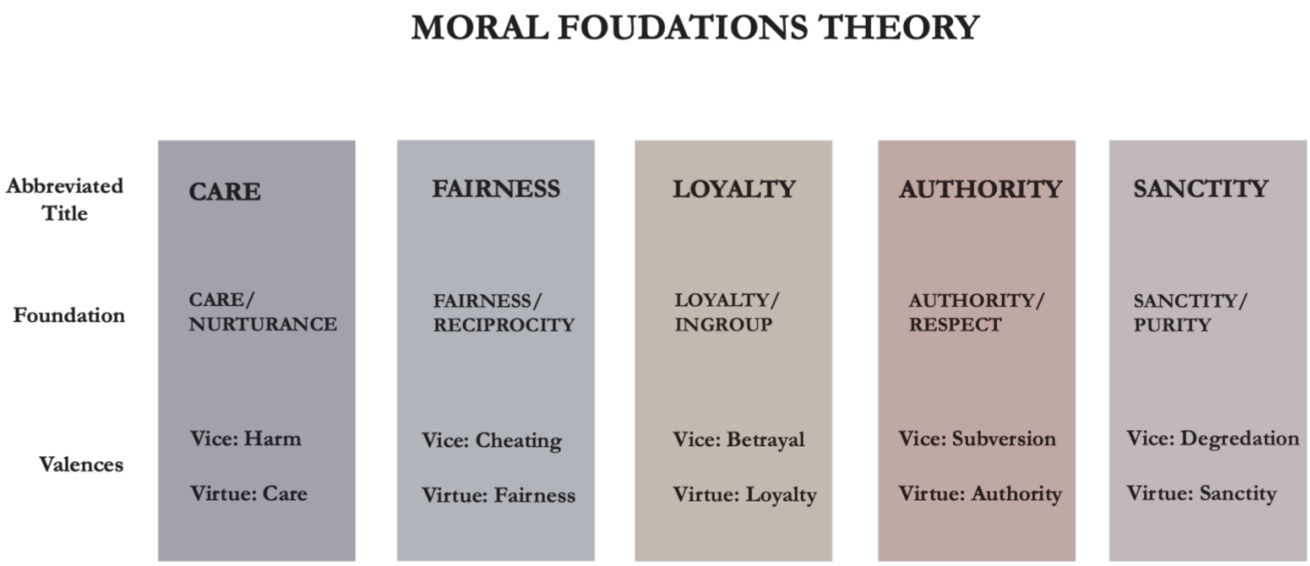


Figure 1 Visualization of the moral foundations theory framework as proposed by Haidt and Joseph (2004)

Research Questions

The present research is structured around the following question: to what extent has the use of moral language served to connect or differentiate political candidates and political parties during recent presidential elections in the U.S.? This question was operationalized through an evaluation of three formal research sub-questions, addressing different aspects of the broader research aim.

The first question addresses a substantial gap in research: while Graham *et al*. (2009) and Feinberg & Willer (2015) found that liberal and conservative individuals were likely to use differing moral language to frame their policy positions, the presence of this partisan divide has never been assessed in the language of campaigning Democratic and Republican politicians. Therefore, the first research question aims to assess whether the differences in moral reasoning found in liberal and conservative individuals — as measured by Graham *et al*. (2009) and Feinberg & Willer (2015) — are reflected in the rhetoric of recent U.S. presidential candidates:

RQ1: To what extent — and along which moral dimensions — do Democratic and Republican presidential candidates tend to diverge in their use of moral language?

The second research question aims to advance the analysis by measuring and visualizing the extent to which competing candidates within the same party discuss the same moral foundations in the same ways, and how this similarity variously binds them together:

RQ2: How are competing candidates within the same party connected to one another through their similar use of individual moral foundations?

The final question aims to assess the ways in which deviation from the moral-rhetorical norms of a party — with an eye toward the persuasive strategic incentives that may underly moral reframing tactics articulated by Feinberg and Willer (2019) — may result in the rhetorical isolation of individual candidates:

RQ3: To what extent did candidates deviate from the moral-rhetorical norms of their party by using language from moral foundations associated with an opposing ideology, and to what extent did this deviation co-occur with an isolated network position?

RESEARCH DESIGN & METHODOLOGY

This chapter contains three sections. The first section will rationalize the chosen methodological approach and situate each implemented methodology in the relevant literature. The second section will describe methodological procedure as it was implemented in this research, detailing both dictionary application and the construction of the networks. The third section will conclude with a brief discussion of research ethics.

Methodological Background

This section will first outline the methodological rationale for this research before discussing and evaluating each of the two selected methodologies: dictionary analysis and text network analysis. This section will conclude with a note about quantitative and qualitative hybridity.

Methodological Rationale

The dual nature of the research questions outlined in this paper — concerned with both the comprehensive measurement of moral language in a large body of political text and the evaluation of its role as a connective agent in campaign discourse — lends itself to methodological pluralism. As a result, this paper employs a mixed-methods approach, aiming to first extract moral language from candidate rhetoric and then effectively visualize how that extracted language connects individual candidates. The former was accomplished using automated text analysis by word counting, or "dictionary analysis", and the latter was accomplished through a novel implementation of text network analysis methods. Both of these methods were selected out of necessity: traditional research methods in communications and political science were insufficient in the context of this research, both in terms of their ability to address the abundance of textual data available and their ability elucidate granular linguistic trends. For further discussion rationalizing the rejection of traditional methods, see **Appendix III**.

The selected methods belong to a family of methods which treat text as data. The main difference between traditional methods like quantitative content analysis (Krippendorff, 1980) and discourse analysis (Fairclough, 2015) and "text as data" methods can be summarized as follows: while content and discourse analysis treat text as something to be read, digested and summarized, "text as data" methods treat text as data to be processed and analysed using quantitive tools, without necessarily being read at all (Benoit, 2020). These methods enable the processing of vastly larger amounts of data, and can offer more comprehensive and computationally intensive analysis. Of course, these methods face limitations as well, most often relating to the fact that a degree of distance is necessarily created between the researcher and their texts, at some stages minimizing human interpretation of speech and making analysis of context difficult (the specific limitations for the methodologies chosen for this research will be described in more detail in the following sections). Nonetheless, "text as data" analysis can yield insights into trends and patterns in political rhetoric not observable by any other method, and constitute the most effective approach for this research. The following sections will explain and contextualize the two "text as data" methodologies selected and further rationalize their use.

Dictionary Analysis

Perhaps the largest difficulty in examining naturally-occurring moral language at the word level is that it is extremely difficult to measure at scale. The "gold-standard" for such analysis is manual human coding (Wang & Inbar, 2020); however, for corpora consisting of hundreds of thousands, or even millions of documents, human coding becomes an impossibility. Automated text analysis by word counting — or "dictionary" methods — can offer a powerful solution (Tausczik & Pennebaker, 2009). In their simplest form, these dictionary methods involve the collection of a group of words whose meaning is determined to be representative of a latent quality of interest (Benoit, 2020). An application of the dictionary would involve the counting of these words across texts of interest, potentially yielding insights about the relative sentiment of a document. Most usefully, these dictionaries can be applied to massive amounts of textual data with relative ease, exceeding what even the most determined human researcher could assess.

However, while efficient, dictionary methods are also limited in important ways. For example, even an extremely well-defined dictionary will struggle to capture a concept in all possible contexts. No researcher can be familiar with all possible idiolects and sociolects; language use changes depending on age group, ethnic group, socioeconomic class, and many other covariates (Louwerse, 2004). This means that each dictionary is in some way biased towards the speech patterns and epistemic positionality of the researcher(s). Another issue is temporality: the rapid shift and evolution of natural language means that both the specific words associated with a concept and the very meanings of the words themselves evolve. This means that a dictionary that worked well at one point might fail entirely when applied years later (Garten *et al.*, 2017).

In addition, the interdependence of language means that a simple list of words can only cover narrow aspects of a concept without introducing error: the same word can, for example, imply positive or negative sentiment depending on the context, making its inclusion in a discrete dictionary key based just one of its meanings occasionally problematic (Wang & Inbar, 2020). For example, a "long" line at the grocery store is a hassle, but a "long" life is desirable. A dictionary is unable to distinguish between these different word senses. This problem expands when one considers polysemy, or words which have multiple meanings (Benoit, 2020): "a cold beer is good, but a cold therapist is probably best avoided".

In the present research, these limitations were assessed and mitigated through a series of methodological choices, customized filtering parameters, and validation steps. These processes are outlined in detail in sections **3.2.3** & **3.2.4**.

Text Network Analysis

In addition to a dictionary analysis, this paper will employ a novel methodology combining the fields of network analysis and natural language processing to construct text networks displaying spatial relationships between political candidates based on their use of moral language. Although it has long been understood that language is inherently networked (Lamb & Newell, 1966), the construction of language networks at the word level from large corpora was previously infeasible. Recently, however, such networks have become an area of interest, especially in corpus linguistic studies (Mehler, 2008). A note situating text network analysis within the broader field of network analysis can be found in **Appendix III.**

Text networks have proven effective in leveraging new digital data sources, and have incentivized the increasing application of mixed-methods approaches which channel text data analytics into networked structures to be analyzed using network analysis metrics (Light, 2014). These methods combining natural language processing and network analysis have been used in social science research to examine how advocacy organizations stimulate conversations on social media (Bail, 2016), and to visualize semantic relationships between philosophical ideas (Drieger, 2013). Text networks have also successfully been used in political communications research, most notably in an analysis of lexical shifts in U.S. State of the Union addresses (Rule *et al*., 2015). The application of text network analysis in this study thus aims to contribute to a nascent body of research using networks to map individual word relationships, the position of words within discursive categories of political speech, the relationship between words and political communicators, and the rhetorical relationships between the communicators themselves.

Methods of network analysis also face their share of limitations: their form is often determined by layout algorithms, which are typically difficult to adjust to the users specifications. Furthermore, these layout algorithms are stochastic, meaning that they are constructed using probability distributions. This means that re-spatializing the same network with the same algorithm can produce slightly different results, making interpretation difficult (Krzywinski *et al*., 2011). This also makes it difficult to compare two different networks, even if they were both spatialized using the same algorithm. Finally, network layouts are often extremely sensitive, meaning that the removal of single "edges" or "nodes" can cause meaningful shifts in the overall layout (Krzywinski *et al*., 2011). Explicit guidance for interpretation of all networks presented in this study is offered in section **4.1**. However, in spite of their limitations, methods of text network analysis offer a significant advantage in the context of the present research: the ability to examine rhetorical trends, patterns, and relationships through an understanding of political language as inherently networked data.

Quantitative and Qualitative Hybridity

It is important to note that both dictionary and network analysis methods are "hybrid" in their approach: while computationally intensive, they also contain a significant interpretive dimension. Dictionary methods, while entirely mechanistic in their computation of word and pattern frequencies, require human judgment from the dictionary author, who must qualitatively determine which words will be representative of the latent meaning being measured. Even after the application of the dictionary to the corpus of interest, human validation and "tuning" of the dictionary is essential. Benoit (2020) notes that it is "only through a careful, qualitative process of inspection of the word matches in context that adjustments can be made to a dictionary and the results can be trusted as valid" (p. 16). In the present study, these tuning decisions are detailed at length in sections **3.2.5** and **3.2.6.**

Methods of text network analysis, on the other hand, are entirely quantitive in their construction: matrices of word co-occurrences are computed, edge lists are induced, and the resulting networks are spatialized according to fully automatic algorithms. However, the meaning of the networks themselves are not always obvious, and the burden is thus placed on the researcher to qualitatively draw out meaning from them. The potential for these text network approaches, therefore, rests squarely "at the nexus of new computational methods and in-depth, qualitative strategies" (Light, 2014). The present research will report and discuss findings accordingly, allowing for a researcher-led, qualitative discussion of the findings alongside empirical analysis.

Methodological Procedure

This section will contain all methodological steps undertaken for the present research. These include corpus selection, corpus construction, dictionary selection, validation, *tf-idf* weighting, filtering and lexical extraction, network construction, and research ethics.

Corpus Selection

Of interest for this research were all tweets published by presidential candidates during the 2016 and 2020 U.S. presidential primaries. In total, 10 Democratic and 17 Republican candidates ran for the their respective party nomination in 2016, and 29 Democrats challenged

Trump during the 2020 Democratic primary. However, in an effort to filter the dataset to campaigns whose rhetoric was likely more substantial, developed, and relevant, candidates were included only if they participated in at least two official primary debates hosted by their national party (either the DNC or the RNC).

As a result of this filtering, 14 candidates were eliminated from the dataset. In total, 3 Democratic and 15 Republican campaigns were included during the 2016 election cycle, and 21 Democratic campaigns were included for the 2020 cycle. Altogether, 39 unique campaigns were assessed, including 24 Democratic campaigns and 15 Republican campaigns spanning the course of the two most recent presidential elections.

Corpus Construction

The complete dataset of tweets published by the campaign account of each of the 39 candidates was collected using Twitter's Academic v2 API endpoints, starting from the day of campaign announcement until the day of campaign suspension[3] for both 2016 and 2020 presidential elections ($N = 139{,}401$). Tweet collection was done in R using the `academictwitteR` package (Barrie & Chun-ting Ho, 2021).

3 In the case of 2016 primary winners Donald Trump and Hillary Clinton and 2020 primary winner Joe Biden, the date on which they became the presumptive nominee was used as the effective end date of the campaign, as from this point their campaign rhetoric may have shifted as they oriented themselves toward the general election; see Appendix I.

Primary	Candidates	Avg. Campaign Length (days)	Tweets	Tweets per Campaign per Day
2016 Republican	15	244	40,607	10
2016 Democratic	3	350	15,520	14
2020 Democratic	21	301	83,274	12
Average	13	298	46,467	12
Total	39	-	139,412	-

Table 1 Summary statistics for the full corpus of candidate primary tweets, including the average number of tweets per campaign per day.

Tweets for each candidate were then concatenated and pasted into a plain text document. All tweets were cleaned through the removal of hashtags, Twitter handles, emojis, and punctuation; all characters were converted to lowercase (Denny & Spirling, 2018). At this stage, any candidate who tweeted 1.5 standard deviations less than the average candidate in their party and election cycle was eliminated as an outlier: this resulted in the elimination of 1 Republican and 2 Democratic candidates. All further data wrangling was done in R using the `quanteda` package for textual analysis (Benoit *et al.*, 2018).

Dictionary Selection

In order to extract and measure use of moral language in candidate tweets, the Moral Foundations Dictionary (MFD) 2.0 (Frimer *et al.*, 2017) was implemented. There are a number of dictionaries constructed specifically for the measurement of moral language in bodies of text: notable others include the original MFD (Graham *et al.*, 2009), the DDR MFD (Garten *et* *al*., 2017), and the eMFD (Hopp *et al*., 2020). However, the MFD 2.0 provided a number of distinct advantages over its peers.

First, it is one of the newest available, and is itself a robust update to a previous dictionary, the original MFD (Graham *et al*., 2009). This update serves to mitigate the temporal concerns associated with changing sociolects as outlined previously. Perhaps most importantly, this dictionary has been used and validated extensively on Twitter (Dehghani *et al*. 2016; Brady *et al*. 2017; Garten *et al*. 2018; Hoover *et al*. 2018; Mooijman *et al*. 2018). The MFD 2.0 has also been validated successfully beyond Twitter data (Graham *et al*, 2009; Clifford & Jerit, 2013; Fulgoni *et al*., 2016; Leidner & Castano 2012; Sagi & Dehghani 2014; Lewis *et al*. 2017; Weber *et al*. 2018; Long & Eveland 2018; Wheeler *et al*., 2019). The extensive testing mitigates concerns related to polysemy, word-sense disambiguation[4], and breadth.

The MFD 2.0 still faces limitations: terms can only be included in a single moral category, when in reality they might have an affiliation with more than one: for example, "worship", included in the "sanctity" category, might also have connotations of respect and submission, which might indicate an additional association with the "authority" moral foundation. A more advanced dictionary might allow for a word-to-moral-foundation contribution score (where "worship" might have a "sanctity" contribution score of 0.8, and an "authority" contribution score of 0.2). Another drawback is the nature of the dictionary construction: while still standard practice and considered by many to be the gold standard, the terms in the dictionary were generated qualitatively by a group of academics and psychologists. This inevitably biases the terms included towards the dialects and epistemic positionality of the researchers, who were certainly not a representative demographic sample of language users. A more advanced dictionary would likely use crowd-sourcing methods for the generation of the terms (Hopp *et al*., 2020) or distributional methods using massive global corpora and only a handful of qualitatively selected "seed" terms (Garten *et al*., 2017).

4 Moreover, entries in the dictionary are not stemmed, meaning that specific word inflections are separate dictionary entries. This improves accuracy, as in a stemmed dictionary, "happ*" (where "*" is a wildcard character) might return "happy" and "happiness" but would also include "happen" and "happenstance".

Still, the extensive validation and track record of successful applications to Twitter discourse made the MFD 2.0 the optimal choice for this project. In total, the dictionary contains 2,233 unique words across each of five moral foundations (care, fairness, loyalty, authority, and sanctity), with an average of 420 words per category. Each foundation is also sub-divided, containing a positive and negative valence category for each of the five moral foundations (care/harm, fairness/cheating, loyalty/betrayal, authority/subversion and sanctity/degradation). For the full list of dictionary terms, see Frimer *et al*. (2017).

Validation

Performance of a dictionary on a new domain and a new dataset is not guaranteed, making validation essential (Grimmer & Stewart, 2013). However, validation for dictionary methods is a challenge: the granularity of their outputs is such that human coders are unable to produce the same measures reliably (Krosnick *et al*., 1999). This means that it is "essentially impossible to derive gold-standard evaluations of dictionaries based on human coding of documents" (Grimmer & Stewart, 2013: 275).

Consequently, to initially validate both the functionality of dictionary and the robustness and consistency of the data set, the MFD 2.0 was applied to the candidate tweets unaltered and the distribution was assessed in terms of its congruence with Zipf's law. Zipf's law states that in a body of naturally occurring language, the most frequently used term will be used twice as frequently as the second-most used term, three times as often as the third-most used term, etc. In other words, in natural language, the term rank-frequency distribution is an inverse relationship (Powers, 1998).

An assessment of the data's congruence with Zipf's law therefore serves two purposes. First, it offers a means of assessing whether the dictionary succeeded in capturing a balanced and holistic selection of naturally occurring moral language: in a corpus of millions of words, and a dictionary containing thousands of words, the relationship predicted by Zipf's law should apply. Second, it serves to assess whether the individual candidate distributions of moral language were were consistent and robust, free from outlier values. For example, if certain candidates used automated social media tools which tweeted the same message over and over, this would be reflected in their rank-frequency distribution.

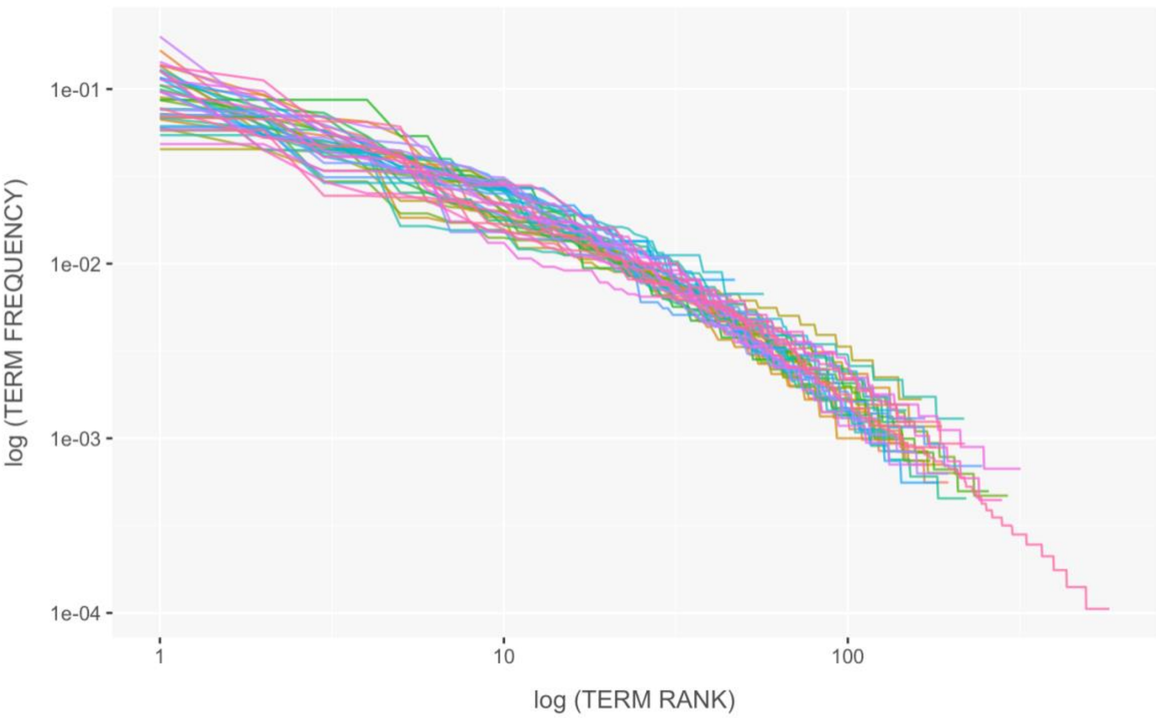


Figure 2 Rank-frequency distribution of the moral terms by candidate. Terms were extracted from over 139,000 tweets from 39 U.S. presidential campaigns. Both axis are scaled logarithmically, making a clear congruence with Zipf's law easily visible across all candidates, 2016-2020.

On a linear scale, Zipf's law takes the form of a power law distribution, but through a log transformation, the relationship becomes negative and linear. The log transformed plot shown in **Figure 2** clearly displays a negative linear relationship, confirming both that the dictionary is complete enough to be working effectively (capturing a representative and "natural" sample of language), and that distributions of moral language used by each candidate were robust and consistent across both political party, election phase, and election year.

TF-IDF Weighting

To more precisely facilitate the application of the dictionary to the specific domain and data set relevant for this research, and to most effectively address the stated research questions, the MFD 2.0 was filtered and customized through a series of tuning steps. The goal of this tuning process was to remove terms in the dictionary unlikely to be informative about partisan deviations in use of moral language because of their frequent and consistent use across all candidates — in other words, language that is simply "par for the course" in U.S. presidential election campaigning. The removal of this generic language allows for the isolation of moral terms that are used by some candidates, but not others. This aids in more accurately addressing *RQ1*, as it can then be clearly assessed whether any "partisan" moral language appears to be randomly distributed across moral foundations and candidates, or if partisanship seems to correspond with deviations towards particular moral foundations.

In order to identify language used in high proportion by both Republican and Democratic candidates, a global document-feature matrix was created[5], where both Democratic and Republican candidates were represented by a document containing moral language extracted using the un-altered MFD 2.0. A minimum term frequency threshold was then applied to every document in the matrix, eliminating from each one the terms not used at least three times[6]. This document-feature matrix was then weighted according to a term frequency-inverse document frequency (*tf-idf*) weighting scheme, often used in natural language

5 Iterative tests concluded that as long as candidates were upsampled such that half were Democrat and half were Republican, the exact candidates added to the global document feature matrix did not have an impact on the moral terms that were ultimately removed during the filtering process explained in 3.2.6. More detail and validation tables are available in Appendix IV.

6 This step makes the tf-idf weighting more robust: if a word was used 50 times by one candidate, but a single time by 15 others, a tf-idf weighing scheme would give this term a document frequency score that is fairly high. Setting the threshold at three occurrences means that document frequency will only be calculated based on a more robust measure of repeated word uses.

processing and information science to identify meaningful terms within documents (Qaiser & Ali, 2018). The weighting was applied using the formula

$$w_{i,j} = tf_{i,j} \times log(\frac{N}{df_i})$$

where tf is the number of occurrences of term i in document j, and df is the number of documents containing term i, and N is the total number of documents. This weighting scheme has the effect of calculating new frequency scores for each term in the document-feature matrix, where words are down-weighted if they occur with high frequency across many documents and up-weighted if they occur unevenly across documents. *Tf-idf* weighting schemes also value raw frequency: in this context, that means the (often higher-profile) candidates who maintained a campaign over a longer period have more weight in determining what constitutes "Democratic" or "Republican" language; candidates with short, non-communicative campaigns will have their words weighted less. This is useful, as it means that lower tier candidates, while included, will not be able to outweigh the "mainstream" political discourse of a given election cycle. All terms in the dictionary were then rank-ordered by their newly computed *tf-idf* frequency. Terms ranked near the bottom of this list can be assessed as the least differentiating amongst candidates, and terms near the top are the most differentiating. **Figure 3** shows the distribution of *tf-idf* frequency scores across all moral terms.

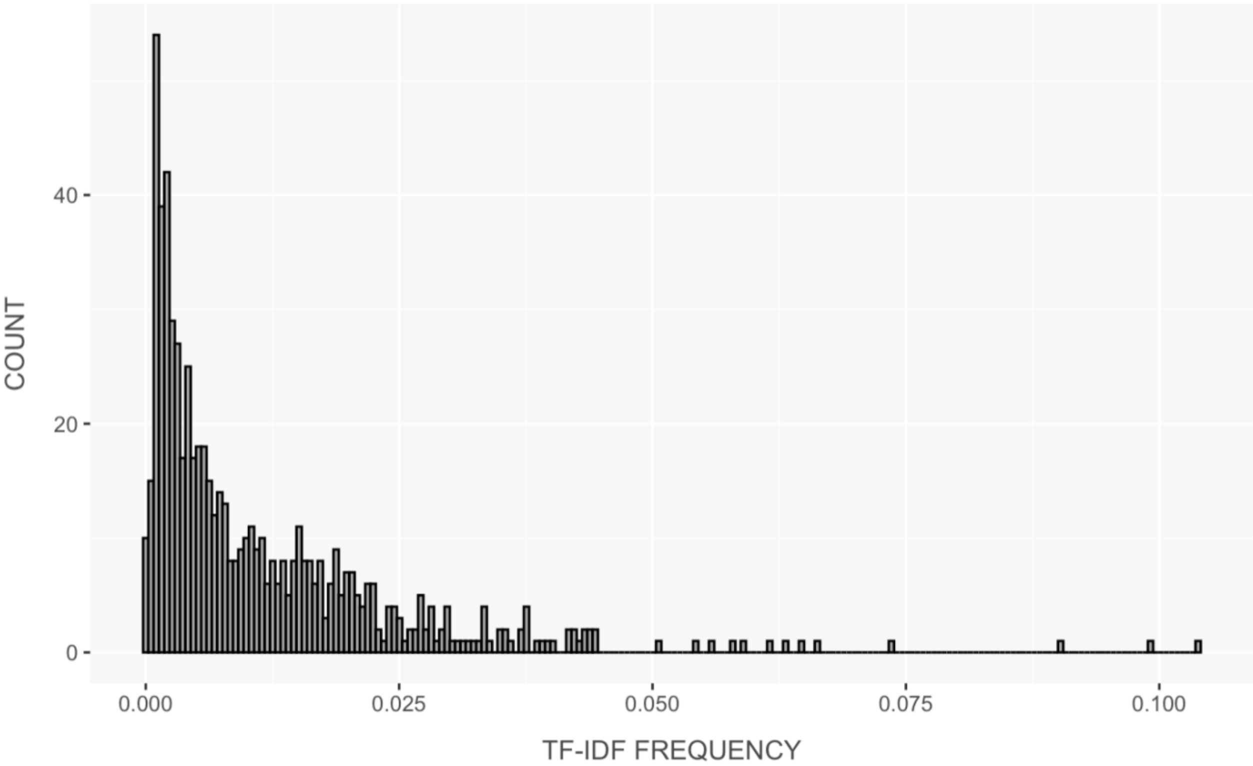


Figure 3 Distribution of 657 moral terms based on their weighted *tf-idf* frequency score. Terms with a lower *tf-idf* frequency score occur often and evenly across all candidates; terms with a higher *tf-idf* frequency score occur less often and more unevenly across candidates.

Filtering & Lexical Extraction

In order to determine where to set the filtering threshold for "generic" language, Zipf's law was one again implemented. **Figure 2**, used earlier to validate the dictionary, was generated again, this time using the new *tf-idf* frequencies. The result in **Figure 4** shows an intuitive threshold where *tf-idf* term frequency rapidly drops off and falls to 0.

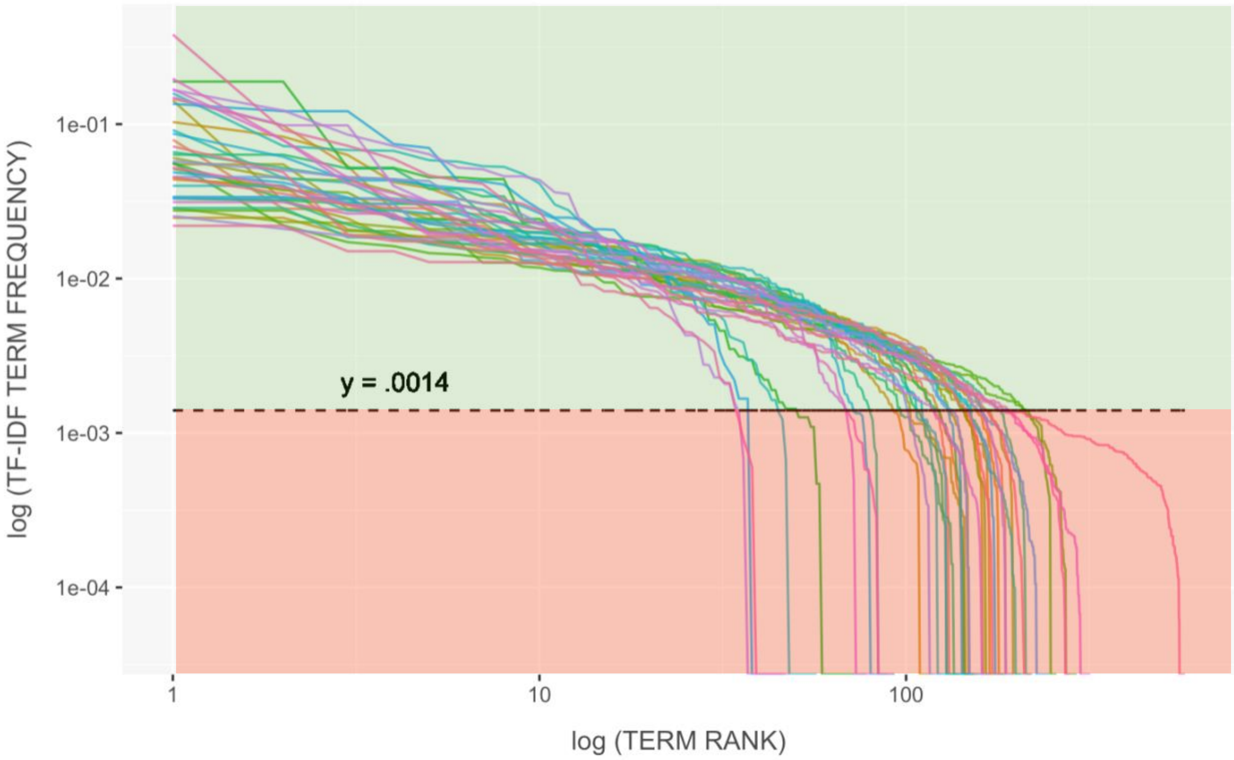


Figure 4 Rank-frequency distribution of moral terms by candidate, weighted by *tf-idf*, and scaled logarithmically. The resulting graph shows a clear threshold where *tf-idf* frequency begins to fall off rapidly across candidates.

A *tf-idf* frequency score of .0014 was intuitively selected as the cutoff point for the terms in the document feature matrix. 83 terms with a tf-idf frequency of less than .0014 — the least informative words — were eliminated. The very least informative of these were "president", "presidential", "family", "love", "country", "leadership", "leaders", "protect", and "fight". The remaining 574 terms — representing not just the moral language used by each candidate, but the *meaningful, non-generic* moral language used by each candidate — were used to construct a new dictionary and were implemented in all subsequent analysis.

In total, the weighting, filtering, and tuning of the dictionary can be summarized through the following steps:

- Out of 1,167 distinct moral terms used by the 42 candidates, 510 (43%) were filtered out by the application of a minimum term frequency threshold of 3, meaning that they were never used by any candidate more than 3 times.
- Out of the remaining 657 terms, 83 (14%) were eliminated as "generic" language by the *tf-idf* weighting process.
- The final custom dictionary consisted of 574 distinct moral terms, and be found in **Appendix VI.** An examination of the effect of the weighting process, including a comparison of pre- and post-weighting proportions of moral language used by each candidate, can also be found in **Appendix VII.**

Network Construction

Two different types of networks were constructed for this analysis. One type further addresses *RQ1* & *RQ3*, and aims to reveal both the community structure of partisan moral discourse & the spatial relationships between individual Democratic and Republican candidates. The second type of network aims to address *RQ2*, and therefore attempts to illustrate how individual candidates within the same party are connected to each other based on their similar use of individual moral foundations. The networks were constructed as follows:

Network Type A: Two-mode network connecting Democratic and Republican candidates to moral language they used on Twitter during their campaigns. These networks were constructed using a combination of R, Gephi[7] and Cortext Manager[8] according to the following steps:

I. The initial incidence matrix M was defined by the number of times moral term i appeared in document of aggregated candidate tweets j.
II. A weighted network was thus constructed such that every time a candidate used a moral word, an edge was drawn between that candidate node and a node representing that moral term. Candidate nodes were never directly linked; their connectedness only occurred through use of the same moral term. Edges in the resulting network were undirected.

7 Gephi is an open-source tool for network analysis (Bastian *et al*., 2009).

8 CorText Manager (https://www.cortext.net) is an online platform built for natural language processing tasks.

III. The network was spatialized with the force-directed Yifan Hu layout (Hu, 2005), which has shown to be especially effective in visualizing smaller bipartite networks.

IV. A Louvain community detection algorithm (Blondel *et al.*, 2008) was then applied to the network to identify clusters of candidates.

Network Type B: A one-mode network connecting candidates to each other through their similar use of individual moral foundations. This type of network was constructed as follows:

I. For each candidate, five documents were created, with one containing the total extracted moral language for each moral foundation. (For example, Biden care, Biden fairness, Biden sanctity, etc.)

II. The pairwise cosine similarity was then calculated for each pair of candidates over each of the five moral foundations subcategory (For example, Biden care and Sanders care, Trump authority & Cruz authority). This process yielded five cosine similarity scores for each combination of two candidates. Cosine similarity was calculated with the vector notation formula:

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\| \mathbf{A} \| \| \mathbf{B} \|}$$

Cosine similarity was selected over other document similarity measures (e.g. Euclidean distance) because it computes similarity based on the proportion of similar words and word frequencies. Other document similarity measures (e.g. Euclidean distance) would have been inappropriately influenced by document length, giving higher similarity scores to candidates who used similar raw quantities of moral language. Given the uneven campaign lengths and tweeting habits across candidates, this was not desirable. Additional validation measures taken to ensure consistency and robustness at this stage can be found in **Appendix VIII**.

III. The resulting pairwise cosine similarity scores were then re-interpreted as weighted edges and used to construct a perfectly dense network, with each candidate connected to every other candidate by five parallel edges of varying weights (with each edge representing a different moral foundation).

IV. Edges in the network were then filtered such that only edges exceeding a given weight threshold θ were conserved. This threshold was computed based on two criteria: the cohesiveness of the final network and the total number of edges. Specifically, the lowest possible θ was selected which produced a network with no disconnected components using the fewest total number of edges. This filtering also allowed for a visual representation of only the most significant inter-candidate relationships, and aided in subsequent analysis by rendering a cleaner network which most clearly displayed the most relevant network structures.

V. Edges were colored by the moral foundation they represented; edges weights were re-scaled to more effectively visualize contrast. Node size was scaled with betweenness centrality.

VI. The networks were spatialized with the force-directed Yifan Hu layout (Hu, 2005).

These methods of network construction are novel, and constructed specifically for the present research. As a result, further visualization of their construction is helpful. **Figure 5** provides a helpful visual aid and provides easy-to-follow examples showing the construction process for both networks.

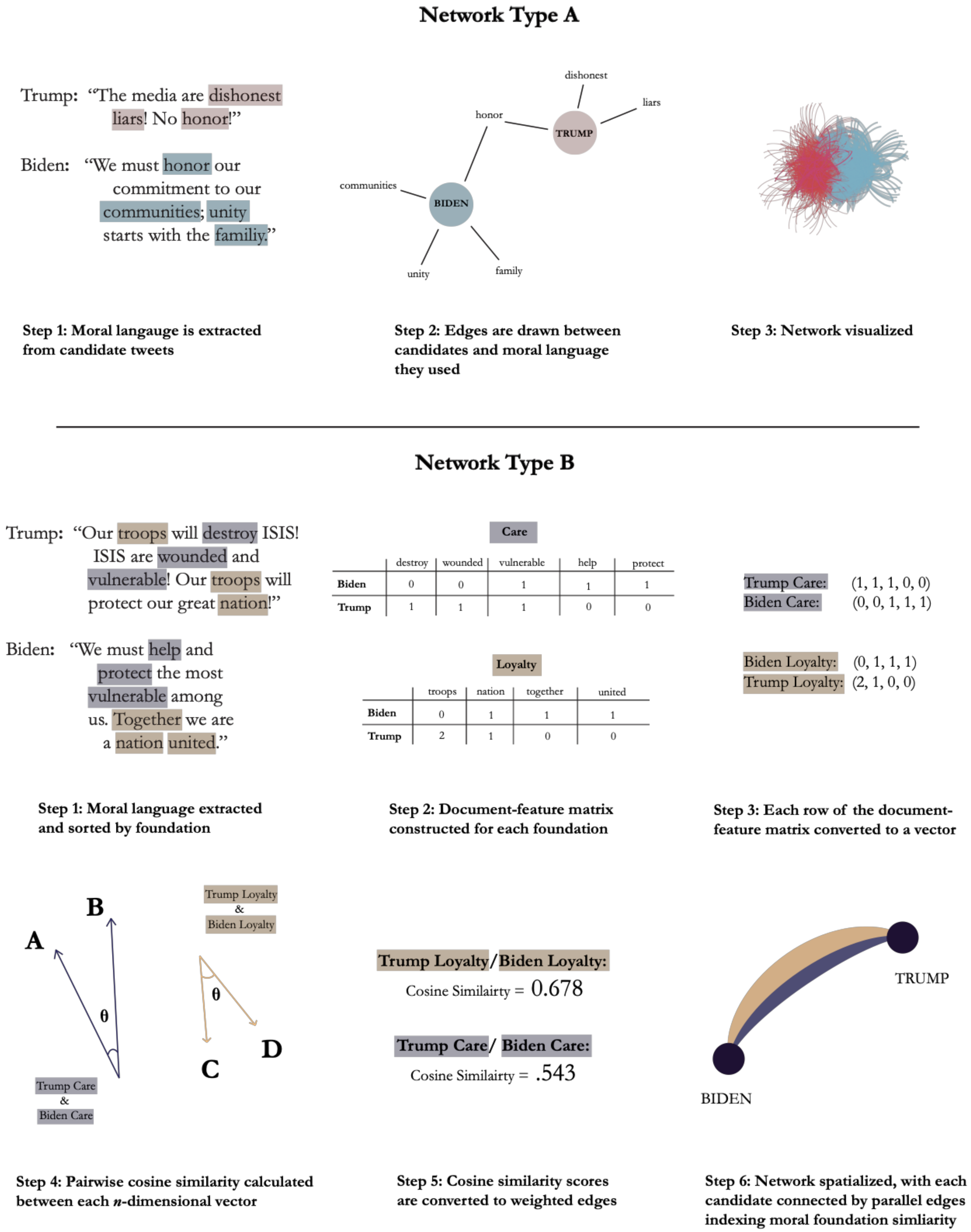


Figure 5 The stages of construction for each network type; illustrations are examples only

Research Ethics

Conducting research using Twitter data poses unique ethical concerns, most often relating to the privacy of individual Twitter users. Tweets were not anonymized in this research, as in their guide to conducting ethical social media research, Townsend & Wallace (2016) emphasize that public figures including politicians constitute "an exception" to privacy and anonymity rules, as they are attempting to reach as wide an audience as possible and thus "aiming for broad readership" (Townsend & Wallace, 2016: 8). These issues are further mitigated by the fact that in all cases the Twitter data used in this analysis is presented in aggregated from.

RESULTS & INTERPRETATION

This chapter, structured in four sections, will present the results of the present research. The first section will offer some guidance for how the results of this study might be interpreted; the proceeding sections will then present the results pertaining to each research question in turn. Wherever noted, additional information can be found in the appendices.

Interpretation

Networks can be difficult to interpret; therefore, explicit guidance regarding precisely what can and cannot be surmised from them is useful in facilitating an informed presentation of the findings. Due to the stochastic nature of force-directed layout algorithms, the precise position of candidate nodes cannot be easily compared across different networks. The spatialized positions of nodes in a network also cannot be interpreted as a precise mathematical spectrum: the layouts use physics-based algorithms which introduce small amounts random variation with each layout. However, all structural properties of networks, such as centrality, betweenness, degree, and modularity, will not change regardless of the layout algorithm applied, nor will they shift across repeated network spatializations. In this analysis, any comparisons made between different networks will thus be validated using the aforementioned network statistics, and not conducted based solely a visual assessment of the network structure.

Within a contained network layout, candidates with proximate spatial positions can be interpreted as having used more similar moral words at more similar frequencies on their campaigns; more distant candidates can likewise be interpreted as having used increasingly different moral words, and at increasingly differing frequencies. The macro-trends present in spatialized candidate positions are robust to different layout algorithms and filtering parameters. Conclusions drawn from the networks constructed for this analysis are also especially robust as they were constructed based on a comprehensive and complete dataset of candidate tweets, and not a sample.

Finally, interpretation of the results is enhanced by an understanding of the language associated with each moral foundation: **Figure 6** shows the most used moral terms by all candidates in the sample. It was constructed using the entire data set, using a combination of moral language espoused by both Democrats and Republicans, and serves as an explicit visualization of the most significant political words associated with each moral category. The full list of dictionary terms can be found in **Appendix VI**.

Figure 6 Word clouds of the extracted moral language most used by 39 U.S. presidential candidates on Twitter, 2016-2020. Word clouds generated from a corpus of 139,412 tweets; word size scaled with the square root of its occurrences.

Partisan Divides in Moral Expression

An evaluation of RQ1 requires the separate examination of two elements: first, the extent to which Democratic and Republican presidential candidates tended to diverge in their use of moral language on Twitter, and second, the moral dimensions along which any such divergence took place.

To examine the first element — the extent to which Democratic and Republican presidential candidates actually diverged in their use of moral language online (or whether they even diverged at all) — two text networks were constructed from the data set. One network was constructed using the moral language generated by the 3 Democratic and 14 Republican candidates who competed during the 2016 election. The second network was constructed using all candidates from both 2016 and 2020 primaries.

Given the asymmetric nature of the data set (which contained two Democratic primaries and a single Republican primary), the construction of two separate networks served to make use of the full data set while also mitigating concerns which might have been present if conclusions were drawn based solely on either one of the networks individually. The network containing only 2016 candidates, for example, serves to mitigate temporal concerns that would arise when comparing 2016 Republican candidates to 2020 Democratic candidates: namely, that any detected difference in use of moral language would be due to temporal context, and not partisanship. Conversely, the network aggregating both 2016 and 2020 candidates serves to confirm — to the extent that the data is able — that any detected difference in use of moral language between 2016 Democrats and 2016 Republicans is not an isolated occurrence, but rather a single data point in a continued trend of moral-rhetorical divergence over time[9].

9 For the purposes of interpretation, the "aggregated" network visualizes what the rhetorical relationships between the 2020 Democratic candidates and the 2016 Republican candidate would have been, had they all been running against each other in 2016.

In total, the 2016 network contained 516[10] nodes and 2,018 weighted edges, while the aggregated 2016-2020 network contained 537 nodes and 4,841 weighted edges. The Louvain resolution community detection algorithm (Lambiotte *et al.*, 2009) (resolution parameter = 2.19) detected two communities of candidate nodes, overlapping exactly with partisanship affiliation.

Figure 7 and **Figure 8** were effective in revealing the "rhetorical distance" existing between the two political parties examined in the study. The networks display a bi-communal modularity structure across both candidates and election cycles, where community membership of each node exactly correlated to the partisan affiliation of the candidate it represents. In both cases, Democratic and Republican candidate nodes were clearly polarized, suggesting two distinct categories of moral speech during primary elections: one used by Democratic candidates and one used by Republican candidates. These results comprehensively address the first element of RQ1, suggesting that Democratic and Republican presidential candidates diverged significantly and consistently in their use of moral language online during recent presidential primaries.

10 While there were 574 moral terms included in the custom dictionary, linguistic pre-processing was applied during the construction of the edge lists which aggregated word inflections into a single root form, enhancing clarity in the network but reducing the total number of terms to 516.

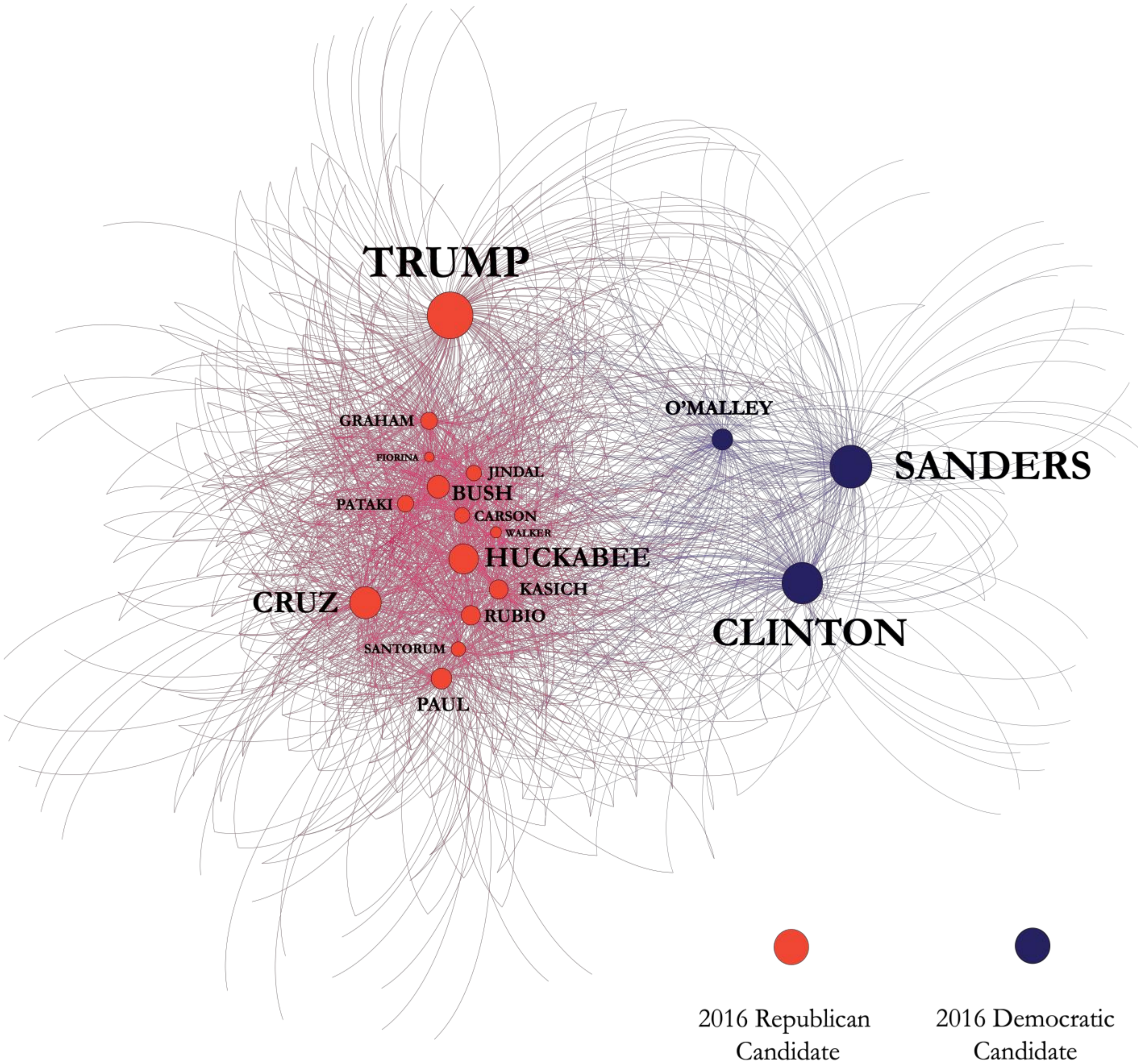


Figure 7 Bipartite text network displaying the moral-rhetorical community structure of the 2016 U.S. presidential primaries, based on a frequency analysis of 574 moral terms used by 17 Democratic and Republican candidates on Twitter. Nodes were colored using a Louvain community detection algorithm, which detected two communities perfectly reflecting partisan affiliation. Candidates are connected to each other through their use of the same moral language. Word nodes were removed to enhance readability, leaving spatialized candidate positions. Node and label sizes scale with betweenness centrality. Edges are colored by their candidate source node.

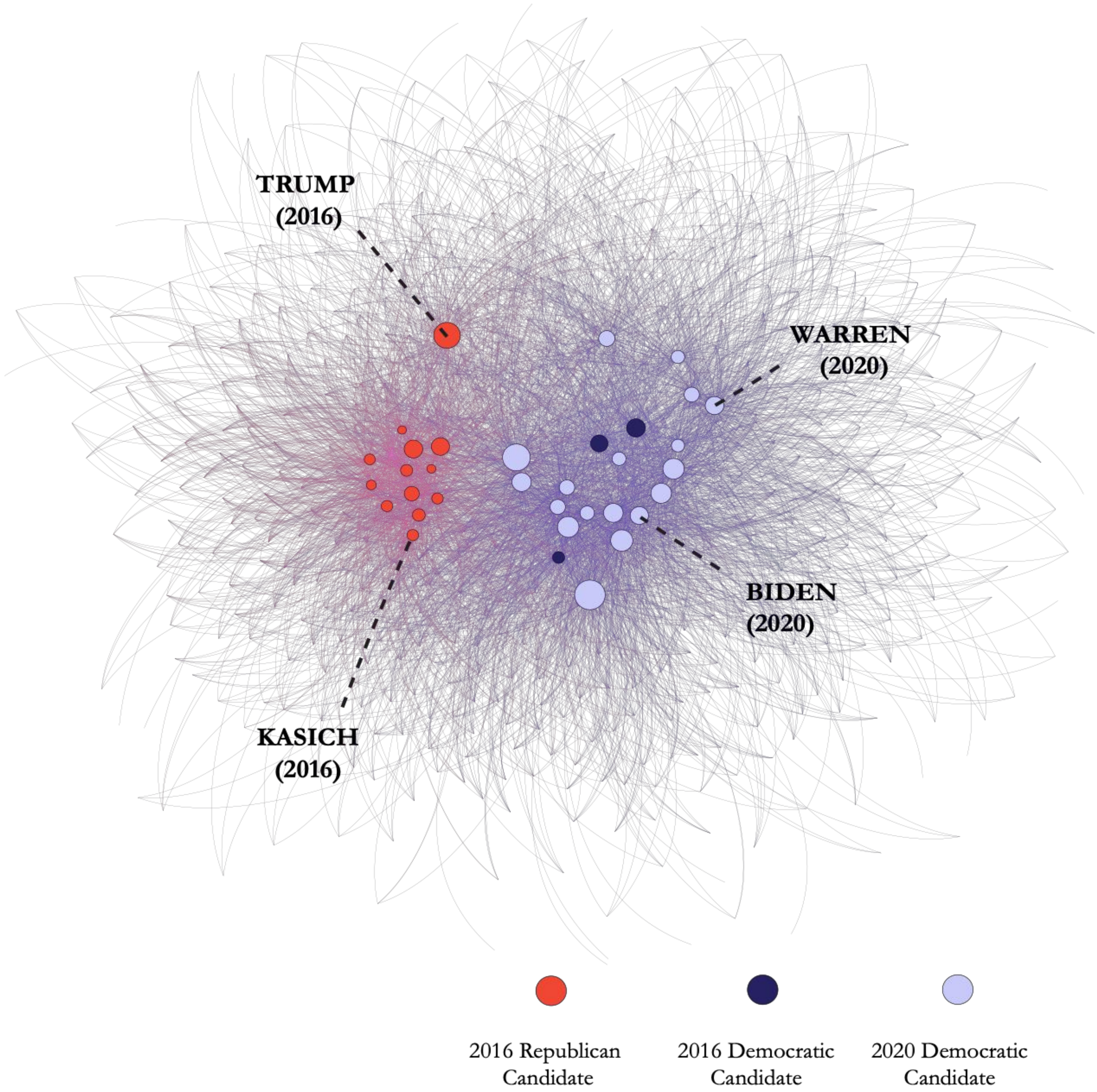


Figure 8 Bipartite text network displaying the moral-rhetorical community structure across both the 2016 and 2020 U.S. presidential primaries, based on a frequency analysis of 574 moral terms used by 39 Democratic and Republican candidates on Twitter. Candidates nodes are connected to each other through their use of the same moral language and colored by partisan affiliation; word nodes have been removed to enhance readability, leaving spatialized candidate positions. Node sizes are scaled by betweenness centrality. Edges connected to Democratic candidates are blue and edges connected to Republican candidates are pink. Some candidate labels were removed to enhance readability.

To examine the second element of *RQ1* — assessing the specific moral dimensions along which this divergence took place — the custom dictionary was applied to each primary corpus. **Figure 9** displays the differing use of each moral foundation by Democratic and Republican candidates; **Table 2** contains the raw proportions of language used in each case.

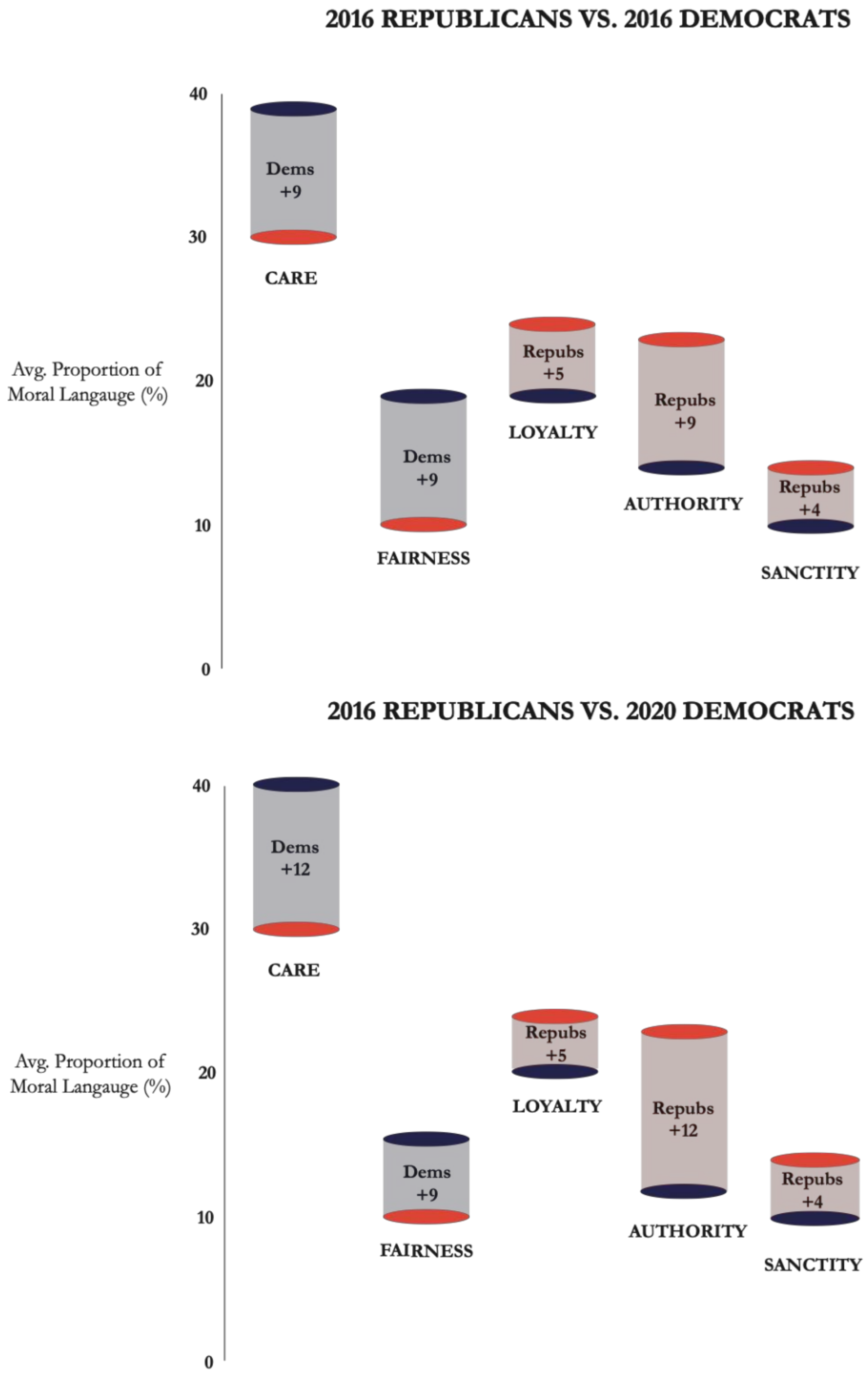


Figure 9 Difference in use of moral foundations between Democratic and Republican primary candidates on Twitter. Top plot compares 2016 Republican candidates and 2016 Democratic candidates; bottom plot compares 2016 Republican candidates and 2020 Democratic candidates.

	Share of Moral Language by Moral Foundation				
Primary	Care	Fairness	Loyalty	Authority	Sanctity
2016 Republican	0.31	0.10	0.24	0.23	0.14
2016 Democratic	0.39	0.19	0.19	0.14	0.10
2020 Democratic	0.42	0.15	0.22	0.11	0.11

Table 2 Average proportion of moral language used on Twitter in each primary, by moral foundation

The results indicate that 2016 and 2020 Democrats used more care and fairness language than did 2016 Republicans; 2016 Republicans used more loyalty, authority, and sanctity language than both 2016 and 2020 Democrats.

Notably, a comparison of the 2016 and 2020 Democratic candidates finds that 2020 Democrats used just 3% more care language and 3% less authority language than their 2016 counterparts, with the distributions of moral language being otherwise identical. This suggests remarkable consistency in use of moral rhetoric by Democratic candidates from primary election to primary election.

These results address the second element of *RQ1*, indicating that Republican and Democratic candidates diverged in their use of moral language: Democratic candidates use more care and fairness language while Republicans favor authority, loyalty, and sanctity language.

Intra-Foundation Similarity

Addressing *RQ2*, this section will examine the extent to which candidates within a party share a moral-rhetorical "intra-foundation similarity" — defined here as a similar pattern of word selection and word use, *within* a given moral foundation — and to what extent these patterns of similar moral expression recur across candidates. Critically, it will examine the connective nature of these shared moral vocabularies, and assess their role in the definition and construction of the moral-rhetorical norms for a political party during a given primary election.

To answer this question, custom "moral similarity" networks were constructed for each primary (as outlined in section **3.2.6**). Specifically, these networks were designed to illustrate how the similar use of particular moral foundations connect individual candidates during a primary. In other words, these networks visualize —across two political parties and three primary elections — the moral foundations which tended to be discussed in the most similar ways.

Figure 10 and **Figure 11** display the results for both 2016 and 2020 primaries; **Figure 12** then compares their unlabelled network structure, emphasizing the varying connective role played by each moral foundation in each party primary. **Table 3** displays the average pairwise cosine similarity between the moral language used by each candidate in each moral foundation for each primary.

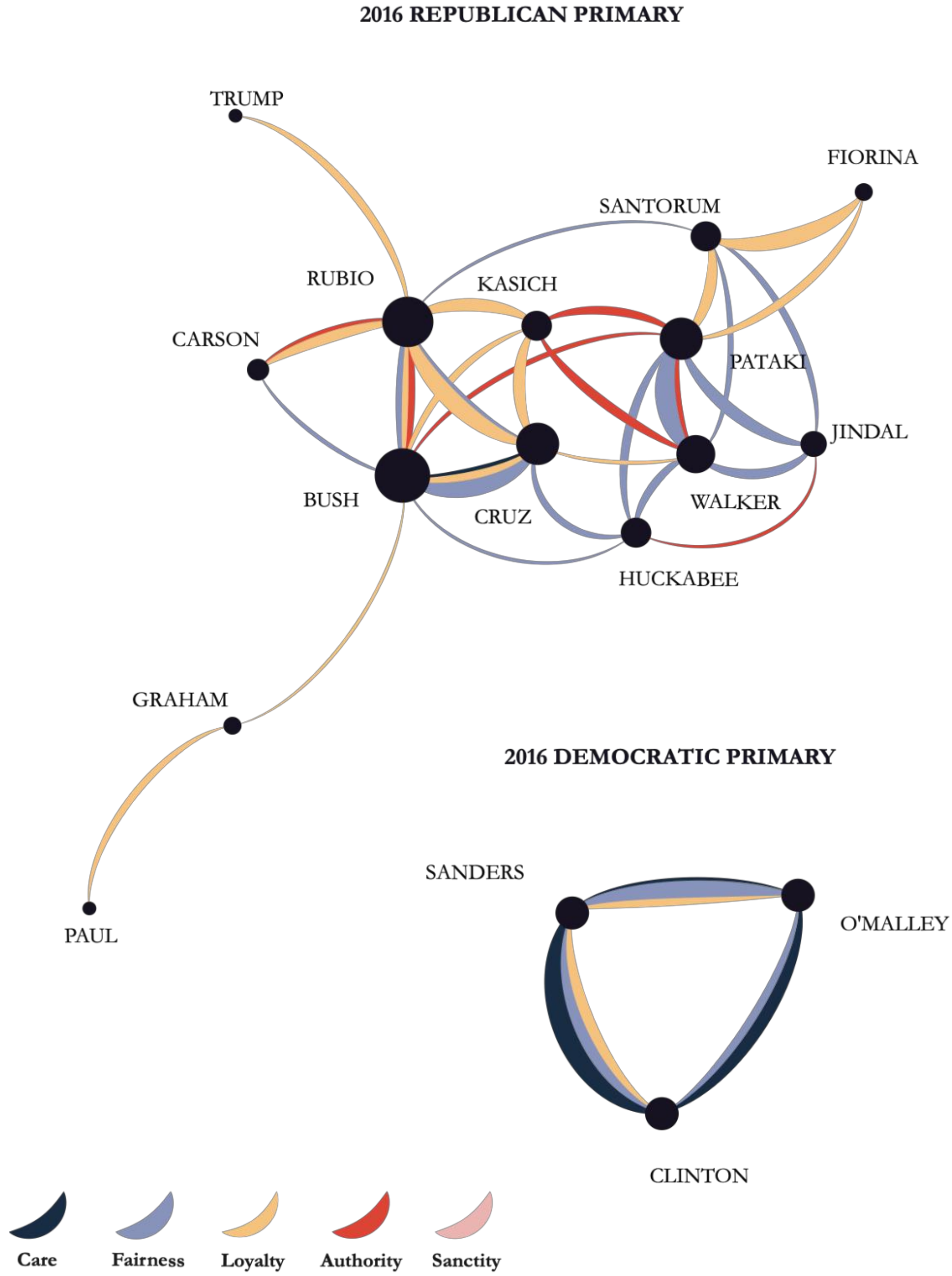


Figure 10 Moral foundation network illustrating how candidates during the 2016 Republican and Democratic primaries were connected through their similar use of individual moral foundations. Edges weights index strength of similarity; edge colors indicate the moral foundation connecting them; node size was scaled with betweenness centrality.

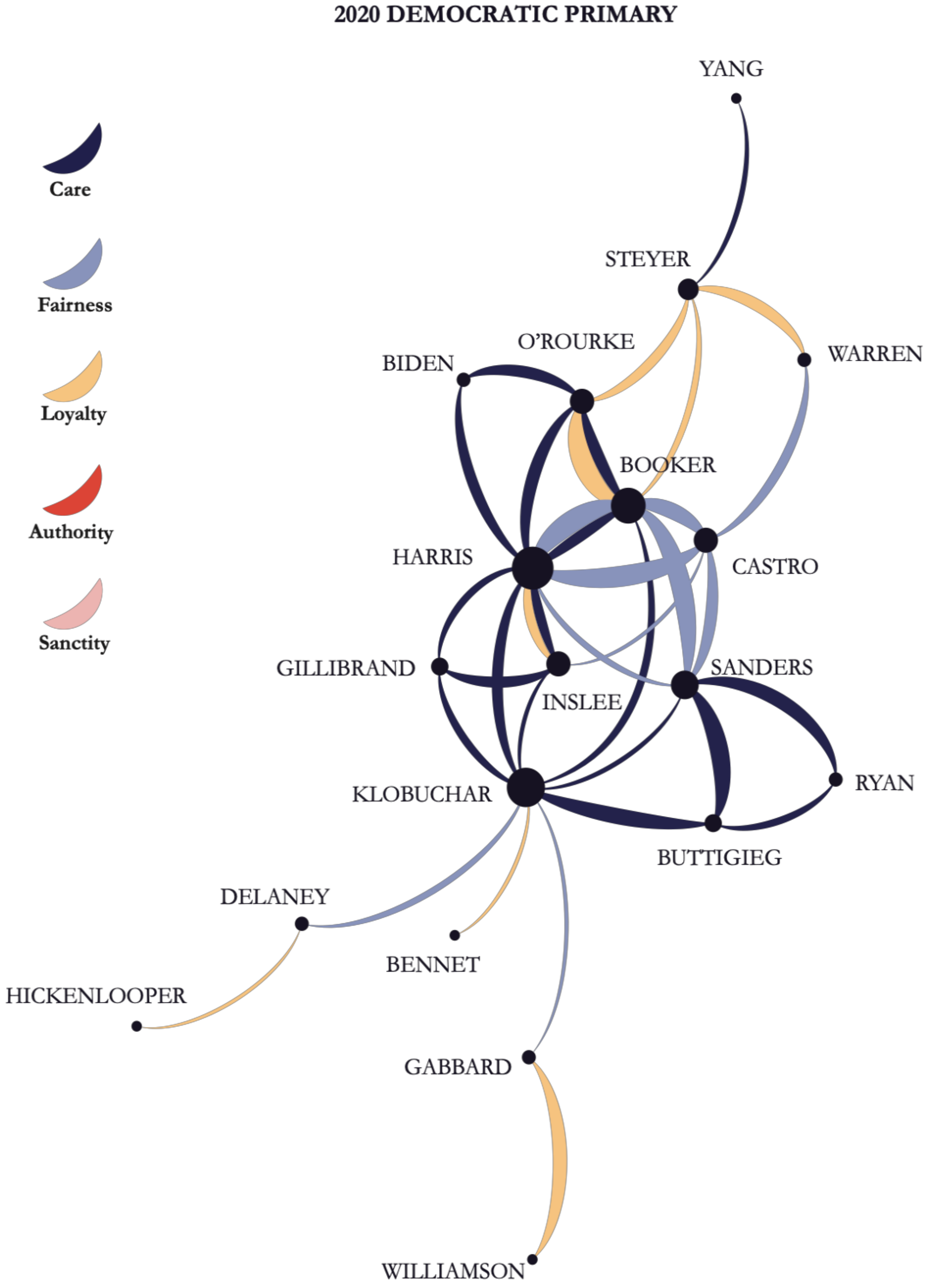


Figure 11 Moral foundation network illustrating how candidates during the 2020 Democratic primary were connected through their similar use of individual moral foundations. Edges weights index strength of similarity; edge colors indicate the moral foundation connecting them; node size was scaled with betweenness centrality.

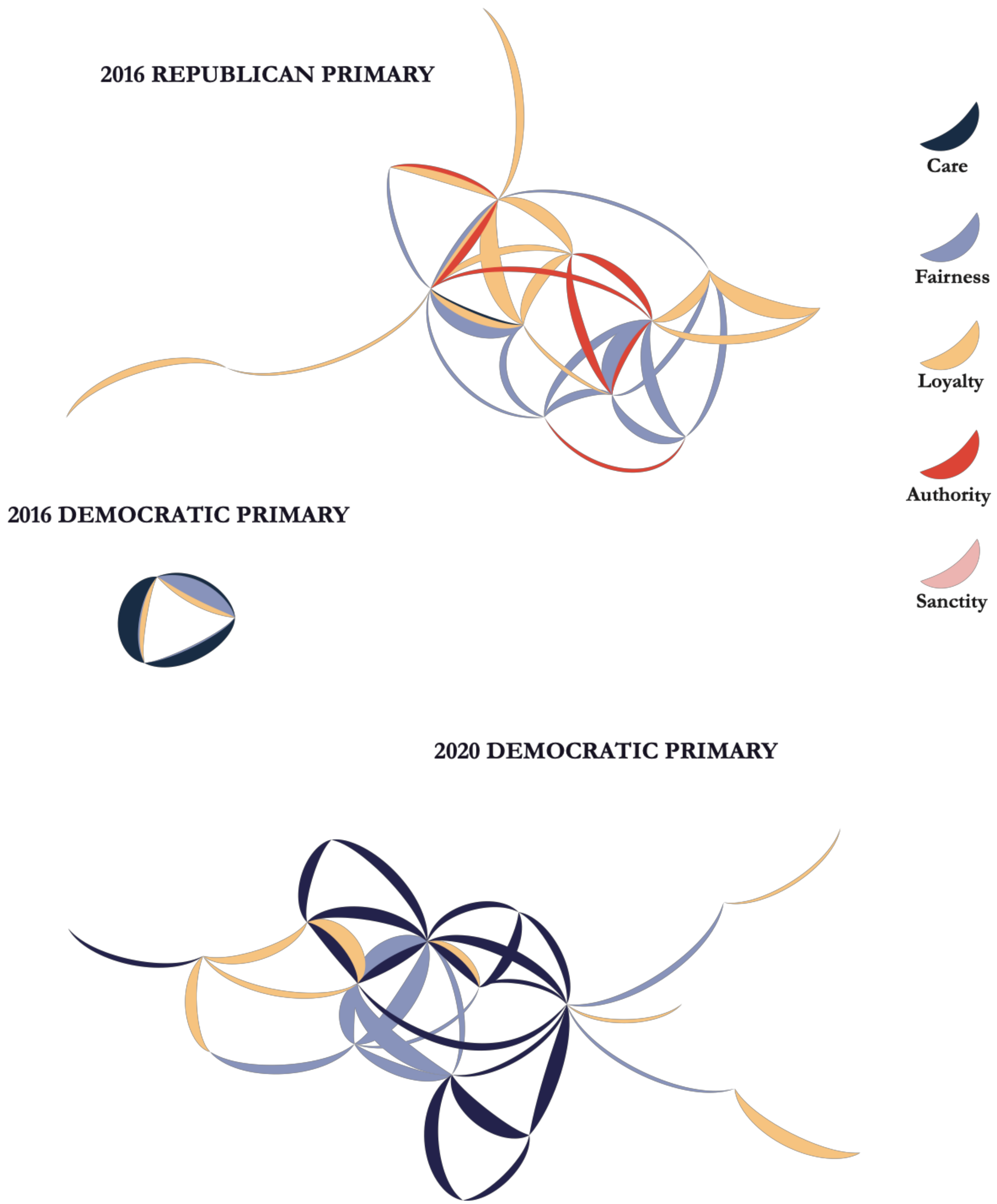


Figure 12 Network skeletons for each of three moral similarity networks. Nodes and candidate labels have been removed to highlight trends.

	Avg. Pairwise Cosine Similarity		
Moral Foundation	2016 Repubs	2016 Dems	2020 Dems
Care	0.563	0.782	0.815
Fairness	0.618	0.784	0.751
Loyalty	0.558	0.782	0.754
Authority	0.511	0.341	0.558
Sanctity	0.439	0.512	0.487
Std. Dev	**0.067**	**0.204**	**0.142**

Table 3 Average pairwise cosine similarity for each moral foundation during each primary election, 2016-2020

The results indicate that Democrats and Republicans differed in their intra-foundation similarity, with Democrats in both 2016 and 2020 connected through the use of similar care and fairness language, and Republicans in 2016 instead connected by similar loyalty, authority, and fairness language. The findings show that just as Democratic candidates talk about care and fairness more, they also do so in highly similar ways, selecting the same care words and using them in the same proportions. Conversely, the results also show that while Democrats used less loyalty, authority, and sanctity language, when candidates did use such language they were more varied in their approach to the specific words they selected and at what frequencies they used them. Republicans, for their part, also achieved higher intra-foundation similarity for the moral foundations they used most often, discussing loyalty and authority language in similar ways. Notably, fairness language was also fairly consistent across Republican candidates.

The average pairwise cosine similarity scores calculated across all candidates in each primary reported in **Table 3** offer comprehensive statistics which offer further insight. The results show that in aggregate, 2016 Republican candidates consistently used different moral words and in different proportions. Taken as a whole, these results suggest that to varying degrees, and especially for Democrats, the most used moral foundations in each party are also discussed with the least variance[11].

Deviation and Moral-Rhetorical Outsiders

RQ3 asks how individual candidates deviated from the moral-rhetorical norms of their party, and to what extent these deviations distanced them from their competitors. To address this question, instances were identified in which candidates deviated from their party norms by using significant proportions of moral language from foundations associated with the opposing party[12]. Their positioning *vis-a-vis* other candidates in a rhetorical network was then examined to determine whether this moral-rhetorical divergence co-occurred with a network position closer to candidates from the opposing party.

In **Figure 13** and **Figure 14**, the results of the dictionary application are displayed at the party and candidate level, making clear both the rhetorical choices made by individual candidates, but also how those choices contribute to — or deviate from — intra-party norms. Both figures display the proportion of moral language used by each candidate as either a positive or negative deviation from 20%, the proportion of moral language that would be used if candidates used each foundation equally. For example, if a candidate recorded a positive 10% deviation for the "care" moral foundation, this would indicate that in total, 30% of their moral language used on Twitter was "care" language (10% more than would be expected if each candidate was equally likely to select words from each moral foundation).

11 These findings are also unrelated to the number of terms in the dictionary for this analysis; for example, if there were fewest care words in the dictionary, candidates would have fewer words from which to choose, resulting in higher cosine similarity scores. This was not the case: in fact, care contained the largest number of words out of any foundation in the custom dictionary.

12 This definition of a deviation was constructed to in congruence with the tenants of moral reframing outlined by Feinberg & Willer (2019).

The results were analyzed such that significant deviations were identified. For the purpose of this analysis, a significant result was noted if a Democratic candidate used a proportion of loyalty, authority, or sanctity language *equal to or greater than* the average proportion of said language used by 2016 Republican candidates shown in **Table 3**. Likewise, a result was noted if a Republican candidate used a proportion of "care" or "fairness" language that was equal to or greater than the average proportion of said language used by 2016 Democratic candidates. Tables with exact deviation data for all 39 individual candidates across all five moral foundations can be found in **Appendix VII**.

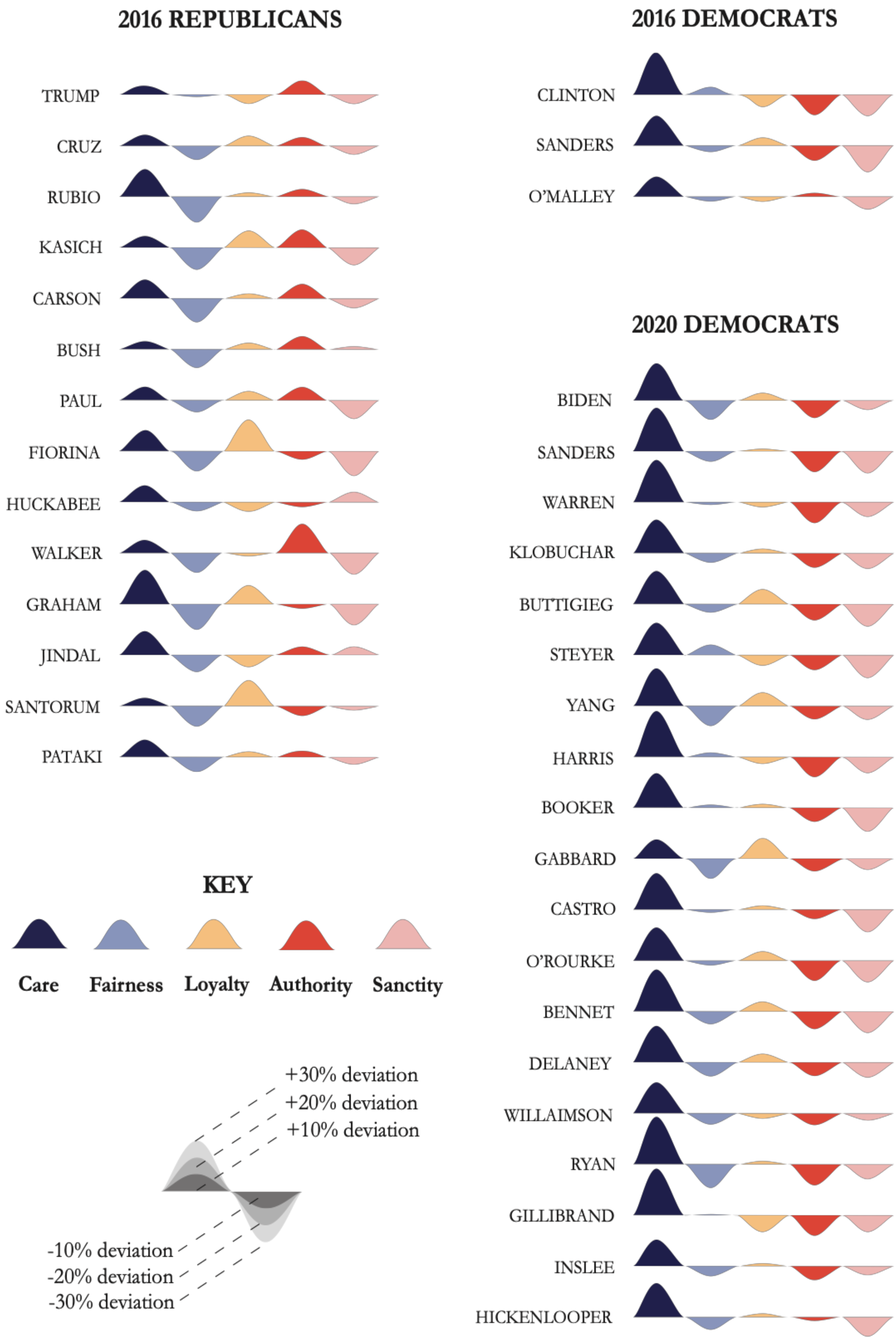


Figure 13 Deviation in proportion of moral foundation used by each candidate.

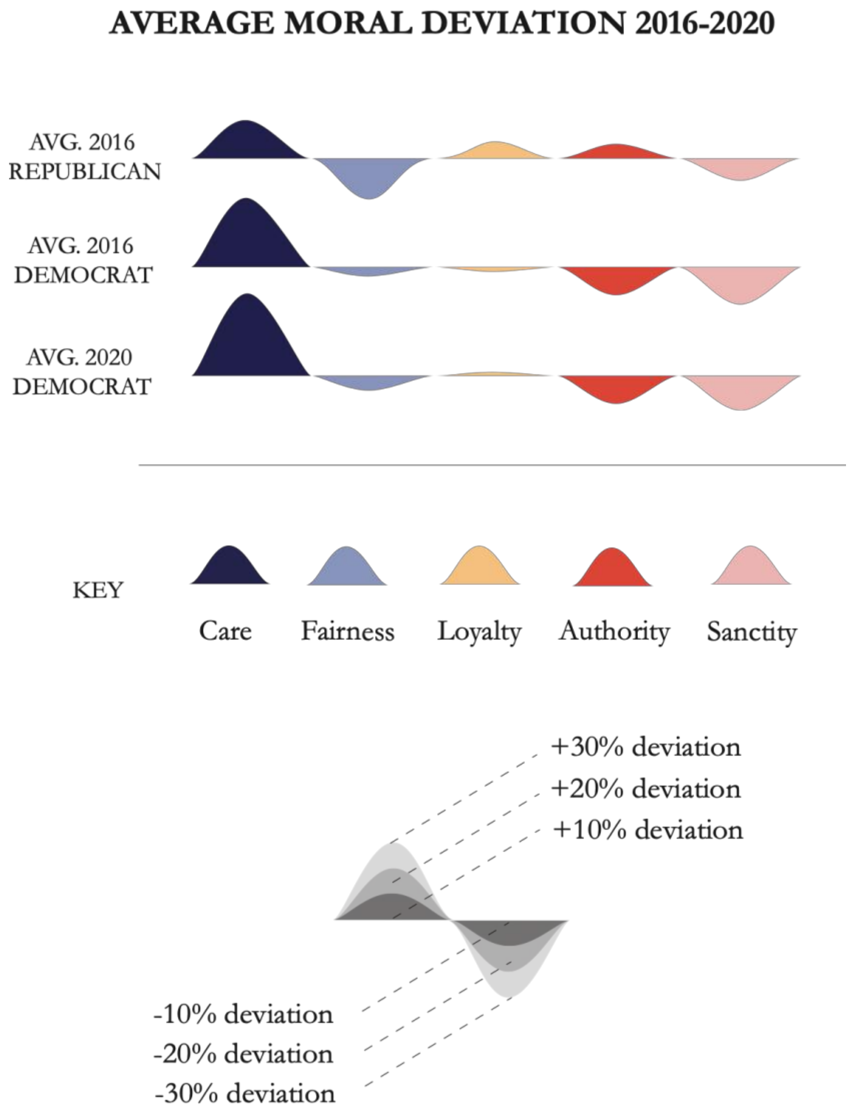


Figure 14 Average deviation in proportion of moral foundation used by each party.

In total, seven significant deviations were identified for further analysis. **Table 4** expresses these deviations in terms of the party average for that primary.

Candidate	Foundation	Deviation from Party Average (%)
Marianne Williamson	Sanctity	+7.0
Andrew Yang	Loyalty	+6.5
Donald Trump (2016)	Fairness	+10.2
Pete Buttigieg	Loyalty	+6.6
Tulsi Gabbard	Loyalty, Sanctity	+11.7, +3.9
Joe Biden	Sanctity	+3.8

Table 4 Significant deviations, expressed a deviation from the party average during that election cycle.

To address the second element of *RQ3* — namely, how specific moral deviations distanced individual candidates from their competitors — their position on the network displayed in **Figure 8** was identified.

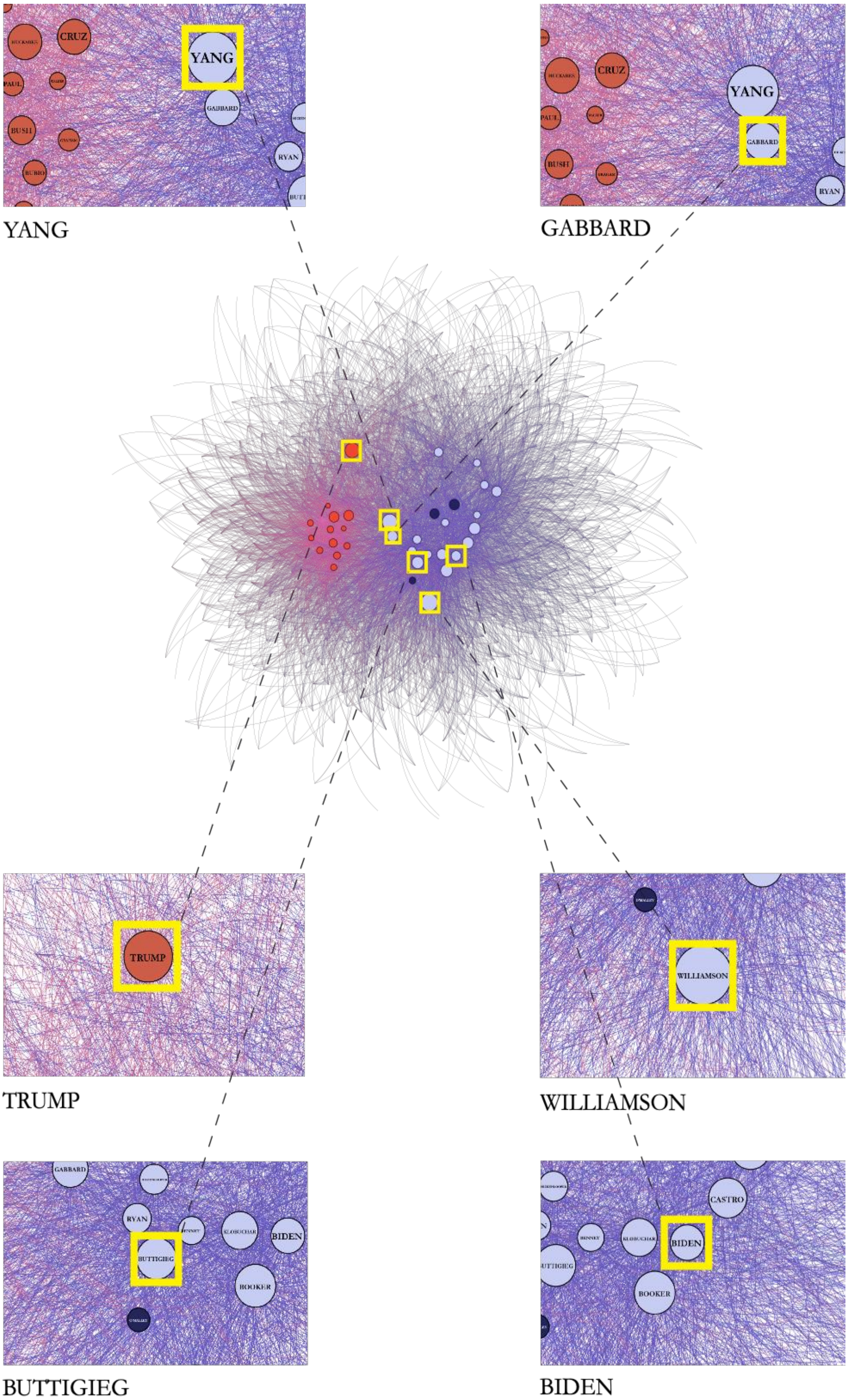


Figure 15 The network position of six candidates who used a significant proportion of moral language associated with a moral foundation endorsed by the opposing party.

The extent to which these candidates were isolated can be validated through an analysis of their betweenness centrality scores, a statistic in network analysis used to determine how much influence a node has over the flow of information in a network, which can be found in **Table 5**. In this context, a large betweenness centrality measure indicates that a candidate used higher proportions of words which were used by few or no other candidates.

Candidate	Betweeness Centrality
Marianne Williamson	10262.0
Andrew Yang	9048.7
Donald Trump (2016)	8638.5
Pete Buttigieg	6448.6
Tulsi Gabbard	5546.2
Bernie Sanders (2016)	5380.9
Joe Biden	5096.2
Average for non-deviating candidates:	3579.2

Table 5 Betweenness centrality scores for seven candidates who were identified as moral-rhetorical deviants.

As shown both by a visual appraisal of the network and validated through a calculation of their betweenness centrality scores, it is clear that candidates who deviated from norms of moral language use within their parties often acted as network "gatekeepers" to clusters of less-used moral words. The network layout also suggests that these deviations do not necessarily have consistent impacts on candidate positions, with candidates deviating in a number of different ways. Donald Trump and Marianne Williamson deviated away from their party, but not necessarily towards the opposing party, while Andrew Yang and Tulsi Gabbard deviated unmistakably towards their Republican counterparts. Interestingly, in spite of their moral-rhetorical deviations, Pete Buttigieg and Joe Biden appeared able to retain central network positions amidst their Democratic peers, even as they increased their use of loyalty and sanctity language.

DISCUSSION

The findings of this study reveal that Democratic and Republican presidential candidates used disparate moral vocabularies on Twitter during recent primary elections and appealed to voters by emphasizing different dimensions of moral reasoning. The results also illustrate the moral foundations that tended to be discussed in the most similar ways within each party, revealing that uniform, unvaried use of care and fairness language connected Democrats in 2016 and 2020, while 2016 Republicans offered a more fragmented moral-rhetorical network. Finally, the results find that while several candidates deviated significantly from the rhetorical norms of their party, this deviation could co-occur with rhetorical isolation — as it did for Donald Trump in 2016 — or, interestingly, an insular network position — as it did for Joe Biden in 2020. The following chapter will discuss these findings before addressing limitations and proposing areas of interest for future research.

Moral Language and Polarization in Digital Networks

The results pertaining to *RQ1* reveal that Democratic and Republican candidates diverged significantly in the moral language they used on Twitter during both 2016 and 2020 primary elections. This divergence was consistent and extreme, with two distinct clusters of candidates clearly visible and exhibiting no overlap. Interestingly, a comparison of two Democratic primaries taking place four years apart showed very little variance in the proportion of each moral foundation used, suggesting that usage patterns of moral language may be deeply entrenched in the norms of Democratic campaigning, and that moral foundations endorsed by political parties remain constant across elections. Beyond the distinct and constant separation between the parties, the results also find that the parties diverged along the moral dimensions hypothesized by Haidt & Graham (2007) and Graham *et al*., (2009), with Democratic candidates using more care and fairness language and Republican candidates using more loyalty, authority, and sanctity language. This establishes that the moral framings used by Democratic and Republican candidates during recent primaries will likely be less persuasive to individuals from opposing parties (Feinberg & Willer, 2019).

These findings also bear significant implications for diffusion of campaign rhetoric online. Research on diffusion of moralized language in digital social networks has found that the presence of moral-emotional language in political messages is associated with increased diffusion within — but not across — partisan communities (Brady *et al*., 2017). The findings of the present research might explain why: if political messages contain two distinct categories of moral-emotional language, and a political message is more widely shared by individuals who endorse the moral-emotional language it contains, then a polarized network would quickly develop. Under these conditions, the ways in which Democratic and Republicans have tended to express their moral campaign rhetoric might be a contributor the polarization of online social networks. To the extent that online campaign rhetoric contains moral language which fails to diverge from the standard moral-rhetorical norms of a party, voters may be increasingly likely to be exposed to moral arguments in a manner consistent with digital-political echo chambers.

Intra-Foundation Similarity and Party Ideology

The results of *RQ2* reveal the degree to which candidates within the same party engaged in similar patters of moral expression and display the moral foundations which played the largest role in rhetorically connecting competing candidates during each primary. They also offered evidence that the most popular moral foundations in a given ideology tend to be expressed in the most similar ways. Critically, results offer a visual means of assessing — along a moral dimension — existing theories of party ideology present during the 2016 and 2020 primaries.

Candidates during the 2016 Republican primary appeared to be less rhetorically unified in their expressions of moral language than were Democrats in 2016 and 2020. This suggests that perhaps a moral fracture was yet another dimension of the GOP "identity crisis" (Lemann, 2020) which has not been often discussed. The most central candidates (candidates with the highest node degree) in the 2016 Republican moral similarity network were Jeb Bush ($k = 11$), Marco Rubio ($k = 10$), and Ted Cruz ($k = 8$). These findings map onto existing theories of fractured party ideology present during the 2016 Republican primary, and emphasize candidates who were expected to play the largest role in defining the party discourse: Noel (2016) described Bush as the establishment favorite, Cruz as the ideological favorite, and then, when neither proved acceptable to the other faction, Rubio as the proposed alternative (Cassidy *et al*., 2015; Kruse & White, 2016; Noel, 2016). The central position of each of these candidates on a moral-rhetorical network is notable.

However, while the moral network identifies the centrality of these three candidates in the rhetorical primary, it also reveals that they tended to use extremely similar moral language to one another, especially along dimensions of loyalty, authority, and fairness. This suggests that "ideologue" and "establishment" GOP candidates as identified by existing literature (Noel, 2016), while differing in their reputations and policy stances, were remarkably less differentiated in their moral expression. This may suggest why more divergent candidates, such as Ben Carson and Donald Trump, ended up successfully activating new segments of the electorate during the 2016 primaries. In fact, Trump and Carson were extremely peripheral on the network, indicating that they did not use the moral language common amongst other Republican candidates, consistent with research labeling them as ideological outsiders during the primary (Noel, 2016). The findings therefore seem to suggest that Trump's top competitors presented similar and undifferentiated moral rhetoric, just as the larger party lacked a coherent moral dimension along which to express a united party message. This might illustrate further evidence of the "shattered" nature of the party in 2016, but along a moral dimension (Goldmacher *et al*., 2016).

In 2016 and 2020 candidates from both parties, but especially Democrats — were to a greater extent undifferentiated in their expression of the moral foundations they used the most frequently. These findings build on the established partisan differences in usage of moral foundations evaluated in *RQ1*, further showing that that an exceedingly similar use of "popular" moral foundations within a party can act as a connective agent during a primary election, rhetorically binding competing candidates to one another. This finding is surprising, given that increased use of a moral foundation offers candidates increased chances to diverge from one another in terms of the specific language that they use.

These findings reveal in an empirical, visual fashion the moral-rhetorical "core" of the Democratic party during these primaries. The candidates with the largest degree during the 2020 Democratic primary — Harris ($k = 10$) , Booker ($k = 9$), and Klobuchar ($k = 9$) — occupy the most central positions in the network, playing the largest role in constructing the "generic" moral discourse for the 2020 Democratic primary. To voters, these candidates are likely to sound like the most "typical" Democrats, as their moral language follows the most generic use patterns. All three were senators, well-connected to the party establishment, and neither the most moderate nor the most progressive candidates running (Herndon, 2019). Thus, the network may have identified a plausible "center" to the party.

Candidate Positionality in Moral-Rhetorical Networks

The results pertaining to *RQ3* reveal the possibility of making empirical each candidate's position in "moral-rhetorical space". The results found that on the rare occasion that a candidate diverged from their party norms in a manner consistent with the moral reframing propositions outlined by Feinberg and Willer (2019), the extent to which they were rhetorically isolated varied significantly. Candidates broadly fell into three groups: candidates who deviated sharply from both parties by using entirely different moral language, candidates who deviated towards the opposing party by using that party's moral language, and candidates whose position remained central within their own party because while they used more language related to a moral foundation endorsed by the opposing party, they did so by selecting language uniquely, though scarcely, used by their own party.

Trump: Political and Moral-Rhetorical Outsider

The findings of the present research establish that Trump's status as a political "outsider" in 2016 (Seers, 2016; Stevenson, 2019; SkyNewsAustralia, 2020) corresponded to meaningful differences in his moral-rhetorical style *vis-à-vis* other candidates, making him a moral-rhetorical "outsider" as well. In fact, while Trump deviated significantly from the moral rhetorical norms of his party by using large proportions of fairness language, the fairness language he used was also divergent from that used by Democrats. **Table 6** offers an example of how his fairness vocabulary largely diverged from both parties:

TRUMP 2016		
Trump Fairness	**2016 Repubs Fairness**	**2016 Dems Fairness**
Dishonest	*Rights*	*Rights*
Fair	*Law*	*Justice*
Biased	*Justice*	*Equality*
Law	*Laws*	*Equal*
Lying	*Fair*	*Law*
Liar	*Trust*	*Inequality*
Lied	*Lying*	*Fair*
Unfair	*Trusted*	*Racism*
Honest	*Integrity*	*Laws*
Trust	*Honest*	*Discrimination*

Table 6 Top ten most-used moral words in the category in which Trump recorded a significant moral-rhetorical deviation (fairness) compared with the top ten terms for the rest of the 2016 Republican party and the 2016 Democratic party. Highlighted words indicate a term unique to Trump.

Yang, Gabbard & the Alt-Right

The two Democratic candidates identified on the network as having the network positions closest to the community of 2016 Republican candidates — Tulsi Gabbard and Andrew Yang — were also the only two candidates who received significant support from the alt-right during the 2020 Democratic primary. Gabbard's campaign in particular was notable for its strange bedfellows: Steve Bannon, President Trump's former chief strategist, was "impressed" with her political talent; Richard Spencer, the white nationalist leader, plainly stated that he would vote for her; and former Republican presidential candidate Ron Paul, a staunch libertarian, also offered his support (Lerer, 2019). Yang, for his part, also received vocal support from white supremacists (Breland, 2019; Budryk, 2019) and on alt-right message boards (Bort, 2019). **Table 7** shows exemplifies how whereas Trump devoted from both parties, even a list of ten words begins to display overlap between the moral language used by Yang and Gabbard and the moral language uniquely favored by 2016 Republicans. These candidates may have begun to realize the potential of moral language to elicit support from the opposing party, even in a political environment characterized by an extremely high levels of polarization.

GABBARD 2020

Gabbard Loyalty	2020 Dems Loyalty	2016 Repubs	Gabbard Sanctity	2020 Dems Sanctity	2016 Repubs Sanctity
War	*Together*	*Joining*	*Clean*	*Clean*	*Religious*
United	*Communities*	*Nation*	*Corruption*	*Corruption*	*Prayers*
Together	*Community*	*Together*	*Wasted*	*Food*	*God*
Groups	*Nation*	*War*	*Corrupt*	*Drug*	*Faith*
Countries	*War*	*United*	*Waste*	*Epidemic*	*Marriage*
Nation	*United*	*Wife*	*Religious*	*Dignity*	*Bless*
Fellow	*Companies*	*Group*	*Drugs*	*Corrupt*	*Church*
Communities	*Joining*	*Allies*	*God*	*Drugs*	*Praying*
Sacrifice	*Coalition*	*Community*	*Drugs*	*Faith*	*Prayer*
Unite	*Group*	*Coalition*	*Wasting*	*Sexual*	*Drug*

YANG 2020

Yang Loyalty	2020 Dems Loyalty	2016 Repubs
Together	*Together*	*Joining*
Communities	*Communities*	*Nation*
Group	*Community*	*Together*
Companies	*Nation*	*War*
Community	*War*	*United*
Wife	*United*	*Wife*
Joining	*Companies*	*Group*
War	*Joining*	*Allies*
United	*Coalition*	*Community*
Countries	*Group*	*Coalition*

Table 7 Top ten most-used moral words in the categories in which Gabbard and Yang recorded a significant moral-rhetorical deviation, compared with the top ten terms for their party and the opposing party. Words highlighted in green indicate a term unique to the candidate; words highlighted in blue indicate a term shared only with the opposing party.

Buttigieg, Biden, and "Liberal" Loyalty

Two candidates were able to use significant proportions of sanctity and loyalty language and retain central network positions amongst their fellow Democrats: Joe Biden and Pete Buttigieg. They achieved this because their most distinctive sanctity and loyalty language was not language also used by 2016 Republicans: as **Table 8** shows, Joe Biden's framing of the 2020 election as "a battle for the *soul* of the nation" was a classic appeal to a preservation of purity and sanctity (Dodman, 2020), but Republican candidates were not discussing "soul". Pete Buttigieg, who described his campaign as an effort create a sense of "belonging" in America, used classic language associated with group loyalty (Rodrigo, 2019), but managed to do so in a unique way, as Republican candidates rarely used the same appeals[13]. **Table 8** also illustrates how they may have used fewer "unique" terms, opting instead to use "Democratic" language from usual moral foundations, but in significantly higher proportions.

[13] George Pataki was the only Republican to do so, and he did so only three times.

BIDEN 2020			BUTTIGIEG 2020		
Biden Sanctity	2020 Dems Sanctity	2016 Repubs Sanctity	Buttigieg Loyalty	2020 Dems Loyalty	2016 Repubs Loyalty
Soul	*Clean*	*Religious*	*Together*	*Together*	*Joining*
Dignity	*Corruption*	*Prayers*	*Communities*	*Communities*	*Nation*
Epidemic	*Food*	*God*	*Community*	*Community*	*Together*
Clean	*Drug*	*Faith*	*Nation*	*Nation*	*War*
Sexual	*Epidemic*	*Marriage*	*War*	*War*	*United*
Prayers	*Dignity*	*Bless*	*Belonging*	*United*	*Wife*
Corrupt	*Corrupt*	*Church*	*United*	*Companies*	*Group*
Faith	*Drugs*	*Praying*	*Belong*	*Joining*	*Allies*
Corruption	*Faith*	*Prayer*	*Coalition*	*Coalition*	*Community*
Drug	*Sexual*	*Drug*	*Joining*	*Group*	*Coalition*

Table 8 Top ten most-used moral words in the category in which Biden and Buttigieg recorded a significant moral-rhetorical deviation, compared with the top ten terms for their party and the opposing party. Words highlighted in green indicate a term unique to the candidate; words highlighted in blue indicate a term shared only with the opposing party.

The success of each of their campaigns — with Biden winning the nomination and Buttigieg claiming an upset victory in the Iowa caucuses — suggests that it is possible for candidates to use unique patterns of moral language while not being positioned as party outsiders.

Political-Strategic Implications

These findings raise questions about the strategic role of moral language in politics. How ought parties and candidates to conceptualize both the polarized, partisan state of moral language on campaigns and the persuasive potential of its strategic use? Under frameworks of political marketing management, parties would likely be advised to emphasize the construction of winning coalitions using moral language as a strategic tool to position political products in the minds of political consumers (Scammell, 1999). However, as widely noted by political scientists, these political marketing frameworks tend to be inconsistent with models of democratic elections (Henneberg, 2009).

The role of moral language on campaigns could instead be understood through a lens of behavioral economics, using notions of "nudge" or "libertarian" paternalism (Thaler, 2005). Through an understanding of the ways in which voters are influenced by moral language frames, candidates and parties might structure their rhetoric in a way that leads to the most "informed" choices by political choice-makers. In a moral-rhetorical context, this means exposing voters to different moral framings of issues and candidates, and allowing them to select the ones they most prefer, without restricting their political options, or changing their incentives through the alteration of policy positions. Moral words in political rhetoric might therefore be more effectively viewed as "nudges" (Thaler, 2005) toward better political choices. Whereas a classical example of a "nudge" might be placing fruits and vegetables at eye level in a grocery store to encourage consumers to more fully consider those options, in the context of political campaigns such a nudge might involve placing specific moral language in more prominent positions in party and campaign rhetoric. Politicians and parties — rather than institutions seeking political profit — might be therefore be thought of as "choice architects", aware of different ways in which choices can be presented to consumers and the effects these can have on voter behavior.

Limitations & Future Research

An unavoidable feature of U.S. primary elections is their temporal asymmetry: party primaries do not always take place on the same election years, and do not always contain similar numbers of candidates. Depending on a number of factors, such as incumbency and political context, they may not take place at all. As a result, this research was limited by the necessity to use 2016 Republican primary ideology as a benchmark with which to compare both 2016 and 2020 Democrats. This limitation was especially significant considering the uniquely fractured nature of the 2016 primary, making it unclear to what extent the moral rhetoric measured during that primary was different from the Republican status quo. These results should therefore be validated against future Republican primaries.

Similarly, the 2016 Democratic primary only contained three serious candidates, a comparatively smaller number than 2016 Republicans, and generated a smaller number of tweets. The present research argues that by including the 2020 Democratic candidates as well — and thus constructing a counterfactual scenario where the 2020 Democrats actually ran as a part of the 2016 primaries — these concerns were mitigated. Still, this is no guarantee that the consistent patterns of moral expression identified between 2016 and 2020 Democrats were not at least partly coincidental. These results should therefore be validated against future Democratic primaries.

It is also important to note that Twitter is just one of many ways in which candidates express moral rhetoric on a campaign: alternate mediums such as debates, interviews, and speeches (not to mention other social media platforms) serve as additional purveyors of a candidate's message. Interpretation of the networks in this study should therefore be informed by an understanding of how the interactions between all of these mediums might constitute a "network of networks". The "hive plot" model for analyzing complex large-scale networks (Krzywinski *et al*., 2011) describes this phenomenon, and is illustrated in **Figure 16**. Critically, it illustrates how individual layers of a network might be structured according to different or even opposing logics. Applied to the present research, this illustrates how moral language present on Twitter might act as a connective tie to the alternative network layers represented by other mediums, but its context may differ in each case. The hive model serves as an effective illustration of the dangers present in generalizing results based on a single "layer" of a network, making clear that the present findings might not remain consistent across all media.

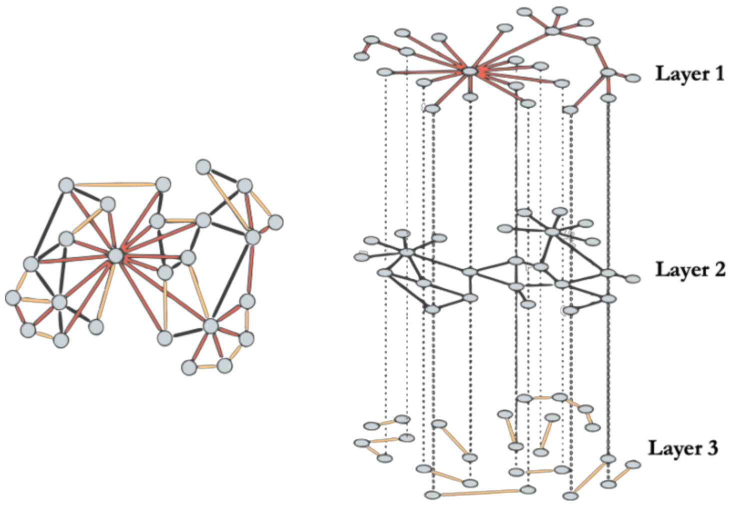


Figure 16 "Hive plot" model (Krzywinski, *et al*., 2011) for analysing multi-layered networks. The right view represents how layers can be threaded by various networking connections. The left illustrates a look through the layers.

Implementations of hive plot models might also form the basis of future research in this field: they offer a meaningful step toward a multi-media model of candidate rhetoric, providing not only a validation of these results but also insights into the ways in which moral landscapes evolve across platforms and mediums. Unlike standard network layouts, hive plots use coordinate-based systems for tracking node locations, allowing different networks to be more effectively compared, overlaid, and combined, thus mitigating some interpretations concerns associated with traditional networks (Krzywinski, *et al*., 2011). Future research should therefore explore the capability of hive plots to construct larger, more comprehensive, and more robust models of candidate rhetoric across mediums and elections.

While this paper serves as a baseline — conducting fundamental comparisons and proposing a networked model for moral rhetoric — future research should explore the promise of these methods for narrower, more specific inquiries. For example, the use of moral language does not always indicate a specific, intentional framing of a policy or issue in along a specific moral dimension. Therefore, to evaluate the extent to which candidates are actively engaging in moral framing or "reframing" (Feinberg & Willer, 2019) requires a closer, more fine-grained analysis. Future research should therefore assess to what extent candidates are engaged in moral re-framing, the issues being reframed, and how these frames are constructed.

The present research illustrates Trump's status as a moral-rhetorical outsider during the 2016 primaries, displaying the promise of text network analysis for displaying relationships between the rhetoric of speakers and discursive categories, across time and rhetorical space. Future research should therefore use the 2024 Republican primaries to examine the degree to which the GOP has been "trumpified" by examining the positioning of 2024 candidates compared to their 2016 counterparts, and evaluating the extent they have rhetorically converged on or diverged from Trump's 2016 position. Similarly, the case of Tulsi Gabbard, Andrew Yang, and the alt-right should also be further examined through a comparison of their 2020 primary rhetoric and the rhetoric existing on white supremacist and al-right message boards. Such an analysis might offer important clues as to the words and moral intuitions underpinning alt-right movements, and lead to more accurate conception of their intersection and activation within popular mainstream politics.

CONCLUSION

This research contributes to an understanding of the ways in which morality and politics intersect, illustrating that just as moral convictions play a critical role in constructing the political attitudes of voters, moral language plays a critical role in connecting and differentiating political candidates and political parties during presidential elections in the U.S. Specifically, this research reveals that Democratic and Republican candidates in 2016 and 2020 emphasized different dimensions of moral reasoning during their campaigns, with conservative candidates using moral language related to in-group loyalty and respect for social hierarchies, and liberal candidates using more language related to the careful and just treatment of individuals. Highlighting the connected and contemporaneous nature of primary competition, this research has also illustrated how candidates were connected to one another through their use of moral language, and displayed the ways in which similar moral vocabularies connect and define moral-rhetorical norms in a primary election. Further, this research has allowed for the empirical representation of moral-rhetorical candidate positioning, allowing for nuanced analysis of the way in which use of moral language may isolate a candidate from their peers — as it did Trump in 2016 — or insulate a candidate amongst them, as it did Joe Biden in 2020.

The patterns of moral expression measured in this study illustrate another dimension along which political discourse in the U.S. has been polarized, and bear significant implications for political campaigns, impacting the way voters engage with campaign messaging, respond to political issues, and form opinions about political candidates. This research therefore secondarily contributes by illustrating the promise of networked approaches to the study of politics and political language, and demonstrates their effectiveness in answering research questions related to campaign rhetoric and candidate positioning. Mapping the moral language used by political candidates in this way can do much to shed light on the emotional underpinnings of a chaotic and expansive national discourse, revealing the ways in which the democratic process of selecting a president has been shaped by — and may impact upon — the moral convictions of individual citizens.

SUPPLEMENTARY MATERIALS

Replication data from this study, including all code used for the weighting, tuning, and application of the dictionaries, as well as data collection and the construction of the networks, is published online at https://github.com/kobihackenburg/LSE-Dissertation.

REFERENCES

Aaker, D. A., & Shansby, J. (1982). Positioning your product. *Business Horizons*, 25(3), 56-62.

Acree, B. D., Gross, J. H., Smith, N. A., Sim, Y., & Boydstun, A. E. (2018). Etch-a-sketching: Evaluating the post-primary rhetorical moderation hypothesis. *American Politics Research*, 48(1), 99-131.

Adams, A., & McCorkindale, T. (2013). Dialogue and transparency: A content analysis of how the 2012 presidential candidates used twitter. *Public Relations Review*, 39(4), 357-359.

Andrews, A. C., Clawson, R. A., Gramig, B. M., & Raymond, L. (2016). Finding the right value: Framing effects on domain experts. *Political Psychology*, 38(2), 261-278.

Anstead, N. (2017). The idea of austerity in British politics, 2003–2013. *Political Studies*, 66(2), 287-305.

Auter, Z. J., & Fine, J. A. (2017). Social media Campaigning: Mobilization and fundraising on Facebook. *Social Science Quarterly*, 99(1), 185-200.

Bail, C. A. (2016). Combining natural language processing and network analysis to examine how advocacy organizations stimulate conversation on social media. *Proceedings of the National Academy of Sciences*, 113(42), 11823-11828.

Baines, P. R., Lewis, B. R., & Ingham, B. (1999). Exploring the positioning process in political campaigning. *Journal of Communication Management*, 3(4), 325-336.

Barabási, A., & Albert, R. (2011). Emergence of scaling in random networks. *The Structure and Dynamics of Networks*, 349-352.

Barrie, C., & Chun-ting Ho, J. (2021). AcademictwitteR: An R package to access the Twitter Academic Research Product Track v2 API endpoint. *Journal of Open Source Software*, 6(62), 3272.

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. Third International AAAI Conference on Weblogs and Social Media.

Benoit, K. (2020). Text as data: An overview. *The SAGE Handbook of Research Methods in Political Science and International Relations*, 461-497.

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). Quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, 3(30), 774.

Blondel, V. D., Guillaume, J., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10).

Bode, L., & Dalrymple, K. E. (2014). Politics in 140 characters or Less: Campaign Communication, Network interaction, and political participation on Twitter. *Journal of Political Marketing*, 15(4), 311-332.

Borgatti, S., Mehra, A., Brass, D., & Labianca, G. (2009). Network Analysis in the Social Sciences. *Science*, 323(5916), 892-895.

Bort, R. (2019, March 20). What is going on with Andrew YANG'S CANDIDACY?, URL: https://www.rollingstone.com/politics/politics-news/yang-gang-explained-810872/ [Last consulted July 28, 2021]

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, 114(28), 7313-7318.

Breland, A. (2019, April 10). Here's why Andrew Yang's alt-right supporters think he's the 2020 candidate for white nationalists, URL: https://www.motherjones.com/politics/2019/04/andrew-yang-4chan-alt-right/ [Last consulted July 28, 2021]

Bright, J., Hale, S., Ganesh, B., Bulovsky, A., Margetts, H., & Howard, P. (2019). Does campaigning on social media make a Difference? Evidence from candidate use of Twitter during the 2015 and 2017 U.k. Elections. *Communication Research*, 47(7), 988-1009.

Bryant, M. (2019, July 31). Outsider Marianne Williamson moves beyond the Fringe in democratic debate, URL: https://www.theguardian.com/us-news/2019/jul/31/marianne-williamson-democratic-debate-election-2020 [Last consulted August 02, 2021]

Budryk, Z. (2019, April 15). Yang disavows support from White nationalists: 'it's been a point of confusion', URL: https://thehill.com/homenews/campaign/438879-yang-disavows-support-from-white-nationalists-its-been-a-point-of-confusion [Last consulted July 28, 2021]

Butts, C. T. (2009). Revisiting the foundations of network analysis. *Science*, 325(5939), 414-416.

Caprara, G. V., Schwartz, S., Capanna, C., Vecchione, M., & Barbaranelli, C. (2006). Personality and politics: Values, traits, and political choice. *Political Psychology*, 27(1), 1-28.

Cassidy, J., Sorkin, A., & Lizza, R. (2015, October 25). Jeb and Hillary: A tale of Two Establishment Favorites, URL: https://www.newyorker.com/news/john-cassidy/jeb-and-hillary-a-tale-of-two-frontrunners [Last consulted August 13, 2021]

Christof, R. (2010). Aristotle's Rhetoric. In *Stanford Encyclopedia of Philosophy*. Palo Alto, CA: Metaphysics Research Lab, Stanford University.

Clifford, S., & Jerit, J. (2013). How words do the work of Politics: Moral Foundations theory and the debate OVER stem cell research. *The Journal of Politics*, 75(3), 659-671.

Condor, S., Tileagă, C., & Billig, M. (2013). Political rhetoric. *Oxford Handbooks Online*.

Conway, B. A., Kenski, K., & Wang, D. (2013). Twitter use by presidential primary candidates during the 2012 campaign. *American Behavioral Scientist*, 57(11), 1596-1610.

Conway, B. A., Kenski, K., & Wang, D. (2015). The rise of twitter in the political campaign: Searching for intermedia agenda-setting effects in the presidential primary. *Journal of Computer-Mediated Communication*, 20(4), 363-380.

Corasaniti, N., & Haberman, M. (2016, August 09). Donald Trump Suggests 'Second Amendment People' could act AGAINST Hillary Clinton, URL: https://www.nytimes.com/2016/08/10/us/politics/donald-trump-hillary-clinton.html [Last consulted August 08, 2021]

Dehghani, M., Johnson, K., Hoover, J., Sagi, E., Garten, J., Parmar, N., . . . Graham, J. (2016). Supplemental material for PURITY HOMOPHILY in social networks. *Journal of Experimental Psychology*: General.

Dehghani, M., Sagae, K., Sachdeva, S., & Gratch, J. (2014). Analyzing political rhetoric in conservative and LIBERAL weblogs related to the construction of The "Ground Zero Mosque". *Journal of Information Technology & Politics*, 11(1), 1-14.

Denny, M. J., & Spirling, A. (2018). Text preprocessing for UNSUPERVISED learning: Why it matters, when It MISLEADS, and what to do about it. *Political Analysis*, 26(2), 168-189.

Dimitrova, D. V., & Matthes, J. (2018). Social media in political campaigning around the WORLD: Theoretical and Methodological Challenges. *Journalism & Mass Communication Quarterly*, 95(2), 333-342.

Dodman, B. (2020, November 07). In the 'battle for the soul' of a nation, Joe Biden offers an antidote to Trump, URL: https://www.france24.com/en/americas/20201107-joe-biden-in-the-battle-for-the-soul-of-a-nation-democrat-offers-antidote-to-trump [Last consulted August 02, 2021]

Drieger, P. (2013, June 17). Semantic network analysis as a method for visual text analytics, URL: https://www.sciencedirect.com/science/article/pii/S1877042813010227 [Last consulted July 15, 2021]

Ekins, E., & Haidt, J. (2016, February 05). Donald Trump supporters think about MORALITY differently than other voters. Here's how, URL: https://www.vox.com/2016/2/5/10918164/donald-trump-morality [Last consulted August 05, 2021]

Ellsworth, J. W. (1965). Rationality and Campaigning: A content analysis of the 1960 presidential CAMPAIGN DEBATES. *Western Political Quarterly*, 18(4), 794-802.

Enli, G. (2017). Twitter as arena for the authentic Outsider: Exploring the social media campaigns of Trump and Clinton in the 2016 US presidential election. *European Journal of Communication*, 32(1), 50-61.

Esser, F., & Strömbäck, J. (2014). *Mediatization of Politics: Understanding the transformation of western democracies.* Basingstoke, Hampshire: Palgrave Macmillan.

Feinberg, M., & Willer, R. (2015). From Gulf to bridge: When do moral arguments facilitate political influence? *Personality and Social Psychology Bulletin*, 41(12), 1665-1681.

Feinberg, M., & Willer, R. (2019). Moral reframing: A technique for effective and persuasive communication across political divides. *Social and Personality Psychology Compass*, 13(12).

Feinberg, M., Willer, R., Antonenko, O., & John, O. P. (2012). Liberating reason from the passions. *Psychological Science*, 23(7), 788-795.

Feld, S. L., Merrill, S., & Grofman, B. (2014). Modeling the effects of Changing Issue Salience in Two-Party Competition. *Public Choice*, 158(3-4), 465-482.

Frimer, J., Haidt, J., Graham, J., Dehghani, M., & Boghrati, R. (2017, December 17). Moral Foundations Dictionaries for Linguistic Analyses, 2.0, URL: http://www.jeremyfrimer.com/uploads/2/1/2/7/21278832/summary.pdf [Last consulted 2021]

Fulgoni, D., Carpenter, J., Ungar, L., & Preoţiuc-Pietro, D. (2016). An Empirical Exploration of Moral Foundations Theory in Partisan News Sources. *ACL Anthology*.

Galioto, K., Siders, D., & DESIDERIO, B. (2019, April 10). Buttigieg attracting praise from an unexpected audience – conservatives, URL: https://www.politico.com/story/2019/04/10/pete-buttigieg-conservative-praise-1264075 [Last consulted July 28, 2021]

Garten, J., Hoover, J., Johnson, K. M., Boghrati, R., Iskiwitch, C., & Dehghani, M. (2017). Dictionaries and distributions: Combining expert knowledge and large scale textual data content analysis. *Behavior Research Methods*, 50(1), 344-361.

Goldmacher, S., Cheney, K., & Collins, E. (2016, February 24). Trump shatters the Republican Party, URL: https://www.politico.com/story/2016/02/trump-shatters-the-republican-party-219711 [Last consulted August 02, 2021]

Goodman, J. (2020, October 16). US election 2020: 'Rigged' VOTES, body doubles and other false claims, URL: https://www.bbc.co.uk/news/54562611 [Last consulted August 05, 2021]

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, 96(5), 1029-1046.

Graham, J., Haidt, J., Iyer, R., & Koleva, S. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, 101(2), 366-385.

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S., & Ditto, P. (2013, April 08). Moral foundations theory: The pragmatic validity of moral pluralism, URL: https://www.sciencedirect.com/science/article/pii/B9780124072367000024?via%3Dihub [Last consulted April 24, 2021]

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3), 267-297.

Haidt, J. (2012). *The Righteous Mind: Why good people are divided by politics and religion*. NYC, NY: Pantheon Books.

Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that Liberals may not recognize. *Social Justice Research*, 20(1), 98-116.

Haidt, J., & Hersh, M. A. (2001). Sexual morality: The cultures and emotions of conservatives and liberals. *Journal of Applied Social Psychology*, 31(1), 191-221.

Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, 133(4), 55-66.

Hart, R. P. (2009). *Campaign talk: Why elections are good for us*. Princeton, NJ: Princeton University Press.

Henneberg, S. C. (2009). Political marketing and theories of democracy. *Political Marketing: Theory and Concepts*, 74-95.

Herndon, A. (2019, September 10). Kamala Harris: Who she is and what she stands for, URL: https://www.nytimes.com/interactive/2020/us/elections/kamala-harris.html [Last consulted August 14, 2021]

Higgins, M., & Lakoff, G. (1998). Moral politics: What conservatives know that liberals don't. *Language*, 74(2), 425.

Hoover, J., Johnson, K., Boghrati, R., Graham, J., & Dehghani, M. (2018). Moral framing and charitable donation: Integrating exploratory social media analyses and confirmatory experimentation. *Collabra: Psychology*, 4(1).

Hopp, F. R., Fisher, J. T., Cornell, D., Huskey, R., & Weber, R. (2020). The extended moral foundations dictionary (emfd): Development and applications of a crowd-sourced approach to extracting moral intuitions from text.

Hu, Y. (2005). Efficient, high-quality force-directed graph drawing. *Mathematica Journal*, 10(1), 37-71, URL: http://asus.myds.me:6543/paper/ktall/37%20-%201984%20-%20Efficient,%20High-Quality%20Force-Directed%20Graph%20Drawing.pdf [Last consulted July 18, 2021]

Håkansson, N. (1997). Election rhetoric of political parties. *Javnost - The Public*, 4(3), 81-95.

Jacob B. Hirsh, C. (2010). Compassionate liberals and Polite Conservatives: Associations of agreeableness with political ideology and moral values - Jacob B. Hirsh, Colin G. Deyoung, Xiaowen Xu, Jordan B. Peterson, 2010, URL: https://journals.sagepub.com/doi/10.1177/0146167210366854 [Last consulted April 24, 2021]

Janoff-Bulman, R. (2009). To provide or protect: Motivational bases of political liberalism and conservatism. *Psychological Inquiry*, 20(2-3), 120-128.

Jerit, J. (2004). Survival of the fittest: Rhetoric during the course of an election campaign. *Political Psychology*, 25(4), 563-575.

Jungherr, A. (2015). Twitter use in election campaigns: A systematic literature review. *Journal of Information Technology & Politics*, 13(1), 72-91.

Keneally, M. (2016, August 10). Clinton Responds to Trump's 2nd Amendment Comments: 'Words Matter', URL: https://abcnews.go.com/Politics/clinton-responds-trumps-2nd-amendment-comments-words-matter/story?id=41278547 [Last consulted August 08, 2021]

Khazan, O. (2018, April 24). People voted for trump because they were anxious, not poor, URL: https://www.theatlantic.com/science/archive/2018/04/existential-anxiety-not-poverty-motivates-trump-support/558674/ [Last consulted August 08, 2021]

Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1), 205395171452848.

Koc-Michalska, K., Lilleker, D. G., Smith, A., & Weissmann, D. (2016). The normalization of online campaigning in the web.2.0 era. *European Journal of Communication*, 31(3), 331-350.

Krebs, V. E. (2002). Mapping Networks of Terrorist Cells. *Connections*, 24(3), 43-52.

Kreiss, D. (2014). Seizing the moment: The Presidential campaigns 'use of Twitter during the 2012 electoral cycle. *New Media & Society*, 18(8), 1473-1490.

Krippendorff, K. (1980). *Content analysis an introduction to its methodology*. Beverly Hills, CA: SAGE.

Krosnick, J. A., Lavrakas, P. J., & Kim, N. (1999). Survey research. *Handbook of Research Methods in Social and Personality Psychology*, 404-442.

Kruikemeier, S. (2014). How political candidates use Twitter and the impact on votes. *Computers in Human Behavior*, 34, 131-139.

Kruse, M., & B. White, J. (2016, February 19). How Marco slew HIS 'MENTOR', URL: https://www.politico.com/magazine/story/2016/02/marco-rubio-slew-jeb-bush-mentor-2016-213652/ [Last consulted August 13, 2021]

Krzywinski, M., Birol, I., Jones, S. J., & Marra, M. A. (2011). Hive plots--rational approach to visualizing networks. *Briefings in Bioinformatics*, 13(5), 627-644.

Lakoff, G. (2002). Moral politics: How liberals and conservatives think.

LaMarre, H., & Suzuki-Lambrecht, Y. (2013, August 20). Tweeting democracy? Examining twitter as an online public relations strategy for congressional campaigns', URL: https://www.sciencedirect.com/science/article/pii/S0363811113000982 [Last consulted August 03, 2021]

Lamb, S. M., & Newell, L. E. (1966). *Outline of stratificational grammar*. Washington, D.C., DC: Georgetown University Press.

Lambiotte, R., Delvenne, J., & Barahona, M. (2009, October 09). Laplacian dynamics AND MULTISCALE modular structure in networks, URL: https://arxiv.org/abs/0812.1770 [Last consulted July 25, 2021]

Lamont, M., Park, B. Y., & Ayala-Hurtado, E. (2017). Trump's electoral speeches and his appeal to the American white working class. *The British Journal of Sociology*, 68.

Leidner, B., & Castano, E. (2011). Morality shifting in the context of intergroup violence. *European Journal of Social Psychology*, 42(1), 82-91.

Lemann, N. (2020, October 22). The republican identity crisis after trump, URL: https://www.newyorker.com/magazine/2020/11/02/the-republican-identity-crisis-after-trump [Last consulted August 02, 2021]

Lerer, L. (2019, October 12). What, exactly, is Tulsi Gabbard up to?, URL: https://www.nytimes.com/2019/10/12/us/politics/tulsi-gabbard.html [Last consulted July 28, 2021]

Lewis, R. (2017). Large-scale patterns of entertainment gratifications in linguistic content of u.s. films, URL: https://www.tandfonline.com/doi/full/10.1080/10510974.2017.1340903 [Last consulted July 15, 2021]

Light, R. (2014). From words to networks and back: Digital text ..., URL: https://cpb-us-e1.wpmucdn.com/blogs.uoregon.edu/dist/6/3554/files/2017/12/sc_speeches_postprint-1wixq61.pdf [Last consulted July 15, 2021]

Lilleker, D. G., Koc-Michalska, K., Schweitzer, E. J., Jacunski, M., Jackson, N., & Vedel, T. (2011). Informing, engaging, mobilizing or interacting: Searching for a European model of web campaigning. European Journal of Communication, 26(3), 195-213.

Long, J. A., & Eveland, W. P. (2018). Entertainment use and political ideology: Linking worldviews to media content. *Communication Research*, 48(4), 479-500.

Louwerse, M. M. (2004). Semantic variation in Idiolect and SOCIOLECT: Corpus linguistic evidence from literary texts. *Computers and the Humanities*, 38(2), 207-221.

Mehler, A. (2008). Large text networks as an object of corpus linguistic studies. Corpus Linguistics. *An International Handbook of the Science of Language and Society.*

Mogelson, L. (2021, January 15). Among the insurrectionists, URL: https://www.newyorker.com/magazine/2021/01/25/among-the-insurrectionists [Last consulted August 05, 2021}

Mooijman, M., Hoover, J., Lin, Y., Ji, H., & Dehghani, M. (2018). Moralization in social networks and the emergence of violence during protests. *Nature Human Behaviour*, 2(6), 389-396.

Morgan, G. S., Skitka, L. J., & Wisneski, D. C. (2010). Moral and religious convictions and intentions to vote in the 2008 presidential election. *Analyses of Social Issues and Public Policy*, 10(1), 307-320.

Noel, H. (2016). Ideological factions in the Republican and Democratic parties. *The ANNALS of the American Academy of Political and Social Science*, 667(1), 166-188.

Pain, P., & Masullo Chen, G. (2019). The president is in: Public opinion and the presidential use of twitter. *Social Media + Society*, 5(2)

Petracca, M. P. (1991). The rational choice approach to politics: A challenge to democratic theory. *The Review of Politics*, 53(2), 289-319.

Phillips, A. (2019, April 26). The 2016 GOP field is the most fractured in recent memory by a lot, URL: https://www.washingtonpost.com/news/the-fix/wp/2015/07/20/the-2016-gop-field-is-the-most-fractured-in-recent-memory-by-a-lot/ [Last consulted August 02, 2021]

Powers, D. M. (1998). Applications and explanations of zipf's law, URL: https://aclanthology.org/W98-1218.pdf [Last consulted July 14, 2021]

Qaiser, S., & Ali, R. (2018). Text mining: Use of tf-idf to examine the relevance of words to documents. *International Journal of Computer Applications*, 181(1), 25-29.

Ramos, J. (2003). Using TF-IDF to Determine Word relevance in Document Queries, URL: http://www.liber.ufpe.br/tg/modules/tg/docs/ramos.pdf [Last consulted July 18, 2021]

Rodrigo, C. (2019, May 12). Buttigieg: Dems EXACERBATING 'crisis of belonging in this country' with 'identity politics', URL: https://thehill.com/homenews/campaign/443280-buttigieg-crisis-of-belonging-in-this-country-exacerbated-by-identity [Last consulted August 02, 2021]

Rule, A., Cointet, J., & Bearman, P. S. (2015). Lexical shifts, substantive changes, and continuity in state of the union discourse, 1790–2014. *Proceedings of the National Academy of Sciences*, 112(35), 10837-10844.

Sagi, E., & Dehghani, M. (2013). Measuring moral rhetoric in text. *Social Science Computer Review*, 32(2), 132-144.

Scammell, M. (1999). Political marketing: Lessons for political science. *Political Studies*, 47(4), 718-739.

Schein, C., & Gray, K. (2017). The theory of dyadic morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review*, 22(1), 32-70.

Seers, G. (2016). Donald Trump – the rise of the political outsider, URL: https://mediahouseinternational.com/donald-trump-the-rise-of-the-political-outsider/ [Last consulted April 06, 2021]

Skitka, L. J., & Mullen, E. (2002). The dark side of moral conviction. *Analyses of Social Issues and Public Policy*, 2(1), 35-41.

SkyNewsAustralia. (2020, October 21). Trump: Thanks for electing an 'outsider as president' instead of 'depraved political establishment', URL: https://www.youtube.com/watch?v=r7yDo_VfyJc [Last consulted April 06, 2021]

Smith, G. (2005). Positioning political parties: The 2005 UK general election. *Journal of Marketing Management*, 21(9-10), 1135-1149.

Sommer, W. (2019, March 13). Presidential hopeful Andrew Yang is a hit with the alt-right, URL: https://www.thedailybeast.com/yang-gang-season [Last consulted July 28, 2021]

Stevenson, C. (2019, July 15). Trump's racist attacks are all part of his 'OUTSIDER' strategy - and if we're not careful it will work again, URL: https://www.independent.co.uk/news/world/americas/us-election/trump-twitter-comments-racist-aoc-ilhan-omar-2020-democrats-a9005921.html [Last consulted April 06, 2021]

Stolee, G., & Caton, S. (2018). Twitter, Trump, and the Base: A shift to a new form of presidential talk? *Signs and Society*, 6(1), 147-165.

Strömbäck, J., & Kiousis, S. (2014). Strategic political communication in election campaigns. *Political Communication*.

Suciu, P. (2020, October 26). Social media could determine the outcome of the 2020 Election, URL: https://www.forbes.com/sites/petersuciu/2020/10/26/social-media-could-determine-the-outcome-of-the-2020-election/ [Last consulted August 03, 2021]

Tausczik, Y. R., & Pennebaker, J. W. (2009). The psychological meaning of WORDS: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1), 24-54.

Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, 78(5), 853-870.

Thaler, R. (2005). Libertarian paternalism. *Laws of Fear*, 175-203.

Thomas, S. (1994). Artifactual study in the analysis of culture: A defense of content analysis in a Postmodern age - SARI thomas, 1994, URL: https://journals.sagepub.com/doi/abs/10.1177/009365094021006002 [Last consulted July 17, 2021]

Townsend, L., & Wallace, C. (2016). Social media research: A guide to ethics by Townsend and ..., URL: https://ahrecs.com/resources/social-media-research-a-guide-to-ethics-by-townsend-and-wallace-guidance-dr-leanne-townsend-prof-claire-wallace-2016/ [Last consulted August 4, 2021]

Vergeer, M., Hermans, L., & Sams, S. (2011). Is the voter only a TWEET away? Micro blogging during the 2009 European Parliament election campaign in the Netherlands. First Monday.

Voelkel, J. G., & Willer, R. (2019). Resolving the Progressive Paradox: Conservative Value framing of progressive economic POLICIES INCREASES candidate support. SSRN *Electronic Journal*.

Wang, S. Y., & Inbar, Y. (2020). Moral language use by U.S. political elites.

Ward, M. D., Stovel, K., & Sacks, A. (2011). Network analysis and political science. *Annual Review of Political Science*, 14(1), 245-264.

Watts, D. J. (2004). The New Science of Networks. *The Annual Review of Sociology*, 30, 243-270.

Weber, R., Mangus, J. M., Huskey, R., Hopp, F. R., Amir, O., Swanson, R., . . . Tamborini, R. (2018). Extracting latent moral information from text narratives: Relevance, challenges, and solutions. *Communication Methods and Measures*, 12(2-3), 119-139.

Wheeler, M. A., McGrath, M. J., & Haslam, N. (2019). Twentieth century MORALITY: The rise and fall of moral concepts from 1900 to 2007. *PLOS ONE*, 14(2).

APPENDICES

A. Supplemental Analysis

Moral-Rhetorical Moderation During General Elections

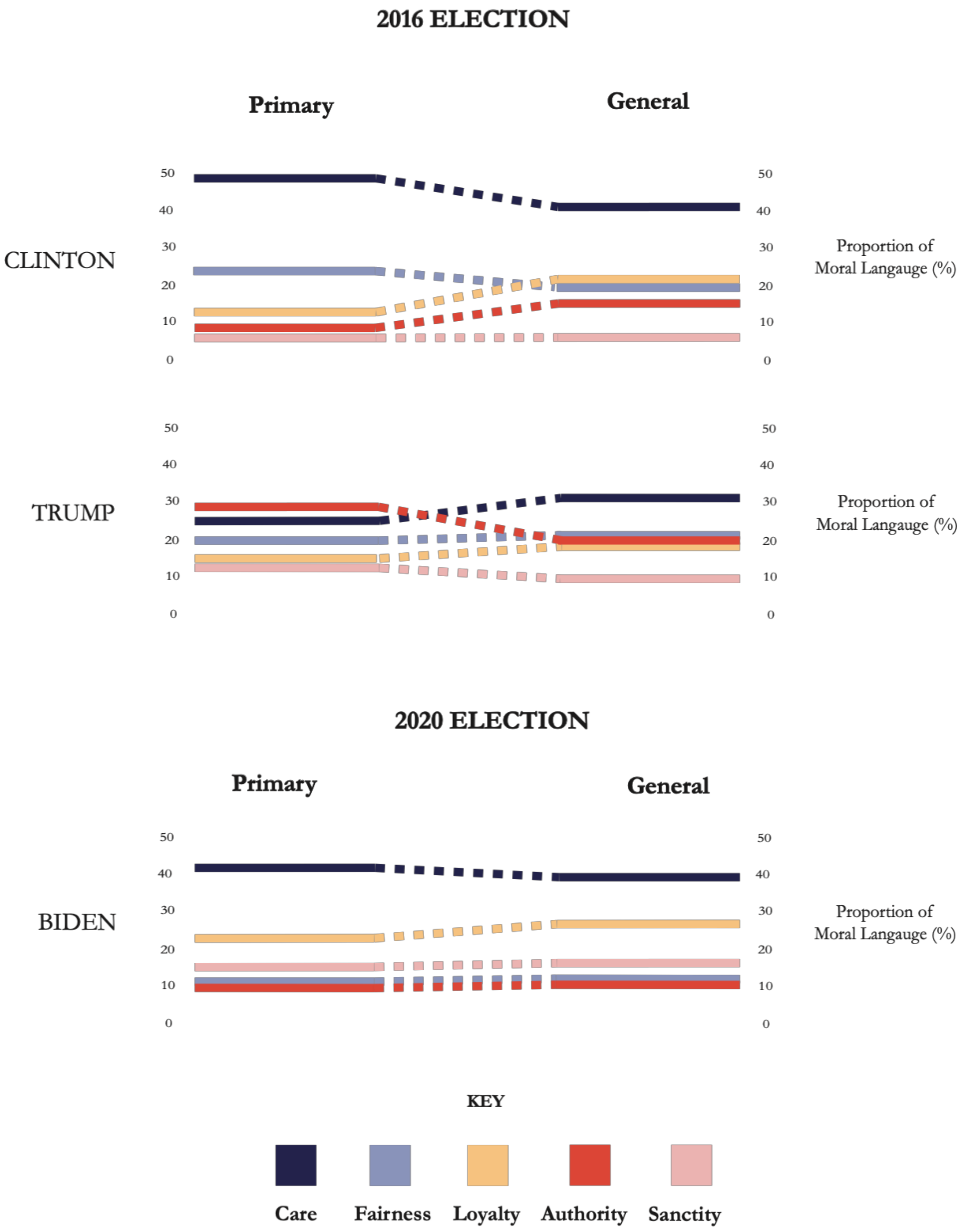
A supplemental analysis was conducted to examine the post-primary rhetorical moderation hypothesis (Acree *et al*., 2018) along a moral dimension. Given the persuasive effects of moral rhetoric outlined by Feinberg (2019), it was hypothesized that the post-primary rhetorical moderation hypothesis (source) might be measurable, if candidates were aware of this effect. It was hypothesized that primary winners would attempt to modulate their moral rhetoric as they orient themselves toward the general election, with Democratic nominees increasing their use of loyalty, authority, and sanctity language and Republican nominees increasing their use of care and fairness language. To conduct this analysis, a supplementary data set including tweets from all general election candidates was assembled (n = 17,921). This included Donald Trump and Hillary Clinton in 2016, and Donald Trump and Joe Biden in 2020.

Nominee	# of Days from Nomination to Election	Tweets	Tweets per Campaign per Day
2016 Clinton	152	4,110	27
2016 Trump	184	2,233	12
2020 Biden	205	2,286	11
2020 Trump	232	8,662	37
Average	193	4,323	22
Total	-	17,291	-

Appendix Table 1 Summary statistics for the full corpus of candidate general election tweets.

The tweets were then analyzed using the same dictionary applied elsewhere in this research. **Appendix Figure 1** shows the change in proportion of moral language used by each nominee across primary and general election stages. This analysis included the three most recent primary winners: Donald Trump and Hillary Clinton in 2016, and Joe Biden in 2020[14].

14 As Donald Trump did not face a primary challenge in 2020, his general election tweets could not be compared.



Appendix Figure 1 Change in proportion of moral foundation language used by the three most recent U.S. presidential primary winners across primary and general election stages.

Candidates in 2016 seemed to largely modulate their rhetoric as hypothesized: after the conclusion of the primary, Hillary Clinton decreased her use of Care (-7%) and Fairness (-5%) language while increasing her use of Loyalty (+8%) and Authority (+6%) language. Her use of Sanctity language remained basically constant. Conversely, after being declared the Republican nominee Donald Trump increased his use of Care (+7%) and Fairness (+1%) language while decreasing his use of Authority (-5%) and Sanctity (-4%) language. However, he also slightly increased his use of Loyalty (+3%). Joe Biden a followed similar pattern after winning the Democratic primary in 2020, but only by the slimmest of margins: he decreased his use of Care language (-4%) and increased his Loyalty language (+2%), but his proportion of Fairness and Authority language was constant, and he decreased his use of Sanctity by 1%.

These results generally show support for the existence of a post-primary rhetorical moderation hypothesis along a moral dimension, although it's not clear if candidates were doing so knowingly or not. In the majority of cases, candidates modulated their rhetoric along the expected moral dimensions implied by Feinberg (2019), albeit mildly. The results were consistent enough to suggest that candidates might be instinctively aware of the persuasive potential of altering their moral language, but mild enough to suggest that such efforts were not a campaign priority.

Topics Discussed on Twitter by 2016 & 2020 U.S. Primary Candidates

A supplemental analysis was also conducted to assess the issues which candidates discussed on Twitter during the 2016 and 2020 elections. To achieve this, text networks were constructed using the text of all candidates in each primary. Containing only three candidates, the 2016 Democratic primary did not offer enough Twitter data for this analysis and was therefore excluded from the analysis. Since these networks were concerned with the substance of what was discussed, nouns and noun phrases were extracted from the tweets. Multi-terms (or *n*-grams) were then extracted using chi-square collocation detection. For this analysis, monograms, bi-grams, and tri-grams were included. The networks were then constructed to display the discursive categories present on Twitter according to the following steps:

Network Construction

Following a modified version of the framework laid out by Rule *et al*. (2015), the construction of the semantic network took place over the same stages expressed previously in this paper. Instances will only be noted where the steps diverge. Most notable was that edges were drawn between nodes based on the co-occurrence of the terms.

I. The initial co-occurence matrix M was defined by the number of times term i and term j appeared together in the same tweet.

II. Using the initial matrix, a proximity score was then calculated to measure similarity for each pair of terms. This proximity score allowed for the construction of a network where terms are nodes and edge weights express the similarity of the terms. Since a one-mode, term-term network such as this one constitutes a homogenous network structure, indirect measures of similarity are preferable (Weeds & Weir, 2005). Accordingly, a distributional measure of similarity was implemented which calculated the relatedness of each term pair based on their respective context (Weeds & Weir, 2005):

$$S(i,j) = \frac{\sum_{k \neq i,j, I(i,j)>0} min(I(i,k), I(j,k))}{\sum_{k \neq i,j, I(i,j)>0} I(i,k)},$$

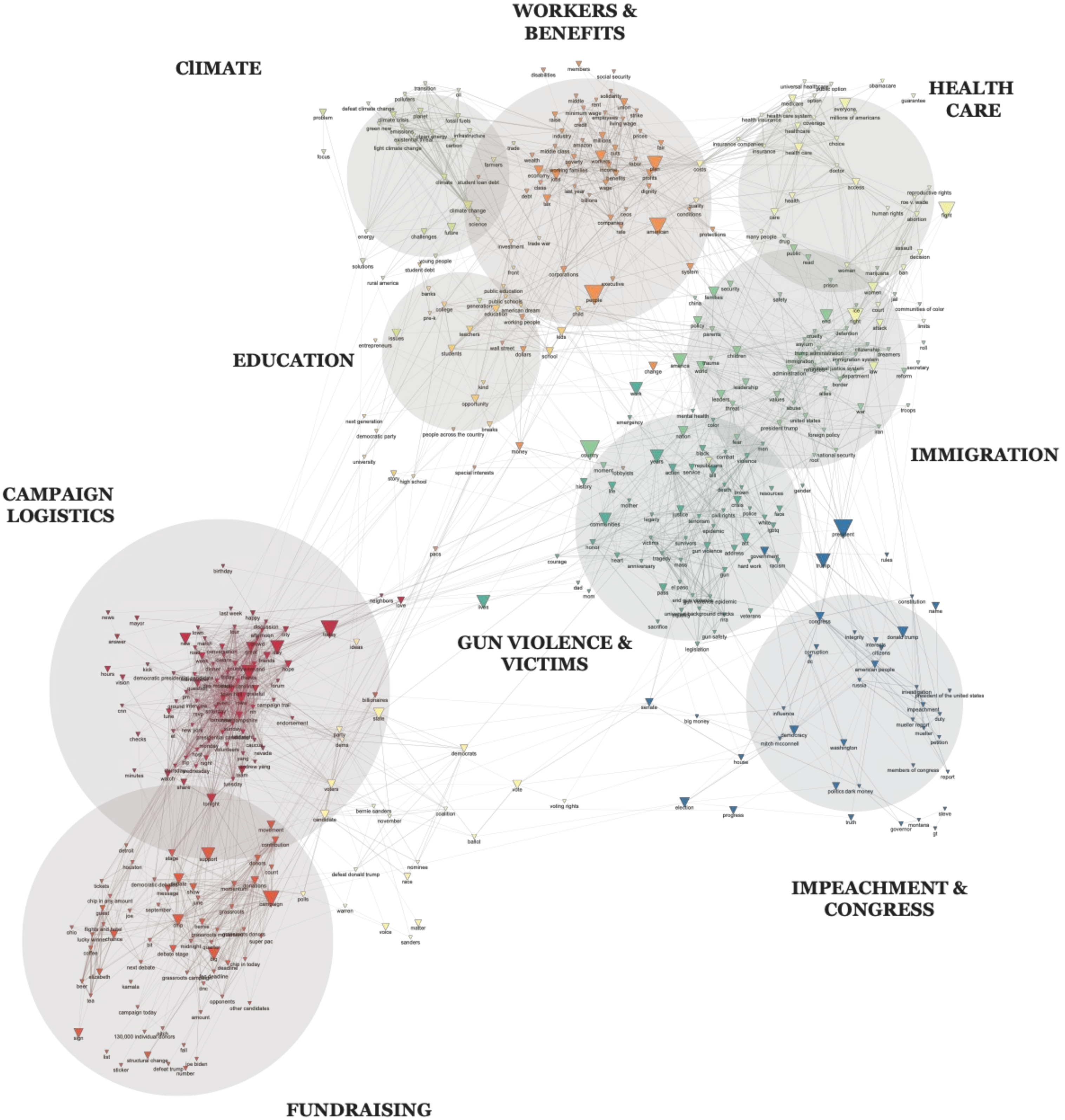
where $I(j,k)$ is the pointwise mutual information between two terms j and k, and a context word k is considered related to word j if their co-occurrence is higher than would be expected in an uncorrelated distribution (Rule *et al.*, 2015). In other words, rather than mapping only raw co-occurrences in the corpus, a distance-weighted averaging model such as the one used here is able to predict unseen co-occurrences of words by combining estimates for co-occurrences of similar words. This has the effect of smoothing the distribution and improving probability estimates (Lee, 1999).

III. Edges in the network were then filtered based on the same two criteria expressed earlier in this paper: the cohesiveness of the final network and the total number of edges. The lowest possible θ was selected which still produced a network with no disconnected components using the fewest total number of edges.

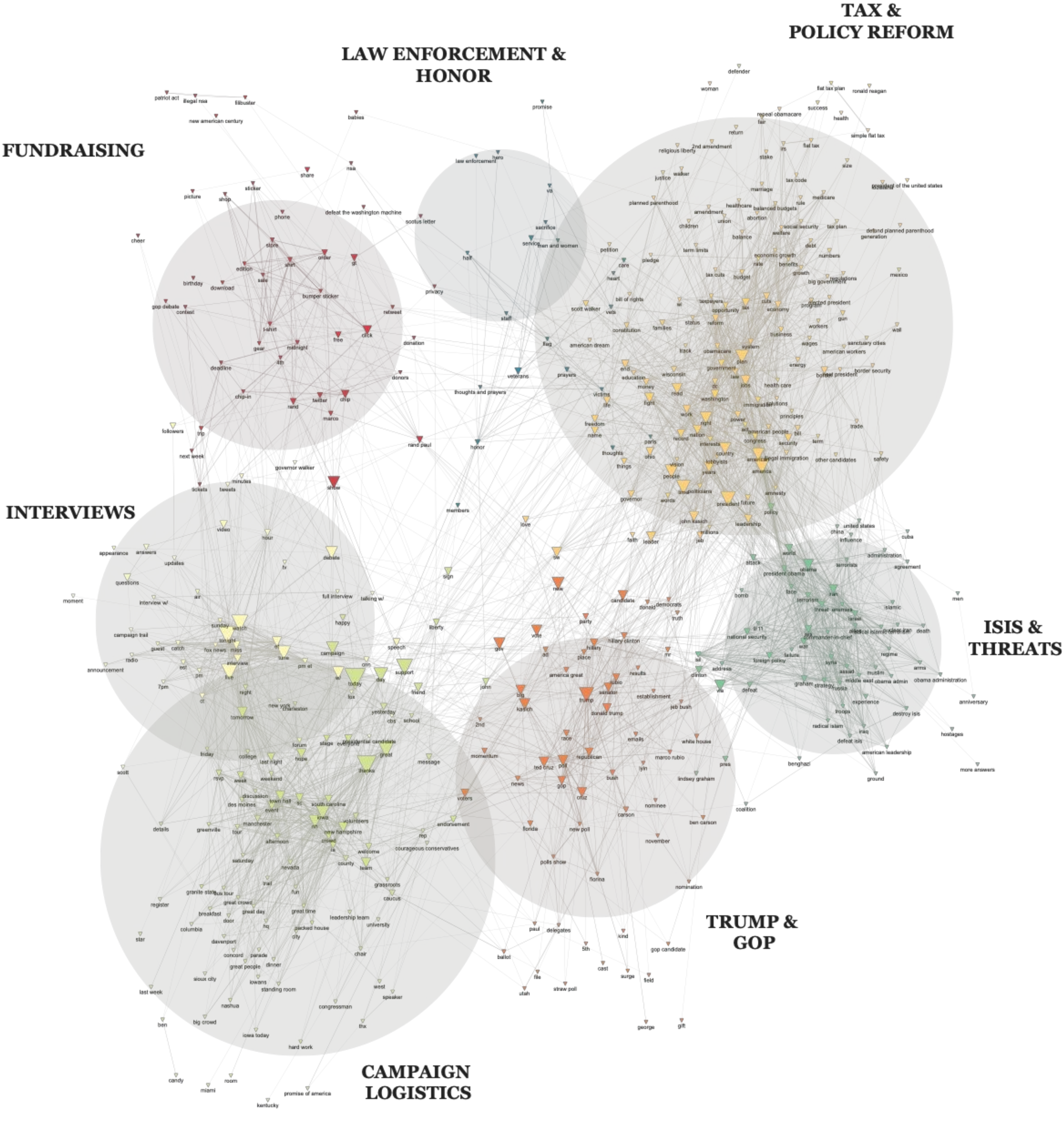
IV. A Louvain community detection algorithm (Blondel *et al.*, 2008) was then applied to the network to identify clusters of similar terms. These clusters are the targets of the analysis, and can be interpreted as discursive categories discussed on Twitter.

V. Finally, clusters were tagged with names based on the names of the two nodes with the highest node-to-cluster contribution score. This node-to-cluster contribution score was calculated following the steps established by (Rule *et al.*, 2015), and was defined as the weighted ratio of edges connecting terms to the same cluster. This score measures the extent to which the area around a word contains words that have been assigned to the same cluster; therefore, the final visualization of the network provides not just a list of terms within a cluster, but also a visualization of the inner structure of each cluster and the connections between clusters. Node sizes scale with the sum of their co-occurrences. Circles were centered on top of clusters as a visual aid; circle area was scaled with the number of words included in the cluster.

Networks were constructed for the 2016 Republican and 2020 Democratic primary. They are displayed in **Appendix Figure 2** and **Appendix Figure 3**.



Appendix Figure 2 Global text network showing discursive categories present on Twitter during the 2020 U.S. Democratic primary. Network generated from a corpus containing each of the 83,274 tweets generated by the 21 Democratic candidates who participated in at least two official primary debates.



Appendix Figure 3 Global text network showing discursive categories present on Twitter during the 2016 U.S. Republican primary. Network generated from a corpus containing each of the 40,607 tweets generated by the 15 Republican candidates who participated in at least two official primary debates.

B. Methodology Appendix

III. Methodological Background

Text as Data

In political eras of the past, the vast majority of campaign communication went unrecorded. This made studying candidate rhetoric difficult and largely defined the methodological approaches available to researchers. Textual records from a political campaign, for example, might include transcripts from a handful of speeches (for example: Ellsworth, 1965). Recently, however, developing technologies have facilitated the mass transcription and storage of political speech in digital formats which invite structured analysis. The migration of politicians and journalists on to social media platforms like Twitter has furthered this trend, removing even the need for transcription services or digitization of records. Consequently, the field of political science now faces "a genuine embarrassment of riches" with regard to textual data (Benoit, 2020, p. 1). This revolution in the availability of political rhetoric as textual data has broadened the scope of questions that can be investigated empirically, as well as the range of political actors to which they can be applied (Benoit, 2020).

Traditional methodological approaches are ill-equipped to manage this changing research landscape. Standard methods of quantitative content analysis, for example, involve the human coding of texts into researcher-defined categories (Krippendorff, 1980). While useful for research questions involving trends in article headlines or media portrayals of certain groups or issues (Thomas, 1994), content analysis are often limited by the amount of text the researcher is able to manually process. Additionally, because coding categories are researcher-defined, it can be difficult to use these methods to uncover latent patterns or unexpected trends in data (Benoit, 2018). Finally, unless hybrid methods of computer-assisted text analysis are employed alongside standard content analysis practices (Anstead, 2018), they typically lack the computational capacity to shed light on subtler, word-level patterns in text. Methods of "discourse analysis" face even greater challenges in this regard: the depth of analysis required to connect the substance of political texts to patterns of knowledge and power in social structures makes processing even a handful of speeches in this manner a daunting task (Fairclough, 2015).

Text Networks

Constructing networks from text is a novel approach: historically, network analysis been most often used to study social interactions and community relationships in the fields of biology, sociology, and anthropology (Borgatti *et al*., 2009), or in other social sciences (Ward *et al*., 2011) for example to study hierarchies and actor connections in political groups (Krebs, 2002). Analyses of these "social" networks tend to to focus on people, communities, and organisms, and often result in networks of a relatively small scale (Butts, 2009). However, disruptive innovations to established research procedures[15] have in recent times facilitated the construction of networks from vastly larger and more complex data sets, containing millions of nodes rather than tens, hundreds or even thousands (Watts, 2004; Barabási & Albert, 2011; Kitchin, 2014). This computational capacity to manage dense, large-scale networks has allowed for an application of these methods to research questions related to the use of natural language.

IV. TF-IDF Weighting Notes

Filtering out hyper-frequent, hyper-generic moral language also serves to "tune" the dictionary to the specific context of this research. The MFD 2.0 was not specifically designed to be implemented on a corpus containing only American presidential campaign rhetoric, and as a result there are words in the dictionary that take on less significance. For example, the word "president" is in the dictionary under the "authority" key. Obviously, counting "president" as an indicator for authoritarian moral reasoning is nonsensical in the context of a presidential campaign, where candidates of all ideologies will use the word frequently. This filtering therefore also serves to remove language that is uninformative for this research context.

During the construction of the aggregated candidate document-feature matrix, iterative tests concluded that as long as candidates were upsampled such that half were Democrat and half were Republican, the exact candidates added to the global document feature matrix (e.g. just 2016 election candidates, both 2016 and 2020 election candidates, primary candidates, both primary and general election candidates) did not have an impact on the moral terms that were ultimately removed during the filtering process explained in **3.2.6**. This suggests that the filtering succeeded in eliminating terms that were consistent not just across partisan affiliation but also election year and election phase. See validation tables below:

2016 Repub & Top 9 2020 Dems		**All 2016 Candidates**		**All 2016 & 2020 Candidates**	
Term	*TF-IDF Score*	*Term*	*TF-IDF Score*	*Term*	*TF-IDF Score*
president	0	president	0	president	0
together	0	together	0	together	0
safe	0	safe	0	safe	0
family	0	family	0	family	0
presidential	0	presidential	0	presidential	0

15 For example: increases in computing power, the emergence of "big data", and new analytics capabilities.

love	0	love	0	love	0
country	0	country	0	country	0
leader	0	leader	0	leadership	0
leadership	0	leadership	0	leaders	0
leaders	0	leaders	0	nation	0
nation	0	nation	0	protect	0
protect	0	protect	0	fight	0
fight	0	trust	0	law	0
law	0	fight	0	families	0
attacks	0	law	0		
families	0	attacks	0		
		families	0		

VI. Filtered Moral Foundations Dictionary

Every word in the list below was used at least three times by a single candidate and achieved a *tf-idf* frequency score of <.0014.

Care Virtue

healthcare, child, safety, loves, benefits, health, heal, sharing, healing, shared, helped, shares, compassion, protecting, protection, mother, healthy, help, patient, helps, relief, feed, vulnerable, benefit, protected, compassionate childhood, helping, patients, charity, cares, condolences, generous, safe, mothers, loved, healthier, protects, safely, kindness, nurses, hospital, loving, caring, care, mommy, childcare, empathy, share, humane, helpful, nursing, protective, wounded, mercy, rescue, comfort, wounds, comforted, hospitality, nurse, charitable, lover, generosity, relieve, feeds, feeding, wound, generously, healed, rescuing, healers, hug, safeguard

Care Vice

violence, destroy, kill, assault, threats, fights, killing, cruelty, cruel, pain, destroyed, fighter, suffering, vulnerable, killed, rape, destroying, hurt, damage, attacked, threatens, die, hurting, destruction, victims, violent, attack, hunger, harm, fighting, threaten, murdered, hurts, persecution, punch, fighters, bully, bullying, murder, threatening, attacks, threat, harassment, injured, suffer, killer, wounded, bullies, rapists, suffers, victim, genocidal, destroys, killers, abuses, crying, threatened, abusers, abused, wounds, suffered, endanger, harmful, abusing, exploit, distressed, brutality, damaging, exploiting, torture, harms, kills, tribulation, injury, endangered, punches, agony, wound, endangering, bullied, assaulting, needy, carnage, harsh, harmed, assassination, genocide, exploitation

Fairness Virtue

equality, rights, equal, justice, integrity, reparations law, equity, fair, honesty, trust, laws, honest, trusted, justices, lawyer, fairness, lawyers, equitable, compensation parity, equalizers karma, compensate, trusting, repay, justify, retaliate,

Fairness Vice

dishonest, fraud, biased, racism, racist, discrimination, lying, inequality, lied, liar, sexist, unfair, scam, stealing, injustice, disproportionately, crooked, hypocrite, theft, cheating, bias, fraudulent, unjust, hypocrisy, liars, oppression, injustices, segregation, stole, hypocrites, freeloading, bigoted, betrayal, sexism, cheat, segregated, prejudice, robbing, exploit, racists, discriminate, exploiting, inequities, betray, betrayed, misleading, disparity, crooks, discriminating, deception, discriminated, cheated, defrauded, unequal, stolen, steal, scammed, cheats, biases, crook, exploitation, scams, imbalances, robbed

Loyalty Virtue

communities, community, homeland, patriot, coalition, troops, companies, wife, unite, joining, groups, followers, allies, patriots, countries, belonging, unity, fellow, pledge, company, nation, nations, group, united, ally, sacrifice, belong, organization tribal, pledged, solidarity, war, uniting, collective, belongs, corps, sacrifices, together, organizations, unites, allegiance, loyal, uniter, collectively, familiar, insider, follower, sacrificed, allied, pledges, player, insiders, coalitions, tribe, enlist, indivisible, tribes, loyalty, pledging, cohorts, fellowship, cult, troop

Loyalty Vice

enemy, enemies, outsider, disloyal, traitor, rebels, betray, betrayed, treason

Authority Virtue

bosses, commander, governor, chief, police, governors, dictate, regulations, ruling, submit, respected, order, traditional, willing, captains, protecting, protection, duty, pope, father, authority, ranked, control, govern, controlled, institutions, policing, fathers, respect, servant, presidents, punish, command, protects, manager, ceo, tradition, dominate, honored, governing, regulation, worship, bully, allegiance, elderly, bullying, arrested, proper, managers, honor, dominating, institution, rank, dictators, institutional, commandments, admiral, honoring, submitted, submissions, captain, dictator, guide, dictating, oligarchy, bullies, punitive, boss, leader, respects, dictated, arrest, ordering, ordered, noble, authorities, dominant, polite, submission, obey, submitting, dominates, honors, respecting, punished, punishment, submits, arrests, chiefs, guiding, punishing, dictates, acquiesce, ranking, dean, bullied, slaves, honorable, governed, servants, elders, dominated, principal, mentor, permission, matriarch, comply, respectfully

Authority Vice

illegal, illegals, chaos, refuse, refuses, anarchists, lawless, refused, orders, unlawful, refusing, dissidents, rioters, disrespect, riots, rebels, overthrow, traditions, rioting, anarchy, uprising, overthrown, disrespects, treason, disorder

Sanctity Virtue

god, marriage, religious, clean, lord, bless, food, prayer, sanctity, soul, church, dignity, christian, bible, christians, pray, praying, pastor, prayers, spiritual, blessed, pope, yogi, faith, religion, body, decency, sacred, blood, mary, holy, pure, catholics, married, worship, faithful, gods, religions, biblical, blessings, spirituality, blessing, jesus, churches, catholic, immunity, noble, nuns, synagogue, christ, enshrined, cleaning, atone, immune, divinity, foods, purity, cleaner, marry, wholesome, souls, righteous, angel, dignified, mosque, pristine, temple

Sanctity Vice

corrupt, corruption, pandemic, epidemic, hell, virus, drug, waste, drugs, sexual, addiction, spreading, degrade, dirty, disgusting, damn, wasted, corrupting, horrific, wasting, disease, rot, sin, horrifying, rotten, dirt, horror, *corrupted, trash, degrading, viral, shit, diseases, contaminants, garbage, plague, alcoholism, sleazy, swear, plagued, stain, plagues, disgusted, fester, godless, damning, pandemics, addicted, incest, swore, contamination, infection, abhor, infected*

VII. Extended Results

Full tables with all raw and weighted proportions of moral language for each candidate. Raw proportions used the unfiltered MFD 2.0; weighted proportions used the TF-IDF weighted version.

Raw Proportions:

Raw Proportion of Moral Language: 2016 Democratic Primary

Candidate	care.virtue	care.vice	fairness.virtue	fairness.vice	loyalty.virtue	loyalty.vice	authority.virtue	authority.vice	sanctity.virtue	sanctity.vice	Care	Fairness	Loyalty	Authority	Sanctity	Prop. Virtue	Prop Vice
Hillary Clinton	0.2498239	0.1509744	0.1258511	0.03968068	0.2099084	0.00164358	0.1784456	0.007513501	0.03615872	0.02418408	0.4007983	0.16553178	0.21155198	0.1859591	0.0603428	0.80018772	0.22399624
Bernie Sanders	0.2243103	0.08876449	0.07497001	0.03718513	0.3570572	0.00039984	0.1497401	0.005797681	0.03018792	0.04338265	0.31307479	0.11215514	0.35745704	0.15553778	0.07357057	0.83626553	0.17552979
Martin O'Malley	0.2021818	0.07418182	0.1076364	0.02036364	0.2210909	0.00145455	0.2887273	0.002181818	0.08363636	0.01381818	0.27636362	0.12800004	0.22254545	0.29090912	0.09745454	0.90327276	0.112
Average	**0.22543867**	**0.10464024**	**0.10281917**	**0.03240982**	**0.2626855**	**0.00116599**	**0.205637667**	**0.005164333**	**0.049994333**	**0.027128303**	**0.3300789**	**0.13522899**	**0.26385149**	**0.210802**	**0.07712264**	**0.84657534**	**0.17050868**

Raw Proportion of Moral Language: 2020 Democratic Primary

Candidate	care.virtue	care.vice	fairness.virtue	fairness.vice	loyalty.virtue	loyalty.vice	authority.virtue	authority.vice	sanctity.virtue	sanctity.vice	Care	Fairness	Loyalty	Authority	Sanctity	Prop. Virtue	Prop Vice
Joe Biden	0.2174379	0.1275683	0.06932285	0.01893872	0.251206	0.00303734	0.2226193	0.005538681	0.06271217	0.04609612	0.3450062	0.08826157	0.25424334	0.22815798	0.10880829	0.82329822	0.20117916
Bernie Sanders	0.2951413	0.1222211	0.08349261	0.03480535	0.257117	0.00321899	0.1289609	0.008449854	0.02846796	0.05230862	0.4173624	0.11829796	0.26033599	0.13741075	0.08077658	0.79317977	0.22100392
Elizabeth Warren	0.2144479	0.2272446	0.08820089	0.0372893	0.2282078	0.000344	0.1375301	0.005985552	0.02187822	0.05304438	0.4416925	0.12549019	0.2285518	0.14351565	0.0749226	0.69026491	0.32390783
Amy Klobuchar	0.2263856	0.1074681	0.08665105	0.0156128	0.2466823	0.00156128	0.2534478	0.007285974	0.03148582	0.04553734	0.3338537	0.10226385	0.24824358	0.26073377	0.07702316	0.84465257	0.17746549
Pete Buttigieg	0.2348873	0.08760246	0.08606557	0.03150615	0.3007172	0.00204918	0.2005635	0.006915984	0.03714139	0.01895492	0.32248976	0.11757172	0.30276638	0.20747948	0.05609631	0.85937496	0.14702869
Andrew Yang	0.2742857	0.07183673	0.04707483	0.01605442	0.2778231	0.0029932	0.2234014	0.001632653	0.04680272	0.04435374	0.34612243	0.06312925	0.2808163	0.22503405	0.09115646	0.86938775	0.13687074
Tom Steyer	0.2558654	0.1164232	0.0894201	0.09871625	0.1646746	0.00575476	0.2246569	0.008410801	0.02567508	0.03098716	0.3722886	0.18813635	0.17042936	0.2330677	0.05666224	0.76029208	0.26029217
Kamala Harris	0.2292484	0.1804041	0.1139701	0.02936473	0.2025004	0.0014537	0.1955226	0.006105539	0.03445268	0.02500363	0.4096525	0.14333483	0.2039541	0.20162814	0.05945631	0.77569418	0.2423317
Cory Booker	0.2180451	0.1510394	0.1163202	0.03162318	0.2494471	0.00088456	0.1508182	0.005307386	0.0495356	0.036046	0.3690845	0.14794338	0.25033166	0.15612559	0.0855816	0.7841662	0.22490053
Tulsi Gabbard	0.2104466	0.09386828	0.04542014	0.01400454	0.3141559	0.00264951	0.2293717	0.008705526	0.05185466	0.05563967	0.30431488	0.05942468	0.31680541	0.23807723	0.10749433	0.851249	0.17486752
Julian Castro	0.1954691	0.1521652	0.09462711	0.03348035	0.249599	0.00080192	0.2323577	0.009021652	0.03428228	0.02085004	0.3476343	0.12810746	0.25040092	0.24137935	0.05513232	0.80633519	0.21631917
Beto O'Rourke	0.1863199	0.1442822	0.1054507	0.03651585	0.3174207	0.0007125	0.1453509	0.009084432	0.0320627	0.03651585	0.3306021	0.14196655	0.3181332	0.15443533	0.06857855	0.7866049	0.22711084
Michael Bloomberg	0.2295831	0.1524843	0.06681896	0.01827527	0.2250143	0.0005711	0.2415762	0.006282125	0.03940605	0.04283267	0.3820674	0.08509423	0.2255854	0.24785833	0.08223872	0.80239861	0.22044547
Michael Bennet	0.2413793	0.08497537	0.06311576	0.02339901	0.3118842	0.00092365	0.221367	0.007697044	0.02463054	0.03109606	0.32635467	0.08651477	0.31280785	0.22906404	0.0557266	0.8623768	0.14809113
John Delaney	0.2373658	0.0936513	0.06901453	0.02226785	0.2736892	0.00268478	0.2163613	0.008843967	0.057012	0.0315856	0.3310171	0.09128238	0.27637398	0.22520527	0.0885976	0.85344283	0.15903349
Marianne Williamson	0.2880926	0.09734763	0.0736456	0.03301354	0.1901806	0.00112867	0.1707111	0.008747178	0.1052483	0.04373589	0.38544023	0.10665914	0.19130927	0.17945828	0.14898419	0.8278782	0.18397291
Tim Ryan	0.3138745	0.1015594	0.03478609	0.0159936	0.267493	0.00039984	0.1823271	0.0019992	0.06077569	0.03478609	0.4154339	0.05077969	0.26789284	0.1843263	0.09556178	0.85925638	0.15473813
Bill de Blasio	0.2485929	0.1444653	0.06660413	0.04221388	0.2579737	0.00375235	0.1697936	0.00750469	0.04315197	0.03095685	0.3930582	0.10881801	0.26172605	0.17729829	0.07410882	0.7861163	0.22889307
Kirsten Gillibrand	0.2407643	0.1987261	0.111465	0.02101911	0.160828	0	0.2073248	0.005095541	0.02770701	0.04426752	0.4394904	0.13248411	0.160828	0.21242034	0.07197453	0.74808911	0.26910827
Jay Inslee	0.1873947	0.1358276	0.08122683	0.03404112	0.2244692	0.00067408	0.2339063	0.009100101	0.08527132	0.02831143	0.3232221	0.11526795	0.22514328	0.2430064	0.11358275	0.81226835	0.20795413
John Hickenlooper	0.2232885	0.1241203	0.0806142	0.01983365	0.2354447	0.00191939	0.2725528	0.004478567	0.03454894	0.02175304	0.3474088	0.10044785	0.23736409	0.27703137	0.05630198	0.84644914	0.17210494
Steve Bullock	0.1415424	0.1612187	0.1142494	0.01110758	0.2392891	0.00095208	0.2748334	0.005077753	0.01142494	0.05426849	0.3027611	0.12535698	0.24024118	0.27991115	0.06569343	0.78133924	0.2326246
Eric Swalwell	0.2282769	0.2371134	0.0736377	0.02135493	0.1745214	0.00368189	0.2091311	0.01546392	0.0353461	0.03681885	0.4653903	0.09499263	0.17820329	0.22459502	0.07216495	0.7209132	0.31443299
Average	**0.23209283**	**0.13537447**	**0.080921493**	**0.0287144**	**0.24436254**	**0.00183255**	**0.206281987**	**0.007075397**	**0.042646702**	**0.037641303**	**0.36746731**	**0.10963589**	**0.2461951**	**0.21335738**	**0.080288**	**0.80630556**	**0.21063813**
Top 9 Average	**0.24063829**	**0.13242311**	**0.086724244**	**0.03487899**	**0.24204172**	**0.00236633**	**0.193057856**	**0.00618138**	**0.037572404**	**0.03914799**	**0.3730614**	**0.12160323**	**0.24440806**	**0.19923924**	**0.07672039**	**0.80003452**	**0.2149978**

Raw Proportion of Moral Language: 2016 Republican Primary

Candidate	care.virtue	care.vice	fairness.virtue	fairness.vice	loyalty.virtue	loyalty.vice	authority.virtue	authority.vice	sanctity.virtue	sanctity.vice	Care	Fairness	Loyalty	Authority	Sanctity	Prop. Virtue	Prop Vice
Donald Trump	0.1629213	0.09550562	0.05176565	0.08868379	0.1729535	0.00842697	0.2728732	0.05216693	0.05497592	0.04895666	0.25842692	0.14044944	0.18138047	0.32504013	0.10393258	0.71548957	0.29373997
Ted Cruz	0.1244186	0.1148256	0.06860465	0.01511628	0.2194767	0.01046512	0.3218023	0.02180233	0.1043605	0.01627907	0.2392442	0.08372093	0.22994182	0.34360463	0.12063957	0.83866275	0.1784884
John Kasich	0.1867794	0.1014604	0.02997694	0.00538048	0.2313605	0.00461184	0.3428132	0.006917756	0.08685626	0.007686395	0.2882398	0.03535742	0.23597234	0.34973096	0.09454266	0.8777863	0.12605687
Marco Rubio	0.1755408	0.1023295	0.01497504	0.00249584	0.2512479	0.00499168	0.3793677	0.004159734	0.04908486	0.02579035	0.2778703	0.01747088	0.25623958	0.38352743	0.07487521	0.8702163	0.13976711
Ben Carson	0.1810585	0.08077994	0.06128134	0.00928505	0.2896936	0.00928505	0.2553389	0.01207057	0.1253482	0.01114206	0.26183844	0.07056639	0.29897865	0.26740947	0.13649026	0.91272054	0.12256267
Jeb Bush	0.1407307	0.1278755	0.0338295	0.0067659	0.2050068	0.00405954	0.394452	0.01894452	0.0744249	0.02097429	0.2686062	0.0405954	0.20906634	0.41339652	0.09539919	0.8484439	0.17861975
Chris Christie	0.1608696	0.1217391	0.06086957	0.00434783	0.2347826	0	0.2434783	0	0.09130435	0.1043478	0.2826087	0.0652174	0.2347826	0.2434783	0.19565215	0.79130442	0.23043473
Rand Paul	0.1820491	0.1003387	0.09229467	0.01397121	0.2464014	0.00677392	0.2548688	0.05165114	0.03175275	0.03302286	0.2823878	0.10626588	0.25317532	0.30651994	0.06477561	0.80736672	0.20575783
Carly Fiorina	0.2053292	0.09404389	0.05015674	0.00940439	0.3040752	0.01097179	0.2805643	0.01097179	0.03134796	0.01253918	0.29937309	0.05956113	0.31504699	0.29153609	0.04388714	0.8714734	0.13793104
Mike Huckabee	0.1405325	0.1849112	0.07988166	0.02810651	0.1434911	0.00961539	0.2588757	0.02366864	0.1442308	0.02662722	0.3254437	0.10798817	0.15310649	0.28254434	0.17085802	0.76701176	0.27292896
Scott Walker	0.1383855	0.1136738	0.06919275	0.00988468	0.1911038	0.00823723	0.416804	0.01647446	0.03789127	0.01812191	0.2520593	0.07907743	0.19934103	0.43327846	0.05601318	0.85337732	0.16639208
Lindsey Graham	0.1154434	0.2385321	0.03287462	0.00840979	0.2568807	0.01223242	0.2775229	0.005351682	0.04357798	0.02140673	0.3539755	0.04128441	0.26911312	0.28287458	0.06498471	0.7262996	0.28593272
Bobby Jindal	0.1952462	0.1086587	0.05602716	0.02037351	0.1375212	0.01018676	0.3089983	0.02716469	0.1375212	0.01697793	0.3039049	0.07640067	0.14770796	0.33616299	0.15449913	0.83531406	0.18336159
Rick Santorum	0.1252144	0.1252144	0.04459691	0.01372213	0.329331	0.01715266	0.2161235	0.01543739	0.1132075	0.01886792	0.2504288	0.05831904	0.34648366	0.23156089	0.13207542	0.82847331	0.1903945
George Pataki	0.08914729	0.1802326	0.07945736	0.00581395	0.1937984	0.01744186	0.3333333	0.007751938	0.09108527	0.02713178	0.26937989	0.08527131	0.21124026	0.34108524	0.11821705	0.78682162	0.23837213
Rick Perry	0.1681416	0.08849558	0.01769912	0	0.2300885	0.00884956	0.3451327	0.02654867	0.1327434	0.02654867	0.25663718	0.01769912	0.23893806	0.37168137	0.15929207	0.89380532	0.15044248
Average	**0.15573801**	**0.12366354**	**0.05271773**	**0.01511008**	**0.22732581**	**0.00895636**	**0.306396819**	**0.01881764**	**0.08435707**	**0.027276302**	**0.27940155**	**0.06782781**	**0.23628217**	**0.32521446**	**0.11163337**	**0.82653543**	**0.19382393**
Top 9 Average	**0.16885524**	**0.10432203**	**0.051528233**	**0.01727231**	**0.23944424**	**0.00662066**	**0.305062078**	**0.019853863**	**0.072161744**	**0.031193185**	**0.27317727**	**0.06880054**	**0.2460649**	**0.32491594**	**0.10335493**	**0.83705154**	**0.17926204**

Raw Proportion of Moral Language: 2016 & 2020 General Election

Candidate	care.virtue	care.vice	fairness.virtue	fairness.vice	loyalty.virtue	loyalty.vice	authority.virtue	authority.vice	sanctity.virtue	sanctity.vice	Care	Fairness	Loyalty	Authority	Sanctity	Prop. Virtue	Prop Vice
Hillary Clinton 2016	0.2498239	0.1509744	0.1258511	0.03968068	0.2099084	0.00164358	0.1784456	0.007513501	0.03615872	0.02418408	0.4007983	0.16553178	0.21155198	0.1859591	0.0603428	0.80018772	0.22399624
Donald Trump 2016	0.1226081	0.1119773	0.04181432	0.2445074	0.1736357	0.00708717	0.2161588	0.02267895	0.02976612	0.03827073	0.2345854	0.28632172	0.18072287	0.23883775	0.06803685	0.58398304	0.42452155
Joe Biden 2020	0.2106653	0.08975713	0.0657339	0.01663147	0.2521119	0.00184794	0.2552798	0.01187962	0.05491024	0.05517423	0.30042243	0.08236537	0.25395984	0.26715942	0.11008447	0.83870114	0.17529039
Donald Trump 2020	0.1637329	0.09513831	0.05965353	0.04875664	0.1739313	0.00922045	0.3023191	0.04316848	0.04708019	0.07166806	0.25887121	0.10841017	0.18315175	0.34548758	0.11874825	0.74671702	0.26795194
Average	**0.18670755**	**0.11196179**	**0.073263213**	**0.08739405**	**0.20239683**	**0.00494979**	**0.238050825**	**0.021310138**	**0.041978818**	**0.047324275**	**0.29866934**	**0.16065726**	**0.20734661**	**0.25936096**	**0.08930309**	**0.74239723**	**0.27294003**
Dem. Average	**0.2302446**	**0.12036577**	**0.0957925**	**0.02815608**	**0.23101015**	**0.00174576**	**0.2168627**	**0.009696561**	**0.04553448**	**0.039679155**	**0.35061037**	**0.12394858**	**0.23275591**	**0.22655926**	**0.08521364**	**0.81944443**	**0.19964332**
Repub. Average	**0.1431705**	**0.10355781**	**0.050733925**	**0.14663202**	**0.1737835**	**0.00815381**	**0.25923895**	**0.032923715**	**0.038423155**	**0.054969395**	**0.24672831**	**0.19736595**	**0.18193731**	**0.29216267**	**0.09339255**	**0.66535003**	**0.34623675**

Weighted Proportions:

Weighted Proportion of Moral Language: 2016 Democratic Primary

Candidate	care.virtue	care.vice	fairness.virtue	fairness.vice	loyalty.virtue	loyalty.vice	authority.virtue	authority.vice	sanctity.virtue	sanctity.vice	Care	Fairness	Loyalty	Authority	Sanctity	Prop. Virtue	Prop Vice
Hillary Clinton	0.309854	0.168981	0.1804944	0.054521	0.1283441	0.002032	0.08262784	0.01015916	0.05045716	0.03250931	0.478835	0.235015	0.130376	0.092787	0.082966	0.751778	0.268202
Bernie Sanders	0.292724	0.08475	0.1031503	0.048787	0.2732088	0.000279	0.1115138	0.007805966	0.04153889	0.05492055	0.377474	0.151938	0.273488	0.11932	0.096459	0.822135	0.196543
Martin O'Malley	0.247716	0.069036	0.1492386	0.024365	0.177665	0.00203	0.2142132	0	0.1116751	0.01827411	0.316751	0.173604	0.179695	0.214213	0.129949	0.900508	0.113706
Average	**0.283431**	**0.107589**	**0.144294433**	**0.042558**	**0.19307263**	**0.001447**	**0.13611828**	**0.005988375**	**0.06789038**	**0.035234657**	**0.39102**	**0.186852**	**0.19452**	**0.142107**	**0.103125**	**0.824807**	**0.192817**

Weighted Proportion of Moral Language: 2020 Democratic Primary

Candidate	care.virtue	care.vice	fairness.virtue	fairness.vice	loyalty.virtue	loyalty.vice	authority.virtue	authority.vice	sanctity.virtue	sanctity.vice	Care	Fairness	Loyalty	Authority	Sanctity	Prop. Virtue	Prop Vice
Joe Biden	0.272222	0.144697	0.09722222	0.025	0.2330808	0.003283	0.09040404	0.007070707	0.08510101	0.06035354	0.416919	0.122222	0.236364	0.097475	0.145455	0.77803	0.240404
Bernie Sanders	0.358743	0.101375	0.108186	0.043091	0.2043222	0.004191	0.07845449	0.01087099	0.03588736	0.0508186	0.460118	0.151277	0.208513	0.089325	0.086706	0.785593	0.210347
Elizabeth Warren	0.280085	0.171216	0.1345725	0.055231	0.1850239	0.000531	0.07392459	0.009028147	0.0330324	0.07923526	0.451301	0.189804	0.185555	0.082953	0.112268	0.706638	0.315242
Amy Klobuchar	0.296224	0.112495	0.1280654	0.021798	0.2207084	0.001946	0.1171662	0.01089918	0.0420397	0.06656286	0.408719	0.149864	0.222655	0.128065	0.108603	0.804204	0.213702
Pete Buttigieg	0.301765	0.095067	0.1188333	0.041772	0.2772776	0.001801	0.09326611	0.007922218	0.04933381	0.02520706	0.396831	0.160605	0.279078	0.101188	0.074541	0.840475	0.171768
Andrew Yang	0.312784	0.086885	0.0666115	0.021928	0.266446	0.003724	0.1232933	0.00206868	0.06702524	0.06123293	0.399669	0.08854	0.27017	0.125362	0.128258	0.83616	0.175838
Tom Steyer	0.317225	0.107916	0.1185888	0.129558	0.1411207	0.007708	0.1082123	0.01126594	0.03261192	0.0391343	0.425141	0.248147	0.148829	0.119478	0.071746	0.717759	0.295583
Kamala Harris	0.302534	0.172663	0.1664892	0.040026	0.1675538	0.000852	0.07579306	0.008728976	0.04960613	0.034703	0.475197	0.206515	0.168405	0.084522	0.084309	0.761976	0.256973
Cory Booker	0.2556	0.150353	0.1595581	0.042037	0.2108009	0.001227	0.07456275	0.006443694	0.06443694	0.0475606	0.405953	0.201596	0.212028	0.081006	0.111998	0.764959	0.247622
Tulsi Gabbard	0.204878	0.109485	0.06341463	0.020054	0.3284553	0.003794	0.1333333	0.009756098	0.07154472	0.07533875	0.314363	0.083469	0.332249	0.143089	0.146883	0.801626	0.218428
Julian Castro	0.243889	0.175227	0.1282615	0.041472	0.2199945	0.001099	0.1354024	0.01180994	0.04449327	0.02609173	0.419116	0.169734	0.221093	0.147212	0.070585	0.772041	0.255699
Beto O'Rourke	0.230309	0.175474	0.1468096	0.050349	0.2465105	0.000997	0.06256231	0.01221336	0.04112662	0.04785643	0.405783	0.197159	0.247508	0.074776	0.088983	0.727318	0.286889
Michael Bloomberg						OUTLIER											
Michael Bennet	0.346862	0.094856	0.09532798	0.03445	0.2548372	0.001416	0.09155262	0.01179802	0.0363379	0.04624823	0.441718	0.129778	0.256253	0.103351	0.082586	0.824917	0.188768
John Delaney	0.308078	0.107555	0.09274984	0.02678	0.2379708	0.00283	0.1143044	0.01088613	0.07424341	0.0385369	0.415633	0.11953	0.240801	0.125191	0.11278	0.827346	0.186588
Marianne Williamson	0.277198	0.112146	0.09538003	0.038376	0.1833085	0	0.1244411	0.009687034	0.1285395	0.04880775	0.389344	0.133756	0.183309	0.134128	0.177347	0.808867	0.209016
Tim Ryan	0.408988	0.116041	0.04948805	0.021047	0.2178612	0.000569	0.07224118	0.002844141	0.08191126	0.04607509	0.525029	0.070535	0.21843	0.075085	0.127986	0.830489	0.186576
Bill de Blasio						OUTLIER											
Kirsten Gillibrand	0.334956	0.187926	0.1694255	0.031646	0.1051607	0	0.07692308	0.006815969	0.04089581	0.06377799	0.522882	0.201071	0.105161	0.083739	0.104674	0.727361	0.290166
Jay Inslee	0.21912	0.145291	0.1126361	0.04638	0.2096545	0.000947	0.1183152	0.01183152	0.1173687	0.03549456	0.364411	0.159016	0.210601	0.130147	0.152863	0.777094	0.239943
John Hickenlooper	0.262976	0.140138	0.1081315	0.025952	0.2128028	0.00173	0.183391	0.005190311	0.04238754	0.02595156	0.403114	0.134083	0.214533	0.188581	0.068339	0.809689	0.198962
Steve Bullock	0.202172	0.115822	0.1845915	0.018097	0.2088935	0.001034	0.1685626	0.00672182	0.01706308	0.08583247	0.317994	0.202689	0.209928	0.175284	0.102896	0.781282	0.227508
Eric Swalwell	0.248958	0.289583	0.1010417	0.026042	0.146875	0.004167	0.08645833	0.01875	0.04791667	0.04791667	0.538542	0.127083	0.151042	0.105208	0.095833	0.63125	0.386458
Average	**0.285027**	**0.138677**	**0.116446902**	**0.038147**	**0.21326947**	**0.002088**	**0.104884017**	**0.009171565**	**0.05728109**	**0.050130299**	**0.423704**	**0.154594**	**0.215357**	**0.114056**	**0.107411**	**0.776908**	**0.238213**
Top 9 Average	**0.299687**	**0.126963**	**0.122014113**	**0.046716**	**0.21181492**	**0.002807**	**0.092786316**	**0.008255392**	**0.05100828**	**0.05164535**	**0.42665**	**0.16873**	**0.214622**	**0.101042**	**0.102654**	**0.77731**	**0.236386**

Weighted Proportion of Moral Language: 2016 Republican Primary

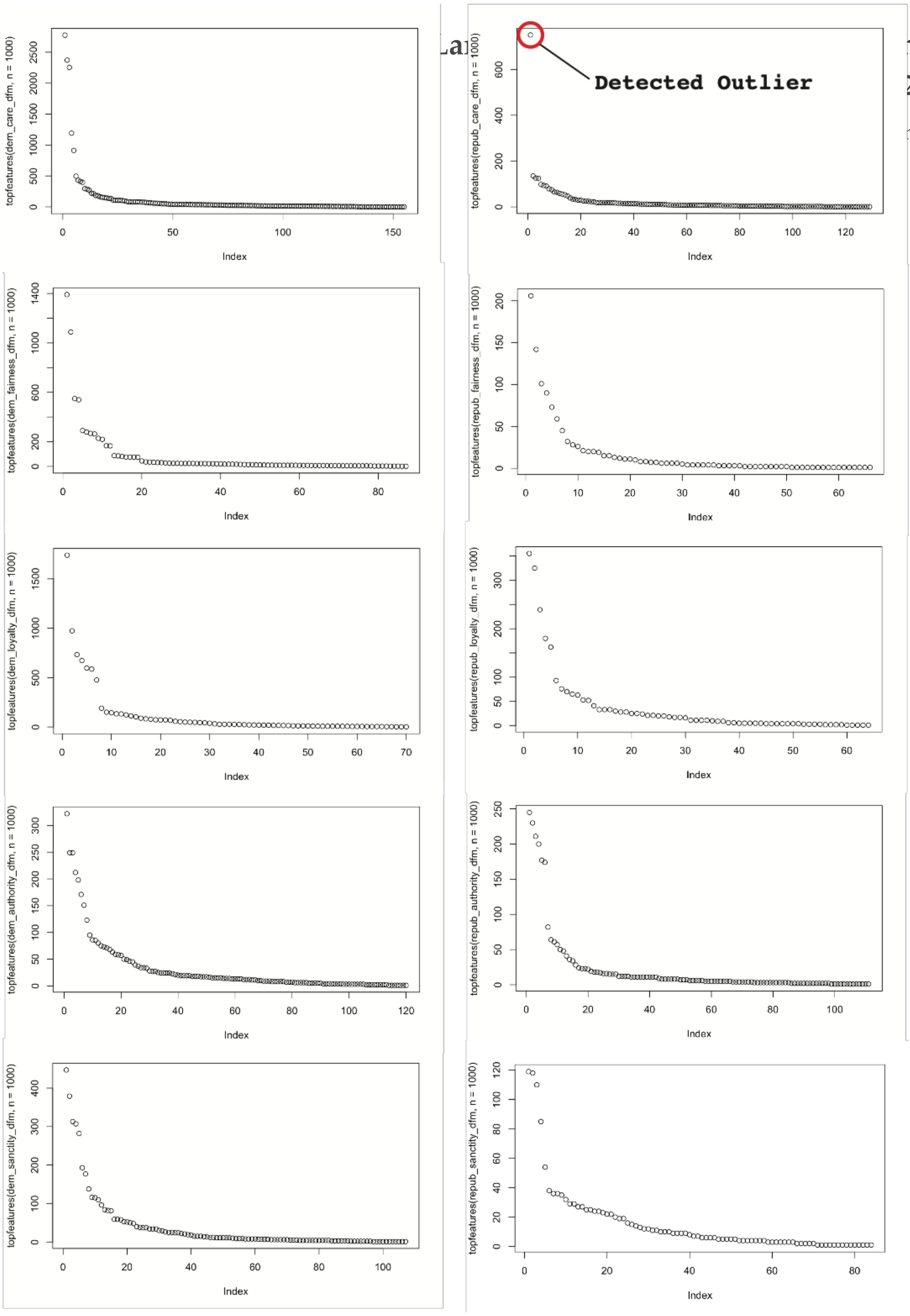
Candidate	care.virtue	care.vice	fairness.virtue	fairness.vice	loyalty.virtue	loyalty.vice	authority.virtue	authority.vice	sanctity.virtue	sanctity.vice	Care	Fairness	Loyalty	Authority	Sanctity	Prop. Virtue	Prop Vice
Donald Trump	0.13528	0.117402	0.07449344	0.125149	0.135876	0.011323	0.1966627	0.07449344	0.07687723	0.06257449	0.252682	0.199642	0.147199	0.271156	0.139452	0.619189	0.390942
Ted Cruz	0.153782	0.107397	0.09694944	0.019641	0.2377768	0.014626	0.1905558	0.02966987	0.1429168	0.02173005	0.261179	0.11659	0.252403	0.220226	0.164647	0.821981	0.193063
John Kasich	0.162382	0.109608	0.02435724	0.00406	0.2909337	0.008119	0.2922869	0.00541272	0.07577808	0.03924222	0.271989	0.028417	0.299053	0.2977	0.11502	0.845738	0.166441
Marco Rubio	0.250605	0.119855	0.04600484	0.007264	0.2082324	0.007264	0.2288136	0.01089588	0.12954	0.01210654	0.37046	0.053269	0.215496	0.239709	0.141647	0.863196	0.157385
Ben Carson	0.18344	0.090446	0.07898089	0.008917	0.3057325	0.010191	0.1528662	0.01656051	0.1630573	0.01146497	0.273885	0.087898	0.315924	0.169427	0.174522	0.884076	0.13758
Jeb Bush	0.178534	0.154091	0.05100956	0.009564	0.2082891	0.006376	0.2497343	0.02763018	0.1105207	0.03188098	0.332625	0.060574	0.214665	0.277364	0.142402	0.798087	0.229543
Chris Christie						OUTLIER											
Rand Paul	0.220678	0.068371	0.1200667	0.01612	0.2390217	0.007226	0.1962201	0.06670372	0.03835464	0.04113396	0.28905	0.136187	0.246248	0.262924	0.079489	0.814341	0.199555
Carly Fiorina	0.269316	0.064018	0.06843267	0.013245	0.3642384	0.013245	0.1456954	0.01545254	0.0419426	0.01103753	0.333333	0.081678	0.377483	0.161148	0.05298	0.889625	0.116998
Mike Huckabee	0.155886	0.149523	0.1134677	0.037116	0.1314952	0.012725	0.1569459	0.03075292	0.2036055	0.03181336	0.305408	0.150583	0.144221	0.187699	0.235419	0.7614	0.26193
Scott Walker	0.162353	0.124706	0.09647059	0.009412	0.1741176	0.011765	0.3388235	0.02352941	0.05411765	0.02352941	0.287059	0.105882	0.185882	0.362353	0.077647	0.825882	0.192941
Lindsey Graham	0.132505	0.281574	0.04347826	0.008282	0.2815735	0.014493	0.1625259	0.00621118	0.05486542	0.02898551	0.414079	0.05176	0.296066	0.168737	0.083851	0.674948	0.339545
Bobby Jindal	0.24228	0.099762	0.0783848	0.026128	0.1235154	0.014252	0.192399	0.03800475	0.1852732	0.02137767	0.342043	0.104513	0.137767	0.230404	0.206651	0.821853	0.199525
Rick Santorum	0.148781	0.097561	0.06341463	0.019512	0.3268293	0.02439	0.1292683	0.02195122	0.1609756	0.02682927	0.246341	0.082927	0.35122	0.15122	0.187805	0.829268	0.190244
George Pataki	0.097493	0.21727	0.1114206	0.005571	0.2005571	0.022284	0.2172702	0.01114206	0.1225627	0.03064067	0.314763	0.116992	0.222841	0.228412	0.153203	0.749304	0.286908
Rick Perry						OUTLIER											
Average	**0.178094**	**0.128684**	**0.076209383**	**0.022141**	**0.23058491**	**0.012734**	**0.203576271**	**0.027029314**	**0.11145624**	**0.028167616**	**0.306778**	**0.098351**	**0.243319**	**0.230606**	**0.139624**	**0.799921**	**0.218757**
Top 9 Average	**0.194252**	**0.103898**	**0.070036848**	**0.025495**	**0.24876258**	**0.009796**	**0.206604375**	**0.030852358**	**0.09737342**	**0.028896343**	**0.29815**	**0.095532**	**0.258559**	**0.237457**	**0.12627**	**0.817029**	**0.198938**

Weighted Proportion of Moral Language: 2016 & 2020 General Election

Candidate	care.virtue	care.vice	fairness.virtue	fairness.vice	loyalty.virtue	loyalty.vice	authority.virtue	authority.vice	sanctity.virtue	sanctity.vice	Care	Fairness	Loyalty	Authority	Sanctity	Prop. Virtue	Prop Vice
Hillary Clinton 2016	0.253514	0.154054	0.1227027	0.064865	0.2064865	0.002162	0.132973	0.01837838	0.04486486	0.0227027	0.407568	0.187568	0.208649	0.151351	0.067568	0.760541	0.262162
Donald Trump 2016	0.161712	0.155767	0.07015458	0.135553	0.1700357	0.010702	0.1783591	0.03804994	0.04399524	0.06064209	0.317479	0.205707	0.180737	0.216409	0.104637	0.624257	0.400713
Joe Biden 2020	0.275029	0.101675	0.09583171	0.021815	0.2559408	0.002727	0.08648228	0.01636151	0.07713284	0.07791196	0.376704	0.117647	0.258668	0.102844	0.155045	0.790417	0.220491
Donald Trump 2020	0.186727	0.111133	0.08207343	0.064598	0.1590418	0.01237	0.1992931	0.05438838	0.05988612	0.09346161	0.29786	0.146672	0.171412	0.253681	0.153348	0.687021	0.335951
Average	**0.219245**	**0.130657**	**0.092690605**	**0.071708**	**0.1978762**	**0.00699**	**0.14927687**	**0.031794553**	**0.05646977**	**0.06367959**	**0.349903**	**0.164399**	**0.204866**	**0.181071**	**0.120149**	**0.715559**	**0.304829**
Dem. Average	**0.264271**	**0.127865**	**0.109267205**	**0.04334**	**0.23121365**	**0.002445**	**0.10972764**	**0.017369945**	**0.06099885**	**0.05030733**	**0.392136**	**0.152607**	**0.233658**	**0.127098**	**0.111306**	**0.775479**	**0.241327**
Repub. Average	**0.17422**	**0.13345**	**0.076114005**	**0.100076**	**0.16453875**	**0.011536**	**0.1888261**	**0.04621916**	**0.05194068**	**0.07705185**	**0.307669**	**0.17619**	**0.176074**	**0.235045**	**0.128993**	**0.655639**	**0.368332**

VIII. Cosine Similarity Validation

To make sure that the cosine similarity scores calculated for the construction of the cosine similarity network were robust, distributions for each vice and valence were calculated to determine that distributions were not being skewed by single outlier terms, thus biassing the cosine similarities. Cosine similarities were found to be extremely robust; only one outlier term, "help", was found to be an outlier for the 2016 Republican care virtue category; "help" was subsequently removed from both Democratic and Republican corpora. The term distributions used for validation can be found in the preceding **Appendix Figure 4**.



Appendix Figure 4 Plots illustrating term distributions used for the calculation of pairwise cosine similarities across moral foundations. A single outlier term was identified and removed from both Democratic and Republican corpora to strengthen the validity of the findings.

Media@LSE MSc Dissertations Series

The Media@LSE MSc Dissertations Series presents high quality MSc Dissertations which received a mark of 75% and above (Distinction).

Selected dissertations are published electronically as PDF files, subject to review and approval by the Editors.

ISSN: 1474-1938/1946