UNDERSTANDING THE ROLES OF HUMANS, ALGORITHMS, AND CYBORGS IN POLITICAL POLARIZATION

by

Huyen Thi Thanh Le

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Computer Science in the Graduate College of The University of Iowa

December 2019

Thesis committee: Zubair Shafiq, Thesis Supervisor
Padmini Srinivasan
James Cremer
Juan Pablo Hourcade
Kang Zhao
ACKNOWLEDGEMENTS

First of all, I would like to thank my advisor, professor Zubair Shafiq, very much for supporting me these past five years. He gave me great guidance on various research problems. I think I could not finish my PhD successfully without his guidance.

I would like to thank professor Padmini Srinivasan who co-advised me in most of my research. I also would like to thank my collaborators in different research problems, including professor Bob Boynton, professor Brian Ekdale, professor Timothy Havens, professor Andrew High, Yelena Mejova, and Raven Maragh.

I would like to thank my thesis committee members, professor Juan Pablo Hourcade, professor Kang Zhao, and professor James Cremer for their valuable feedback and comments on my research.

I am thankful to have an opportunity to work with great colleagues from the Internet Research Lab, department of Computer Science. I would like to thank Shehroze Farooqi, Adnan Ahmed, Hammad Mazhar, Umar Iqbal, and John Cook for their help and support.

Last but not least, I would like to have special thanks to my family, especially my husband Thanh Chu, for their love and support.
ABSTRACT

Recent studies revealed more increasingly political polarization in the distribution and consumption of political news. Political polarization demonstrates the disagreement between people aligned with different ideologies or political parties (e.g. left vs. right, Democrats vs. Republicans). Increasingly political polarization can have negative effects on our society; for example, extreme cases influenced by the left/right ideology can lead to massive bombing or shooting incidents. Thus, through four different research streams this thesis will help people to understand the roles of humans, algorithms, and cyborgs in political polarization.

In terms of humans, prior research has shown that people are mainly consuming news conforming to their pre-existing beliefs (a.k.a. selective exposure). People also prefer to have homophilous social interactions. Both of these lead to political polarization. Thus, to help inform the political slant of news people consume, we develop a lightweight and scalable news slant measurement using Twitter. Moreover, utilizing this method to estimate each Twitter user as a Republican or Democrat, we analyze political discourse on Twitter communications in the combination of three aspects including political affiliation, personality perception, and policy discussion around several presidential candidates from both parties during the 2016 U.S. presidential election.

In terms of algorithms, researchers have recently started to question whether algorithms create distinct personalized experiences for users. While selective exposure
requires deliberate acts of media choice, algorithmic personalization interprets past behavior as precedent for future preference. In other words, algorithmic personalization can intensify selective exposure beyond a person’s choice, resulting in a vicious cycle that can contribute to an increasingly polarized society. Thus, it is important to study the roles of personalization algorithms employed by search engines and social media in reinforcing pre-existing biases. To this end, we examine the personalization of Google News Search based on the users’ browsing history, especially when it comes from the users with different political biases. We discover that in fact Google News Search personalized its returned search results, and surprisingly, did reinforce the presumed political biases of the users.

In terms of cyborgs, there have been numerous reports of widespread misinformation campaigns during the 2016 U.S. presidential election. Of particular notes, there are reports which identified the efforts to manipulate social media (e.g. Twitter, Facebook) by the Russian state-sponsored accounts. These external manipulations by cyborgs cause significant pressure on social media services to mitigate spam, abuse, and political polarization on their platforms. Specifically, Twitter publicly acknowledged the exploitation of their platform and has since conducted aggressive cleanups to suspend the involved accounts. To shed light on Twitter’s countermeasures, we conduct a postmortem analysis of about one million Twitter accounts who engaged in the 2016 U.S. presidential election but were later suspended by Twitter.
PUBLIC ABSTRACT

Recent studies revealed more increasingly political polarization in the distribution and consumption of political news. Political polarization demonstrates the disagreement between people aligned with different ideologies or political parties (e.g. left vs. right, Democrats vs. Republicans). Increasingly political polarization can have negative effects on our society; for example, extreme cases influenced by the left/right ideology can lead to massive bombing or shooting incidents. Thus, through four different research streams this thesis will help people to understand the roles of humans, algorithms, and cyborgs in political polarization.

In terms of humans, prior research has shown that people are mainly consuming news conforming to their pre-existing beliefs. People also prefer to have homophilous social interactions. Both of these lead to political polarization. Thus, to help inform the political slant of news people consume, we develop a lightweight and scalable news slant measurement using Twitter. Moreover, utilizing this method to estimate each Twitter user as a Republican or Democrat, we analyze political discourse on Twitter communications in the combination of three aspects including political affiliation, personality perception, and policy discussion around several main candidates from both parties during the 2016 U.S. presidential election.

In terms of algorithms, researchers have recently started to question whether algorithms create distinct personalized experiences for users, which can unintentionally contribute to a more polarized society. Thus, it is important to study the roles of
personalization algorithms employed by search engines and social media in reinforcing pre-existing biases. To this end, we examine the personalization of Google News Search based on the users’ browsing history, especially when it comes from the users with different political biases.

In terms of cyborgs, there have been numerous reports of widespread misinformation campaigns during the 2016 U.S. presidential election. Of particular notes, there are reports which identified the efforts to manipulate social media (e.g. Twitter, Facebook) by the Russian state-sponsored accounts. These external manipulations by cyborgs cause significant pressure on social media services to mitigate spam, abuse, and political polarization on their platforms. Specifically, Twitter publicly acknowledged the exploitation of their platform and has since conducted aggressive cleanups to suspend the involved accounts. To shed light on Twitter’s countermeasures, we conduct a postmortem analysis of about one million Twitter accounts who engaged in the 2016 U.S. presidential election but were later suspended by Twitter.
TABLE OF CONTENTS

LIST OF TABLES ......................................................... ix
LIST OF FIGURES ....................................................... xi

CHAPTER

1 INTRODUCTION ...................................................... 1

2 UNDERSTANDING THE ROLES OF HUMANS IN POLITICAL POLARIZATION .............................................................. 7
   2.1 Scalable News Slant Measurement Using Twitter .................. 7
      2.1.1 Background, Motivation, and Research Statement .......... 7
      2.1.2 Related Work ............................................... 8
      2.1.3 Methodology ............................................... 10
         2.1.3.1 Landmark Selection .................................... 10
         2.1.3.2 Slant Estimation ...................................... 12
      2.1.4 Results .................................................... 14
         2.1.4.1 Evaluation ............................................. 14
         2.1.4.2 Impact of Landmark Selection ......................... 15
         2.1.4.3 Slant of Neutral Articles .............................. 16
      2.1.5 Summary .................................................... 17
   2.2 Revisiting The American Voter on Twitter ....................... 18
      2.2.1 Background, Motivation, and Research Statement .......... 18
      2.2.2 Related Work ............................................... 20
         2.2.2.1 The American Voter .................................... 20
         2.2.2.2 Party .................................................. 21
         2.2.2.3 Personality ............................................ 23
         2.2.2.4 Policy .................................................. 24
         2.2.2.5 Election Prediction .................................... 25
      2.2.3 Methodology ............................................... 26
         2.2.3.1 Party .................................................. 27
         2.2.3.2 Personality ............................................ 28
         2.2.3.3 Policy .................................................. 29
         2.2.3.4 Sentiment .............................................. 30
      2.2.4 Dataset .................................................... 31
         2.2.4.1 Data Collection ........................................ 31
         2.2.4.2 Data Statistics ........................................ 32
      2.2.5 Results .................................................... 36
         2.2.5.1 Party .................................................. 36
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Confusion matrix of the match between crowdsourced news slant measured by [47] and our measured slant scores when the threshold is symmetric ($b = -a = 0.37$).</td>
<td>17</td>
</tr>
<tr>
<td>2.2</td>
<td>Confusion matrix of the match between crowdsourced news slant measured by [47] and our measured slant scores when the threshold is asymmetric ($a = -0.48, b = 0.28$).</td>
<td>17</td>
</tr>
<tr>
<td>2.3</td>
<td>Statistics of candidate tweet collections from November, 15 2015 to February 29, 2016. Close to half the users and tweets in the dataset were posting on Trump. Trump’s numbers reflect a 3X lead in tweets and users compared to Clinton and Sanders. Negative sentiment tweets significantly outnumber positive and neutral ones. The sentiment charged nature of the dialog is also indicated by the low prevalence of sentiment neutral tweets (about 8%). Same party tweets outnumber competing party tweets for seven out of ten candidates. Democrats tweeted more about Trump than about Clinton or Sanders. Republicans were more interested in Clinton than in Sanders.</td>
<td>34</td>
</tr>
<tr>
<td>2.4</td>
<td>Statistics of tweets conveying personality perceptions and their breakdown across party affiliation and sentiment.</td>
<td>38</td>
</tr>
<tr>
<td>2.5</td>
<td>Statistics of tweets discussing different policies, and their breakdown across party affiliation and sentiment.</td>
<td>42</td>
</tr>
<tr>
<td>2.6</td>
<td>List of statistically significant variables ($p$-value &lt; 0.001) in our regression model for Democratic candidates. The features are sorted in descending order with respect to estimated coefficient values. Adjusted $R^2 = 0.908$.</td>
<td>47</td>
</tr>
<tr>
<td>2.7</td>
<td>List of statistically significant variables ($p$-value &lt; 0.001) in our regression model for Republican candidates. The features are sorted in descending order with respect to estimated coefficient values. Adjusted $R^2 = 0.883$.</td>
<td>48</td>
</tr>
<tr>
<td>3.1</td>
<td>Statistics of pro-immigration and anti-immigration Twitter accounts.</td>
<td>60</td>
</tr>
</tbody>
</table>
3.2 Personalization (quantified using difference and edit distance) for 50 search terms. Note that \( E(P, A) \), \( E(P, C) \), and \( E(A, C) \) are pairwise edit distance among pro-immigration, anti-immigration, and control profiles.  

3.3 Average personalization for all 50 search terms according to top-k (\( k \in \{10, 20, ..., 100\} \)) ranked search results. We normalize edit distance (as \( k \)-edit distance) to make a fair comparison across different \( k \) values.  

3.4 Political bias for nine top-personalized search terms. Note that \( B(P) \), \( B(A) \), and \( B(P) - B(A) \) respectively are political bias of personalized search results for the pro-immigration and anti-immigration profiles, and the difference between their political bias. Table is sorted in ascending order of \( B(P) - B(A) \) values. A negative \( B(P) - B(A) \) value indicates the pro-immigration profile received more Democratic-leaning personalized search results, while a positive \( B(P) - B(A) \) value indicates the pro-immigration profile received more Republican-leaning personalized search results.  

4.1 Statistics of suspended users as of February 2018 in our tweet collection around Clinton and Trump during nearly four months, from June 01 to August 01 & from September 09 to November 08, 2016 (except September 18-20, October 31, and November 1).  

4.2 Statistics of retweet and mention network from suspended and regular users.  

4.3 Statistics of six representative communities for qualitative analysis.
LIST OF FIGURES

Figure

1.1 Factors leading to political polarization. ............................................. 2

2.1 Visualization of follower graph similarity among Democratic and Republi-
can landmark users. The value of a cell is the percentage overlap between
followers of two landmarks. Brighter colors represent more similarity and
darker colors represent less similarity. ......................................................... 11

2.2 Proposed method to measure political slant of news articles. ............... 12

2.3 Impact of landmark selection on slant measurement accuracy and coverage. 15

2.4 Time series of tweets for candidates. The vertical black lines indicate
the following major election events during the data collection period. (1) DebR1: CNN Republican debate, (2) DebR2: Fox Business Republican
debate, (3) DebD2: NBC News Democratic debate, (4) DebR3: Fox News
Republican debate, (5) IAD: Democratic Iowa caucus, (6) IAR: Repub-
lican Iowa caucus, (7) DebD3: MSNBC Democratic debate, (8) DebR4:
ABC News Republican debate, (9) NHD: Democratic New Hampshire pri-
mary, (10) NHR: Republican New Hampshire primary, (11) DebD4: PBS
Democratic debate, (12) DebR5: CBS Republican debate, (13) NVD:
Democratic Nevada caucus, (14) SCR: Republican South Carolina pri-
mary, (15) NVR: Republican Nevada caucus, (16) DebR6: CNN Republi-
can debate, (17) SCD: Democratic South Carolina primary. .................. 33

2.5 Scores for four most discussed personality dimensions. Scores are in the
range [-1,+1]. -1 (+1) indicates the personality is viewed as absent (present)
with high confidence. The number of tweets for absence (presence) of each
personality dimension are provided on the left (right) of the bar plot. The
net score is included in the parenthesis. The bar plots are shown for can-
didates receiving a minimum of 100 tweets on a dimension. .................... 39

2.6 Number of tweets for the six most discussed policies with party affiliation
downbreak. The blue and red regions indicate tweets from Democrats and
Republicans, respectively. The white region represents tweets from the
remaining users. .................................................. 43
2.7 Proportion of tweet volume about personality dimensions of Clinton (above) and Trump (below) from users identified as Democrats (blue circles), Republican (red squares), and Other (grey triangles). 50

3.1 Our sock puppet auditing system to measure political personalization of Google News search. 58

3.2 Top-10 first-party domains for hyperlinks crawled from pro- and anti-immigration Twitter accounts. 62

3.3 Top-30 third-party domains identified on the hyperlinks crawled from pro- and anti-immigration Twitter accounts. 63

4.1 CDFs of suspended and regular communities in five dimensions. The vertical line places at the threshold as 50% and 200 times. From this vertical line to the right in each plot are communities having high value for that respective dimension. Specifically, the percentage of communities having high values = 1 - Yvalue of the intersection between the CDF and the vertical line. 88

4.2 The illustrative suspended communities in terms of burstiness: Vote!Black LivesMatter Community. 93

4.3 Two decals from Ebay URLs posted by EbayAds community. 94

4.4 Two pictures opposing Clinton from tumblr.com posted by NoEthicsNoOffice community. 95
CHAPTER 1
INTRODUCTION

In society, polarization can happen in multiple aspects such as economics (middle class vs. upper class), race (black vs. white), and politics (Democrat vs. Republican). Especially, in U.S. where two political parties (Democrat vs. Republican) compete over many issues, political polarization is an interesting and important matter to study. Political polarization is when an individual’s political stance is defined by his/her ideology or political party (e.g. left vs. right, Democrats vs. Republicans). For example, given a policy issue such as immigration, people who identify as Democrats will be pro-immigration, which aligns with the Democratic party’s stance while people who identify as Republicans will be anti-immigration, which aligns with the Republican party’s stance. Thus, political polarization highlights the disagreement between people who align with different ideologies or political parties.

Researchers from Pew Research Center reported that American society is increasingly becoming more politically polarized [53]. Researchers did two independent surveys on nationally representative Democrats and Republicans, specifically asked for their ratings on 10 political items such as “how do you feel about governments efficiency?” As a result, the score distribution showed the increasing disagreement between Democrats and Republicans over two decades. This increasing disagreement is even beyond dislike when one party views the other party as “a threat to the nation’s well being.” As a consequence, increasing political polarization can have negative effects on our society when extreme cases influenced by the left/right ideology can lead
to bombing or shooting incidents. For example, there was a bombing incident in Florida in 2018 [196]. The suspect was arrested for sending explosive packages to at least a dozen Democrats who are well-known critics of president Trump. The investigation revealed many pictures in the suspect’s van showing that he was influenced by right wing ideology and pro-Trump. These extreme and negative dangers are reasons why it is very important to study the factors that lead to political polarization.

From our research, we recognize that three factors including humans, algorithms, and cyborgs\(^1\) play significant roles in political polarization (Figure 1.1). While humans’ innate instincts directly lead to political polarization, algorithms can accidentally intensify the political polarization and cyborgs can intentionally worsen the political polarization. In terms of humans’ roles, political polarization is partly explained by the selective exposure theory [170], which suggests that users tend to be

\(^1\)which are human-assisted bots or bot-assisted humans and can be understood as bad actors that try to manipulate the interactions between humans and algorithms.

![Figure 1.1: Factors leading to political polarization.](image)
attracted to, consume, and share news that conforms with their ideological beliefs. It is also, in part, explained by user preference for homophilous social interactions [117]. Both of these are due to humans’ innate instincts. In terms of algorithms’ roles, researchers have recently started to question whether algorithms create distinct personalized experiences for users, leading to so-called “filter bubbles” or “echo chambers” [108, 151, 33]. While selective exposure requires deliberate acts of media choice, algorithmic personalization interprets past behavior as precedent for future preference. In other words, algorithmic personalization can intensify selective exposure beyond a person’s choice, resulting in a vicious cycle that can contribute to an increasingly politically polarized society. In terms of cyborgs’ roles, there have been numerous reports of widespread misinformation campaigns and political propaganda caused by cyborgs during elections. Of particular note, there are reports that have identified the efforts of paid social media trolls by Russia’s Internet Research Agency (RU-IRA) to manipulate social media platforms (e.g. Twitter, Facebook) in the 2016 U.S. presidential election [105, 16]. These external manipulations by cyborgs cause significant pressure on social media services to mitigate spam, abuse, and political polarization on their platforms.

This thesis explores different research streams (including measuring the political slant of individual news articles, revisiting the American voters in presidential campaigns, measuring political personalization on Google News Search, and analyzing suspended Twitter account activities) to help people better understand the roles of humans, algorithms, and cyborgs in political polarization.
• **Understanding the Roles of Humans in Political Polarization:** To help discover how imbalanced the political news that people might consume and share may be, we develop a method to measure ideological slant of individual news articles by monitoring their consumption on online social media. Specifically, we measure an article’s slant by analyzing the users tweeting about the article and evaluating their connectivity to a set of Republican and Democratic landmark users. Our method takes advantage of the massive amount of news consumption data available on online social media to measure news slant on a large scale. This approach - while seemingly simple - is powerful enough to effectively gauge article slant while overcoming the scalability limitations of prior research. This work was published in The 2017 International AAAI Conference on Web and Social Media (ICWSM). Furthermore, utilizing this method to estimate each Twitter user as a Republican or Democrat, we analyze political discourse in the 2016 U.S. presidential campaigns on Twitter in the combination of three aspects including political affiliation, personality perception, and policy discussion around several main candidates from both parties. Thus, while our work derives in a principled way from *The American Voter* - an established and long line of election research in political science, our contribution is to move it forward from survey research to the realm of online social media by applying computational methods for tracking political discourse about party, personality, and policy on Twitter. This work was published in The 2017 ACM Conference on Hyper-text and Social Media (HyperText) and The 2017 ACM SIGCHI Conference on

- **Understanding the Roles of Algorithms in Political Polarization:** It is important to study how personalization algorithms employed by search engines and social media reinforce pre-existing biases. To this end, we examine the personalization of Google News Search based on the users’ browsing history, especially when it comes from users with different political biases. Specifically, while controlling for other factors, we train a pair of fresh browser profiles by visiting websites that reflect pro-immigration and anti-immigration stances. We then execute search queries on Google News related to a variety of political topics. We analyze the search results to quantify the magnitude and direction of personalization. We discover that in fact, Google News Search personalized its returned search results, and surprisingly, did reinforce the presumed political bias of the users. These findings not only set the baseline for search personalization based on political biases in browsing history, but also contribute to the broader understanding of selective exposure and algorithmic personalization. This work was published in The 2019 Web Conference (WWW).

- **Understanding the Roles of Cyborgs in Political Polarization:** Aligning with Twitter’s efforts to remove spam, fake, and political propaganda accounts in order to make their platform become a healthier environment for upcoming elections, we do a postmortem of our 2016 election-related tweet collection to shed light on activities of suspended Twitter accounts. To systematically analyze the coordinated activities of these suspended accounts, we first group them
into communities based on their retweet/mention network and then analyze different characteristics such as popular tweeters, domains, and hashtags. The results show that suspended and regular communities exhibit significant differences in terms of popular tweeters and hashtags. Our qualitative analysis also shows that suspended communities are heterogeneous in terms of their characteristics (e.g. pro-Trump vs. pro-Clinton, bursty posting vs. constant posting). Furthermore, we find that accounts suspended by Twitter’s new countermeasures are closely connected to the original suspended communities. This work was published in The 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).

The rest of this thesis is organized as follows: Chapter 2 presents details of our work in “Scalable News Slant Measurement Using Twitter” and “Revisiting The American Voter on Twitter” to help understand the roles of humans in political polarization. Chapter 3 describes our work in “Measuring Political Personalization on Google News Search” to help understand the roles of algorithms in political polarization. Chapter 4 shows our work in “A Postmortem of Suspended Twitter Accounts in the 2016 U.S. Presidential Election” to help understand the roles of cyborgs in political polarization. Finally, Chapter 5 concludes the thesis and outlines some directions for future work.
CHAPTER 2
UNDERSTANDING THE ROLES OF HUMANS IN POLITICAL POLARIZATION

2.1 Scalable News Slant Measurement Using Twitter

2.1.1 Background, Motivation, and Research Statement

People increasingly rely on online social media to consume, share, and discuss news. According to the Pew Research Center’s Journalism Project, more than 50% of Facebook and Twitter users consume news on social media sites [143]. Ideological polarization of news outlets is extensively reported in prior literature [106]. Recent studies on Facebook [32] and Twitter [60] have revealed polarization in the consumption and distribution of political news on social media sites. Political news polarization in online social media is partly explained by the selective exposure theory [170] which suggests that users tend to be attracted to, consume, and share news that conform with their ideological beliefs. It is also, in part, explained by user preference for homophilous social interactions [117].

Set in this context, we present a method to measure ideological slant of individual news articles by monitoring their consumption on online social media. Specifically, our proposed method measures an article’s slant by analyzing the users tweeting about the article and analyzing their connectivity to a set of Republican and Democratic landmark users. Landmark users are well-recognized individuals whose slants (or political affiliations) are known with high confidence. A news article is assessed to have Republican slant if it is shared more by users who follow Republican landmarks than
users who follow Democratic landmarks. Similarly, a news article is assessed to have Democratic slant if it is shared more by users who follow Democratic landmarks than users who follow Republican landmarks. Our method exploits the massive amount of news consumption data available on online social media to measure news slant at a large scale. This approach - while seemingly simple - is powerful enough to effectively gauge article slant while overcoming scalability limitations of prior research.

Prior work used crowdsourcing to label political slant of news articles [47]. However, large-scale news slant estimation using crowdsourcing requires the availability and monetary compensation for a large number of politically informed crowd workers. We compare our news article slant estimates with prior work [47] that relies on crowdsourcing. The comparison shows that our method can accurately estimate article-level slant for 80% Democratic-leaning news articles and 76% Republican-leaning news articles.

2.1.2 Related Work

Prior methods for estimating slant can be broadly divided into two categories: content-based methods and audience-based methods.

Content-based methods, as the name implies, address the problem more directly by analyzing the content of news articles. For example, Groseclose et al. [87] measured media slant by monitoring the relative citation frequency of various policy groups by news outlets and the members of Congress. As another example, Gentzkow et al. [80] parsed congressional speeches to identify partisan phrases that are more
commonly used by Republican or Democratic members of Congress. They then analyzed text content in newspaper articles and quantified each newspaper’s political slant by measuring the relative use of partisan phrases. More recently, researchers have used crowdsourcing to estimate slant of individual news articles. Budak et al. [47] used crowd workers on Amazon Mechanical Turk (AMT) to identify slant of individual news articles. Unfortunately, content-based methods that utilize crowdsourcing for slant measurement do not scale well because they require non-trivial monetary compensation to crowd workers for large-scale labeling. In contrast, our proposed method does not require explicit user participation to measure slant of news articles.

Audience-based methods, in contrast, rely on the idea that news readers have their own ideological biases which are reflected in their news consumption and sharing behavior [170]. These methods analyze data about news consumption and sharing to indirectly measure slant of news articles. The widespread usage of online social media allows audience-based approaches to operate at a large scale. Our proposed slant measurement method falls in this latter category. Prior audience-based methods [80, 25, 200, 145, 83, 122] aim to measure slant of individual users or news outlets in aggregate. Unlike prior audience-based methods which are dependent on manually labeled ground truth [209] or self-reported user political alignments [32], our method requires only a small set of pre-labeled landmarks. Our proposed method leverages a small number of landmarks to accurately estimate slant of individual news articles at scale.
2.1.3 Methodology

We want to measure the political slant of news articles (towards Republicans or Democrats) by observing the patterns of sharing amongst users in online social media and the connectivity of these users to a set of landmark individuals. At a high level, our idea is that if a news article is tweeted/retweeted by more Democrats than Republicans, it is likely to have a Democratic slant. It is the opposite for articles likely to have a Republican slant. This strategy is similar to Bakshy et al. [32], who used the ratio of number of Democrats to Republicans for estimating slant on Facebook. The key difference is that they relied on self-reported political affiliations. However, only 9% of adult Facebook users in the U.S. self-report their political affiliation. Since such data is very limited, as explained next, we propose to estimate user affiliation or slant with a different strategy involving landmark users.

2.1.3.1 Landmark Selection

A landmark user is one whose slant, here political party affiliation, is well recognized. We manually identify a set of 30 popular Democrats (e.g., Rachel Maddow) and 30 popular Republicans (e.g., Sean Hannity) on Twitter as our “landmark” users. We curated these landmarks in consultation with political scientists. We chose several journalists as landmarks because they have large followings on Twitter. On average, each Democratic landmark user has 223,656 followers and each Republican landmark user has 277,671 followers. The large following on Twitter provides us sufficient coverage to quantify slant of news articles. Our coverage analysis showed that
Figure 2.1: Visualization of follower graph similarity among Democratic and Republican landmark users. The value of a cell is the percentage overlap between followers of two landmarks. Brighter colors represent more similarity and darker colors represent less similarity.

approximately 95% of news articles in our data set have more than 10 connections to one of the 60 landmarks (median is 115). Thus, our analysis shows that 60 landmarks (each with a large Twitter following) provide us reasonable coverage.

Next, to validate our selection of landmarks, we examine the overlap in followers for pairs of landmarks. We compute the follower graph similarity (i.e., percentage overlap between their followers) for all pairs of landmarks. Figure 2.1(a) shows the percentage follower overlap between Democratic landmarks, Figure 2.1(b) shows the percentage follower overlap between Republican landmarks, and Figure 2.1(c) shows the percentage follower overlap across Democratic and Republican landmarks. The colorbar denotes the percentage follower overlap between two landmarks. Overall, we note that Figures 2.1(a) and (b) are much brighter (more overlap) than Fig-
More specifically, the average percentage of follower overlap amongst Democrats is 32.4% and amongst Republicans is 43.6%. In contrast, the average percentage of follower overlap across Democrats and Republicans is only 8.8%. This pattern demonstrates political polarization on Twitter and is consistent with [60], i.e., Democrats tend to make connections with other Democrats and Republicans tend to make connections with other Republicans, while tending to not do so across party lines. Most importantly for us these results demonstrate that selective exposure theory and homophily in social interactions are exhibited in the follower connections of our landmarks. This result also reaffirms our confidence in the selection of Democratic and Republican landmarks.

2.1.3.2 Slant Estimation

Figure 2.2 illustrates our proposed news slant estimation method. The top tier contains the Democratic and Republican landmarks. The middle tier contains Twitter

Figure 2.2: Proposed method to measure political slant of news articles.
users. The bottom tier represents news articles. Each article-user link indicates that a user tweeted/retweeted a news article. Based on the selective exposure theory [170], we expect users to consume and share news that conform with their ideological beliefs. Each user-landmark link indicates that a user follows a landmark. Again due to the natural preference for homophilous social interactions [117], we expect a large number of links from users to landmarks of the same political affiliation and only a few links to the opposite political affiliation. Thus, we can compute a news article’s political slant by monitoring its sharing activity on Twitter and user connectivity to the landmarks.

To implement our proposed method, we first collect tweets which mention a news article using Twitter’s streaming API. From this sample of tweets, we get a list of all users who have tweeted about the news article. We then count the number of landmark Democrats and Republicans which each user follows. We use Twitter’s REST API to collect the follower lists of landmark Democrats and Republicans. Using the counts of landmark Democrats and Republicans that all users follow, we quantify the political slant of a news article as: \( Slant = \frac{\#\text{Republicans} - \#\text{Democrats}}{\#\text{Republicans} + \#\text{Democrats}} \). We quantify the slant of a news article in the range of -1 to 1, where -1 indicates Democratic slant and +1 indicates Republican slant.
2.1.4 Results

2.1.4.1 Evaluation

Below we compare our article-level slant estimation results with Budak et al. [47]. They used crowdsourcing to label political slant of 10,500 news articles as Republican-leaning, Center, or Democratic-leaning. They recruited two crowd workers to evaluate the political slant of each news article on a five-point scale (“Positive”, “Somewhat Positive”, “Neutral”, “Somewhat Negative”, “Negative”) for both Democrats and Republicans. To mitigate noise and increase reliability of slant labels by crowd workers, we only consider the articles for which the evaluations by two crowd workers match each other. After this filtering, the data set contains 605 Democratic-leaning, 653 Republican-leaning, and 3,837 Neutral news articles. For comparison, we compute the slant score for these Democratic- and Republican-leaning news articles using our proposed method. The results show that our estimated slant scores are well correlated with the article set’s political affiliation. The average estimated slant score for the Democratic-leaning article set is strongly Democratic-leaning (-0.40), while the average estimated slant score for the Republican-leaning set is strongly Republican-leaning (0.44). Furthermore, 80% of Democratic-leaning articles have slant scores smaller than 0 and 76% of Republican-leaning articles have slant scores larger than 0.

\footnote{We thank [47] for sharing their crowd labeled data set.}
Figure 2.3: Impact of landmark selection on slant measurement accuracy and coverage.

2.1.4.2 Impact of Landmark Selection

We next examine the impact of the selected landmarks on the accuracy (with respect to crowdsourcing) and coverage (number of news articles) of slant measurement.

We compute the accuracy for the sets of Democratic- and Republican-leaning articles respectively when the number of chosen landmarks is gradually increasing from 1 to 30. For this experiment, we select landmarks randomly from the full set of 30 landmarks for Democrats and Republicans. Figure 2.3 shows the results for the average and standard deviation of the accuracy and coverage of our slant measurements for 1,000 independent runs. Overall, we note that both average accuracy and coverage increase (while their standard deviations decrease) as we include more landmarks. It is noteworthy that the accuracy and coverage start to plateau when we
reach around 10 landmarks. These results demonstrate the diminishing returns on average slant accuracy and coverage as we include more landmarks. We conclude that our full set of 30 landmarks are sufficient for achieving good accuracy and coverage for both Democratic- and Republican-leaning news articles.

2.1.4.3 Slant of Neutral Articles

Previously we showed that our estimated slant scores are well correlated with each article set’s political affiliation labeled using crowdsourcing.

We now want to categorize news articles into 3 categories of Democratic, Republican, and Neutral according to their measured slant scores. To do so, we need to decide suitable thresholds $a$ and $b$ such that $-1 \leq a < b \leq 1$, and label the news articles with slant scores lying within $(a, b)$ as Neutral, the news articles with slant scores in the range $[-1, a]$ as Democratic-leaning, and the news articles with slant scores in the range $[b, 1]$ as Republican-leaning. We experiment with multiple choices for the thresholds $a$ and $b$. For each choice, we compute a confusion matrix of the match between crowdsourced news article slant by [47] and our measured slant scores. Table 2.1 depicts the confusion matrix for symmetric thresholds $b = -a = 0.37$ which maximize the sum along the diagonal of the confusion matrix. Table 2.2 depicts that confusion matrix for asymmetric threshold $a = -0.48, b = 0.28$ which maximize the sum along the diagonal of the confusion matrix.
Crowdsourced Slant Label | Democratic | Neutral | Republican |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Democratic-leaning</td>
<td>62.5%</td>
<td>26.8%</td>
<td>10.7%</td>
</tr>
<tr>
<td>Neutral</td>
<td>34.1%</td>
<td>46.0%</td>
<td>19.8%</td>
</tr>
<tr>
<td>Republican-leaning</td>
<td>10.7%</td>
<td>23.6%</td>
<td>65.7%</td>
</tr>
</tbody>
</table>

Table 2.1: Confusion matrix of the match between crowdsourced news slant measured by [47] and our measured slant scores when the threshold is symmetric ($b = -a = 0.37$).

Crowdsourced Slant Label | Democratic | Neutral | Republican |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Democratic-leaning</td>
<td>55.9%</td>
<td>31.6%</td>
<td>12.5%</td>
</tr>
<tr>
<td>Neutral</td>
<td>26.5%</td>
<td>50.5%</td>
<td>23.0%</td>
</tr>
<tr>
<td>Republican-leaning</td>
<td>8.4%</td>
<td>22.8%</td>
<td>68.8%</td>
</tr>
</tbody>
</table>

Table 2.2: Confusion matrix of the match between crowdsourced news slant measured by [47] and our measured slant scores when the threshold is asymmetric ($a = -0.48, b = 0.28$).

2.1.5 Summary

We presented a method to scalably measure the political slant of news articles. We estimated the slant of individual news articles by analyzing the social connectivity of users who tweet about them. Our proposed method relies on a small set of Democratic and Republican landmarks on Twitter. The results showed that our method accurately identifies slant of individual news articles.
2.2 Revisiting The American Voter on Twitter

2.2.1 Background, Motivation, and Research Statement

The American Voter [51], published in 1960, demonstrated that the most important factors for voters when choosing their President were: *party or partisanship*, *policy considerations*, and *personality* of the person seeking office. Partisanship – being a Democrat or a Republican – was extremely important and it was the one factor that extended from one election to the next. Though policies shifted across elections, people’s assessment of the importance of specific policies influenced their votes. The third factor was how the candidates were perceived – their personality or character. The American Voter was based on the largest collection of survey research of its day, covering presidential elections from 1948 through 1956. The study became an instant classic and set a research tradition that has extended to this point. The American Voter Revisited [127] replicated the original research in the 2000s and found among other things that once again party, policy, and personality remain the three most important factors in understanding who people vote for in presidential elections.

Our aim is to use the classic The American Voter study as the basis for a principled analysis of Twitter conversations around the 2016 U.S. presidential election. Specifically, we assess partisanship, discussions on policies, and perceptions of candidates’ personality using Twitter communications. While our approach derives in a principled way from The American Voter, our work is novel in that it bridges from traditional survey research to mining large-scale publicly available online social media communications.
Social media has already played a significant role in elections in the U.S. and elsewhere [189, 102]. It was pivotal in the 2016 U.S. presidential election as well [21]. Growing numbers of the general public, especially the younger demographic, follow elections on social media [175, 84]. Given this trend, most candidates and their campaigns are actively trying to attract and engage social media users. Taking candidacy announcement as an example, Ted Cruz was the first mainstream politician to officially announce his candidacy with a tweet. A few weeks later, Hillary Clinton also took advantage of Twitter to announce she was running. Candidates differ in their social media strategies and successes [101]. Donald Trump posted inflammatory tweets to dominate the news cycle and drive up attendance at his rallies. Bernie Sanders’s campaign used #FeelTheBern to gather and rally a large grassroots movement.

The analysis of social media activity in prior literature has primarily focused on counts (tweets, retweets, comments, likes, etc.) and sentiment as indicators of public engagement, reach, and opinion. Content tends to be considered only in outlier conditions such as when a post goes viral. In contrast, our goal is to use computational methods to explore content about policy discussions and personality perceptions in addition to party-specific engagement on Twitter. Our key contributions are the following.

- We study Twitter communications around the 2016 U.S. presidential candidates using 50 million tweets collected from November 15, 2015 to February 29, 2016. This period covers Iowa, New Hampshire, Nevada, and South Carolina caucuses and primaries as well as several debates.
• We implement computational methods for tracking political discourse about party, personality, and policy on Twitter.

• We conduct statistical analysis of electoral polls to show the importance of party, personality, and policy in the modeling of political deliberation.

The unique contribution of our work is in applying computational methods to an established and long line of election research in political science. Specifically, we contribute to the stream of research initiated by *The American Voter* by moving it forward from survey research to the realm of online social media.

2.2.2 Related Work

2.2.2.1 The American Voter

Although published several decades ago, *The American Voter* [51] continues to serve as the baseline for researchers to understand voting behavior. The original study used survey data collected during three U.S. presidential elections (1948, 1952, and 1956). The central argument of the study is the funnel model, which claims that party affiliation shapes voters’ attitude towards policy considerations and personality perception of presidential candidates. The study analyzed how voters form their own party identification and how the psychological attachment between voters and their party determines political attitudes.

Based on the original study, *The American Voter Revisited* [127] in 2008 attempted to understand voting behaviors during the U.S. presidential elections in 2000 and 2004. The authors found that, as compared to the 1950s, more voters identified
with their party affiliations in the 2000s. More importantly, the authors also found that voting outcomes are still explained by party identification, short-term policy issues, and perceptions of candidates. The authors concluded that even though the influence of these factors on voting outcomes may have changed over time, the three-pronged paradigm of party, policy, and personality remains intact.

We ask whether these dimensions are present in current social media data, whether it is possible to track them, and whether they still relate to the politicians’ success at the polls. Below, we discuss prior research on measurement and analysis of party, policy, and personality, especially in online social media.

2.2.2.2 Party

Identification of political affiliations is a well-researched area. Prior political science research primarily relies on interviews to explicitly ask users about their political affiliations. On online social media, however, only a small fraction of users report their political affiliation. For example, less than 10% of U.S.-based adult Facebook users self-report their political affiliation [32]. It is challenging to infer users’ political party affiliations at a large scale using online social media data. Cohen and Ruths [58] showed that it is difficult to infer political orientation for “normal” Twitter users who rarely discuss politics. Some researchers have used machine learning techniques such as LDA (Latent Dirichlet Allocation) [59], SVM (Support Vector Machine) [59, 205, 153], and BDT (Boosted Decision Tree) [153] to infer users’ political affiliations based on profile features (e.g., name, location), linguistic features (e.g.,
tweet text, hashtags), and network features (e.g., followers, retweets, replies). Some researchers have used label propagation techniques where a user's political affiliation is inferred based on whether they post about well-known conservative/liberal issues or follow well-known conservative/liberal users. For example, Zhou et al. [209] inferred political leaning of users on Digg based on how they voted (equivalent to like or share) on news articles with labeled political ideology using Amazon Mechanical Turk (AMT) workers. More recently, Golbeck et al. [83] inferred political leaning of Twitter users based on whether they follow a seed group of well-known political personalities (e.g., Congress members). We use a similar method to infer political affiliation of Twitter users in our work.

Prior political science research has shown that the political party affiliation's impact on voting outcomes may not be straightforward. Using survey data on political figures and events, Bartels [37] applied a Bayesian model to study opinion change with respect to partisan bias. He found that partisanship is not just a simple “running tally”, but rather shapes voters’ attitudes and reactions to politics, resulting in sharp differences in opinions between Democrats and Republicans towards various political events. Other studies have also looked at the marginal and joint impact of political affiliation and other factors on voting preferences (e.g., [40, 69, 97]), which we also examine here.
2.2.2.3 Personality

Perception of politicians’ personality reflects both the campaign’s success in framing their candidate and the values of the electorate projected onto these public personas. Pew Research Center and the Washington Post conducted surveys to study how personalities of political leaders are perceived by the public over time [74, 75]. Through phone interviews, U.S. adult voters were asked for the one word that comes to mind when a politician is named. Comparing Obama and Romney, voters’ most common perception of Obama were Good, Trying, President, and also Failure and Incompetent and for Romney were Honest, Businessman, Rich, Good, and Conservative. However, survey-based personality research cannot rival the scale of online social media. Prior studies on personality perceptions of presidential candidates from online social media communications have relied on broad categorization of sentiment such as positive, negative, and neutral [41, 113, 197, 139]. Tumasjan et al. [189] looked at more detailed sentiment aspects such as anxiety, anger, and sadness for different candidates in the 2009 German national election. More recently, Bhattacharya et al. [43] proposed a method to systematically measure personality traits suggested by The American Voter [51] using a template-driven approach that measures the personality trait on a continuum, measuring either its presence or absence. As state-of-the-art, we use their method in our work.

Further, party affiliation has been repeatedly shown to relate to the rhetoric of the politicians in question, as been shown by [64, 81, 76], who have studied the influence of politicians’ personality traits on their leadership and decision-making
styles. For example, Gallagher et al. [76] found that the personality traits of political leaders shape their choices and their level of consistency in policy making. Further, Benoit [40] showed that Republicans discuss character more, and policy less, than Democrats. While examining debates, television spots, and acceptance addresses from 1948 to 2000, he found that Democrats emphasized more on traits like ‘empathy’ and ‘drive’ while Republicans emphasized more on traits like ‘sincerity’ and ‘morality’. The automated analysis framework we present in this paper provides parallel insights of the politicians’ character as perceived by social media users.

2.2.2.4 Policy

Policy issues have been examined at both the macro-level (nation or public as a unit of analysis) [57, 38] and the micro-level (how individuals define issues) [201, 202]. In order to track policy-related discussions on online social media, researchers typically create a lexicon of relevant terms of each policy and track their occurrences within the content [208, 181]. For example, Zhang et al. [208] manually identified relevant keywords, phrases, and hashtags related to same-sex marriage on Twitter, community wikis, and news articles to predict policy changes on the issue. We follow a similar high-precision approach, engaging political scientists’ domain expertise to build vocabularies for each topic.

Prior political science research has explored the impact of policies on voters’ evaluation of presidential candidates. Benoit [40] investigated presidential elections between 1948 and 2000 by quantitatively analyzing the texts of primary and general
debates, television spots, and acceptance addresses. The key conclusions were that Democrats discuss policy more than Republicans, and that Democrats focus more on issues such as education while Republicans focus more on issues such as national security. Dolan [69] examined the American National Election Study (ANES) election data between 1992-2006 and concluded that while voters may have a different view on abortion than their party, people tend to go with their party affiliation when voting in elections. Finding that more Democrats consider abortion an important issue than Republicans, Dolan showed that the relationship between policy and party is not homogeneous. A similar recent study by Highton [97] of ANES data spanning three presidencies (H.W. Bush: 90-92, Clinton: 92-96, and W. Bush: 00-04) further showed a strong partisanship effect, but one which over time changes with economic and cultural attitudes. In this work we aim to capture the interplay between partisanship and views on policy and on politicians’ characters, however incorporating economic and cultural variables (which is beyond the scope of this work) is an exciting future direction.

2.2.2.5 Election Prediction

Due to the popularity of social media, using Twitter data to track public opinion and specifically to predict presidential elections has been an active research area. There are conflicting results reported in prior literature for election prediction using sentiment analysis of social media communications. In one of the seminal attempts, Tumasjan et al. [189] analyzed the content of more than 100K tweets published prior
to the German national election. They found that the share of tweets for six different
parties closely matched the election results with an error of less than 2%. O’Connor
et al. [148] studied the correlation between public opinion measured from traditional
polls and sentiment measured from Twitter. They found that while Twitter senti-
ment correlates with consumer confidence and presidential job approval polls, there
is not a strong correlation with the 2008 presidential election poll. Gayo-Avello et
al. [79] also evaluated the power of Twitter data in predicting the 2010 U.S. election
outcomes. They also found that Twitter sentiment analysis is not accurate in pre-
diction and its performance is only slightly better than a random classifier. Other
researchers [113, 78, 174, 167] have also highlighted issues with using Twitter to pre-
dict elections, such as the need of methodological justification in terms of accuracy,
the need to produce a true forecast (i.e. issued prior to the election), the need to
control for biases, etc. Later, some researches [41, 68, 55, 49] tried to address these
problems albeit with limited success. Due to these various issues, in this paper we do
not attempt to predict the outcome of the election. Quite the opposite, we provide
a complex view of the electorate’s decision landscape that cannot be reduced to a
single metric.

2.2.3 Methodology

Our aim is to use state-of-the-art computational approaches to track party,
policy, and personality in online social media communications. Each approach is
driven by a lexicon or a seed list, supplied by subject matter experts, and is then
exploited to produce quantitative measurements for further analysis. Additionally, search templates are used for personality.

2.2.3.1 Party

We infer political affiliations of Twitter users using a method (having the same logic of the method in our first work) that is both simple and efficient. This method is supported by the theory of selective exposure [170] which implies that online social media users tend to follow other users with similar beliefs or ideology. Specifically, in the context of American politics, Democrats are more likely to follow other Democrats and Republicans are more likely to follow other Republicans [60]. Thus, a user following more Republicans than Democrats is likely to be affiliated with Republicans. Similarly, a user following more Democrats than Republicans is likely to be affiliated with Democrats. We manually curated sets of 30 well recognized Democrat (e.g., Rachel Maddow) and 30 Republican (e.g., Sean Hannity) accounts on Twitter as the “landmarks.” The curation was done in consultation with political scientists. We intentionally include several journalists with recognized affiliations as landmarks because of their large Twitter followings. On average, each Democratic landmark has 223,656 followers and each Republican landmark has 277,671 followers. Political affiliation is then a function of the number of landmark Democrats and Republicans that each user follows on Twitter: \( \frac{\#\text{Republicans} - \#\text{Democrats}}{\#\text{Republicans} + \#\text{Democrats}} \). The output is in the range of \([-1, 1]\), where -1 indicates Democratic affiliation, +1 indicates Republican affiliation, and 0 indicates Other (independent or alternative). Our method to
infer political affiliation of Twitter users based on whether they follow well-recognized Democratic and Republican landmarks is high precision and low recall. Therefore, the "Other" category may include politically inactive Twitter users whose political affiliation cannot be inferred by our method [58].

2.2.3.2 Personality

To characterize personality perceptions, Bhattacharya et al. [43] used the Adjective Check List (ACL) [85], which has 300 adjectives or traits commonly used to characterize a person's personality. The ACL covers a wide variety of traits such as intelligent, creative, determined, cheerful. Traits are viewed as either positive (e.g., honest), negative (e.g., anxious) or neutral (e.g., jolly). Simonton reduced them to a core set of 110 traits with factor analysis and used these traits to characterize the personality of 39 American presidents [173]. Simonton further consolidated these 110 traits into 14 non-orthogonal personality dimensions, which include moderation, friendliness, intellectual brilliance, machiavellianism, poise and polish, achievement drive, forcefulness, wit, physical attractiveness, pettiness, tidiness, conservatism, inflexibility, and pacifism. Subsets of the 110 traits are given loadings on a continuous scale of [-1,1] that show how traits positively or negatively contribute to a particular personality dimension. For example, the intellectual brilliance dimension is composed of artistic (.84), inventive (.76), curious (.74), intelligent (.64), sophisticated (.62), insightful (.54), wise (.46), dull (-.71), and commonplace (-.41).

In our study, we characterize perceptions on the personalities of the 2016
presidential candidates using Simonton’s 110 traits and 14 personality dimensions. We use a set of forty high-precision search templates [43] to identify tweets expressing these personality perceptions. There are two types of templates, one to retrieve tweets stating that a trait is present and the other to retrieve tweets saying that a trait is absent. An example template is: [P] is [A]? [T]. Here [P] is a variable representing a person name such as Clinton or Hillary Clinton, [A] represents a class of high certainty words (e.g., definitely, very) and [T] is a specific trait such as honest (or its synonyms). The ‘?’ designates item being optional. This template retrieves statements such as ‘Hillary Clinton is certainly smart’ and ‘Hillary Clinton is intelligent’. Another example is [P] is [S] [T], where [S] represents words that are only somewhat certain (e.g., sort of, somewhat, kinda). It retrieves statements such as ‘Hillary Clinton is sort of decisive’ and ‘Hillary Clinton is somewhat friendly’. We consider negation in statements (‘Hillary Clinton is not decisive’) and trait antonyms (‘Hillary Clinton is not unfriendly’). We further manually examine tweets retrieved by the search templates to eliminate false positives due to sarcasm. Using the tweets retrieved by these search templates, we calculate a score for each trait. Since different presidential candidates may accumulate varying numbers of tweets, we normalize this score for the number of tweets discussing the trait. Further details are in [43].

2.2.3.3 Policy

To understand how policy preferences of candidates impact their perception, we track tweets related to 11 different policy categories for each candidate. The list
of policies includes abortion, gay rights, climate change, foreign policy, health care, immigration, gun control, education, economy, veterans, and miscellaneous. While not a complete list by any means, these are some of the key issues discussed in our data, and were identified as most commonly discussed around candidates on Twitter by Rupar et al. [164]. In consultation with political scientists, we compiled the list of keywords for each policy by starting with a few well recognized keywords for each policy, e.g., “pro-life” and “pro-choice” for abortion. We then identified other related keywords with which they co-occurred (e.g., “planned parenthood” was frequently mentioned for abortion). Further, the miscellaneous policy category includes other keywords such as “blacklivesmatter” and “sex discrimination” which could not be neatly included in other policy categories. Given the set of these highly precise keywords for a policy, we extracted all tweets that contain at least one of the keywords. Given the limited size of tweets (140 characters), we can safely assume that the sentiment detected in political tweets con-

2.2.3.4 Sentiment

While our main points of focus are the three factors from *The American Voter*, we also analyze our data using the more common text mining strategy of sentiment analysis. The goal here is to gauge the general negativity or positivity in the discussion surrounding each presidential candidate. Given the limited size of tweets (140 characters), we can safely assume that the sentiment detected in political tweets con-
cerns the entities (in this case, the candidates) mentioned therein [148]. For sentiment analysis, we rely on the SentiWordNet lexicon [30], which assigns positive and negative scores to each synset (set of synonyms) of WordNet (containing around 117K synsets). To this end, we split a tweet’s text as separate sentences, remove symbols such as “< [?]* >”, tokenize, and stem before matching to the SentiWordNet lexicon. To quantify the overall sentiment of each tweet, we use the common approach which is to sum up positive and negative sentiment scores of the matched tokens. If the positive sentiment score is larger than the negative sentiment score, we label the tweet as positive, and similarly for negative. If both scores are equal, we label the tweet as neutral.

2.2.4 Dataset

2.2.4.1 Data Collection

This work uses Twitter collections around 10 major presidential candidates listed in Table 2.3. Two of these candidates are running as Democrats (Clinton and Sanders) and the remaining eight are running as Republicans. For each candidate, we collected tweets posted from November 15, 2015 to February 29, 2016. The data were collected using Twitter’s streaming API with filter keywords (statuses/filter) for each candidate. We used full names of candidates such as “hillary clinton” for Clinton. This API provides all tweets related to the filter keywords, but caps the tweets at 1% of all public tweets. Since more than 500 million tweets per day are posted on Twitter [7], we are set to capture up to five million tweets per day for each candidate.
Note that the highest daily tweet count (for Trump) is less than one million, thus we can safely assume that we are capturing a vast majority of tweets for all candidates. Moreover, for computing Twitter users’ political affiliation, we also used Twitter’s REST API to crawl the follower lists for the sixty landmark accounts.

We decided not to discount retweets versus original tweets when counting tweets. We find retweets to usually be quoted as-is, without modification from the user – our data is 58.15% retweets, and of those 99% are quoted verbatim. It may be the case that retweeting is less powerfully associated with user opinion. It would be an interesting future work to find a suitable discount factor for retweets. We did not explicitly attempt to geographically filter tweets because only a small fraction are GPS located. When examining the GPS located tweets in our data, we find 85.3% to be from the USA and 3.3% from the UK in a distant second place. Our data may include some tweets from international Twitter users, although the lexicons used in our analysis should at least limit them to English language tweets.

2.2.4.2 Data Statistics

As shown in Table 2.3, Trump has a clear lead with about 7 million tweets about him from more than 5 million users. Moreover, about half of the users and tweets were posting about Trump. This is roughly 3X the Twitter attention given to Clinton or Sanders who have similar numbers of tweets and users. Cruz follows with about 6.5 million tweets from close to a million users. The remaining candidates trail behind to varying degrees.
Table 2.3: Statistics of candidate tweet collections from November, 15 2015 to February 29, 2016. Close to half the users and tweets in the dataset were posting on Trump. Trump’s numbers reflect a 3X lead in tweets and users compared to Clinton and Sanders. Negative sentiment tweets significantly outnumber positive and neutral ones. The sentiment charged nature of the dialog is also indicated by the low prevalence of sentiment neutral tweets (about 8%). Same party tweets outnumber competing party tweets for seven out of ten candidates. Democrats tweeted more about Trump than about Clinton or Sanders. Republicans were more interested in Clinton than in Sanders.

Figure 2.4 shows the time series of tweets for all presidential candidates. We observe sharp spikes for both Democratic and Republican candidates. These spikes typically correspond to major election events. Our data collection covers the Iowa, New Hampshire, Nevada and South Carolina caucuses and primaries as well as several debates. Some events happened on the same day for both parties. For example, both Democrats and Republicans held the Iowa caucus on February 1 and New Hampshire primary on February 9. Others happened on different days for each party. For example, the Nevada caucus for Democrats was held on February 20 while that for
Republicans was held on February 23. Different events happened on the same day for both parties as well. For example, the Nevada caucus for Democrats and the South Carolina primary for Republicans were held on February 20.

We observe that all candidates received spikes on major election event days. The magnitudes of spikes differ depending on how well candidates performed at that event. For example, Clinton received 51% more tweets than Sanders on the day she defeated him at the South Carolina primary. As another example, Sanders received 102% more tweets than Clinton when he defeated her in the New Hampshire primary. However, we also note that losing candidates sometimes received more tweets than winning candidates. For example, even though Clinton won the Iowa caucus, Sanders received 8% more tweets. This difference is very small compared to the previous examples possibly because Sanders lost to Clinton by a very small margin. While Trump received more spikes than other Republican candidates, Cruz received the largest spike when he unexpectedly won the Iowa caucus instead of Trump. Cruz received 16% more tweets than Trump on the day of event. As another example, Bush received the second largest spike for the South Carolina primary, even though Rubio and Cruz are ranked second and third respectively after Trump. The spike was because Bush announced that he would quit the race after his disappointing performance in the South Carolina primary. Bush received 166% more tweets than Rubio while only 32% fewer tweets than Trump on the event day.

Besides the spikes for all candidates on days of major election events, candidates sometimes had spikes on specific days based on their own campaign activities.
For example, Trump received many more tweets than other candidates starting December 7 (which then gradually faded after 5-6 days) when he called for a ban on Muslims entering the U.S. [8]. While not obvious in Figure 2.4 due to Trump’s large tweet counts, Kasich had a spike on November 19 when he announced plans to create a new federal agency that would promote Judeo-Christian values [71]. Fiorina also received many more tweets on November 29 when she told the Fox News that Obama is “delusional” for saying that climate change is a major national security threat [10].

Table 2.3 also dissects the tweet data by detected sentiment. Negative tweets significantly outnumber positive tweets for all candidates, but to a different extent for each candidate. For the whole time period, Bush received the highest ratio of negative tweets to positive ones (13.4×), followed by Cruz (7.6×) and Trump (4.3×) while Carson (2.5×), Christie (2.6×), and Sanders (2.7×) received the lowest ratio.

2.2.5 Results

2.2.5.1 Party

The party affiliation serves as a key factor in determining the electorate’s perception of candidates and voting decisions. In the U.S., it is usually dichotomized into Democrats vs. Republicans, although alternative or independent affiliations have been on the rise [54]. For those affiliated, prior literature has shown that only a small fraction of people vote against their party affiliations [51, 127]. Here, we analyze the breakdown of tweet volume and its sentiment with respect to party affiliations of the users as Republicans, Democrats, and the catch all group: Others. In total,
our data contains almost 50 million tweets from more than 11 million Twitter users. It is noteworthy that the total number of tweets from Republicans (6,741,416) and Democrats (6,686,944) are almost equal.

Interestingly, the breakdown of tweets by party in Table 2.3 shows that tweets from users affiliated to the same party outnumber tweets from the competing party for all candidates, except for Fiorina, Christie, and Carson. For example, 23% (1,640,937) tweets mentioning Sanders are from Democrats while 6% (414,679) tweets are from Republicans. As another example, 27% (1,732,862) tweets mentioning Cruz are from Republicans while 13% (821,024) tweets are from Democrats. Overall, Trump received most tweets across all political affiliations as compared to other candidates.

Table 2.3 also shows the presence of some “crosstalk”. While Republican leaning user tweets were mostly about Republican candidates (78%), Democrat leaning users were tweeting slightly more (57%) about Republican candidates than about candidates from their own party. Thus, we find far more “crosstalk” from Democratic leaning users than from Republican leaning users [35], suggesting an aggravated echo-chamber effect on the Republican side. Republican leaning users tweeting about Democrats were far more interested in Clinton (2.6× more) than in Sanders. Whereas when tweeting about Republicans, Democrats mostly talked about Trump, followed Cruz in the distant second.

Table 2.3 also provides the sentiment breakdown of tweets for all party affiliations. As indicated earlier, negative tweets - almost three-fourths of all tweets - dominate. This observation holds for each candidate as well. The highly charged
nature of the Twitter election dialog is also seen in the low numbers of sentiment neutral tweets. Around 8% of the tweets posted by each user group is sentiment neutral.

### 2.2.5.2 Personality

Table 2.4 provides an overview of more than 316 thousand tweets in our data that convey perceptions of candidates’ personality along six out of fourteen personality dimensions. The table is again broken down by party affiliation, Republican leaning or Democrat leaning. The most discussed personality dimension is moderation, which accounts for almost one-third of all personality-related tweets. This is followed by friendliness, machiavellianism, intellectual brilliance, pacifism, and wit. The remaining personality dimensions are discussed infrequently and are therefore left out of the table.

Comparing the types of posts made by users leaning towards each party we find a few differences. For example, Democrats discuss friendliness of candidates more than Republicans. In contrast, Republicans discuss machiavellianism and pacifism of
Figure 2.5: Scores for four most discussed personality dimensions. Scores are in the range [-1, +1]. -1 (+1) indicates the personality is viewed as absent (present) with high confidence. The number of tweets for absence (presence) of each personality dimension are provided on the left (right) of the bar plot. The net score is included in the parenthesis. The bar plots are shown for candidates receiving a minimum of 100 tweets on a dimension.
candidates more than Democrats. Overlaying sentiment, we note that while negative tweets account for about four-fifths of all personality-related tweets overall, Republicans tend to be more negative than Democrats (83% vs. 77%). Finally, as compared to Republicans, Democrats are more positive about pacifism and more negative about machiavellianism and wit.

Figure 2.5 shows the candidate-level breakdown of the four most frequently discussed personality dimensions. We excluded candidates with very few tweets for some personality dimensions from our analysis. For systematic analysis, we divide candidates into high-frequency and low-frequency groups based on their tweet counts for each personality dimension and limit candidate comparisons to each group.

**Moderation.** Clinton and Rubio are perceived as moderate while Trump, Sanders, and Cruz are perceived as not moderate in the high-frequency group. Clinton received the highest score of 0.28 for moderation. For example, Clinton received more than 11 thousand tweets containing phrase “Hillary Clinton Is Calm, Cool and Effective”. Keywords calm and cool are positively aligned with moderation. On the other hand, Trump received -0.14 score for moderation – he received more than a thousand tweets including “Donald Trump is terrible for disrespecting Moslems” (terrible being negatively aligned with moderation). In the low-frequency group, Bush, Carson, and Kasich are perceived as moderate while Fiorina is perceived as not moderate.

**Friendliness.** Sanders is perceived as friendly, while Cruz and Trump not, whereas Clinton is seen as neutral in the high-frequency group. Kasich and Carson are
perceived as friendly while others, led by Fiorina, are perceived as not friendly in the low-frequency group. In particular, Sanders was considered by many as cute (“Bernie Sanders is so cute”), whereas, for instance, Cruz received more than a thousand tweets saying “Cruz is nasty” and “Ted Cruz isn’t likeable”. Although having no direct relation to the political strengths of the candidates, surprisingly, thousands of tweets referred to their personal amiability.

Machiavellianism. On the opposite side we find machiavellianism, associated with being deceitful and unscrupulous. In the high frequency group, Clinton and Cruz are perceived as machiavellian, while Sanders and then Trump are not. For example, Cruz received hundreds of tweets such as “Ted Cruz is untrustworthy” and “Ted Cruz is so dishonest that TV stations won’t run his ads because they’re worried about legal culpability”. On the other hand, Sanders received more than a thousand tweets such as “100% of 18–29 year olds think Bernie Sanders is honest and trustworthy” which result in -0.37 machiavellian score for him.

Intellectual Brilliance. In the high-frequency group, Trump and Cruz and both are perceived as intellectually brilliant. For example, Trump received more than six thousand tweets such as “Donald Trump Is Smart To Remind Voters Of Clinton Drama”. All other candidates are in the low frequency group and most of them (Clinton followed by Bush, Sanders, and then Fiorina) are perceived as intellectually brilliant. Only Kasich and Christie are perceived as not intellectually brilliant. Kasich, for example, even received tweets like “John Kasich is a very stupid man” contributing to his low score.
Table 2.5: Statistics of tweets discussing different policies, and their breakdown across party affiliation and sentiment.

2.2.5.3 Policy

Table 2.5 presents an overview of the more than six million tweets about policy in our data. The table is limited to the six most discussed policies and the remaining five policies are summarized under ‘Other Policies’. Foreign policy (33%), immigration (23%), healthcare (14%), and the economy (13%) are the most discussed policies. Democrat leaning users express most interest in foreign policy compared to immigration which ranks second in drawing their attention (11% difference). In contrast, the two policies draw equal attention from Republican leaning users. Republicans tweet more about abortion, health care and immigration (3-6% difference) than Democrats while Democrats tweet more about gay rights, foreign policy, and economy (2-6% difference) than Republicans. Overlaying sentiment, we again note that negative tweets account for about four-fifths of all policy-related tweets. While Republicans are more positive than Democrats for gun control, veterans, and gay rights (6-8% difference) the order is reversed for abortion (6% difference).

Figure 2.6 shows the candidate-level breakdown of the six most frequently
Figure 2.6: Number of tweets for the six most discussed policies with party affiliation breakdown. The blue and red regions indicate tweets from Democrats and Republicans, respectively. The white region represents tweets from the remaining users.
discussed policies. We note that Clinton, Sanders, Trump, Cruz, and Rubio received much more policy-related tweets than the remaining candidates. Thus, we limit our analysis to these top candidates.

**Foreign policy.** Trump is by far the most discussed candidate for foreign policy, with almost twice relevant tweets than Clinton in the second place. The most mentioned foreign policy keyword is *ISIS* for both candidates, with Trump having about 200 thousand tweets and Clinton having about 150 thousand matching tweets. However, after *ISIS*, the most mentioned keyword for Trump is *Paris* (78 thousand tweets) and that for Clinton is *Benghazi* (75 thousand tweets). This observation can be explained by Clinton’s email scandal as part of hearings conducted by the House Select Committee on Benghazi.

**Immigration.** Trump is the most discussed candidate for immigration, followed by Cruz, Rubio, and Clinton. In general, for these candidates three most mentioned keywords are *refugee* and *immigrants*. However, immigration-related tweets for Rubio mostly had the keyword *amnesty*. This observation can be explained by Rubio’s change his stance on amnesty. While the other Republican candidates were consistently against amnesty, Rubio changed his stance on amnesty when he ran for the presidency.

**Healthcare.** Trump and Cruz received most healthcare related tweets, followed by Sanders and Clinton. The most mentioned keyword in healthcare is *ACA* (Affordable Care Act) which outnumbers *ObamaCare*.

**Economy.** Democratic candidates are most discussed with respect to the
economy, with Sanders leading Clinton. The most mentioned keywords in economy related tweets are *Wall Street, tax, social security,* and *poverty.*

**Abortion.** Three most discussed candidates about abortion are Cruz, Trump, and Clinton – they received roughly equal number of tweets. There most mentioned keyword about abortion is *Planned Parenthood.* Each of the top three candidates received about 30 thousand tweets containing *Planned Parenthood* with much more tweets from Republicans than Democrats.

**Gay rights.** Republican candidates are the most discussed about gay rights, with Cruz in the lead followed by Trump and Rubio. The most mentioned keyword in gay rights is *gays,* with Cruz and Trump each receiving more than 50 thousand tweets. Other popular keywords include *gay marriage, same-sex marriage,* and *marriage equality.* All candidates received more tweets about gay rights from Democrats than Republicans.

### 2.2.6 Impact of Party, Personality, and Policy on Voting Outcomes

Heretofore, we showed that the factors of party, personality, and policy outlined in *The American Voter* can be tracked in social media using computational methods, and that the output of these methods is directly interpretable in the context of the ongoing presidential campaigns. However, to what extent do these signals correspond to the outcomes of electoral polls? To answer this question, by which we may confirm or disconfirm the validity of *The American Voter’s* framework in the social media-based electoral opinion modeling, we quantify the effect each factor has
on electoral outcomes.

Using the methods described above, we build features for the factors of party, personality, and policy:

- 3 features for party (Democrat, Republican, Other),
- 14 features for different personality dimensions, which are separately computed for presence and absence as well as for party affiliations,
- 11 features for different policy categories, which are separately computed for party affiliations,
- raw tweet counts and sentiment-labeled tweet counts,

in total resulting in 184 variables available to regress on poll numbers. Since a large number of polling organizations conduct polls at different time intervals, we rely on the RealClearPolitics (RCP) [9] aggregation of reputed national polls. Specifically, we use RCP’s daily average of national polls for all presidential candidates.

First, we preprocess this set of 184 variables for use in regression analysis. We log-transform all volumetric variables (tweet counts) to alleviate the skew. Also, many of the features are interdependent, making any model built with such variables suffer from multi-collinearity. Thus, we use a stepwise selection process to eliminate features with high VIF (Variance Inflation Factor) and AIC (Akaike Information Criterion) scores. The process results in 72 variables for the Democratic candidates and 61 for the Republican candidates.

Due to data scarcity, we are not able to separately run regressions for individual candidates (however such models would be very interesting for individualized
Table 2.6: List of statistically significant variables \((p\text{-value} < 0.001)\) in our regression model for Democratic candidates. The features are sorted in descending order with respect to estimated coefficient values. Adjusted \(R^2 = 0.908\)

<table>
<thead>
<tr>
<th>Feature</th>
<th>(\beta)</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tweets from Republicans on Abortion</td>
<td>1.48</td>
<td>0.20</td>
</tr>
<tr>
<td>Tweets from Republicans on Veterans</td>
<td>1.33</td>
<td>0.3</td>
</tr>
<tr>
<td>Tweets on presence of Pacifism</td>
<td>1.30</td>
<td>0.17</td>
</tr>
<tr>
<td>Tweets from Democrats on Gay rights</td>
<td>1.18</td>
<td>0.24</td>
</tr>
<tr>
<td>Tweets from Republicans on Healthcare</td>
<td>1.05</td>
<td>0.28</td>
</tr>
<tr>
<td>Tweets on presence of Machiavellianism</td>
<td>0.85</td>
<td>0.19</td>
</tr>
<tr>
<td>Tweets from Democrats on Gun control</td>
<td>0.58</td>
<td>0.16</td>
</tr>
<tr>
<td>Tweets from Democrats on Abortion</td>
<td>-0.96</td>
<td>0.18</td>
</tr>
<tr>
<td>Tweets from Democrats on Miscellaneous</td>
<td>-1.14</td>
<td>0.23</td>
</tr>
<tr>
<td>Tweets from Republicans on Climate change</td>
<td>-1.55</td>
<td>0.21</td>
</tr>
<tr>
<td>Tweets from Others on Veterans</td>
<td>-1.65</td>
<td>0.30</td>
</tr>
<tr>
<td>Score of presence of Achievement Drive</td>
<td>-2.78</td>
<td>0.79</td>
</tr>
<tr>
<td>Tweets from Republicans on absence of Intellectual Brilliance</td>
<td>-3.37</td>
<td>0.96</td>
</tr>
</tbody>
</table>

The model for Democratic candidates (Clinton and Sanders), which uses the 24 variables, explains 0.908 proportion of variance (adjusted \(R^2 = 0.908\)).\(^2\) Table 2.6 lists the subset of 13 variables with \(p\text{-value} < 0.001\). The model includes 4 personality-based variables and 9 policy-based variables, most of which have integrated the party aspect. For personality, we note that tweets on presence of pacifism and machiavellianism have a positive impact on poll numbers. Interestingly, tweets from Republicans on absence of Intellectual Brilliance have a negative impact on poll

\(^2\)Note that the goal of our regression model here is descriptive. Therefore, we tolerate some degree of over-fitting to the data.
Table 2.7: List of statistically significant variables ($p$-value < 0.001) in our regression model for Republican candidates. The features are sorted in descending order with respect to estimated coefficient values. Adjusted $R^2 = 0.883$

<table>
<thead>
<tr>
<th>Feature</th>
<th>$\beta$</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tweets from Others on Miscellaneous</td>
<td>1.35</td>
<td>0.21</td>
</tr>
<tr>
<td>Tweets from Others on Climate Change</td>
<td>1.27</td>
<td>0.21</td>
</tr>
<tr>
<td>Tweets on absence of Conservatism</td>
<td>1.17</td>
<td>0.22</td>
</tr>
<tr>
<td>Tweets from Republicans on Foreign policy</td>
<td>1.12</td>
<td>0.14</td>
</tr>
<tr>
<td>Tweets from Others on Gun control</td>
<td>1.10</td>
<td>0.17</td>
</tr>
<tr>
<td>Tweets from Republicans on Abortion</td>
<td>0.95</td>
<td>0.16</td>
</tr>
<tr>
<td>Tweets from Democrats on Gay rights</td>
<td>0.54</td>
<td>0.13</td>
</tr>
<tr>
<td>Tweets from Others on Foreign policy</td>
<td>-0.58</td>
<td>0.16</td>
</tr>
<tr>
<td>Tweets on presence of Pacifism</td>
<td>-1.07</td>
<td>0.15</td>
</tr>
<tr>
<td>Tweets on absence of Forcefulness</td>
<td>-1.04</td>
<td>0.31</td>
</tr>
<tr>
<td>Tweets from Democrats on Gun control</td>
<td>-1.09</td>
<td>0.21</td>
</tr>
<tr>
<td>Tweets from Democrats on Climate change</td>
<td>-1.15</td>
<td>0.20</td>
</tr>
<tr>
<td>Tweets from Democrats on absence of Moderation</td>
<td>-2.65</td>
<td>0.80</td>
</tr>
</tbody>
</table>

For policy, the impact of party aspect is more obvious on policy-related variables with four from each party. It is noteworthy that tweets from Republicans on abortion have a positive impact, whereas tweets from Democrats on abortion have a negative impact on poll numbers. Similarly, tweets from Republicans on veterans have a positive impact whereas those from Others have a negative impact on poll numbers of Clinton and Sanders.

The model for Republican candidates, which uses the 28 variables, explains 0.883 proportion of variance (adjusted $R^2 = 0.883$). Table 2.7 lists the subset of 13 variables with $p$-value < 0.001. The model also includes 4 personality-based variables and 9 policy-based variables, most of which have integrated the party aspect. For personality, we note that tweets on absence of conservatism has the most positive impact on poll numbers, reflecting the preference for political outsiders. In contrast, tweets on presence of pacifism as well as on absence of forcefulness have a negative
impact. It is interesting to note that tweets from Democrats on absence of moderation has the most negative impact on poll numbers. For policy, we note that tweets from Others on climate change and gun control have a positive impact, whereas those from Democrats have a negative impact on poll numbers. Moreover, tweets from Republicans on foreign policy have a positive impact, whereas those from Others have a negative impact on poll numbers.

The comparison of these models for both parties reveals several key similarities and differences. Among the most significant variables with $p$-value $< 0.001$, both models include 4 personality-based variables and 9 policy-based variables. Furthermore, both models include policy-based variables about abortion, gay rights, gun control, and climate change. The interesting differences between these models are related to the sign of variable coefficients. For example, tweets on presence of pacifism have a positive impact ($\beta = 1.30$) on poll numbers of Democratic candidates while they have a negative impact ($\beta = -1.07$) on poll numbers of Republican candidates. Similarly, tweets from Democrats on gun control have a positive impact ($\beta = 0.58$) on poll number of Democratic candidates while they have a negative impact ($\beta = -1.09$) on poll numbers of Republican candidates.

2.2.7 Discussion

In our exploration of *The American Voter* factors on Twitter, we find that all three factors still play an important role in the electoral discourse. The interplay between these factors is especially fruitful for further analysis and validation of estab-
lished political theories. For instance, in Figure 2.7 we plot the proportion of tweets from users of different political affiliations about personality dimensions of the two eventual nominees – Clinton (above) and Trump (below). We can see that the personality dimension of moderation is much more emphasized for Clinton, however it is only the Other (perhaps more independent) and Republican users who emphasize her pacifism (which includes keywords such as *weak*). Further, only users marked to be Republican emphasize Trump’s conservatism, whereas the Other group - potentially the independent vote - not at all.

Figure 2.7: Proportion of tweet volume about personality dimensions of Clinton (above) and Trump (below) from users identified as Democrats (blue circles), Republican (red squares), and Other (grey triangles).
Further examining partisanship, we were able to identify 572,316 and 893,048 of the 11 million users as Republican leaning and Democrat leaning respectively. Although the seed list used in our approach limits the reach of our classification, the Twitter conversation is dominated by a vast majority that is not following these established opinion leaders. Even among those we detected as interested in partisan opinion leaders, we observed considerable crosstalk about other party’s candidates. However, when combined with the overall negative tone, this possibly indicates the presence of partisanship for those already aligned with a party, as found, for example by [17].

Moreover, our study corroborates findings reported in prior literature and also highlights some new trends. For example, Benoit [40] reported that Republicans emphasized more on personality traits like ‘sincerity’ while Democrats emphasized more on ‘drive’. In corroboration, we also find that Republicans discuss the personality dimension of machiavellianism (which includes sincere trait) more than Democrats, who discuss the personality dimension of achievement drive more than Republicans. Benoit [40] also reported that Democrats focus more on policy issues such as education while Republicans focus more on national security. Our analysis shows that Democrats not only discuss education more than Republicans, but also issues such as gay rights and climate change. On the other hand, Republicans discuss policy issues such abortion and veterans more than Democrats.

Methods presented in this work were able to uncover more than 6 million policy-related tweets and 316 thousand personality-related tweets, but we caution the
reader not to focus on raw volumes, as they depend on the sensitivity of the lexicons used wherein. Instead, cross-policy and cross-personality trait analyses would present more insight, as presented in this work. The crafting of more accurate and up-to-date lexicons is a task involving subject-matter experts, and thus can be only partially automated, however attempts have recently been made to automatically augment lexicons, such as in crisis situations [149].

2.2.8 Summary

The multi-factor framework of *The American Voter* brings a structure to the rich political discourse on social media. Our use of computational methods allows the evolution from survey-based electoral research to encompass the automated analysis of new rich sources of user-generated media. Our findings reveal the partisan divides in the perceptions of policy positions and personality traits of the presidential candidates. Further, using statistical analysis of *party*, *personality*, and *policy* factors we show the importance of each one in the modeling of electoral deliberation.

No social media study would be complete without the standard cautionary statement that social media analysis is not meant to replace the traditional survey techniques, having neither comparable participant sampling, nor the affordances of a (sometimes long-form) survey. However, the fact that political figures are increasingly using Twitter to reach their supporters indicates its increasing importance, at least in the eyes of the politicians themselves. This study establishes a baseline from which the views of the candidates can be studied, views which we fully expect to change
not only in this election cycle but also in future elections.
CHAPTER 3
UNDERSTANDING THE ROLES OF ALGORITHMS IN POLITICAL POLARIZATION

3.1 Background, Motivation, and Research Statement

Researchers have long been concerned about the effect of selective exposure—seeking information that reinforces preexisting beliefs while avoiding other information—on democratic societies [183, 210]. Through selective exposure, people with racial and political biases seek out content that conforms with their preexisting beliefs. For example, selective exposure to conservative news is associated with support for strict immigration policies [211], and people who score high on a scale of modern racism are more likely to view non-traditional Internet sources as credible sources of news [140].

More recently, researchers have started to question whether algorithms create distinct personalized experiences for users, leading to so-called “filter bubbles” or “echo chambers” [108, 151, 33]. Algorithms reflect the societies in which they are produced, so it is unsurprising that they are encoded with racial and political biases [77]. For example, personalization algorithms have been found to discriminate against women [67] and people of color [147].

While selective exposure requires deliberate acts of media choice, algorithmic personalization interprets past behavior as precedent for future preference. In other words, algorithmic personalization can intensify selective exposure beyond a person’s choice, resulting in a vicious cycle that can contribute to an increasingly polarized
society. According to Pew Research, Americans are more ideologically polarized now than at any point in the last two decades [53]. This increase in polarization coincides with the rise of search engines and social media sites as primary sources of information. Thus, it is important to study the role of personalization algorithms employed by search engines and social media sites in reinforcing pre-existing biases.

Prior research has reported significant personalization in online advertising based on a user’s browsing history. Personalization in online advertising [34, 198, 67, 176] is not surprising because behavioral targeting is its key design feature. To target ads, advertisers employ sophisticated online tracking techniques for profiling users’ browsing habits across the Web [72]. Naturally, we would expect search engines to leverage browsing history information obtained by online tracking for search personalization as well. However, prior studies [203, 91, 116] have reported significant search personalization mainly based on location, not browsing history.

In fact, when a person looks up information on a search engine, the search does not occur in a vacuum. Before any particular search is made, the person has already lived an active digital life—reading news stories, liking pages on online social media, and posting comments on blogs—activities that are tracked across the Web to train personalization algorithms. In the current divisive political climate [53], people might be more likely to access content online that conforms to their political ideologies, but little is known about whether accessing certain political content impacts personalization of other elements of a person’s online life. Set in this context, we ask: If a person’s digital life demonstrates a political bias, does that person receive
personalized search results and do those search results reflect that person’s bias?

Thus, in this research we would like to study search personalization in Google News using a “sock puppet” auditing system [166] in which we use computer programs to impersonate users with different political ideologies. Our results show that Google can infer browsing histories of our trained profiles through its pervasive online tracking network and can leverage it to personalize search results based on recent changes in its privacy policy [6, 27]. To the best of our knowledge, this work presents the first empirical evidence of political personalization on Google News based on browsing history. The results not only set the baseline for search personalization based on political biases in browsing history, but also contribute to the broader understanding of selective exposure and algorithmic personalization.

3.2 Related Work

A long line of research has looked at web search personalization. Hannak et al. [91] examined several features such as browser user-agent, user’s ethnicity, browsing history, and search query that Google can use to personalize search results. However, they only found significant personalization based on Google account login status and IP geolocation. In a follow-up [92], the author extended this study to Bing and DuckDuckGo. They found that, on average, 15.8% search results on Bing are different due to personalization based on the same features as on Google. They found no significant personalization on DuckDuckGo.

To further examine the role of location in triggering personalization, Xing
et al. [203] used a browser extension called Bobble and reconfirmed that a user’s location (inferred by geolocating IP addresses) is the most dominant trigger factor for personalization. Furthermore, Kliman-Silver et al. [116] examined the relationship between distance and location-based personalization, and the impact of location-based personalization on different types of queries. They found that the differences between search results grow as the physical distance between locations of the users increases.

Puschmann [155, 156] analyzed crowdsourced Google search results from approximately 4,000 users for 16 terms related to German political parties and candidates [13]. He found that search results are personalized primarily based on location, language, and time. He concluded that a small fraction of personalized search results, which could not be explained based on the aforementioned factors, may be trigged by latent (or non-observable) factors such as browsing history and Google account information that they did not collect from users.

Robertson et al. [161] audited personalization on Google Search by asking crowd workers to install a browser extension which collected Google search results for a variety of political search queries in standard and incognito modes. They found significant personalization based on Google account login status, self-reported usage of Google/Alphabet services such as YouTube, and self-reported rating of Trump but not self-reported political party affiliation. In a follow-up [160], the authors reported that personalized search results have minimal differences in terms of political bias.

In short, prior research has found Google search personalization primarily due
to geolocation and account login status. In contrast to prior work, we believe that our sock puppet auditing system can help reveal the impact of browsing histories, which are explicitly designed to reflect different political ideologies, on search personalization.

### 3.3 Methodology

Building on existing research focused on tracking and personalization, the overarching goal of our study is to examine the extent to which a user’s web browsing history affects algorithmic personalization in web search. We seek to understand both whether algorithmic personalization exists and whether it reinforces or dampens the preferences displayed in web browsing history. To this end, we design and implement a sock puppet auditing system to train fresh browser profiles by visiting hyperlinks that embody distinct political discourses and then compare Google News search results for identical politically oriented search terms. Figure 3.1 depicts the four main components of our system that are discussed below.

Figure 3.1: Our sock puppet auditing system to measure political personalization of Google News search.
3.3.1 Identifying Pro- and Anti-immigration Hyperlinks

We aim to evaluate Google News personalization based on browsing histories that reflect different stances on the topic of immigration. We focus on immigration because it was a key campaign issue in the recent 2016 U.S. presidential election [26, 122]. Our goal was to find a list of websites that reflect pro- and anti-immigration stances in the context of U.S. politics. To this end, we rely on two popular Twitter accounts that reflect both discourse communities. Discourse communities are groups of people that come together, whether physically or virtually, for the purpose of building community through shared goals and forms of speech [185]. For the anti-immigration stance, we use @wginfonetorg, which belongs to a popular anti-immigrant website whitegenocide.info that is dedicated to “fighting the crime of white genocide.” For the pro-immigration stance, we use @DefineAmerican, which belongs to a popular pro-immigrant website defineamerican.com that is dedicated to “shift the conversation about immigrants, identity, and citizenship in a changing America.” Both of these Twitter accounts are very active and fairly popular. As listed in Table 3.1, @wginfonetorg has posted more than 63K tweets and has 9.4K followers while @DefineAmerican has posted more than 17K tweets and has 30.3K followers. Because both accounts frequently share content from other websites that support their respective views on immigration, we collected the hyperlinks posted on the timelines of these Twitter accounts over the duration of two weeks in March 2017. We treat these two sets of hyperlinks as representations of pro- and anti-immigration discourses.
Table 3.1: Statistics of pro-immigration and anti-immigration Twitter accounts.

<table>
<thead>
<tr>
<th>Screen Name</th>
<th>Favorites</th>
<th>Followers</th>
<th>Followings</th>
<th>Tweets</th>
<th>Account Creation</th>
</tr>
</thead>
<tbody>
<tr>
<td>DefineAmerican</td>
<td>10,043</td>
<td>30,253</td>
<td>1,776</td>
<td>17,350</td>
<td>May 2011</td>
</tr>
<tr>
<td>wginfo.netorg</td>
<td>1,253</td>
<td>9,424</td>
<td>223</td>
<td>63,916</td>
<td>April 2014</td>
</tr>
</tbody>
</table>

3.3.2 Training Browser Profiles

Using the two sets of hyperlinks collected from pro- and anti-immigration Twitter accounts, we now train fresh browser profiles. Specifically, to train the pro-immigration browser profile, we install a fresh copy of Firefox web browser and open hyperlinks from the timeline of the pro-immigration Twitter account. Similarly, to train the anti-immigration browser profile, we install a fresh copy of Firefox web browser and open hyperlinks from the timeline of the anti-immigration Twitter account. We take several steps to imitate a real user during browser profile training. First, we limit the number of hyperlinks that we use for training per day to 50. To this end, we randomly sample a subset of 50 hyperlinks from the hyperlinks posted on the timelines of both Twitter accounts. Second, we add a random delay averaged at five minutes between opening consecutive hyperlinks during training. Finally, we conduct the training between 9 am and 11 pm local time to reflect diurnal user activity.

It is noteworthy that we do not create or log-in to a Firefox account during training because it can save bookmarks, passwords, browsing history, and cookie information. We also do not create or log-in to a Google account that may be used by Google to link our account information with browsing history. Therefore, the only way for Google to know about our browsing history during training would be via
third-party tracking using its own advertising/analytics network (e.g., DoubleClick and Google Analytics) via cookies or browser fingerprinting. As discussed in related work, prior research has reported on the extensive third-party tracking capabilities of Google.

Our analysis of the set of hyperlinks used for training browser profiles shows that pro-immigration training hyperlinks belong to more mainstream news outlets such as Washington Post while anti-immigration training hyperlinks belong to more alternative news sites such as Breitbart and user-generated content sites such as YouTube and WordPress (Figure 3.2). We find that several Google owned domains are top third-parties on these training hyperlinks. Figure 3.3 plots the popularity distribution of third-party domains on the pro- and anti-immigration training hyperlink sets. More specifically, doubleclick.net is the top third-party tracker on both anti- and pro-immigration training hyperlink sets. Thus, we conclude that Google is able to learn the browsing histories of our trained pro- and anti-immigration browser profiles and can later use it for personalization of search results.

3.3.3 Searching Google News

After training fresh browser profiles using hyperlinks crawled from pro- and anti-immigration Twitter accounts, we are set to conduct Google News searches. We are interested in testing if web browsing histories that reflect these divergent discourses would result in search personalization along politically partisan lines.¹

¹Note that the extreme anti-immigration position of those concerned about white genocide are not wholly aligned with the Republican Party’s nor are the political views of our
In other words, we want to test whether a user who consistently consumes anti-immigration content would receive more or less Republican-leaning news stories in the search results. We execute searches from three sets of browser profiles: (1) pro-immigration, (2) anti-immigration, and (3) control. The control browser profile, which is not trained (i.e., by visiting hyperlinks), provides us a “neutral” perspective to judge whether personalization occurs for pro- and/or anti-immigration profiles. We execute Google News searches using different search terms about a wide variety of

Figure 3.2: Top-10 first-party domains for hyperlinks crawled from pro- and anti-immigration Twitter accounts.

pro-immigration profile identical to the Democratic Party platform. These two profiles are not equally polarized even in terms of immigration position. In fact, after our experiment is finished, we discover that the anti-immigration profile was suspended since it violated Twitter rules by involving into extremely hateful speeches.
political topics. We decide to search for news related to the top-10 most discussed policy issues on Twitter [122] during the 2016 U.S. presidential election campaign.

We use five different search terms for each of these policy issues: Immigration, Foreign
Policy, Healthcare, Economy, Abortion, Gay Rights, Gun Control, Climate Change, Education, and Veterans. For each search term, we repeat the search process every day over the duration of one week.

To ensure that the search results returned by Google are not impacted by anything other than the browsing histories, our search process is designed as follows. First, we use Selenium WebDriver [3] to automatically conduct searches using the Google News web interface instead of Google's search API. We configure Google News to return up to 100 search results. Second, we use a wait period of 11 minutes between consecutive searches to eliminate any carry-over effect [91]. Third, we conduct searches from pro-immigration, anti-immigration, and control browser profiles in a synchronized fashion to eliminate any temporal effects. Fourth, each browser profile conducts search queries on a separate Amazon EC2 cloud instance to avoid any interdependencies between different profiles. Fifth, we use a static DNS entry for Google News to ensure that our search queries are routed to the same datacenter. Finally, to eliminate potential noise in search results due to A/B testing [150], we train four separate profiles for pro-immigration, anti-immigration, and control browser profiles. In the absence of A/B testing, we would expect the same search results for these four profiles. When there is A/B testing, we can mitigate its effect by eliminating search result differences. To this end, we compute the pairwise intersection of search results among the four profiles. We then use search results from a randomly selected profile.

---

Note that while each Amazon cloud instance has a different IP address, they are all belong to the same /24 IP prefix range geolocated in Oregon. Therefore, we do not expect any IP geolocation based search personalization as has been reported in prior work [91, 116].
from the pair with the maximum intersection.

3.3.4 Quantifying Search Personalization

We use trained browser profiles for both pro- and anti-immigration stances as well as a control browser profile to execute Google News searches. We can quantify personalization by comparing the presence and ranking of search results across different browser profiles. First, we quantify personalization simply in terms of the differences in search results. Let $P$ and $A$ respectively denote the lists of search results for the pro- and anti-immigration browser profiles. Let $C$ denote the list of search results for the control browser profile. We can measure the differences in search results for the control browser profile. We can measure the differences in search results by comparing $P$, $A$, and $C$ in a pairwise manner. To this end, let $P - A$ represent the set of unique search results for the pro-immigration profile that do not appear for the anti-immigration profile. Let $P - C$ and $A - C$ respectively represent the set of unique search results for the pro-immigration and anti-immigration profiles with respect to the control. We use $|P - A|$, $|P - C|$, and $|A - C|$ to quantify the differences in search results in the range of $[0\%,100\%]$, where $0\%$ indicates no personalized search results and $100\%$ indicates all 100 search results are personalized. Note that since we get 100 search results for each search term, we have $|P| = |A| = 100$. In this case, the difference in search results is symmetric in terms of quantities (i.e. $|P - A| = |A - P|$) although it may be not be symmetric as the set (i.e. $P - A \neq A - P$). Second, we quantify personalization while taking into account the ranking of search results across different browser profiles. To this end, we measure edit distance among $P$, $A$, and
C in a pairwise manner. Specifically, we compute the Damerau-Levenshtein distance [66] which represents the number of deletions, insertions, or substitutions needed to make a pair of lists identical. If two lists are identical, the edit distance is 0. The greater the edit distance, the more different the lists are in terms of their membership and ordering. Note that both the difference and edit distance are calculated based on search results as URLs and we do not consider title or text of these URLs.

In addition to quantifying personalization, we also analyze whether personalized search results reflect political bias. To assess political bias, we prefer automatic methods over manual coding because the latter are expensive and time consuming. Thus, we first try to use methods in prior literature [120, 119] that can automatically measure political bias of URLs by analyzing their sharing patterns on Twitter. Since these methods require a large number of tweets containing the news URLs, we cannot use them because very few of the personalized search results (URLs) are frequently tweeted. Therefore, we use mediabiasfactcheck.com which provides the political bias of 1,540 media sources (identified as domains) on a 5-point scale: left, left-center, center, right-center, and right [15]. For further quantification, we convert this 5-point scale to specific scores as: left = -100, left-center = -50, center = 0, right-center = 50, and right = 100. Note that mediabiasfactcheck.com provides the political bias for domains (e.g., nytimes.com) not URLs. Since we cannot automatically measure the political bias of personalized news URLs, we estimate their political bias using their domain’s political bias from mediabiasfactcheck.com.
3.4 Evaluation

Table 3.2 reports the personalization results quantified as pairwise difference and edit distance across pro-immigration ($P$), anti-immigration ($A$), and control ($C$) profiles. In terms of difference, although it varies across different search terms, the average difference between pro-immigration and anti-immigration ($P - A$) is 3.2%, between pro-immigration and control ($P - C$) is 3.9%, and between anti-immigration and control ($A - C$) is 3.8%. Using the standard t-test, we are able to conclude that these average differences are significantly different from zero at 0.0001 significance level. The average edit distance between pro-immigration and anti-immigration is 8.4, between pro-immigration and control is 10.3, and between anti-immigration and control is 10.3. Using the standard t-test, we are again able to conclude that these edit distances are significantly different from zero at 0.0001 significance level. The edit distance values are higher than the difference because edit distance not only considers the difference in two lists of search results but also the changes in relative rankings of the search results. Since difference and edit distance show the same trend in personalization across different search terms, we focus on the difference metric for discussion in the rest of this section.

To analyze personalization across different search terms, we sort search terms in Table 3.2 in descending order of $P - A$ values for each policy category. We focus our attention on search terms for which $P - A$ average difference exceeds 5%. We find that a total of nine search terms meet this criterion: (1) comprehensive immigration reform, (2) uninsured Americans, (3) Medicare for all, (4) national debt, (5) flat tax,
<table>
<thead>
<tr>
<th>Policies</th>
<th>Search Terms</th>
<th>Difference (%)</th>
<th>Edit Distance</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Immigration</td>
<td>Comprehensive immigration reform</td>
<td>6.3</td>
<td>18.2</td>
<td>13.5</td>
</tr>
<tr>
<td></td>
<td>White nationalism</td>
<td>4.3</td>
<td>4.2</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>Dream Act</td>
<td>3.0</td>
<td>5.8</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>Anchor babies</td>
<td>2.5</td>
<td>4.3</td>
<td>6.8</td>
</tr>
<tr>
<td></td>
<td>Illegal immigrants</td>
<td>1.2</td>
<td>2.3</td>
<td>1.2</td>
</tr>
<tr>
<td>Immigration</td>
<td>ISIS</td>
<td>1.7</td>
<td>0.8</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>Benghazi</td>
<td>1.3</td>
<td>5.7</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>Syria war</td>
<td>1.0</td>
<td>1.2</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>Iran deal</td>
<td>0.5</td>
<td>0.5</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>Aleppo</td>
<td>0.0</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>Foreign Policy</td>
<td>uninsured Americans</td>
<td><strong>7.2</strong></td>
<td><strong>11.3</strong></td>
<td><strong>10.5</strong></td>
</tr>
<tr>
<td></td>
<td>Medicare for all</td>
<td>6.2</td>
<td>10.0</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td>Health insurance</td>
<td>4.3</td>
<td>2.0</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>Affordable Care Act</td>
<td>2.7</td>
<td>5.0</td>
<td>6.8</td>
</tr>
<tr>
<td>Healthcare</td>
<td></td>
<td>0.3</td>
<td>4.3</td>
<td>4.3</td>
</tr>
<tr>
<td>Economy</td>
<td>National debt</td>
<td>12.0</td>
<td>16.5</td>
<td>12.2</td>
</tr>
<tr>
<td></td>
<td>Flat tax</td>
<td>6.7</td>
<td>6.2</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>NAFTA</td>
<td>2.8</td>
<td>3.3</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>Wall street</td>
<td>2.8</td>
<td>0.3</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>Federal budget</td>
<td>1.5</td>
<td>1.3</td>
<td>0.7</td>
</tr>
<tr>
<td>Abortion</td>
<td>Pro-life</td>
<td>5.7</td>
<td>2.3</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>Planned parenthood</td>
<td>3.5</td>
<td>7.0</td>
<td>5.4</td>
</tr>
<tr>
<td></td>
<td>Roe v. Wade</td>
<td>2.8</td>
<td>2.2</td>
<td>4.3</td>
</tr>
<tr>
<td></td>
<td>Pro-choice</td>
<td>1.7</td>
<td>3.0</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td>Women’s rights</td>
<td>0.5</td>
<td>2.8</td>
<td>2.8</td>
</tr>
<tr>
<td>Gay Rights</td>
<td>LGBT</td>
<td>4.3</td>
<td>3.2</td>
<td>7.0</td>
</tr>
<tr>
<td></td>
<td>Traditional marriage</td>
<td>4.2</td>
<td>2.8</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>Gay couple</td>
<td>3.8</td>
<td>3.3</td>
<td>4.3</td>
</tr>
<tr>
<td></td>
<td>Marriage equality</td>
<td>3.0</td>
<td>2.8</td>
<td>4.3</td>
</tr>
<tr>
<td></td>
<td>Same-sex marriage</td>
<td>2.0</td>
<td>1.8</td>
<td>1.7</td>
</tr>
<tr>
<td>Gun Control</td>
<td>Gun license</td>
<td>4.3</td>
<td>2.2</td>
<td>5.7</td>
</tr>
<tr>
<td></td>
<td>Background checks</td>
<td>4.3</td>
<td>2.2</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>NRA</td>
<td>3.0</td>
<td>0.3</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>Gun control</td>
<td>2.2</td>
<td>3.8</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>Gun accessibility</td>
<td>0.7</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>Climate Change</td>
<td>Paris climate agreement</td>
<td>7.2</td>
<td>7.2</td>
<td>8.8</td>
</tr>
<tr>
<td></td>
<td>Carbon footprint</td>
<td>5.2</td>
<td>5.4</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>Climate debate</td>
<td>4.3</td>
<td>3.0</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>Greenhouse gases</td>
<td>2.7</td>
<td>2.2</td>
<td>8.8</td>
</tr>
<tr>
<td></td>
<td>Global warming</td>
<td>0.2</td>
<td>2.8</td>
<td>2.7</td>
</tr>
<tr>
<td>Education</td>
<td>No Child Left Behind</td>
<td>5.0</td>
<td>4.7</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>Department of Education</td>
<td>3.3</td>
<td>2.7</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>College affordability</td>
<td>2.8</td>
<td>3.0</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>Race to the Top</td>
<td>1.0</td>
<td>1.8</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>Free community college</td>
<td>0.3</td>
<td>2.7</td>
<td>2.8</td>
</tr>
<tr>
<td>Veterans</td>
<td>support our veterans</td>
<td><strong>7.2</strong></td>
<td><strong>5.8</strong></td>
<td><strong>5.5</strong></td>
</tr>
<tr>
<td></td>
<td>Veterans affairs</td>
<td>3.5</td>
<td>3.7</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>Veterans</td>
<td>1.2</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>Veteran benefits</td>
<td>0.8</td>
<td>1.2</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>PTSD</td>
<td>0.5</td>
<td>3.3</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>3.2</strong></td>
<td><strong>3.9</strong></td>
<td><strong>3.8</strong></td>
</tr>
</tbody>
</table>

Table 3.2: Personalization (quantified using difference and edit distance) for 50 search terms. Note that $E(P, A), E(P, C),$ and $E(A, C)$ are pairwise edit distance among pro-immigration, anti-immigration, and control profiles.
(6) pro-life, (7) Paris climate agreement, (8) carbon footprint, and (9) support our veterans.

The comparison of pro-immigration versus anti-immigration and control profiles for these search terms reveals interesting insights. First, we observe that five of them (comprehensive immigration reform, uninsured Americans, national debt, Paris climate agreement, and support our veterans) also have high \( P - C \) and \( A - C \) differences. For example, national debt search term has 12.0\% \( P - A \) difference, 16.5\% \( P - C \) difference, and 12.2\% \( A - C \) difference. This shows that both pro- and anti-immigration profiles receive personalized search results that are different from each other as well as different from the control. Second, we observe that three of them (flat tax, medicare for all, and carbon footprint) have high \( P - C \) difference but low \( A - C \) difference. For example, flat tax search term has 6.7\% \( P - A \) difference, 6.2\% \( P - C \) difference, and 1.0\% \( A - C \) difference. This shows that the pro-immigration profile receives personalized search results that are different from both the control and anti-immigration profile. Third, we observe that only one of them (pro-life) has a high \( A - C \) difference but low \( P - C \) difference. Specifically, pro-life search term has 5.7\% \( P - A \) difference, 2.3\% \( P - C \) difference, and 4.0\% \( A - C \) difference. This shows that the anti-immigration profile receives personalized search results that are different from both the control and pro-immigration profile.

Note that there are search terms for which \( P - A \) difference is not high (e.g., < 1.5\%), but it does not necessarily mean that there is no personalization because their differences to control (both \( P - C \) and \( A - C \)) may be high. This is because both
Table 3.3: Average personalization for all 50 search terms according to top-k (k \in \{10,20,\ldots,100\}) ranked search results. We normalize edit distance (as k-edit distance) to make a fair comparison across different k values.

Pro- and anti-immigration profiles receive similar personalized search results with respect to the control so they are not much different from each other. For example, Benghazi search term has only 1.3% $P - A$ difference, but 5.7% $P - C$ difference and 6.0% $A - C$ difference. We also note that minor changes in queries can trigger big differences among search results. For example, veterans affairs search term has 3.5% $P - A$ difference that are about three times those of veterans and veteran benefits search terms, which are 1.2% and 0.8% respectively.

Next, we study the rank of personalized search results to gauge whether they disproportionately appear at the top or bottom of the list of results. Table 3.3 reports the average personalization for all search terms based on the top-k (k\in\{10,20,\ldots,100\}) search results. We note that average $P - A$ difference started at 2.3% for k=10, slightly increases for increasingly k values, and reaches 3.2% for k=100. We observe a similar trend for $P - C$ and $A - C$ differences as well as in terms of edit distance.
Table 3.4: Political bias for nine top-personalized search terms. Note that $B(P)$, $B(A)$, and $B(P) - B(A)$ respectively are political bias of personalized search results for the pro-immigration and anti-immigration profiles, and the difference between their political bias. Table is sorted in ascending order of $B(P) - B(A)$ values. A negative $B(P) - B(A)$ value indicates the pro-immigration profile received more Democratic-leaning personalized search results, while a positive $B(P) - B(A)$ value indicates the pro-immigration profile received more Republican-leaning personalized search results.

Note that we normalize edit distance as k-edit distance (edit distance/k) for a fair comparison across different k values. Overall, while personalization slightly increases at bottom ranks, we conclude that personalization remains substantial for top ranked search results. Since personalization (in terms of both difference and edit distance) for several search terms exceeds the average by multiple factors, we gather that there exists significant personalization even in top-10 search results.

We further analyze political bias of the personalized search results for the nine most personalized search terms in Table 3.2. Using mediabiasfactcheck.com, we

---

3In fact, we also analyze political bias of the personalized search results for all 41 remaining search terms. Using mediabiasfactcheck.com, we are able to estimate the political bias of 626 news stories out of 1,235 (51%) personalized results for these search terms.
are able to estimate the political bias of 373 news stories out of 792 (47%) personalized results for these nine most personalized search terms. Table 3.4 reports the average political bias of personalized search results for the pro- and anti-immigration profiles, and the difference between their political bias averages. The average political bias of pro-immigration profile is -8.6 which is Democratic-leaning and that of anti-immigration profile is 1.3 which is Republican-leaning (negative values represent Democratic-leaning and positive values represent Republican-leaning). In other words, the pro-immigration profile receives more Democratic-leaning personalized search results than that of the anti-immigration profile. We use the Kolmogorov-Smirnov test [136] to compare the political bias distributions of personalized search results for pro- and anti-immigration profiles. We are able to reject the null hypothesis that both distributions are the same at the 0.05 significance level. Thus, we conclude that the search terms receiving most personalization tend to get personalized results that reinforce the presumed partisanship. Note that political bias of personalized search results varies across different search terms. Personalization reinforces the presumed partisanship for six out of nine (carbon footprint, comprehensive immigration reform, Paris climate agreement, pro-life, support our veterans, and uninsured Americans) search terms that received most personalization. For the remaining search terms, it counters the presumed partisanship.

However, we find that the difference is not significant between two profiles. Specifically, the average political bias of pro-immigration profile is -5.45 while that of anti-immigration profile is -5.26. Thus, we note that further research is needed to better understand political personalization on specific topics.
3.5 Discussion

We are the first to report political personalization on Google News search based on browsing history. The explanation of why we found evidence of significant web search personalization while past work did not could be the combination of three following reasons.

First, our research is different from prior work in terms of the methodology in training browser profiles. While previous work [91, 90] trained browser profiles to reflect different demographics, we trained browser profiles to reflect different political stances. Specifically, both Hannak et al. [91] and Haim et al. [90] trained browser profiles to represent different demographic groups such as gender, age, income, lifestyle, and ethnicity. In contrast, we trained browser profiles to explicitly reflect opposing views on the topic of immigration rather than different demographics. Specifically, we trained browser profiles using news stories posted by Twitter accounts who clearly demonstrated distinct political stances on the topic of immigration.

The second reason could be the difference in search terms that were used to test personalization. Prior literature reported that different search terms can trigger different magnitudes of personalization [91, 119]. As compared to both Hannak et al. [91] and Haim et al. [90], which used search terms covering a variety of topics popular at the time, we used search terms related to the training topic of immigration. Specifically, [90] used general search terms such as “Germany”. Hannak et. al. [91] used general search terms about topics such as news sources and literature. In contrast, after training browser profiles reflecting opposing views on the topic of
immigration, we used immigration-related search terms such as “comprehensive immigration reform” and “illegal immigrants”. We also used search terms about other relevant political topics [122] such as foreign policy.

Last but not least, the changing nature of Google’s personalization algorithm could be another reason. Google is known to continuously tinker personalization algorithms as well as update their data sources over time. For example, Google changed its privacy policy in 2012 [6] and more recently in 2016 [27] allowing them to combine user data collected across all of its services (e.g., Search, Gmail, Google Analytics, DoubleClick) for targeted advertising and content personalization. Thus, Google can now more effectively personalize search results based on a user’s browsing history inferred from its third-party analytics and tracking network. Since personalization algorithms are continuously being tweaked, we plan to longitudinally study personalization for different search terms over an extended period of time as part of our future work.

Prior work using controlled experiments [91, 90] as well as surveys of real users [155, 156] has been unable to find evidence of search personalization on Google News based on browsing history. While simulating behaviors of real users completely will always be impossible, controlled experiments enable us to isolate the effect of browsing history on search personalization. Thus, in this work, we studied search personalization on Google News in a controlled setting by using browser profiles that were trained to reflect strongly divergent opinions on the topic of immigration. Using this strategy, we find evidence of significant search personalization on Google News.
This finding creates an opportunity to conduct further research on other factors that can trigger search personalization.

3.6 Summary

In short, while controlling for other factors, we train a pair of fresh browser profiles by visiting websites that reflect pro-immigration and anti-immigration stances. We then execute search queries on Google News related to a variety of political topics. We analyze the search results to quantify the magnitude and direction of personalization. While the magnitude and direction of personalization varies depending on the search term, overall we are able to conclude that profiles trained by browsing websites reflecting distinct political positions indeed receive significant personalization that tends to reinforce the presumed partisanship. Our findings provide further empirical evidence for the underlying causes of filter bubbles or echo chambers. Given the heightened political tensions in the United States, search returns reinforcing people’s pre-existing biases could further insulate, rather than unite, individuals.
CHAPTER 4
UNDERSTANDING THE ROLES OF CYBORGS IN POLITICAL
POLARIZATION

4.1 Background, Motivation, and Research Statement

Social media usage in elections. Social media sites such as Twitter and Facebook have played a significant role in elections across the world (e.g., U.S. [68, 122], Australia [48], Britain [28], India [18]). Social media sites are used by people to discuss election campaigns as well as by politicians to directly reach out to the electorate. A recent survey by the Pew Research Center shows that two-thirds of U.S. social media users discuss political issues on these sites [70]. Moreover, about a quarter of U.S. adults directly relied on social media sites to keep up with the 2016 presidential election campaigns of Donald Trump and Hillary Clinton [172].

Social media manipulation and countermeasures during elections. Unfortunately, the open nature of social media platforms also makes them susceptible to manipulation. Prior research has extensively reported on the widespread nature of spam and other types of abuses in popular online social networks [86, 182, 187, 204]. It is perhaps unsurprising that social media sites have been targeted during elections beyond “vanilla” spam. There were reports of social media misinformation campaigns targeting various political candidates going as far back as the 2010 U.S. midterm election [158]. There have also been numerous reports of widespread misinformation campaigns during the 2016 U.S. presidential election [130, 42], including reports by the Office of the Director of National Intelligence [105] and the Senate Intelligence
Committee [16] concluding Russian state-sponsored misinformation campaigns. Popular social media sites such as Twitter [191] and Facebook [178] publicly acknowledged the exploitation of their platforms during the 2016 U.S. presidential election by state-sponsored attackers. They have since announced a number of “cleanups” [163, 191, 179, 73], which have resulted in suspension of millions of accounts [188, 107].

**Analysis of social media manipulation during elections.** There is significant interest in understanding the countermeasures deployed by social media sites to counter spam and misinformation campaigns specifically targeting elections. One line of research has specifically focused on characterizing state-sponsored misinformation campaigns during the 2016 U.S. presidential election. For example, researchers showed that RU-IRA (Russian Internet Research Agency) Twitter accounts systematically manipulated political discourse [130], were able to reach a substantial number of Twitter users [206], and produced content that had a mostly conservative agenda [31]. Another line of research has more broadly focused on measuring the impact of countermeasures deployed by Twitter and Facebook. For example, researchers found that popular Twitter accounts in aggregate lost over half a billion followers due to a recent Twitter “purge,” in which former president Obama lost about 2 million followers and president Trump lost about a half million followers [50].

**Gaps in prior work.** There are two main gaps in prior research that we hope to address in this work. First, prior research studying social media manipulation during the 2016 U.S. presidential election is mostly limited to analyzing a few thousand Russian or Iranian state-sponsored accounts publicly disclosed by social media oper-
ators [206, 31, 29, 207]. We argue that social media manipulation during the 2016 U.S. presidential election was likely more diverse and at a much bigger scale. Second, while prior research has studied the impact of new countermeasures deployed by social media platforms, there is dearth of research on understanding the inner-working of these new countermeasures. We argue that better understanding the newly deployed countermeasures can shed light into their potential blind spots and lead to development of more effective solutions.

Proposed research. To address the first limitation, we propose to retrospectively analyze the activities of suspended Twitter accounts that engaged in political discourse during the 2016 U.S. presidential election. We believe that a postmortem analysis of targeted cleanups (identified using suspended accounts) provides a valuable ground truth that can be leveraged to study a wide variety of social media manipulation during the 2016 U.S. presidential election at scale. To address the second limitation, we propose to utilize two sets of suspended accounts (identified about a year apart) before and after Twitter announced new countermeasures against spam [163, 94] to analyze Twitter’s new countermeasures. We believe that an examination of how the newly suspended accounts connect to the older suspended accounts can shed light on the inner-working of Twitter’s countermeasures.

In this work, we identify nearly a million suspended Twitter accounts that engaged with the presidential election campaigns of Donald Trump and Hillary Clinton over the duration of four months leading up to the 2016 U.S. presidential election. Then, to systematically analyze the coordinated behavior of a million suspended ac-
counts, we group them into different communities based on their retweet and mention activities. Next, in order to examine the characteristics of suspended communities, we transfer measures for individual-level features into community-level features along five dimensions: dominant poster (responsible for posting most tweets), dominant retweet content producer (responsible for producing content for other users to retweet), burstiness (temporal bumps of tweets), dominant domain (which was tweeted most frequently), and dominant hashtag (which was tweeted most frequently).

**Key findings.** We summarize our key observations as follows:

- By systematically comparing characteristics of suspended account communities versus regular (not suspended) communities, we find differences across all five defined dimensions, but most significantly for the dominant poster and dominant hashtag dimensions. Through community-level analysis along different dimensions, we hope to provide insights to social media platforms as well as the broader research community for developing effective methods of detecting malicious accounts based on their group-level activities.

- By qualitatively analyzing suspended communities along different dimensions, we find that each of the five proposed dimensions is useful in identifying heterogeneous themes across suspended communities. We find communities that contain a large number accounts engaged in state-sponsored propaganda, those engaged in selling of political merchandise, as well as pornographic materials. This demonstrates that our analysis of targeted cleanups by Twitter through suspended accounts provided a general ground truth to study political spam
and misinformation campaigns during the 2016 U.S. presidential election.

- By analyzing suspended accounts before and after Twitter’s deployment of new countermeasures, we find that more than 90% of the newly suspended accounts have direct connections to the communities of suspended accounts that were detected earlier. Moreover, a large fraction of the newly suspended account retweet or mention the top retweet content producers of old suspended communities. These findings suggest that Twitter’s new countermeasures are targeting accounts that are linked to previously suspended accounts. It also suggests that our community based methodology may potentially be used to identify more users that are possibly eligible for suspension.

4.2 Related Work

There are a variety of works related to the 2016 U.S. presidential election on multiple topics, such as policy discussion [164, 26, 122], political disinformation [130, 111, 21], Russian trolls [206, 31, 29] and bot activities [42, 31, 159]. Overall, they have shown that the 2016 U.S. presidential election has been manipulated by state-sponsored propaganda as well as distorted by rumors and social bots.

In our work, we focus on the topic of suspended accounts on Twitter. Although this topic is limited to a very small dataset (e.g. Russian trolls) in election-related analysis, it has received large attention from research communities in terms of spam and fake account detection [182, 187, 129, 123, 195]. One of the most similar works to ours is by Thomas et. al. [187], who analyzed Twitter suspended accounts as
spammers. The authors examined these accounts as a whole on a number of properties such as active duration, tweet rates, relationships, domain usage and compared to non-spam accounts in some cases. However, unlike their work, since our dataset is within four months before the 2016 U.S. presidential election day and directly related to two main candidates Donald Trump and Hillary Clinton, we suspect there can be different themes causing user suspension other than spam. In fact, the suspended percentage in our data collection is 9.5%, nearly triple theirs (3.3%). Moreover, we group suspended users into communities and analyze their activity at the community level rather than as a whole.

More recently, Volkova et al. [195] built classifiers to distinguish deleted and suspended accounts from active ones in three different languages. The authors found that neural network models trained on text and network features produce the highest performance for most of tasks, despite the fact that the network features they used are very simple (e.g. number of mentions). Thus, although our goal is not to predict suspended accounts, by analyzing suspended communities we can provide more insights for researchers to develop a more effective method of predicting suspended accounts based on their community activities.

4.3 Methodology

4.3.1 Data Collection

During the 2016 U.S. presidential election, as a part of our previous studies on political discourse [122, 121], we collected tweets around two major presidential
candidates Clinton and Trump. Specifically, we used Twitter’s streaming API with filter keywords as full names of candidates (e.g. “hillary clinton” for Clinton) to collect tweets for each candidate. Since this API caps the tweets at 1% of all public tweets and there are more than 500 million tweets posted per day on Twitter [7], we are set to capture up to five million tweets per day for each candidate. However, since in our data collection the highest daily tweet count (for Trump) was less than two million, we were still able to capture the vast majority of tweets for both candidates.

In this work, we are interested in analyzing Twitter suspended accounts during the 2016 U.S. presidential election. Thus, in February 2018, using Tweepy - a Python library to access the Twitter API - we were able to examine how many accounts were suspended in our previous tweet collection around Clinton and Trump. Specifically, Twitter API returns the request of the user status with the response code of 63 if the user was suspended. Table 4.1 shows the statistics of our data during nearly four months, from June 01 to November 08, 2016 (except for most of August and some individual days due to the crash in our collecting process). In total, there were 912,979 accounts suspended out of 9,572,020 which made the suspended percentage to be 9.5% on average.

### 4.3.2 Suspended Communities

Motivated by the lack of using network activities in understanding user suspension, we decide to analyze suspended accounts based on their retweeting and mentioning activities. We decide to use retweet and mention activities because of
Table 4.1: Statistics of suspended users as of February 2018 in our tweet collection around Clinton and Trump during nearly four months, from June 01 to August 01 & from September 09 to November 08, 2016 (except September 18-20, October 31, and November 1).

<table>
<thead>
<tr>
<th></th>
<th>unique</th>
<th>suspended</th>
</tr>
</thead>
<tbody>
<tr>
<td>Users</td>
<td>9,572,020</td>
<td>912,979</td>
</tr>
<tr>
<td>Tweets</td>
<td>90,652,722</td>
<td>9,218,751</td>
</tr>
</tbody>
</table>

their representativeness for tweeting interactions. Specifically, retweet and mention activities among accounts demonstrates explicit engagement, especially since users can retweet/mention another user who they do not follow. In the case of follower network the level of engagement may be regarded as weaker and more passive. Besides, we are also unable to have their follower network since these are suspended users. To this end, we first build the retweet and mention network starting from these suspended users. Table 4.2 shows the statistics of this network’s directed graph, with the frequency weighted edge coming from the suspended user who retweeted/mentioned to the user who was retweeted/mentioned. The total of nodes or users is nearly one million while the total of edges is more than 14 million.

Next, suspecting that suspended users tend to act together (a.k.a synchronized or coordinated behaviors) or more strongly connected to each other, which was shown for spammers and malicious accounts in previous work [52, 110, 204], we want to analyze suspended accounts at the group level for better interpretation. To this end, we apply Louvain community detection algorithm [45] on the retweet and men-
Table 4.2: Statistics of retweet and mention network from suspended and regular users.

<table>
<thead>
<tr>
<th></th>
<th>Suspended</th>
<th>Regular</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retweet + Mention Network</td>
<td>995,983</td>
<td>6,430,293</td>
</tr>
<tr>
<td>Nodes</td>
<td>14,321,863</td>
<td>38,728,400</td>
</tr>
<tr>
<td>Biggest Community’s Size</td>
<td>13,809</td>
<td>13,537</td>
</tr>
<tr>
<td>Number of Communities</td>
<td>72,864</td>
<td>690,337</td>
</tr>
<tr>
<td>Number of Communities (Size ≥ 10)</td>
<td>9,554</td>
<td>67,734</td>
</tr>
</tbody>
</table>

In general, this algorithm tries to build communities so that connections are much more inside but less outside a community. Specifically, it optimizes modularity value which measures the density of edges inside communities to edges outside communities. Small communities are found at first by optimizing modularity value locally, then each small community is considered as one node and the process is repeated. In this study, we aim to have a manageable size of a community for better analysis so we constrain size of the biggest community at most 15K. To this end, the algorithm results in 72,864 communities with the size of the biggest one as 13,809 (suspended users). Moreover, we do not focus on very small communities (size < 10) so we finally have 9,554 suspended communities to analyze.

Note that the retweet and mention network starting from our collection of suspended users also contains other users who were retweeted/mentioned and could be suspended or regular. However, after communities were finally built based on this retweet and mention network, we study communities from members who are suspended users in our collection. Thus, we call communities as suspended communities.
4.3.3 Community-Level Features

Previous works \[98, 42\] have reported the list of signs that could suggest a Twitter account is bot or automated such as very high posting rate, only a few sources or accounts are retweeted, multiple tweets of the same link. While these signs could be missed for an individual account, synchronized behaviors of users in a community could reveal them. Thus, inspired by these previous observations as well as Twitter rules on account suspension \[190\], we define analogous community-level characteristics to examine evidence of synchronized behaviors at the community level. In essence, we are looking for community-level signals for account suspension. Follows are our five defined dimensions representing distinct community-level features.

1. Dominant Poster: The highest percentage of tweets in a community that are posted by a single user. The higher this percentage, the more dominant a single user.

2. Dominant Retweet Content Producer: The highest percentage of tweets in a community that are retweeted from a single user (tweets can be posted by either one or multiple users). The higher this percentage, the more dominant a single user.

3. Burstiness: The highest times tweets posted in one hour are higher than the average tweets per hour in a community. The higher this percentage, the more bursty tweets were posted.

4. Dominant Domain: The highest percentage of tweets from a community that contain URLs from a single domain. The higher this percentage, the more
dominant a single domain.

5. Dominant Hashtag: The highest percentage of tweets from a community that contain a single hashtag. The higher this percentage, the more dominant a single hashtag.

While the first and second dimensions represent network features of these suspended communities, the third dimension represents temporal features and the two last represent content features. In fact, we also analyze suspended communities on other features such as tweet text and language. However, the findings either mostly overlap with these five dimensions (e.g. tweet text) or are almost meaningless (e.g. language). Thus, in this work we only present our analysis and findings on these five dimensions.

Note that since these proposed dimensions serve as collective means to measure the characteristics of a community, the higher values in these five dimensions the more distinct and possibly suspicious the community activities. Thus, we are interested in communities which have high values in these dimensions. Specifically, we focus on communities having at least 50% of tweets posted by one user, or having at least 50% of tweets retweeted from one user, or having tweets posted in one hour which is at least 200 times the average tweets per hour, or having at least 50% of tweets containing URLs from one domain, or having at least 50% of tweets containing the same hashtag. In other words, we specify these high thresholds for our further analysis. While as an arbitrary choice, the thresholds as 50% and 200 times serve well as moving these values higher will not change the meaning of our analysis and findings. Further details
can be seen in next section Results.

4.4 Results

4.4.1 Suspended and Regular Communities

To serve as a baseline for observing special characteristics of suspended accounts’ activities, we compare the activities of suspended accounts to those of regular accounts. Here regular accounts are users who Twitter API returns the request of user status as “regular”. In total, our data collection have 7,740,693 regular users who posted nearly 48 million tweets. We apply our previous approach of building suspended communities for these regular users. Specifically, we first build the retweet and mention network starting from these regular users. Table 4.2 shows the statistics of this network’s directed graph, with the frequency weighted edge coming from the regular user who retweeted/mentioned to the user who was retweeted/mentioned. The total of nodes or users is nearly 6.5 million while the total of edges is nearly 39 million. We then use the Louvain community detection algorithm with the same constraint for the size of the biggest community (at most 15K). The algorithm results in 690,337 communities with the size of the biggest one as 13,537 (regular users). Finally, we have 67,734 regular communities with size ≥ 10.

We then analyze both suspended and regular communities along our proposed community-level features. Figure 4.1 plots CDFs of suspended and regular communities for these five dimensions. We further use the Kolmogorov-Smirnov (KS) test to investigate whether the distributions of suspended and regular communities in each
Figure 4.1: CDFs of suspended and regular communities in five dimensions. The vertical line places at the threshold as 50% and 200 times. From this vertical line to the right in each plot are communities having high value for that respective dimension. Specifically, the percentage of communities having high values = 1 - Yvalue of the intersection between the CDF and the vertical line.
dimension are significantly different. We are able to reject the null hypothesis that both distributions are the same at the 0.02 significance level for the dominant hashtag dimension and at the 0.0025 significance level for the dominant poster dimension. Thus, among five investigated dimensions we find that suspended and regular communities exhibit significant difference in terms of hashtag dominance in posting content and very significant difference in terms of user dominance in posting behavior.

Moreover, since we are interested in communities which have high values in these five dimensions, we focus on communities contributing to the CDFs from the vertical line to the right in each plot of Figure 4.1. Due to the big differences between the total number of suspended communities (9,554) and regular communities (67,734), we only compare the level of suspended and regular communities having high values in these five dimensions in terms of percentage for a fair comparison. Specifically:

- The feature displaying the largest difference in percentages between suspended and regular communities is the dominant poster dimension. Figure 4.1(a) shows that 38.3% suspended communities have at least 50% of tweets posted by a single user while this type of user dominance occurs only in 3% regular communities. Thus, the ratio of percentages for suspended to regular communities is 12.8.

- The three features with the next highest ratio (i.e. 3 to 4) for suspended versus regular communities are dominant retweet content producer, dominant domain, and dominant hashtag dimensions. For example, Figure 4.1(c) shows that 8.6% suspended communities have at least 50% of tweets containing URLs from one domain while this type of domain dominance occurs only in 2.3% regular com-
• The suspended and regular communities are about the same on burstiness dimension. Figure 4.1(b) shows that 17.2% suspended communities have tweets posted in one hour which is at least 200 times the average tweets per hour while this type of bursty posting occurs in 13.7% regular communities. Thus, the ratio of percentages for suspended to regular communities is only 1.3.

Note that one community can have high values in multiple dimensions. In total, 57.6% suspended communities have high value in one or more of five investigated dimensions while that percentage is lower as 17.6% for regular communities. Thus, the ratio of suspended communities to regular communities which have high value in one or more of five investigated dimensions is 3.3. This suggests that high values in multiple dimensions is a signal for the closer scrutiny on a community. Besides, in terms of users, these 57.6% suspended communities contain 668,573 suspended users which account for nearly 75% of all suspended users. Although these five dimensions cannot be used to explain completely why these users were suspended, it is noticeable that they pertain to 75% of suspended users through their communities.

**Takeaway:** Overall, our results show that suspended and regular communities exhibit differences on all five investigated dimensions. Especially these differences are significant in dominant hashtag dimension and very significant in dominant poster dimension. Moreover, compared to regular communities there are more suspended communities having high values on all five dimensions. Especially this can be up to one order magnitude (12.8 times) for the dominant poster dimension.
Table 4.3: Statistics of six representative communities for qualitative analysis.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Community Name</th>
<th>Users</th>
<th>Tweets</th>
<th>Retweets</th>
<th>%Retweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dominant Poster</td>
<td>Trump-IRA</td>
<td>28</td>
<td>201</td>
<td>172</td>
<td>85.6</td>
</tr>
<tr>
<td>Dominant Retweet Content Producer</td>
<td>GayRights</td>
<td>1,578</td>
<td>2,935</td>
<td>2,850</td>
<td>97.1</td>
</tr>
<tr>
<td>Burstiness</td>
<td>Vote!BlackLivesMatter</td>
<td>1,544</td>
<td>2,313</td>
<td>2,000</td>
<td>86.5</td>
</tr>
<tr>
<td>Dominant Domain</td>
<td>EbayAds</td>
<td>26</td>
<td>77,400</td>
<td>91</td>
<td>0.1</td>
</tr>
<tr>
<td>Dominant Hashtag</td>
<td>NoEthicsNoOffice</td>
<td>89</td>
<td>5,532</td>
<td>407</td>
<td>7.4</td>
</tr>
<tr>
<td>All except Dominant Poster</td>
<td>PoliticalPorn</td>
<td>1,062</td>
<td>2,105</td>
<td>2,105</td>
<td>100.0</td>
</tr>
</tbody>
</table>

4.4.2 Qualitative Analysis

To illustrate what is going on suspended communities’ activities, we next do qualitative analysis on several representative communities which have high values in these five dimensions. Specifically, we first analyze one representative community which has a significant high value in only each dimension. Particularly, for content dimensions such as domain and hashtag, the analyzed community is one of suspended communities who posted unique and suspicious dominant domains and hashtags compared to regular communities. We then analyze one representative community which has high values in multiple dimensions. Table 4.3 shows statistics of these six suspended communities.

4.4.2.1 Dominant Poster

- **Trump-IRA Community**: Most tweets (147 out of 201) are from an account @td21241 (profile name as terri in July, mainah4Trump in most of September, and DeplorableME4Trump for other times in our collection), who retweeted multiple different users (from outside the community) with contents supporting Trump and attacking Clinton. Aligned with this user’s profile description which
only contains #MAGA, top hashtags used in the community include #MAGA, #Trump2016, and #NeverHillary. More interestingly, checking with the list of 3.8K RU-IRA accounts on Twitter released by U.S. congress [171], we notice that 11 out of 28 members in this community are RU-IRA accounts. Overall, we see 295 RU-IRA accounts present in 165 different suspended communities (out of the total 9,554 suspended communities). This community appears to be the second top suspended community in containing the most RU-IRA accounts. Looking closely on the community’s network, it reveals that the dominant poster happened to play a role as filtering messages from external users for some RU-IRA accounts to retweet.

4.4.2.2 Dominant Retweet Content Producer

- **GayRights** Community: Most tweets (2,716 out of 2,935) were posted by an account @DCHomos and retweeted by this community’s members. Aligned with this account’s profile description which contains “all things LGBT+ in DC”, this community discussed mainly about Gay Rights policy, with most tweets including keywords as lgbt, gay, marriage equality. For example, more than 100 retweets with content “hillary clinton mention lgbt rights in her opening statement! “do not reverse marriage equality”” appeared shortly in two hours of October 20. Moreover, the community’s messages show that they support Clinton and oppose Trump. For example, nearly 100 retweets with content “don’t ever forget, republican nominee Donald Trump mocking a disabled re-
porter. #demconvention #demsinphilly” appeared shortly for one hour in July 26 - time of the Democratic National Convention.

### 4.4.2.3 Burstiness

![Figure 4.2: The illustrative suspended communities in terms of burstiness: Vote!Black LivesMatter Community.](image)

- **Vote!BlackLivesMatter Community**: Most tweets (2K out of 2,313) are retweets from many different users, of which 642 are retweets from a user @BernieSanders. The content of this community’s tweets shows that they first supported Sanders but then gave their support to Clinton after Sanders lost to Clinton for the Democratic nomination. Particularly, there is the highest spike in the late morning of September 22 including 862 retweets from a user @NadelParis with the content as “!#vote! @hillaryclinton: #humanrights w. @berniesanders on #freecollege + #retrainpolice!! #blacklivesmatter!” (Figure 4.2).
Figure 4.3: Two decals from Ebay URLs posted by EbayAds community.

4.4.2.4 Dominant Domain

- **EbayAds Community**: Almost URL-posted tweets in this community (75K out of 77K) used Ebay platform (ebay.com) to promote selling stickers such as “CHICKEN TRUMP” or “HILLARY CLINTON LOVE TRUMP HATE” (Figure 4.3). This is community supporting the Democratic party, along with Hillary Clinton and Bernie Sanders. Their top used hashtags contain multiple Democrat-supported hashtags such as #dems, #blm (Black Lives Matter), #toporog (Top Progressives), #ctl (Connect The Left) as well as multiple Trump-opposed hashtags such as #trumpdrseuss, #chickentrump, #dumptrump. It is also noticeable that this community posted mostly original tweets, with the retweet rate of 0.1%.
Figure 4.4: Two pictures opposing Clinton from tumblr.com posted by NoEthicsNoOffice community.

### 4.4.2.5 Dominant Hashtag

- **NoEthicsNoOffice** Community: Their dominant hashtags include #hillaryclinton, #hillary, #clinton, #election, #politics. The reason is that near the end of October, from different users in this community about 3.2K tweets (out of 5.5K) contained these dominant hashtags. Moreover, these hashtags were mostly posted together with URLs from domain tumblr.com which conveyed messages to oppose Hillary Clinton such as “No Ethics No Office” or “She sold out Bernie, next she’ll sell out America” (Figure 4.4).
4.4.2.6 Combination

There is no community having high values on all five dimensions. Thus, our representative community is the one having high values in four dimensions: dominant retweet contain producer, burstiness, dominant domain, and dominant hashtag.

- **PoliticalPorn** Community: Their dominant hashtags include #porn, #xxx, #makeamericagreatagain, #imwither (note that neither support Trump nor Clinton). Their dominant domain is pornsiteflip.com. The reason is that among 2.1K retweets, 1,061 retweets are retweets from a user @forplayxxx with content “donald trump, hillary clinton #porn...#xxx #makeamericagreatagain #imwither http://pornsiteflip.com/2016/08/04/donald-trump-and-hillary-clinton-fucking-bernie-sanders-and-megan-parody/”. Moreover, suspiciously these tweets were shortly tweeted only from 6 to 9am in November 5. It is also noticeable that this is a complete retweet community (retweet 100%).

**Takeaway:** Overall, our qualitative analysis demonstrates that each of five proposed dimensions is very useful in identifying a single theme inside the community which reflects the community’s feature or characteristic. And despite being limited to very small set of suspended communities due to the difficulty nature of qualitative analysis, the results show that suspended communities are heterogeneous in terms of different characteristics. Specifically, some communities post content that is pro-Trump, anti-Clinton (e.g. Trump-IRA) while some are pro-Clinton, anti-Trump (e.g. GayRights) and some are even neither (e.g. Political Porn). The posting behaviors of some communities are bursty with noticeably bumps at certain times (e.g.
Vote!BlackLivesMatter) while others are not. They include URLs pointing to various domains such as ebay.com to sell bumper stickers (e.g. EbayAds) or tumblr.com to spread memes and quotes (e.g. NoEthicsNoOffice). They use hashtags for different purposes such as promoting votes (e.g. Vote!BlackLivesMatter) or spreading porn (e.g. PoliticalPorn). They engage in the discussion of a diverse set of issues ranging from foreign policy to gay rights (e.g. GayRights). Especially, some of communities appear to have connections to Russian trolls (e.g. Trump-IRA).

4.4.3 Twitter’s Suspension Algorithm

Accounts that violate Twitter policies are constantly removed from the platform. We waited for more than one year after the 2016 U.S. presidential election (as of February 2018) to get the suspended account collection as described and analyzed in previous sections. However, since May 2018 Twitter announced a number of cleanups, so-called Twitter “purge” [50], which led to suspension of many more Twitter accounts. This happened because Twitter claimed to develop “new measures to fight abuse and trolls, new policies on hateful conduct and violent extremism, and ... new technology and staff to fight spam and abuse” during May-August 2018 [163, 94]. Thus, in January 2019 we rechecked to see how many more accounts in our election-related tweet collection were suspended. Although these newly suspended accounts may be suspended because of either actions during 2016 or actions taken since the election, we assume these newly suspended accounts might be the result of the new development in Twitter’s suspension algorithm. We want to know whether
these newly suspended users had connections to the original suspended ones. We hope our findings will shed light on Twitter’s new development in their account suspension algorithm.

In fact, our 2019 recheck has found an additional newly suspended accounts in our dataset. Specifically, there are 192,415 suspended accounts that had not been suspended earlier. This increases the percentage of suspended accounts in our dataset from 9.5% to 11.6%. We then determine if these newly suspended accounts have connections to the 9,554 communities of the original suspended accounts. To this end, we calculate the maximum retweet/mention connection of each newly suspended account to the original suspended communities. The result shows that more than 90% of the newly suspended accounts had at least one retweet/mention connection to an original suspended community.

Next, since most of new suspended accounts have direct connections to old suspended communities, we want to examine the strength of these connections. Specifically, we ask whether these new suspended accounts connected to major actors in the original suspended communities or is it that there is no pattern in the connections. To this end, since the direct connections from new suspended users to old suspended communities are retweets or mentions, we examine what percentage of these newly suspended users connected to the top-k retweet content producers of the original communities. Note that when $k = 1$, the top-k retweet content producers actually is the dominant retweet content producer for that community, one of our five proposed dimensions. Besides, although a newly suspended account may be connected to more
than one of the original communities, we consider only the strongest connection (i.e. smallest $k$) for each newly suspended account. The result shows that 72% of the newly suspended users retweeted or mentioned the dominant retweet content producer of an old suspended community. And this percentage increases to 85% with $k = 5$, which means that 85% had a connection to one of the top five active users in the original community.

**Takeaway:** Overall, we find that Twitter’s new development on their suspension algorithm helps to detect more malicious users. However, our analysis on Twitter newly suspended users reveals that more than 90% of them have direct (retweet/mention) connections to communities of suspended users that Twitter detected before. More interestingly, a high percentage (>72%) of these new suspended users retweet or mention the top retweet content producers of the old suspended communities that they connect to.

### 4.5 Summary

In this work, we retrospectively analyze the activities of suspended Twitter accounts that engaged in political discourse during the 2016 U.S. presidential election. By developing community-based method and measures which are new to the field of analyzing suspended accounts, we are able to characterize about a million suspended accounts. In short, we find that (1) suspended communities are different from regular communities in their posting behavior; (2) suspended communities exhibit heterogeneous characteristics; and (3) newly suspended accounts connect tightly to the old
suspended communities.

To the best of our knowledge, we are the first to conduct in-depth postmortem analysis of accounts suspended by Twitter to study their community-level activities as well as assess the effectiveness of Twitter’s new countermeasures. Our community-level analysis of suspended accounts highlights their coordinated behaviors which can be leveraged to develop more effective countermeasures. Although the results in this work are limited to the 2016 U.S. presidential election, our postmortem analysis approach is broadly applicable to any future election-related events for which social media sites are looking to develop more effective countermeasures.
CHAPTER 5
CONCLUSION

This thesis explores four different research streams to help understand the roles of humans, algorithms, and cyborgs in political polarization. Specifically, we conduct two research studies, “Scalable News Slant Measurement Using Twitter” and “Revisiting The American Voter on Twitter,” to help understand the roles of humans in political polarization. An additional study “Measuring Political Personalization on Google News Search” explores the roles of algorithms in political polarization and another study “A Postmortem of Suspended Twitter Accounts in the 2016 U.S. Presidential Election” investigates the roles of cyborgs in political polarization.

In terms of humans’ roles in political polarization, humans’ innate instincts directly lead to political polarization. Specifically, people tend to share and consume news conforming with their pre-existing ideology or political affiliation. People also tend to follow and like people having the same ideology or political affiliation. Thus, to help people know how imbalanced the political news that they might consume and share, in our first research we propose a simple method to effectively measure the political slant of individual news articles at scale. Moreover, in our second research we analyze political discourse on Twitter in the 2016 U.S. presidential election in order to inform people about how they have discussed the election according to their political affiliation.

In terms of algorithms’ roles in political polarization, in our third research we design a sock puppet auditing system that helps to discover political personalization
on Google News Search based on a user’s browsing history. More importantly, the most personalized results tend to reinforce the presumed partisanship. Thus, algorithms designed to create personalized experience can inadvertently intensify political polarization. Since there is a growing concern about the extent to which algorithmic personalization limits people’s exposure to diverse viewpoints, thereby creating “filter bubbles” or “echo chambers”, our research provides further empirical evidence for the underlying causes of “filter bubbles” or “echo chambers” in the aspect of political polarization.

In terms of cyborgs’ roles in political polarization, in our fourth research we retrospectively analyze the activities of suspended Twitter accounts that engaged in political discourse during the 2016 U.S. presidential election. By developing a community-based method and measures that are new to the field of analyzing suspended accounts, we are able to characterize about a million suspended accounts. Specifically, suspended communities exhibit heterogeneous characteristics such as pro-Trump vs. pro-Clinton or the involvement of Russian trolls which are well-known in previous research for campaign/election manipulation. Thus, one of the suspended accounts’ or cyborgs’ negative impacts could be to increase the political polarization.

For the future works, we would like to focus on studying how to mitigate the political polarization. Specifically, for the roles of humans in political polarization, our work of measuring the individual news’ political bias on online social media at scale can help to improve the transparency of news. Future research can study different disclosure methods/designs of news political slant (which can make people more self-
aware of their choices) as well as the impacts of these methods. Moreover, our work of analyzing political discourse can systematically identify the issues that people care about. This information can be used by politicians or policy makers to prioritize their agendas. This information can also help people to have a better election forecasting model that can replace our current laborious poll systems. For the roles of algorithms in political polarization, personalization can create distinct experiences for different users so future studies should be aware of the potential for algorithmic unfairness [177], especially when the unfairness can come from the bias in data [133] or the design of the algorithm itself [82]. Moreover, browsers like Safari or Firefox have built-in functionality to block trackers (thus block or obfuscate users’ privacy data) [115, 180] so researchers should study how these privacy-enhancing technologies can help to reduce the algorithmic personalization. For the roles of cyborgs in political polarization, our analysis on Twitter’s new countermeasures shows that the newly suspended accounts are actually linked to the old suspended communities that we identified before. It suggests our network-based and community-level activities can help to detect cyborgs in a timely and accurate manner. This will be very helpful since detecting cyborgs is much more complex than detecting purely bots or fake accounts.
Bibliography


[36] P. Barford, I. Canadi, D. Krushevskaja, Q. Ma, and S. Muthukrishnan. Ad-
scape: Harvesting and analyzing online display ads. In ACM International
Conference on World Wide Web (WWW), 2014.


[38] F. R. Baumgartner, F. R. tner and B. D. Jones. Agendas and Instability in

Differentials in the Housing Market. National Bureau of Economic Research
(NBER), 2012.


Sentiment and Predict Election Results. In Sentiment Analysis where AI meets
Psychology, 2011.

[42] A. Bessi and E. Ferrara. Social bots distort the 2016 U.S. Presidential election
online discussion. First Monday, 2016.


[113] A. Jungherr, P. Jrgens, and H. Schoen. Why the Pirate Party Won the German 
Election of 2009 or The Trouble With Predictions. Social Science Computer 

[114] F. Karimi. From Guinea to the U.S.: Timeline of first Ebola pa-
new-york-ebola-timeline/, October 2014.

[115] J. Kastrenakes. Firefox will begin blocking trackers by 
firefox-block-trackers-default-enhanced-tracking-protection, 
2019.

[116] C. Kliman-Silver, A. Hannak, D. Lazer, C. Wilson, and A. Mislove. Location, 
Location, Location: The Impact of Geolocation on Web Search Personalization. 
In ACM Internet Measurement Conference (IMC), 2015.


[124] M. Lewis. The Top 25 Conservatives on Twitter, http://aol.it/1OSbuBT.


