Does Congressional Polarization Decline as Election Approaches: Evidence from Twitter Data in USA

Shilpi Mukherjee[*,†]

October 2020

Abstract

In this paper I study the short run political polarization between Republican and Democrat politicians in the House of Representatives before and after the November 2018 midterm election, using Twitter data. I compute various metrics of ideological polarization at weekly intervals using methods such as hashtag analysis, topic modelling, Bayesian Ideal Point Estimation, mention and retweet network analysis. I empirically check for the trends in political polarization during the election cycle at the level of discourse. Different measures of polarization signal different trends in polarization. When polarization is measured by hashtag divergence or topical divergence, it seems to increase as the election approaches. But when polarization is measured by divergence in word distribution, sentiment-augmented topic divergence, or cited-media ideology divergence, it seems to decrease as the election approaches. This pattern is consistent with a divergence in preferred electoral agenda but convergence in agenda-item-specific positioning.

JEL Classification Codes: C55, D72

[*] shilpim@clemson.edu, John E. Walker Department of Economics, Clemson University.

[†] I am deeply grateful to my advisor Dr. Patrick Warren for his constant guidance and motivation throughout the length of this project. I am also thankful to Dr. Tom Lam, the members of the Clemson Public Economics Workshop as well as the IO Workshop for their helpful comments. Clemson Computing and Information Technology is acknowledged for generous computational resources, without which this project would not have been possible. I would also like to thank Mr.Ankur A Sharma for his generous help with technical programming issues.

1 Introduction

Political partisanship has been shown to affect not only preferences and voting behaviour but also other political and economic outcomes. Survey results from Greber and Huber(2009) show that beliefs about current and expected future economic performance are more positive when the respondent's partisanship matches that of the current President. A paper by McConnel at al. show that partisanship can also affect non - political markets such as the labor market and the goods market. Using a large experimental design they show that workers are willing to take a lower pay from co-partisans, suggesting a compensating differential. In the market for goods they show that buyers are nearly twice as likely to engage in a transaction when they and the seller share the same partisanship.

While political partisanship has been shown to affect key outcomes, the theory of political partisanship is mixed. Several papers have modelled a theory of short run political partisanship where parties are only motivated to win elections. In these models where parties are assumed to be only motivated to hold office, their proposed equilibrium policy stands converge before the elections. However, the policy convergence hypothesis has been subjected to a lot of subsequent scrutiny, which has been critical of the hypothesis and predicted policy divergence, at least for Congressional elections in the U.S. This area of research has focused on the short run modelling of behaviour of political parties.

On the other hand, research on long run political polarization in the U.S. has shown that polarization between parties has been on the rise since World War II. Although there have been arguments about whether ideological scores are just moving to pre-WWII era or if there is real ideological polarization, there is some consensus that polarization has been increasing for the last few decades. Figure 1 shows the distance between the mean ideological score for the Democrat and Republican politicians based on the first dimension of the DW-Nominate scores.[1] The first dimension computes the ideology of politicians along the liberal-conservative scale. The figure shows that polarization in the House and the Senate has been increasing since the 1980s, and is at an all time high in the past 125 years.

There also has been some research on how partisanship has evolved in the medium run. These papers mostly make use of text data. The paper by Gentzkow et al. (2019) shows that partisanship is much greater in the recent years compared to the past, and has increased substantially since 1990, following the Contract with America after remaining relatively flat before that. Several articles have also shown that partisan differences have seeped into media language and of late the two parties are using strategically different languages through the role of consultants, focus groups and polls. Lakoff(2003) suggests that represent a consequential change. (Bai(2005),Luntz(2006),Issenberg(2012))suggest that the partisan differences in language that we see today might represent aconsequential change (Lakoff(2003))

Therefore although there has been substantive empirical research showing rising rates of political polarization between Democrats and Republicans in the long run as well as in the medium run. However, due to paucity of data, there has not been substantial empirical research on short run political polarization, even though theoretical research abounds. A potential reason for the dearth of empirical analysis of short-run political polarization is that no good secondary source of data is available for measuring political ideology, except for the DW-Nominate scores. These scores are constructed using roll-call data. Roll-call data is however only available for each session of Congress, and therefore cannot be used to understand ideological movements during the election cycle. Survey data can help us look at polarization in a time series fashion at a high level of granularity, but the collection of survey data at such short intervals is very costly, especially for elites such as Representatives or Senators. To circumvent this problem, I look at short run polarization between Democrat and Republican politicians using their Twitter feeds. The Twitter data, really helps to zoom

[1] The trend looks very similar when we use the Poole and Rosenthal estimate for the first dimension

in on the behaviour of the politicians in the short run, at a high level of granularity.

To quantify the degree of polarization between Republican and Democrat politicians, I compute a number of metrics as measures of expressive ideological polarization using Twitter data. I use the term 'expressive ideology' because I measure their ideological estimates using rhetorical data collected through Twitter and not their actual behaviour. To do this, I collect tweets from incumbent politicians in the 116^{th} House of Representatives, one year before and after the midterm elections conducted on 6^{th} November, 2018 at a weekly interval. I exploit both the linguistic aspects of the tweets, as well as the network structure of the tweets to compute a number of metrics of polarization. For understanding the linguistic aspect, I perform Hashtag Analysis, topic modelling using Latent Dirichlet Allocation(LDA) and Bayesian Ideal Point Estimation to compute various estimates of polarization. I also study the retweet and mention networks of politicians and construct measures of polarization based on these networks.

A brief definition of each of these estimates is given here. Hashtag Similarity measures the number of common hashtags used by Democrats and Republicans conditional on the top hashtags used by them. Inverse of Sentiment Augmented Hashtag Distance ($Inverse\_Score\_std$) is the inverse of the euclidean distance between vectors of fractions of tweets corresponding to a particular hashtag with a negative sentiment by Republicans and by Democrats. The lower the $Inverse\_Score\_std$ the lesser is the polarization. Hellinger Distance and Kullback-Leibrar Divergence measure the distance between topic distributions(distributions of topics used by Democrats and Republicans in their tweets) whereas the Jaccard Distance measures the distance between word distributions. Next, I assign each tweet to the most dominant topic contained in the tweet. Dominant Topic Distance ($Score$) calculates the Euclidean distance between two vectors of fractions devoted to each topic by Democrats and Republicans. Sentiment Augmented Dominant Topic Distance (Sum_Frac_Dis) measures the euclidean distance between the vectors of fractions of tweets that Republicans and Democrats devote to positive, negative and neutral sentiment for each topic. The greater the distance, the higher is the polarization for these distance based measures. I also perform Bayesian Ideal point estimation based on the URL sharing behavior of the politicians to estimate the ideologies of the politicians. For the mention and retweet network analysis I calculate the share of how many times a Democrat mentions/retweets a Republican negatively and vice versa. One can find a detailed discussion on the construction of these measures in Section 4.

Polarization as measured by these metrics varies considerably in the election cycle. To get a sense of what happens close to the election, I zoom into 8 weeks before and after the election to look at what happens as we move into and away from the election. The choice of 8 weeks is made because there are no primaries in this time period and I hope to capture the effect of the upcoming midterm election through these estimates. I report the broad trends in these estimates here. Since, there are only 8 weeks of data, the estimates are not very precise, and most of them are not statistically significant. One way to interpret these results strictly from the perspective of statistical significance would be to say that there is no evidence of decrease or increase in polarization as we approach or move away from the election. This would suggest that the theoretical literature of policy convergence versus divergence do not play out in the Twitter feeds of politicians. However, since we only have one year of data, and if we are willing to take a Bayesian approach, looking at the coefficients does suggest some trends.

Polarization as measured by hashtag similarity and $Inverse\_Score\_std$ increases as we approach the election, falls once the election is over, and increases over the next 8 weeks after the election. The distance between topic distributions increases as we approach the election, falls once the election is over and keeps falling over the next 8 weeks. $Score$ increases as we approach the election, falls after the election, and keeps falling over the next 8 weeks. The Jaccard distance which is the euclidean distance between word distributions falls as we approach the election, falls after the election and increases we move away from the election. $Sum\_Frac\_Dis_{std}$ falls as we approach the election, increases once the election is over and falls again. Ideological difference as measured by the difference between Bayesian Ideal Point Estimates falls as we approach the election, falls once the election is over and then keeps increasing as we move away from the election.

Negative mentions of Republicans by Democrats increases as we move into the election, falls once the election is over and decreases in the 8 weeks following the election. Negative mentions of Democrats by Republicans increases as we move into the election, rises(significantly) once the election is over and decreases for as we move away from the election. Negative retweets of Republicans by Democrats decreases as we move into the election, falls once the election is over(significantly) and increases in the 8 weeks after the election. Negative retweets of Democrats by Republicans increases as we approach the election, decreases after the election and falls thereafter.

I find that there is an implicit and explicit element to polarization. Polarization, as measured by hashtag analysis, topic divergence and retweet and mention networks, increases as we approach the election suggesting that politicians get more polarized in their agenda setting behaviour, although negative retweets of Democrats by Republicans decrease in the last 8 weeks in the approach to the election. However, more implicit measures of polarization, as measured by divergence in words used, sentiment augmented content analysis as well as URL sharing behaviour, decreases as we approach the election. It makes sense to think that politicians are probably trying to appeal to their electoral bases through their agenda setting behaviour in terms of they hashtags that they use and the topics that they talk about. However there is decrease in polarization in terms of the conversations within a topic. This suggests that politicians are trying to appeal to the median voter through the content within a topic. We therefore find that politicians can and do use different aspects of a tweet to talk to different sub populations and while they might seem to diverge in the agendas that they talk about they might be converging within a particular agenda.

My contribution in this paper therefore runs along three dimensions. First, I help to test the policy convergence versus policy divergence hypothesis, using a very rich data set, that allows me to compute ideological scores, from various aspects of the Twitter data. This is the only paper to compute high frequency ideological estimates of polarization in a time series fashion. Second, I add to the methodology of measuring political polarization in the literature using the distance between two topic distributions and sentiment augmented content analysis. These methods can therefore, be replicated in a less developed country fairly easily which might not have official data like roll call votes found in the U.S.[2] Third, I also show that various dimensions of a tweet, such as hashtags, content, networks which can all be used to convey information need to be studied separately as they sometimes can provide competing signals.

The rest of the paper is organized as follows. Section 2 discusses the relevant literature. Section 3 discusses the data. Section 4 discusses the methodology of computation of the various measures. Section 5 talks about the empirical methodology and the results and Section 6 concludes.

2 Background and Literature Review

There have been broadly two ways of looking at the motivations of political parties to contest in an election. The first one assumes that political parties are only interested in holding office, or winning the election. The most seminal result from this assumption, as shown by Downs (1957) assuming rational voters and using Hotelling (1990) model of spatial competition is that parties converge on the policy preferred by the median voter. Comanor (1976) has shown that the median voter theorem holds true under reasonable degrees of

[2] Although the algorithms are scalable, they are costly in terms of computational and manual resources.

skewness of political preferences. Other works such as Ledyard (1984), Coughlin and Nitzan (1981), Hinich (1976) also predict policy convergence of other types, though not always to the preference of the median voter, in the case of office motivated political parties.

Other papers consider that politicians or parties which are sometimes considered synonymous are ideologically motivated. Wittman (1983), Calvert (1985) consider ideologically motivated politicians, but predict that equilibrium policies chosen by the parties, are very close to each other because they assume that parties have binding commitment to their policy platform. Alesina (1988) in his seminal paper shows that in the absence of a binding commitment device, which are mostly absent in elections, in an one shot electoral game, political parties have no incentive to stick to their announced policy level, after the election is over. Rational voters can correctly anticipate this behaviour of the politicians, and therefore expect to have policy divergence which then becomes the equilibrium strategy for the political parties. In an infinitely repeated game if the parties have a reasonably high discount factor, then the they might be able to sustain a convergent policy position. However, if one party believes that the leadership of the other party might change, leading to a low discount factor, then the co-operative equilibrium breaks. The parties revert back to the outcome of the one shot electoral game predicting policy divergence. Osborne and Slivinski (1996), Besley and Coate (1997) develop citizen candidate models, where citizens contest in election and implement their preferred policy on winning. These models also predict political divergence. Glaeser et al. (2005) in their paper talk about another reason for policy divergence. They argue that if promoting extreme party positions, helps in the sorting of voters such that it increases donations and voter turnout, thereby increasing the probability of winning, parties will prefer to diverge their policies.

Therefore, although the theoretical literature on short run polarization is quite mature, the question has not been explored a lot empirically. Lee et al. (2004) is one of the few papers that test for political divergence or convergence. They show using a Regression

Discontinuity Design (RDD) approach, focussing on elections won by a narrow margin, that electoral strength do not make politicians converge in their policy positions, and voters merely elect from the preferred policies of the politicians.

Juxtaposed with the theoretical and empirical modelling of short run political partisanship, there has been another literature which has documented the growing long run polarization between Republicans and Democrats. Aldrich (1995), McCarty et al. (1997), Jacobson (2000), Hetherington (2001), Collie and Mason (2000) have all shown the growing political polarization between Democrats and Republicans in the U.S in the government. Layman et al. (2010) show that Republicans and Democrats have become polarized both in the government and in the electorate through conflict extension along several dimensions.

Independent of the previous literature, there has been a lot of recent research on the link between the advent of social media and greater polarization, not only for politicians but also for general public. Two major explanations have been forwarded. The first uses the idea of the presence of fake news or misinformation and uses the bounded rationality argument by which people use the Bayesian updating rule and converge to wrong beliefs because they cannot distinguish the truth from fake information, Azzimonti and Fernandes (2018). Another explanation uses the idea of echo chambers and the concept of homophily, cognitive dissonance and confirmation bias where people believe and remember information which conforms to their pre-existing beliefs.

The first explanation has been mostly forwarded by economists and a lot of theoretical work has been done at this front, with little empirical work mostly pertaining to the spread of fake news during the 2016 Presidential Election in the U.S. Allcott and Gentzkow (2017) empirically looked at the effect of fake news on social media for the 2016 election. They show among other results that the average American adult saw on the order of one or perhaps several fake news stories in the months around the election, with just over half of those who recalled seeing them believing them and that people are much more likely to believe stories that favor their preferred candidate, especially if they have ideologically segregated social media networks.

The second explanation has been mostly propounded by scholars in departments of communication, computational social science or computer science. Several papers have tried to test the second hypothesis as to whether exposure to social media networks expose people to like-minded perspectives and make them more polarized or do they expose them to more diverse information and depolarize them. Lee et al. (2014) find that participating in social media networks have a positive effect on network heterogeneity in terms of mixing with diverse groups of people. However, they find no direct association between social media network heterogeneity and four different measures of opinion polarization - partisanship, ideology, same-sex marriage, and health care reform. Barberá et al. (2015) finds that social media networks allow one to form weak ties with people who are more politically heterogeneous than their immediate personal networks and thereby reduces political extremism. He uses data from Germany, Spain, and the United States to show that exposure to political diversity has a positive effect on political moderation. On the other hand, the "echo chambers" argument suggests that political polarization may be more salient in the highly fragmented and customized environment created by social media, Sunstein (2018). A number of studies have provided empirical support for this line of reasoning. For instance, Adamic and Glance (2005) find that political blogs tend to link to sites sharing similar ideological views. Gruzd and Roy (2014) demonstrate some support for political polarization on Twitter during the 2011 Canadian Federal Election. Hong (2013) finds evidence that politicians with extreme ideological positions tend to be more successful at campaign fundraising online. Hong and Kim (2016) find that ideologically extreme politicians have more followers than more balanced politicians which they believe supports the idea of a strong polarization in Twitter readership.

My aim in this paper is therefore to look at polarization between politicians in the context of social media over a shorter time period and at a high frequency. According to median voter theorem, politicians should converge in ideology as an election approaches. Since traditional data on politicians ideology is measured by DW-Nominate Score, which is based on Congressional voting records, it is not possible to see the movement of ideological scores over short intervals of time, due to data constraints. To solve this problem, and to understand elite polarization in the context of social media, I use Twitter data from the incumbent Congressional members of the 116th House of Congress and compute their ideological scores using their Twitter data from one year before and after the election and see if their ideological scores converge as the election on November 6, 2018 approaches.

Twitter data is a very good source of data to identify this kind of variation in ideology for a number of reasons. Twitter data helps us gain insight into the political discourse at very short intervals of time, almost week by week. A possible flip side is that Twitter data unlike roll call data, does not give us any access to behavioural insights, in terms of actual votes in favour or against a particular policy. However, the fact that politicians have increasingly taken to Twitter to state their policy positions, lends legitimacy to analyse their Twitter feeds. For example, President Donald Trump sent out 518 tweets (11 deleted) in the first 100 days of his Presidency according to politico.com, meaning that he alone sent out 5 tweets on average per day. And this is not a trait specific to the President. Most politicians have tried to make use of the "Obama Model" to reach the general public (Towner and Dulio (2012)).The way that social media has been used for political campaigning has been documented in a number of research studies (Adams and McCorkindale (2013), Conway et al. (2013), Golbeck et al. (2010), Graham et al. (2013), Grant et al. (2010), Johnson and Perlmutter (2010), Xiong et al. (2019)). These studies testify to the fact that ever since, Twitter came into effect in 2006, it has been increasingly adopted by politicians all over the world to influence their campaign strategy. Social media and traditional media are also found to have a symbiotic relationship in terms of agenda-setting during election campaigns as found in a paper by (Conway et al. (2015)). They investigated the relationship between the twitter feeds of political candidates and parties and the news media output .

Although social media has been considered as an important element of political campaigning, social networking sites (SNSs) such as Twitter are considered to be the favored forms of social media for campaign purposes. SNSs are unique because they allow connections to be displayed openly. However, unlike some SNSs which have privacy controls, Twitter users have mainly public profiles, which do not require bidirectional confirmation of networks (Boyd and Ellison (2007), Vergeer (2015)). This allows it to be used as a broadcast medium, an attribute that is used in political campaigning extensively. This makes Twitter a very natural choice for the question that I attempt to answer.

3 Data

3.1 Twitter Data

The United States of America had a midterm election on November 6, 2018 where 435 seats from the U.S House of Representatives were contested. I collected the official twitter handle of all the incumbents from the US House of Representatives, after the results of the elections were announced. This was done by searching for verified handles for each of the incumbent members. In the presence of more than one verified handle[3] I considered the one which had the link to the Representative's official page in the official website of the U.S House of Representatives. I collected tweets at weekly intervals for each politician in my dataset tweeted by them from 7th November, 2017 to 5th November, 2019, a total of 104 weeks or two years of data using an application called Social Studio[4]. For some of the weeks where

[3]This was true for popular politicians, whose campaign accounts or personal accounts were also verified.

[4]I am thankful to the Social Media Listening Center, Clemson University for providing me access to the Social Studio app.

data was not available through Social Studio, I used the Twitter API.

In my data-set, I have access to the name of Congress person who tweeted, the content of the tweet, the publish date and time of the tweet and whether the tweet was a normal tweet, retweet or quote tweet.

There are 177 Democrats, and 165 Republicans in my data set[5]. Only one Democrat incumbent winner and two Republican incumbent winners did not have an official handle. The number of politicians in my data-set account for 78.62 % of the total number of politicians in the U.S House of Representatives.[6] Figure 2 shows the number of active incumbent politicians over time, with the number being calculated every week. A politician is defined as an active politician even if s/he makes atleast one tweet in the span of the week, under consideration, s/he is counted into the number of active politicians. So, if politician X tweets one message in the 22^{nd} week, but not in the 23^{rd} week he/she would be counted as an active politician in the 22^{nd} week but the not the 23^{rd} week. We have an average number of 146 active Republicans and 168 active Democrats in our data set every week. The higher number of average active Democrats compared to Republicans is attributed to two reasons. First, there are higher number of Democrats in my data-set, but second whereas approximately 94.91 % of the total Democrats are active authors, 88.48 % of Republicans are active authors.

Figure 3 shows the total tweets by Democrats and Republicans over time, computed on a weekly basis. This figure shows that based on absolute numbers Democrats tweet more than Republicans. Republicans post around 1852 tweets every week, whereas Democrats post 3350 tweets every week. To see if this variation comes only from the higher number of Democrats in my sample or if Democrats demonstrate a higher tweeting propensity than

[5]The twitter handles were collected in October 2019, and the Congresspeople who had verified twitter handles then, are included in the data-set.

[6]I only consider incumbents who contested in 2018 and won, so that I can get their tweets from the official handles after the election too. However, there were only three incumbents who contested in the 2018 election and did not win, and hence I argue that since the number is so small, it should not introduce a huge selection bias in my data

Republicans, we refer to Figure 4. Figure 4 shows the average number of tweets by each active incumbent member of the House over time, computed on a weekly basis. This figure shows that even after controlling for the number of active authors, Democrats tweet more on average than Republicans. Whereas a Republican posts 13 tweets on average per week, a Democrat posts 20 tweets on average per week.

Figures 2, 3, 4 also show some broad trends. For example, for all the graphs we see a dip in the numbers during the last week of December and beginning of New Years, when politicians are probably spending time with their families. Another unique thing about the figures is that trend for Democrats and Republicans appear to follow the same pattern in the crests and troughs. Whether this is due to some underlying causal factor or if there is a feedback mechanism between Democrats and Republicans is unclear.

3.2 Competitiveness of Race Data

Along with the twitter data, I also collect the competitiveness of race information for all the congressional districts of the United States for the 104 weeks that are there in my data-set. I collect this data from the Cook Political Report. I scrape the website for the data. The data provides information on whether a district is Solid Republican, Solid Democratic, Likely Republican, Likely Democratic, Lean Republican, Lean Democratic, Republican Toss-Up or Democratic Toss-Up. As the name suggests, solid refers to the safest districts, followed by Likely and Lean whereas Toss-up refers to the most competitive districts. There is 64 weeks of unique data. I match up competitiveness of race data [7] with the 104 weeks of data in my original sample, by assigning the value of race competitiveness of a particular district in a particular week to the closest race competitiveness data available at that time. This allows for a time sensitive data on race-competitiveness. Auter and Fine (2016) use this measure

[7] There are 64 weeks of race competitiveness data in my dataset. There is no competitiveness data between 5th November, 2018 and 12th April 2019. I assign the race competitiveness data of either 5th November,2018 or 12th April, 2019 to the weeks for which I don't have any race competitiveness data depending upon which date is closer to the particular week in question

of competitiveness of an election in their paper on negative campaigning in Facebook. They find that underdog candidates in less-competitive races indulge in negative campaigning in issue attacks, whereas candidates in competitive races are more into personal attack. There is therefore reason to believe that the competitiveness of race in a district will also influence the ideology of Democrats and Republicans in that district.

4 Computation of Ideological Estimates

To understand whether politicians behave differently atleast in the domain of rhetoric in the advent of an election, I analyse the content of the tweet from a linguistic perspective, along with the network structure of the tweeting behaviour of the politicians. With the increase in computational power, and the explosion in unstructured data, text-data analysis is constantly being used to answer a variety of questions, and text data is being considered to be an increasingly important data source, Gentzkow et al. (2019). Gentzkow and Shapiro (2010) use text data to develop an index of media slant to asses the similarity of the language used by a news outlet to that used by a Republican or Democrat. Social scientists have also analyzed text data for understanding polarization specifically. Gentzkow et al. (2019) study the trend in partisanship in the Congress, by analyzing speech from 1873 to 2009. Ash et al. (2017) similarly look at the polarization of U.S Circuit court judges using text data of the court opinions from 1890s to 2010s. Bara et al. (2007) analyses parliamentary debates in U.K to identify the dominant themes in debate and also the difference in discourse between leaders favoring different policy positions.

Some other studies have specifically used Twitter to understand polarization. Demszky et al. (2019) use the natural language processing framework to understand political polarization in Twitter, in the context of 21 mass shootings in the USA. Monti et al. (2013) model political disaffection using Italian Twitter data by employing sentiment analysis.

Although the use of text data is increasing over time, the field itself is in its early stages and is still evolving. There are a number of different methods available that scholars have used in the past. In this paper to analyze the linguistic aspect of the tweets, I perform Hashtag analysis, topic modelling, sentiment analysis as well as Bayesian Ideal Point estimation in order to create metrics of ideological polarization. In assessing the network structure of the tweeting behaviour of the politicians, I study the retweet network as well as the mention network complemented with sentiment analysis of the tweets.

4.1 Hashtag Analysis

To start off with the computation of metrics of polarization I look at the similarity in hashtags used. To compute hashtag similarity for each week. Hashtag similarity is defined as the number of common hashtags used by Democrats and Republicans conditional on the top hashtags used by them. To compute the hashtag similarity, I proceed in the following way. First, I extract the top 40 hashtags used by Republicans in a week. Let us denote this set of hashtags by R_{40}. Second I extract the top 40 hashtags used by Democrats in a week.[8] Let us denote this set of hashtags by D_{40}. I then compute the number of similar hashtags between the sets R_{40} and D_{40}. Let's denote it by $Hashtag_{40}$ In other words,

$$Hashtag_{40} = n(R_{40} \cap D_{40}), \tag{1}$$

where $n(.)$ denotes the cardinal number. I also compute $Hashtag_{10}$, $Hashtag_{20}$, $Hashtag_{50}$, $Hashtag_{100}$ for robustness checks. These estimates show how the usage of common hashtags used by Republicans and Democrats vary conditional on the top hashtags used by each group. Table 1 shows the top 20 hashtags one week before and after the election. The italicized and underlined words show the common hashtags used both by Republicans and Democrats. In the week before the election there is only one common hashtag whereas in the week after

[8] I convert all the hashtags to lower case because sometimes the same hashtags can be written in different cases,

the election, there are four common hashtags.

Hashtags in Twitter are used as an organic and community-driven method to add context to the data, Wang et al. (2011). They can therefore be thought of as broad topics that the Twitter users are talking about. Some studies have also shown that hashtags are sometimes used as framing devices, Moscato (2016) in guiding the political conversation. Bruns and Burgess (2015) talks about how hashtags have evolved from ad hoc devices in Twitter to tools that can be used to organize movements and also guide the discussion of topics in the platform. The role of hashtags in guiding social and political movements have been studied in a number of situations such as Canadian elections, Arab Springs movement, student protest movement against high fees in Africa, and the recent feminist movement which is best known by the hashtag it used, #MeToo (Langa et al. (2017), Small (2011), Moscato (2016),Huang (2011), Bruns et al. (2014). I therefore start by looking at the similarity of hashtags used by Republican and Democrat politicians over time as they give us the first piece of evidence of the way conversation changes between these two groups as the election approaches and if it changes once the election is over.

Figure 5 shows the trend of these estimates. According to this figure the overlap between hashtags keeps decreasing as the election approaches, and increases after the election. This suggests that Democrats and Republicans increasingly talk about different agendas in the approach to the election. The same pattern is true for all the four trends. I also do the hashtag analysis for tweets which have a negative sentiment and a positive sentiment and get the same trends. This means that irrespective of the sentiment of the tweet, the number of common hashtags used by Republicans and Democrats fall as they approach the election. The figures are presented in the Appendix.

4.1.1 Euclidean Distance between interacted with sentiment

Hashtags are generally very context specific and are used in conveying only a particular sentiment, as shown by the hashtags used in Table 1. While Democrats use #getcovered, #protectourcare and #goptaxscam, Republicans use #taxreform, #taxcutsandjobsact and #maga. However, there can be times when the a particular common hashtag is used positively by representatives from one party but negatively by representatives from the other party. In that case, increase in number of common hashtags might give us a false sense of decreasing rhetorical polarization between the two parties. To tackle this problem I perform sentiment analysis of the tweets containing hashtags. I use the Vader Sentiment analysis module in Python, which is a valence based sentiment analysis module developed especially for micro-blogging sites such as Twitter, Hutto and Gilbert (2014).[9] The package computes the positive, negative and neutral polarity for each tweet. It also gives a compound probability score. If the compound probability is less than -0.5, the tweet is considered to be negative, if it is greater than +0.5, the tweet is considered positive, and if the scores lies between -0.5 and + 0.5, it is considered to be neutral.

To combine the hashtag analysis with the sentiment analysis, I compute the distance between fraction of negative tweets for the Republicans and Democrats out of all the tweets that use a similar hashtag. For example a common hashtag in the top 40 hashtags used by Democrats and Republicans for the week of 1st August to 7th August, 2018 is #smallbusinessweek. I count the fraction of tweets using the hashtag #smallbusinessweek that have a negative sentiment for Democrats and Republicans. I repeat this for all the common hashtags in $Hashtag_{40}$ and calculate the Euclidean distance between those two vectors.

In other words, assume that there are s common topics in the top 40 hashtags used by Republicans and Democrats. Therefore, the length of $Hashtag_{40}$ which we have already

[9] They use a gold standard of lexical features as well as the polarity and intensity of words to compute the sentiment score. They also show that their approach is better than eleven of the common and most widely used Sentiment Analysis methods, and outperforms human accuracy.

defined is s. I now construct two vectors D_{40_s} and R_{40_s}. Let the first element of D_{40_s} be denoted by $D_{40_s}(1)$. Then,

$$D_{40_s}(1) = \left[\frac{\begin{array}{r} n(Tweets\ by\ Democrats\ which\ contain\ the\ first\ hashtag\ in\ top\ 40\ hashtags \\ and\ have\ a\ negative\ sentiment) \end{array}}{n(Tweets\ by\ Democrats\ which\ contain\ the\ first\ hashtag\ in\ top\ 40\ hashtags)} \right], \tag{2}$$

where n denotes the cardinal number. I similarly compute all the elements for D_{40_s}, R_{40_s} and find the euclidean distance between those two vectors. This is denoted by $Score_{40}$, where $Score_{40}$ is defined as follows:

$$Score = d(D_{40_s}, R_{40_s}) = \sqrt{(D_{40_s}(1) - R_{40_s}(1))^2 + (D_{40_s}(2) - R_{40_s}(2))^2 + ... + (D_{40_s}(s) - R_{40_s}(s))^2}. \tag{3}$$

I also compute a standardized score which is denoted by $Score_{40_{std}}$ which is computed as follows

$$Score_{40_{std}} = \frac{Score}{\sqrt{s}}. \tag{4}$$

I similarly also compute $Score_{10_{std}}$, $Score_{20_{std}}$, $Score_{50_{std}}$ and $Score_{100_{std}}$.The reason I focus only on negative sentiments is because hashtags being context specific and have a positive or negative undertone. Hence, it does not make much sense to distinguish between tweets with a positive sentiment and a neutral sentiment in case of hashtags. Therefore, treating positive and neutral tweets as single non-negative category essentially means that we can only calculate the Euclidean distance for the negative category, and a high Euclidean Distance automatically implies that distance in the non-negative category is also high and vice-versa.

Another point to note is that in doing the actual analysis I use the inverse of $Score_{10}$ and $Score_{10_{std}}$ which I refer to as Inv_Score_10 and $Inv\_Score\_10\_std$. This is done because if there are no common hashtags for any of the groups, then distance would be calculated as 0, but that does not make sense because a distance close to 0 implies no polarization whereas 0 common hashtags does not imply the same. To resolve this ambiguity we take the inverse of the score, such that a high score means less polarization and low score means high polarization. When there are no common hashtags the metric is set to a value of 0, as no common hashtags imply the greatest degree of polarization.

Fig 6 shows the trend of the inverse of the standardized scores. The non-standardized scores look the same and are included in the Appendix. This graph shows that as we approach the election, the inverse of the euclidean distance falls, or in other words the euclidean distance increases. This means that conditional on using the same hashtags in the top hashtags that Democrats and Republicans use, they use it with different sentiments as they approach the election implying that polarization in using hashtags increases as election gets closer. The trend after the election is different for top 10 and 20 hashtags versus the others. This might be because there are some hashtags in the top 10 or 20 hashtags which have very different characteristics compared to the hashtags in top 40, 50 or 100. But nevertheless, it is quite clear that the euclidean distance has a clear pattern before the election and it increases as we approach the election.

4.2 Topic Modelling

After looking at the hashtags which are broad level agenda setting items, I look at the content of the tweets directly. This allows us to understand the data even better. To do this, I use the method of topic modelling. Topic modeling is an unsupervised machine learning technique that's capable of scanning a set of documents, detecting word and phrase patterns within them, and automatically clustering word groups and similar expressions that best characterize a set of documents. These documents can be news articles, congressional speeches, parliamentary debates or in my case tweets. Topic Modelling helps us go one step further in looking at the divergence between Republicans and Democrats by looking at the content of the tweets. A number of very prominent and influential studies have been conducted in the field of information retrieval and automatic detection of topics in political speeches(Steyvers et al. (2004), Mamou et al. (2007), Quinn et al. (2010)). Boyd-Graber et al. (2017) shows the recent applications of topic modelling for information retrieval, linguistic understanding, statistical inference and other tasks. Topic modelling has also been used in the domain of social media data. Lucas et al. (2015) analyses how to perform topic modelling for tweets in different languages. For applying topic modelling to my case, I use the content of the tweets sent out by the politicians.

To perform topic modelling, I apply the model of Latent Dirichlet Allocation (LDA) to my corpora of tweets. To apply the LDA model the data needs to be pre-processed in order to be ready for the application of Latent Dirichlet Allocation (LDA). In keeping with the norms of Natural Language Processing (NLP), and the specificity of Twitter data, I remove the special characters such as '@', '#' that are specific to Twitter, and punctuation like the period, comma, semicolon and others. I also remove all the stopwords, words such as 'the', 'in', 'from', etcetera, which are very common in the English language, but devoid of any meaning[10]. I then lemmatize the data, which means that all words in our dataset, (referred to as tokens in the NLP nomenclature), are converted to their base form. For example, am/is/are are all converted to be[11]. I also create bigrams and trigrams to capture words that might be always associated together. For example, the term 'White House' is a good example of a bigram, that could be present in our data-set. We would lose the significance of the term "White House", if we used only unigram model, which would treat 'white' and 'house' as two separate words.

One of the parameters that need to be provided to the LDA model as a parameter is the number of topics in the corpora. There is no perfect objective measure to estimate the

[10] I use the NLTK corpus of stopwords for the English for this step of pre-processing

[11] I also stemmed the data, which is another form of converting the words to their base form, but in context of Twitter data, lemmatization seems to be better at tokenization, compared to stemming.

optimal number of topics for a given corpora, in the literature yet. One of the ways to estimate the right number of topics, is to look at the coherence score, for different number of topics, and select the number of topics, when the coherence score stops increasing[12].

To implement the LDA model for my data, I first run the model on each week of tweets for both Republicans and Democrats combined.[13] I run the LDA model for single digit topic numbers, because using more than those number means that my topics are going to have sparse number of tokens/words, and more than 9 topics seem too many for one week. I calculate the coherence score for each of these models, and choose the optimal number of topics based on the coherence score. I then re-run the model with the optimal number of topics for the Republican tweets and Democrat tweets separately. In the optimal LDA model I assign each tweet by the Republicans and Democrats to one of the topics. For example, if a particular week, had say 4 optimal topics based on the coherence score, all tweets in that week are assigned to one of the 4 topics. I use the topic with the highest percentage in a tweet, to assign that topic to that particular tweet. After doing this I analyze trends in the tweets by the following two methods.

The LDA model, first developed by Blei et al. (2003) has revolutionized the field of information retrieval. LDA is an unsupervised, probabilistic machine learning algorithm that automatically groups words together based on which words occur together more frequently in a corpus of data. Barberá et al. (2018) uses LDA model on tweets by the 113th Congress members, select media outlets, and other groups of people, such as general public, attentive,close party supporters, media and show using a Vector Auto Regression model that politicians are most attentive to issues of close party supporters for setting the agenda. Nardi Jr (2012) uses the LDA model to analyze the text of Supreme Court Decisions in the

[12]Since, the LDA is a probabilistic model each run of the model, generates new values of the coherence score. I set a random seed equal to zero, so that the model can be replicated.

[13]I use the mallet wrapper to run the LDA model because it is considered to be a faster implementation of the LDA model, than the traditional gensim library.

Phillipines Supreme Court. Jacobi et al. (2016) uses the LDA model to study large volumes of journalistic text from The New York Times from 1945 to present. Sokolova et al. (2016) identified election related events from Twitter data using LDA. Ryoo and Bendle (2017) use the LDA model to study the social media strategies of the two campaigns in the 2016 U.S election. The model can be used to infer what percentage of each topic is present in a particular tweet. This helps us understand which topic a particular tweet is about, from a broad level perspective. Figure 7, and Figure 8 shows the first topic in a LDA model fitted over Republican and Democrat politicians' tweets separately with 10 topics, in the month of November to December, 2017. The figure shows that for the first topic mostly deals with taxation and economy, but whereas we see words such as *taxreform*, *economy*, and *american* in the tweets by Republicans, we see words such as *goptaxscam* and *middleclass* by the Democrats.[14]

4.2.1 Computation of Distance Metrics

After obtaining separate distributions for the Democrats and Republicans, I compute measures of similarity and dissimilarity between the two topic probability distributions obtained after training the LDA model separately. There are three measures in the literature which seem to serve our purpose. The Hellinger distance is the analog of measuring the Euclidean distance between two probability distributions and is a symmetric distance measure. [15]

The Kullback–Leibler divergence also known as relative entropy also measures the distance between two probability distributions, but is not symmetric like the Hellinger distance.[16]

[14]This is only for purposes of illustration and the actual models are trained on weekly data, after picking the optimal number of topics using coherence score.

[15]The formula for the Hellinger distance between two probability distributions, P and Q is as follows

$$h(P,Q) = \frac{1}{2}||\sqrt{P} - \sqrt{Q}||^2. \tag{5}$$

[16]For two probability distributions, P and Q, the Kullback-Leibler divergence is as follows.

$$D_{KL}(P||Q) = \int_{-\infty}^{+\infty} p(x) log \frac{p(x)}{q(x} dx, \tag{6}$$

The Jaccard index, which is also known as the Intersection over Union or the Jaccard similarity coefficient is a measure of the overlap between two sets. The Jaccard coefficient measures similarity between finite sample sets, and is defined as the size of the intersection divided by the size of the union of the sample sets. The Jaccard distance is the complement to the Jaccard index. It measures the dissimilarity between two sets, and is obtained by subtracting the Jaccard index from 1. [17]

The distances are computed from November 7, 2017 to Novembwer 5, 2019 at weekly intervals for 104 weeks. Fig 10 shows the trend of these lines. The trends of the distance between the topic distributions as measured by the Hellinger distance Kullback Leibler Divergence are very similar. The divergence between topic distributions increases as we approach the election. The Jaccard distance between the raw text used by Democrats and Republicans, since the Jaccard distance is computed between two vectors. I use the vector of all words together used by Democrats and Republicans, as well as the vectors containing list of words, each list being one tweet. Each of these words are words that remain after the twitter data has been pre-processed. According to Fig 10 the distance between the word distributions fall as we approach the election.

Figure 9 shows the different types of distance metrics used to compute the Euclidean distance between the topic distributions. The trend in the distance of distance measures is the same as that for hashtag analysis. As the election approaches, Republicans and Democrats increasingly talk about different things.The Euclidean distance between the fraction of topics spoken between Democrats and Republicans also increases, but it increases as election approaches with a slight dip in the week before the election, and then again increases. It

where p(x) and q(x) are the density functions for P and Q respectively.

[17]The Jaccard distance is defined as follows:

$$d_J(A,B) = 1 - J(A,B) = 1 - \frac{|A \cap B|}{|A \cup B|}, \tag{7}$$

where A and B are finite samples.

is to note that these shows the relative weight each of these two parties assign to different topics each week.

4.2.2 Euclidean Distance

The LDA model assigns each tweet to multiple topics. For the next part of the analysis, I find the dominant topic for each tweet and assign the tweet to that particular topic. I calculate the euclidean distance between the vectors of fraction of tweets devoted to each topic by the Democrats and Republicans for each week. For example, if there are s topics in a particular week, I obtain two vectors D_s and R_s for that week. Both these vectors have s elements. Let the first element of D_s vector be denoted by $D_s(1)$. Then,

$$D_s(1) = \frac{n(Tweets\ by\ Democrats\ which\ belong\ to\ topic\ 1)}{n(Total\ tweets\ by\ Democrats)}, \tag{8}$$

where n denotes the cardinal number. I similarly compute the all the elements for D_s and R_s and find the euclidean distance between those two vectors. This denote by Score.

$$Score = d(D_s, R_s) = \sqrt{(D_s(1) - R_s(1))^2 + (D_s(2) - R_s(2))^2 + ... + (D_s(s) - R_s(s))^2} \tag{9}$$

These estimates help in understanding the between topic variability in the dominant topics used by Democrats and Republicans. A low value of the *Score* implies that both the parties devote similar weights to the various dominant topics whereas a higher value of *Score* implies that talk about different topics. Fig 10 shows the trend of the *Score* metric. As the election approaches, there is a rise in the value of the score, which again suggests that Democrats and Republicans are talking about different agendas in the approach to the election.

4.2.3 Euclidean Distance Interacted with Sentiments

The Euclidean distance helps us understand the pattern in the usage of topics by Republicans and Democrats by provide no insight into how these topics are being used. To make more sense of the intent of the content used for each topic, I perform sentiment augmented content analysis, I compute two types of metrics where sentiment is interacted with the topic. These measures give us a low value if Democrats and Republicans talk about the same topic with similar sentiments, and gives us a high value if they talk about the same topic using different sentiments. There are two approaches that I take.

For the first type of metric, I compute the fraction of positive, negative and neutral shares for each topic and then compute the euclidean distance for each topic between Republicans and Democrats, and add the distances for all the topics. For example, if there are s topics I compute 6 vectors $Positive_{D_s}$, $Negative_{D_s}$, $Neutral_{D_s}$, $Positive_{R_s}$, $Negative_{R_s}$, $Neutral_{R_S}$ with s elements each. The first element for $Positive_{D_s}$, denoted by $Positive_{D_s}(1)$ is defined as follows,

$$Positive_{D_s}(1) = \frac{n(Tweets\ by\ Democrats\ which\ belong\ to\ topic\ 1\ and\ have\ a\ positive\ sentiment)}{n(Total\ tweets\ by\ Democrats\ which\ belong\ to\ topic\ 1)}, \tag{10}$$

where n denotes the cardinal number. I similarly compute all the vectors. My first measure Sum_Frac_Dis is defined as follows,

$$Sum\_Frac\_Dis = d(Positive_{D_s}, Positive_{R_s}) + d(Negative_{D_s}, Negative_{R_S}) + d(Neutral_{D_s}, Neutral_{R_s}), \tag{11}$$

where $d(.,.)$ denotes the Euclidean distance between two vectors. I also computed a standardized version of Sum_Frac_Dis which is defined as follows,

$$Sum\_Frac\_Dis_{std} = \frac{Sum\_Frac\_Dis}{\sqrt{n}} \tag{12}$$

Second, instead of computing the fractions of positive tweets for a particular topic for each group, I use the intensity of the sentiment to derive the metric for Euclidean distance. Therefore, instead of counting the number of positive, negative or neutral tweets for each group, I compute the intensity of positivity, negativity and neutrality in the tweets. For n topics, I again compute the six vectors $Positive_{D_s}$, $Negative_{D_s}$, $Neutral_{D_s}$, $Positive_{R_s}$, $Negative_{R_s}$, $Neutral_{R_s}$. The first element for $Positive_{D_s}$, denoted by $Positive_{D_s}(1)$ is defined as follows,

$$\begin{aligned} Positive_{D_s}(1) = & \; Mean\ value\ of\ positive\ score\ for\ tweets\ by\ Democrats \\ & \; which\ had\ a\ positive\ sentiment\ and\ belonged\ to\ topic\ 1 \end{aligned} \tag{13}$$

In this case, I compute two measures, Sum_Dis and Sum_Dis_{std}, where Sum_Dis is defined as follows,

$$Sum\_Dis = d(Positive_{D_s}, Positive_{R_s}) + d(Negative_{D_s}, Negative_{R_s}) + d(Neutral_{D_s}, Neutral_{R_s}), \tag{14}$$

where $d(.,.)$ denotes the Euclidean distance between two vectors. I also computed a standardized version of Sum_Dis which is defined as follows,

$$Sum\_Dis_{std} = \frac{Sum\_Dis}{\sqrt{n}} \tag{15}$$

Figure 11 shows the trend of euclidean distance interacted with the sentiment for the fraction of positive, negative and neutral tweets of a particular topic. Unlike the $Score$ metric that we discussed in the last subsection, the value of the Sum_Frac_Dis and $Sum\_Frac\_Dis\_std$ both fall as we approach the election. This suggests that Democrats and Republicans use similar sentiment to talk about common topics more and more as they approach the election. We get the same trends when we use the intensity of the positive, negative or neutral sentiment as opposed to the fractions. The figure is shown in the Appendix.

4.3 Bayesian Ideal Point Estimation

The third way by which I try to capture this short-run polarization is through computation of ideological estimates of the politicians using their tweets through the methods of Bayesian ideal point estimation. This method developed by Eady (2018) uses the URL sharing behaviour to infer the ideology of the politicians. The idea is that politicians will share morel URLs from a media source which is close to them in the ideological spectrum. A Monte Carlo simulation is then performed to infer the ideologies of the politicians.

To estimate the ideologies of the politicians, I extract the URLs from the tweets of the Representatives. The extracted URLs are shortened URLs, and cannot be directly used. The URLs are then expanded into their long form by querying the server using the shortened URLs. I then extract the domain names of the URLs and compute an adjacency matrix of how many times each Congressperson has tweeted any particular website. Retweets are included in this analysis, as retweets also signify the reiteration of the content of the original tweet by the person who is re-tweeting the original tweet. I remove all social media domain names from the adjacency matrix such as google.com, facebook.com, instagram.com and others as these do not have any ideological content of their own.[18] For data sanity purposes, I take only 90 percent of the total shares of news domains, as this helps the data to be devoid of obscure websites, which had only been mentioned once or twice, and other obscure names, which just show up in the adjacency matrix due to technicalities of domain name extraction. After this I employ the empirical strategy used by Eady (2018)[19]. I differ from the strategy used by Eady in that I do not specify whether an individual is a Republican and Democrat and only use the information from the URL sharing behaviour to get my estimates.

[18] I also remove any websites that has "house.gov", in them, because the Representatives just seem to be using Twitter as a platform to broadcast those websites.

[19] He has developed a mediascores package in R available in Github for faster and simpler implementation of the Bayesian Ideal Point Estimation, which I have used for my analysis.

The empirical strategy is implemented as follows:

$$y_{im} \sim \text{NegBin}(\pi_{im}, \omega_i \omega_m) \tag{16}$$

$$\pi_{im} = \exp(\alpha_i + \gamma_m - ||\vartheta_i - \zeta_m||^2), \tag{17}$$

where y_{im} denotes the count of shares of domain m shared by user i; α_i denotes a user-specific intercept which essentially means that some Congresspeople are more active on Twitter, and may indulge in higher URL sharing activity ; γ_m denotes a domain-specific intercept which similarly accounts for the fact that some domains might have a higher probability of being shared than others; and ω_i and ω_m denote user- and domain-specific dispersion parameters respectively. The quantities of interest are denoted by ϑ_i, which represents the ideology of user i, and ζ_m, the ideology of the website being shared m. As the term $-||\vartheta_i - \zeta_m||^2$ in the linear predictor makes clear, the larger the ideological spatial distance between the ideology ϑ_i of the user and the ideology ζ_m of the website, the less likely the user is to share stories from that organization. Priors are placed on the model parameters as follows:

$$\alpha_i \sim \text{Normal}(\mu_\alpha, \sigma_\alpha) \tag{18}$$

$$\gamma_i \sim \text{Normal}(0, \sigma_\gamma), \tag{19}$$

where uniform prior distributions are placed on the hyperparameters μ_α, σ_α and σ_γ. The variance parameters ω_i and ω_m, are given common distributions $\omega_i \sim \text{InvGamma}(\omega_a^{(i)}, \omega_b^{(i)})$ and $\omega_m \sim \text{InvGamma}(\omega_a^{(m)}, \omega_b^{(m)})$, with uniform priors on the hyperparameters. For identification, the parameters representing news media ideology are centered on 0: $\zeta_m \sim \text{Normal}(0, \sigma_\zeta)$. Lastly, the direction of the model needs to be set, such that high values represent ideological liberalism or conservatism. The anchors are fixed by the researcher (e.g. nytimes.com, foxnews.com), such that the ideology of the first media organization defines the low end of the scale (in this example, liberal), and the second, the high end (in this example, conservative).

In my case, I choose the anchors to be domains shared by one of the parties more relatively to the other party. By this mean that, I compute the ratio of the number of times a Democrat has shared a particular domain divided by the number of times it has been tweeted totally, to be the most Democratic domain and similarly for Republicans. I call this the most differential domain for Democrats and Republicans respectively. I also eliminate the domains which have been shared by less than 2 percent of the respective group of politicians as anchor websites, as they would just be fringe websites, and won't be efficient for the Monte Carlo Simulation. One of the upshots of this analysis is that I choose the anchor websites dynamically and objectively based on the tweeting behaviour of the politicians and not subjectively as has been done in the original implementation of the model. The most differential domain by Democrats moves between vox.com, nytimes.com, cnn.com, npr.org and others. The most deferentially highly tweeted domain name by the Republicans lingers between foxnews.com, wsj.com, www.washingtonexaminer and others. I had also replicated the analysis with the most highly tweeted domain, and in that case the Democrat anchor website is overwhelmingly nytimes.com. Since, nytimes.com is not a very extreme right news media organization, I think fixing the anchor website to be the most differentially highly tweeted domain makes more sense. Fig 12 and Fig 13 shows the most popular domain names for the Democrat and Republican politicians respectively. Another point of departure from the traditional model as implemented is that I do not assign separate groups to Democrats and Republicans as I want only the URL sharing behavior to inform their ideological points, and not to get biased by their group identity, as already explained before. Therefore, the estimates in my model are not biased by ex-ante group identity. Figure 14 shows the mean ideological trend for the Democrats and the Republicans calculated using Bayesian Ideal Point Estimation.15 shows the ideological polarization between Republicans and Democrats a year before and after the election.

4.4 Mentions Network Analysis

For the fourth part of ideological metric, I calculate the polarization in the mentions network in my data set. This gives us an idea about the affective polarization within the network. Figure 16 visually shows the polarization in the mention network. As we can see the mention network is heavily polarized. Since, I want to look at change in polarization in a time series fashion, traditional measures of network analysis such as nodes, degree or centrality does not help.

To compute the degree of affective polarization in this network, I compute the shares of how many times Democrats mention other Republicans over the span of 104 weeks, and vice-versa[20]. Figure 17a shows that as election approaches Democrats mention Republicans less positively. There is a slight drop in positive mentions after the election but it keeps increasing for sometime as we move away from the election. Figure 17b however shows that for Republicans we get the opposite result, and they mention Democrats more positively as we approach the election.There is again a drop after the election after which it keeps decreasing for sometime as we move away from the election. Figure 18a and 18b show that both parties mention each other more negatively with the approach of the election. But while there is a drop in negative mentions of Republicans by Democrats, there is a significant increase in negative mentions of Democrats by Republicans after the election. Both the estimates however keep falling as we move away from the election. It is important to note that these are broad level trends and some of these trends do change direction when we zoom in to the last 8 weeks of the election. The results of what happens in the last 8 weeks is discussed in Section 5.

[20]To make the analysis more accurate, I collect the handles of all Democrats and Republicans for the present House and Senate, and also for the previous House, as well as the current and past President. Therefore, the list of Republicans whom Democrats mention are not only Republicans in the current house of Congress, but also include Senators and House members in the 115th House of Congress, with a verified official handle

4.5 Retweet Network Analysis

I compute similar measures of polarization for the retweet network as well. Figure 19a shows that similar to the mention network, as election approaches Democrats mention Republicans less positively. There is a slight drop in positive mentions after the election but it keeps increasing for sometime as we move away from the election. Figure 19b shows that Republicans retweet Democrats more positively in the weeks leading up to the election. There is a decrease in the positive mentions after the election and it keeps increasing as we move away from the election. Figures 20a (a) and 20b (b) show that there negative mentions of out group increases as we approach the election. There is a slight drop in negative mentions of Republicans by Democrats and a slight increase in negative mentions of Democrats by Republicans and negative mentions of the out group decreases as we move away from the election. However, for the negative retweets both Democrats and Republicans retweet each other more negatively as the election approaches as shown by Figure 20a and Figure respectively.

5 Empirical Analysis

5.1 Regression Discontinuity Analysis

After computing the metrics of ideological polarization, I use the non-parametric Regression Discontinuity Design to see if there is a significant discontinuous jump before and after the election. I use weeks to the election as my running variable, with the cut-off at 0 and the metrics that I have already computed as my outcome variables. The RDD set-up works very well in this scenario. Although the election is not an exogenous event as is generally the requirement for a RDD, this works in our favour because we are trying to measure the effect on ideology as soon as the anticipation of an impending election goes away. It is for this reason that I use a non-parametric RDD so as to look at only a narrow window before and after the election. I use a data driven approach to find the right bandwidths for the regression as outlined in Calonico et al. (2014).[21] I perform non-parametric Regression Discontinuity in Time as performed in Davis (2008). I report the results for varying degrees of polynomials. I use the following equation to estimate the non-parametric RDD

$$y = f(t) + \epsilon, \tag{20}$$

where y denotes the various estimates that I have calculated and t denotes time in weeks, and is the running variable.

5.1.1 Results

In the RDD estimates, I only find a significant increase in the number of common hashtags used which shows Republicans and Democrats start talking about similar things at a higher rate just after election compared to that before election as shown in Table (2). None of the other estimates have any significant discontinuities as shown in Tables 3 - 7. There are some effects in the mention network analysis and retweet network analysis. $Mentioned\_Dem\_Negatively$ estimates by Republicans increase right after the election as shown in Table(6). In the retweet network analysis, estimates of $Retweeted\_Rep\_Negatively$ by Republicans decreased right after the election as shown in Table (7).

5.2 OLS Estimates for sub sample

It is clear by looking at the smoothed lines in the graphs of the estimates that they vary considerably depending on where one is in the election cycle. The locally polynomial regression lines shown in the graphs help us get a sense of the way these estimates vary. These estimates therefore give us a sliding window view of what is happening to polarization at any point in time. As is clear polarization is different depending upon where one is in the

[21] I use the rdrobust package in R for the Regression Discontinuity estimation. The package uses a data-driven methodology to select the best bandwidth

election cycle and does not have a long term trend. It therefore does not make much sense to infer the effect of the election by using the entire time series. Another potential challenge in inferring any causality from these graphs is that a lot of events can influence the tweeting behaviour of politicians such as major world wide events or primaries. There could also be potential seasonality effects in the time series. Therefore, to develop an understanding of what is happening just before the election, one needs to focus on a narrow window close to the election.

As a specific case for illustrative purposes, I look at a window of 8 weeks before and after the election. I choose 8 weeks because there are no primaries contested in the last 8 weeks for the federal election and there is no major worldwide event to influence the tweeting behaviour of the politicians. It is therefore reasonable to assume that all tweets in the last 8 weeks of the election cycle will be in reference to the midterm election and this helps us get closer to the effect of the election.

I run a simple linear regression model to estimate the slopes of the estimates before and after the elections. It is important to note that I am not trying to show causality or compute very precise estimates about the magnitude of change but I am more interested in the direction of change. The direction of change sheds light on whether polarization decreases or increases as we approach the election. The equation that I estimate is

$$y_t = \alpha + \beta Week + \epsilon, \tag{21}$$

where y_t represents the various estimates that I create. I estimate two separate slope coefficients one before the election and one after the election. For the pre-election version, Weeks increase from 1 to 8 as we get closer to the election, with the 8^{th} week being closest to the election. In the post-election version, weeks increase as we get further away from the election., with the 8^{th} week being the furthest away from the election. Another way to think about this would be that for the pre-election and post-election specification weeks increases in the positive direction of the time.

5.2.1 Results

The slopes for the hashtag estimates is negative in the last 8 weeks of the election as shown in Table 8. This implies that Democrats and Republicans use increasingly less similar hashtags conditional on the top hashtags used by them. The estimates for $Hashtag_{50}$ and $Hashtag_{100}$ are significant at 5 percent level. The estimates also decrease as the politicians move away from the election although there is a significant increase in the number of common hashtags just after the election as shown in Table 2. The slope of the Inverse Score estimate is also negative, implying inverse score falls and therefore distance increases as we approach the election. This means that as elections get closer, conditional on using the same hashtags politicians from different parties use increasingly different sentiments to talk about those hashtags. The slope of the measure is negative after the election too, but there is a increase in the inverse score as soon as the election is over as shown in Table 2. Both these measures hint at an increase in polarization or movement away from the median voter in the anticipation of the election.

Similarly the the slopes of Hellinger distance and the Kullback-Leibler divergence are positive as shown in Table 9 implying an increase in topic divergence. Even when we focus only on the dominant topics and use the distance between the relative shares that Democrats and Republicans devote to such topics, the slope is positive implying that distance increases as we approach the election as shown in Table 10. The combined evidence suggests that Democrats and Republicans increasingly talk about different agendas with the approach of the election. Therefore, both in case of hashtags which are very context specific framing devices as well as for broad topics Democrats and Republicans grow increasingly divergent with the approach of the election.

However, the results are opposite when we consider the contents of all the tweets through topic modelling. The slope of the Jaccard Distance between word distributions is negative as shown in Table 9. This implies that politicians increasingly use similar words in their tweets. When the topic analysis is augmented with sentiment scores, the euclidean distance score decreases which means that politicians increasingly use the same sentiment to talk about different topics as shown in Table 10. Therefore, the content analysis as implemented through sentiment augmented topic modelling hints at decrease in polarization and movement towards the median voter in anticipation to the election.

The ideological difference between Republicans and Democrats as computed using the Bayesian ideal point estimation also decreases for both competitive districts and non-competitive districts as shown in Table 11. The ideological difference in the competitive districts increases after the election whereas the difference in the non-competitive districts decreases after the election. This also suggests a decrease in polarization and movement towards the median voter as the election approaches.

Negative mentions of Republicans by Democrats increases as we move into the election, falls once the election is over and decreases for sometime and increases again as we move away from the election. Negative mentions of Democrats by Republicans increases as we move into the election, rises(significantly) once the election is over and decreases for sometime and increases again as we move away from the election. Negative retweets of Republicans by Democrats decreases as we move into the election, falls once the election is over(significantly) and increases in the 8 weeks after the election. Negative retweets of Democrats by Republicans increases as we approach the election, decreases after the election and falls thereafter.

The combined pieces of evidences suggest that while Democrats and Republicans become more polarized in their agenda setting behaviour with the approach of the election, they become less polarized in terms of the content shared within a particular agenda. In other words, whereas the between agenda or between topic variability increases with the approach of the election, the within topic variability decreases. One way to think about this would be to say that while politicians are trying to appeal to their extreme electoral bases through their agenda setting behaviour, they try to appeal to the median voter or the swingers by remaining more moderate in the content they share within the diverse agendas.This could also be because while faithful voter bases might be lured by token gestures or the appearance of extremism, more moderate and attentive voters might need more content to win them over.

6 Conclusion

In this paper I collect tweets from incumbent Representatives in the 116th House of Congress at a weekly interval one week before and after the 2018 midterm election. I then use the Twitter data to construct several estimates of political ideology using methods such as hashtag analysis, topic modelling, Bayesian Ideal Point Estimation as well as analyze the mention and retweet networks of the politicians.

There are two ways one can interpret the result from this paper. From a Frequentist perspective, the statistical insignificance of the Regression Discontinuity estimates as well as the sub sample estimates (except for the Hashtag estimates and the some of the mention network estimates) suggests that there is no discernible discontinuity at the election. It also implies that there is no significant change in behaviour at the level of discourse either before or after the election.

However, one could also argue that there is very little variation in the independent variable for the OLS sub sample estimates and therefore it is difficult to get precise estimates with low standard errors. Adopting a Bayesian perspective helps us infer some patterns from the estimates that can help us to inform our priors which can be later validated/rejected through a future project. When we use the Bayesian perspective we find some interesting patterns. Polarization as measured by broad level agenda setting behaviour such as hashtags or topics is found to increase with the approach of the election. However, when we shift our measuring instrument to similarity in words used, sentiment augmented topic analysis or ideological scores inferred from the media sharing activity we find polarization to be decreasing with the approach of the election. This suggests that while there is increasing divergence in preferred electoral agenda setting, there is convergence in agenda specific positioning.

A potential future area of research would be to repeat this analysis for multiple election periods, and to check if the patterns that we find here repeat in multiple election periods or if it is something unique to this election period.

References

Adamic, L. A. and N. Glance (2005). The political blogosphere and the 2004 us election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, pp. 36–43. ACM.

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

Aldrich, J. H. (1995). *Why parties?: The origin and transformation of political parties in America*. University of Chicago Press.

Alesina, A. (1988). Credibility and policy convergence in a two-party system with rational voters. *The American Economic Review 78*(4), 796–805.

Allcott, H. and M. Gentzkow (2017). Social media and fake news in the 2016 election. *Journal of economic perspectives 31*(2), 211–36.

Ash, E., D. L. Chen, and W. Lu (2017). Polarization of us circuit court judges: a machine learning approach. *Available at SSRN 2993009*.

Auter, Z. J. and J. A. Fine (2016). Negative campaigning in the social media age: Attack advertising on facebook. *Political Behavior 38*(4), 999–1020.

Azzimonti, M. and M. Fernandes (2018). Social media networks, fake news, and polarization. Technical report, National Bureau of Economic Research.

Bara, J., A. Weale, and A. Bicquelet (2007). Analysing parliamentary debate with computer assistance. *Swiss Political Science Review 13*(4), 577–605.

Barberá, P., A. Casas, J. Nagler, P. J. EGAN, R. Bonneau, J. T. Jost, and J. A. Tucker (2018). Who leads? who follows? measuring issue attention and agenda setting by legislators and the mass public using social media data. *American Political Science Review*, 1–19.

Barberá, P., J. T. Jost, J. Nagler, J. A. Tucker, and R. Bonneau (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological science 26*(10), 1531–1542.

Besley, T. and S. Coate (1997). An economic model of representative democracy. *The Quarterly Journal of Economics 112*(1), 85–114.

Blei, D. M., A. Y. Ng, and M. I. Jordan (2003). Latent dirichlet allocation. *Journal of machine Learning research 3*(Jan), 993–1022.

Boyd, D. M. and N. B. Ellison (2007). Social network sites: Definition, history, and scholarship. *Journal of computer-mediated Communication 13*(1), 210–230.

Boyd-Graber, J., Y. Hu, D. Mimno, et al. (2017). Applications of topic models. *Foundations and Trends® in Information Retrieval 11*(2-3), 143–296.

Bruns, A. and J. Burgess (2015). Twitter hashtags from ad hoc to calculated publics. *Hashtag publics: The power and politics of discursive networks*, 13–28.

Bruns, A., T. Highfield, and J. Burgess (2014). The arab spring and its social media audiences: English and arabic twitter users and their networks. In *Cyberactivism on the participatory web*, pp. 96–128. Routledge.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust data-driven inference in the regression-discontinuity design. *The Stata Journal 14*(4), 909–946.

Calvert, R. L. (1985). Robustness of the multidimensional voting model: Candidate motivations, uncertainty, and convergence. *American Journal of Political Science*, 69–95.

Collie, M. P. and J. L. Mason (2000). The electoral connection between party and constituency reconsidered: Evidence from the us house of representatives, 1972–1994. *Continuity and change in house elections*, 211–34.

Comanor, W. S. (1976). The median voter rule and the theory of political choice. *Journal of Public Economics 5*(1-2), 169–177.

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

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

Coughlin, P. and S. Nitzan (1981). Directional and local electoral equilibria with probabilistic voting. *Journal of Economic Theory 24*(2), 226–239.

Davis, L. W. (2008). The effect of driving restrictions on air quality in mexico city. *Journal of Political Economy 116*(1), 38–81.

Demszky, D., N. Garg, R. Voigt, J. Zou, M. Gentzkow, J. Shapiro, and D. Jurafsky (2019). Analyzing polarization in social media: Method and application to tweets on 21 mass shootings. *arXiv preprint arXiv:1904.01596*.

Downs, A. (1957). An economic theory of political action in a democracy. *Journal of political economy 65*(2), 135–150.

Gentzkow, M., B. Kelly, and M. Taddy (2019). Text as data. *Journal of Economic Literature 57*(3), 535–74.

Gentzkow, M. and J. M. Shapiro (2010). What drives media slant? evidence from us daily newspapers. *Econometrica 78*(1), 35–71.

Gentzkow, M., J. M. Shapiro, and M. Taddy (2019). Measuring group differences in high-dimensional choices: Method and application to congressional speech. *Econometrica 87*(4), 1307–1340.

Glaeser, E. L., G. A. Ponzetto, and J. M. Shapiro (2005). Strategic extremism: Why republicans and democrats divide on religious values. *The Quarterly journal of economics 120*(4), 1283–1330.

Golbeck, J., J. M. Grimes, and A. Rogers (2010). Twitter use by the us congress. *Journal of the American Society for Information Science and Technology 61*(8), 1612–1621.

Graham, T., M. Broersma, K. Hazelhoff, and G. Van'T Haar (2013). Between broadcasting political messages and interacting with voters: The use of twitter during the 2010 uk general election campaign. *Information, communication & society 16*(5), 692–716.

Grant, W. J., B. Moon, and J. Busby Grant (2010). Digital dialogue? australian politicians' use of the social network tool twitter. *Australian Journal of Political Science 45*(4), 579–604.

Gruzd, A. and J. Roy (2014). Investigating political polarization on twitter: A canadian perspective. *Policy & Internet 6*(1), 28–45.

Hetherington, M. J. (2001). Resurgent mass partisanship: The role of elite polarization. *American Political Science Review 95*(3), 619–631.

Hinich, M. J. (1976). Equilibrium in spatial voting: The median voter result is an artifact.

Hong, S. (2013). Who benefits from twitter? social media and political competition in the us house of representatives. *Government Information Quarterly 30*(4), 464–472.

Hong, S. and S. H. Kim (2016). Political polarization on twitter: Implications for the use of social media in digital governments. *Government Information Quarterly 33*(4), 777–782.

Hotelling, H. (1990). Stability in competition. In *The Collected Economics Articles of Harold Hotelling*, pp. 50–63. Springer.

Huang, C. (2011). Facebook and twitter key to arab spring uprisings: report. In *The National*, Volume 6, pp. 2–3.

Hutto, C. J. and E. Gilbert (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth international AAAI conference on weblogs and social media.*

Jacobi, C., W. Van Atteveldt, and K. Welbers (2016). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism 4*(1), 89–106.

Jacobson, G. C. (2000). Party polarization in national politics: The electoral connection. In *Polarized politics: Congress and the president in a partisan era*, Volume 5, pp. 17–18.

Johnson, T. J. and D. D. Perlmutter (2010). Introduction: the facebook election. *Mass Communication and Society 13*(5), 554–559.

Langa, M., S. Ndelu, Y. Edwin, and M. Vilakazi (2017). # hashtag: An analysis of the# feesmustfall movement at south african universities.

Layman, G. C., T. M. Carsey, J. C. Green, R. Herrera, and R. Cooperman (2010). Activists and conflict extension in american party politics. *American Political Science Review 104*(2), 324–346.

Ledyard, J. O. (1984). The pure theory of large two-candidate elections. *Public choice 44*(1), 7–41.

Lee, D. S., E. Moretti, and M. J. Butler (2004). Do voters affect or elect policies? evidence from the us house. *The Quarterly Journal of Economics 119*(3), 807–859.

Lee, J. K., J. Choi, C. Kim, and Y. Kim (2014). Social media, network heterogeneity, and opinion polarization. *Journal of communication 64*(4), 702–722.

Lucas, C., R. A. Nielsen, M. E. Roberts, B. M. Stewart, A. Storer, and D. Tingley (2015). Computer-assisted text analysis for comparative politics. *Political Analysis 23*(2), 254–277.

Mamou, J., B. Ramabhadran, and O. Siohan (2007). Vocabulary independent spoken term detection. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 615–622.

McCarty, N. M., K. T. Poole, and H. Rosenthal (1997). *Income redistribution and the realignment of American politics*. Aei Pr.

Monti, C., A. Rozza, G. Zappella, M. Zignani, A. Arvidsson, and E. Colleoni (2013). Modelling political disaffection from twitter data. In *Proceedings of the second international workshop on issues of sentiment discovery and opinion mining*, pp. 1–9.

Moscato, D. (2016). Media portrayals of hashtag activism: A framing analysis of canada's# idlenomore movement. *Media and Communication 4*(2), 3.

Nardi Jr, D. J. (2012). "it's only words, and words are all i have": Using latent text analysis to analyze topics in philippine supreme court decisions.

Osborne, M. J. and A. Slivinski (1996). A model of political competition with citizen-candidates. *The Quarterly Journal of Economics 111*(1), 65–96.

Quinn, K. M., B. L. Monroe, M. Colaresi, M. H. Crespin, and D. R. Radev (2010). How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science 54*(1), 209–228.

Ryoo, J. and N. Bendle (2017). Understanding the social media strategies of us primary candidates. *Journal of Political Marketing 16*(3-4), 244–266.

Small, T. A. (2011). What the hashtag? a content analysis of canadian politics on twitter. *Information, communication & society 14*(6), 872–895.

Sokolova, M., K. Huang, S. Matwin, J. Ramisch, V. Sazonova, R. Black, C. Orwa, S. Ochieng, and N. Sambuli (2016). Topic modelling and event identification from twitter textual data. *arXiv preprint arXiv:1608.02519*.

Steyvers, M., P. Smyth, M. Rosen-Zvi, and T. Griffiths (2004). Probabilistic author-topic models for information discovery. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 306–315.

Sunstein, C. R. (2018). *# Republic: Divided democracy in the age of social media*. Princeton University Press.

Towner, T. L. and D. A. Dulio (2012). New media and political marketing in the united states: 2012 and beyond. *Journal of Political Marketing 11*(1-2), 95–119.

Vergeer, M. (2015). Twitter and political campaigning. *Sociology compass 9*(9), 745–760.

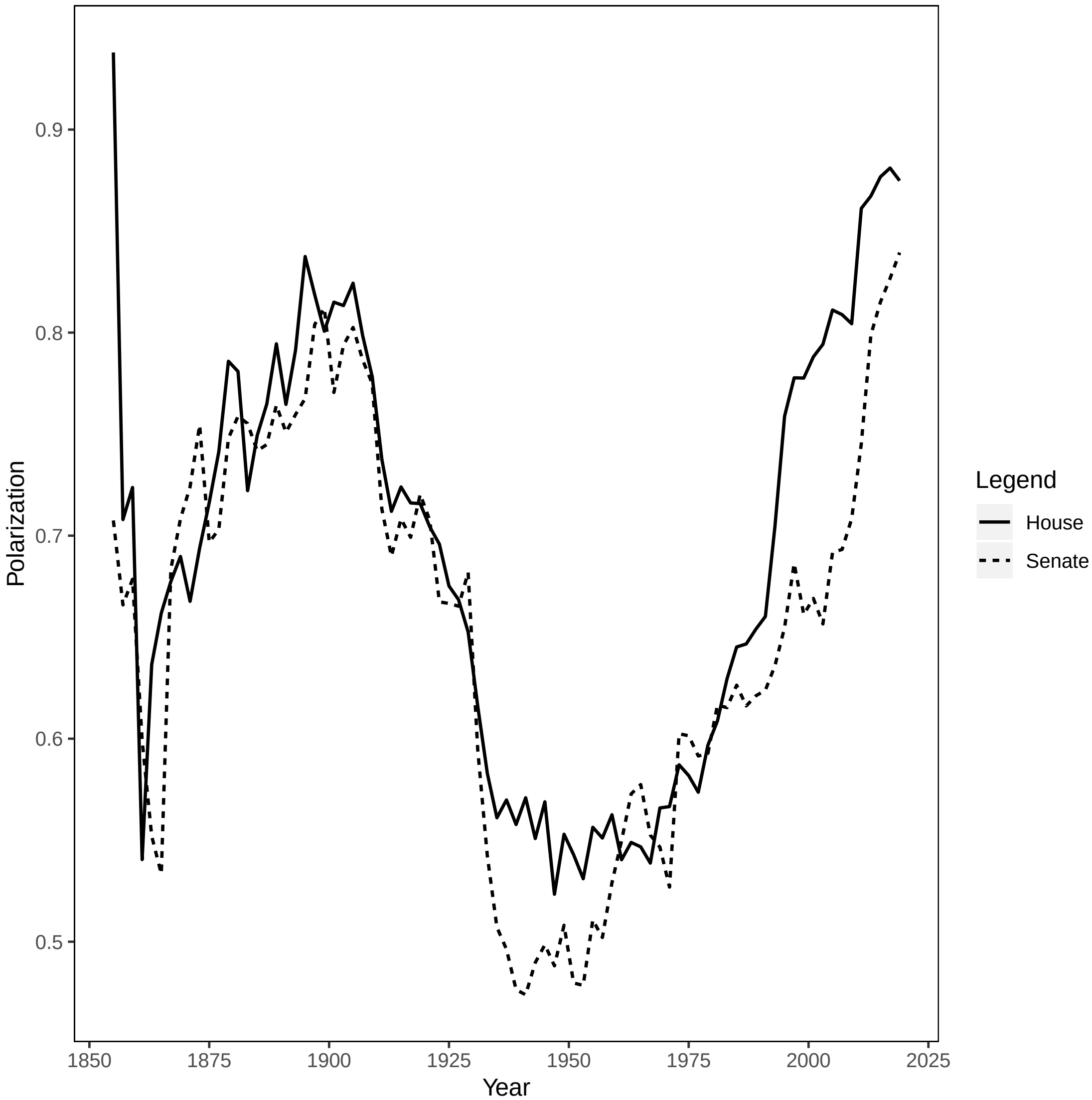
Wang, X., F. Wei, X. Liu, M. Zhou, and M. Zhang (2011). Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pp. 1031–1040.

Wittman, D. (1983). Candidate motivation: A synthesis of alternative theories. *American Political science review 77*(1), 142–157.

Xiong, Y., M. Cho, and B. Boatwright (2019). Hashtag activism and message frames among social movement organizations: Semantic network analysis and thematic analysis of twitter during the# metoo movement. *Public relations review 45*(1), 10–23.

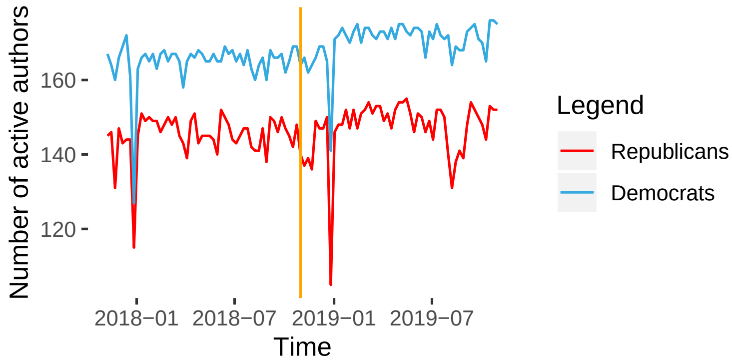
7 Figures and Tables

Figure 1: Polarization as measured by DW-Nominate scores over the years



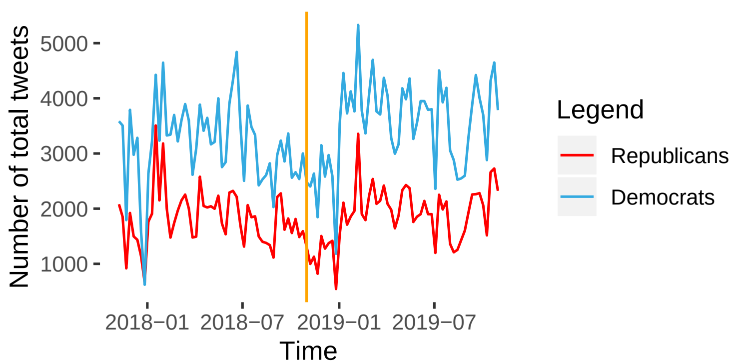
Notes: Difference between mean ideological positions for Republican and Democrat politicians from 1855 to 2019 along dimension 1 of DW-Nominate scores, using data from voteview.com. An almost similar graph is reproduced by the estimates of dimension 1, by Poole and Rosenthal

Figure 2: Total number of active politicians in each week.



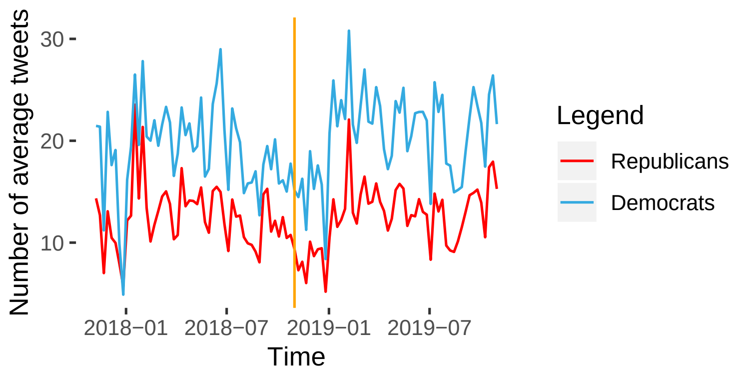
Notes: The number of active authors (authors who have tweeted atleast one tweet in a week) are counted for each week. The orange line indicates the election.

Figure 3: Total number of tweets in each week.



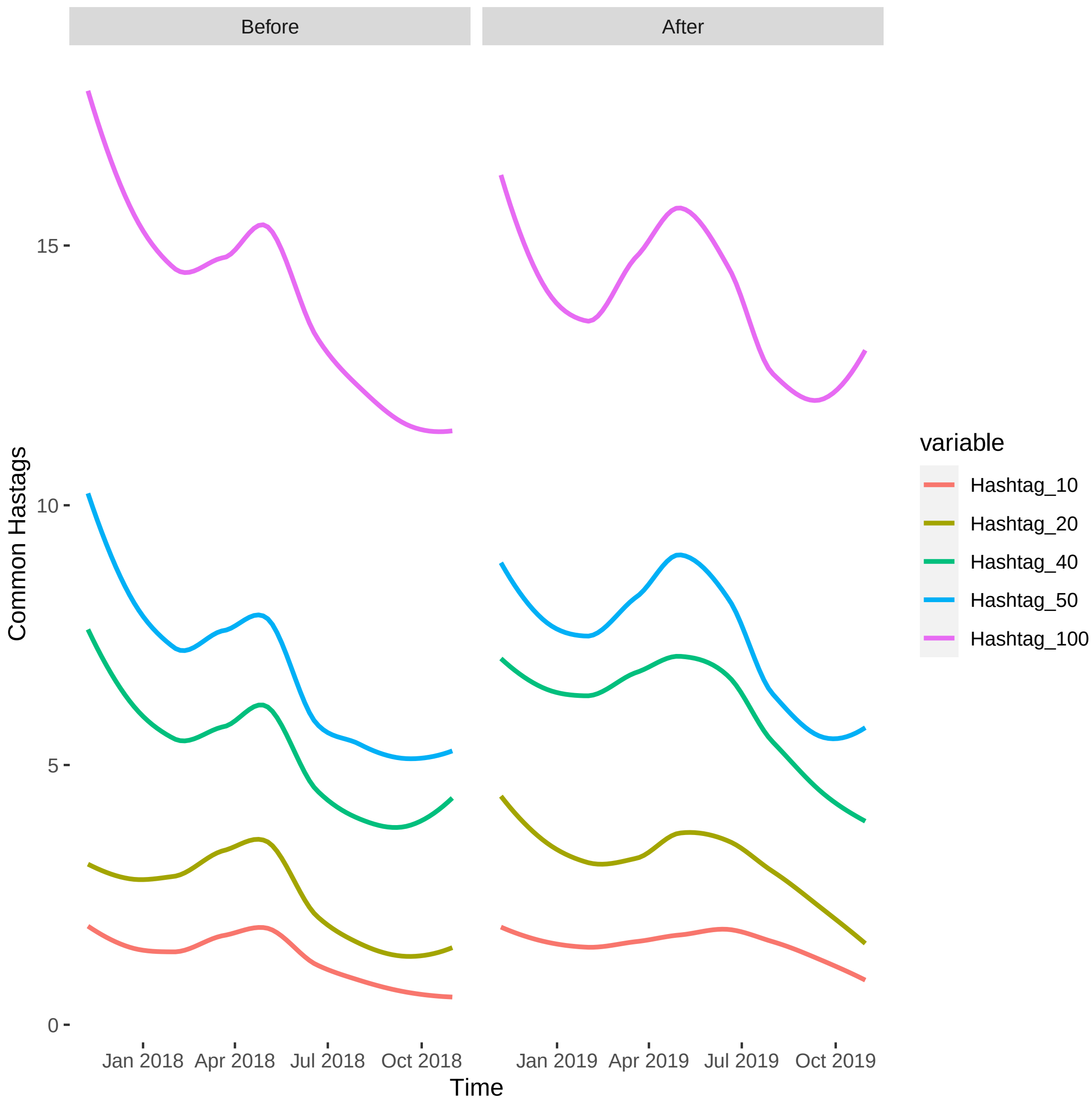
Notes: The total number of tweets are plotted for each. The orange line indicates the election.

Figure 4: Number of average tweets in each week.



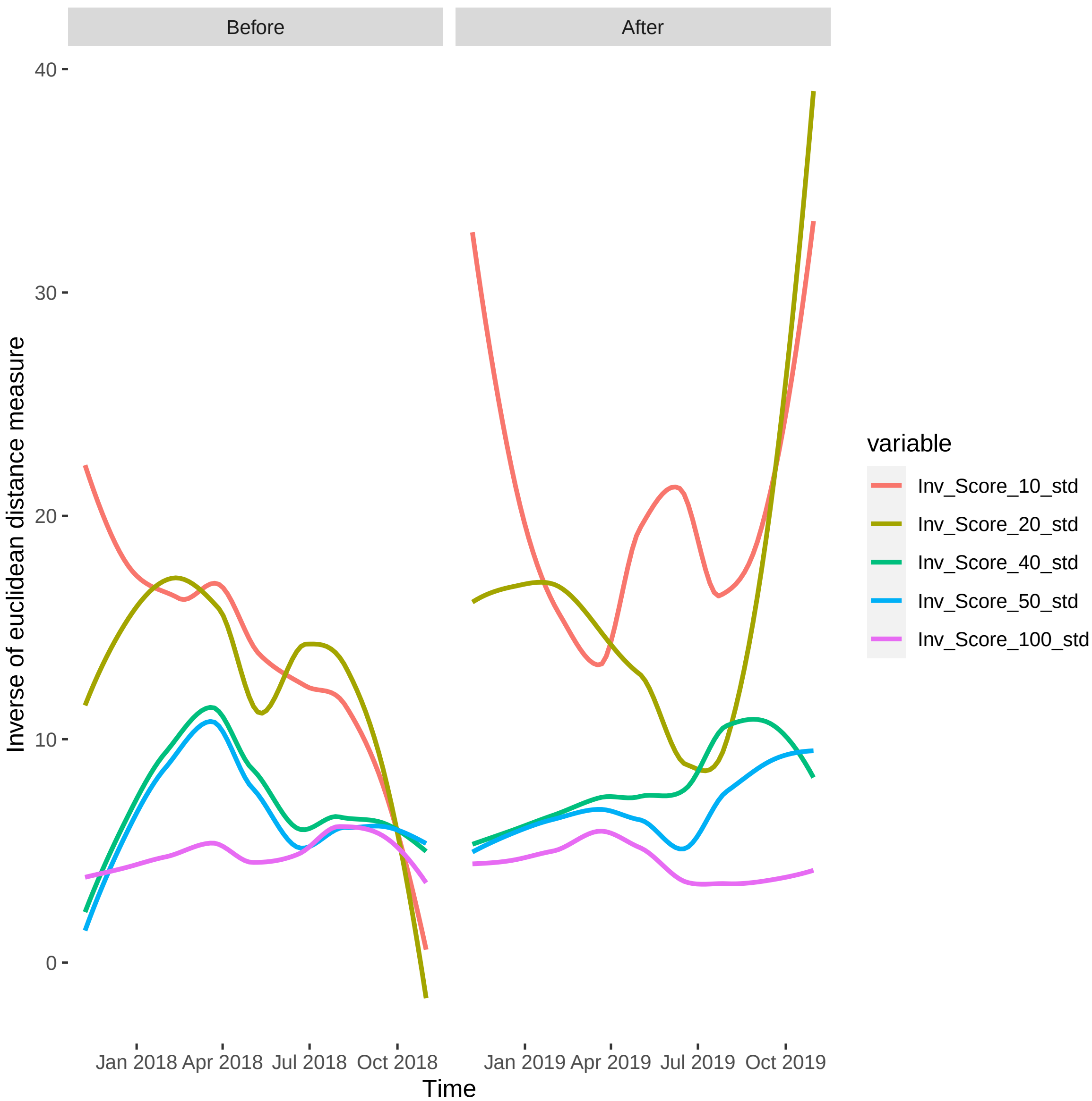
Notes: The number of tweets sent out on average each week. The numbers plotted in this graph is obtained by dividing the total number of tweets by the number of active authors. The orange line indicates the election.

Figure 5: Similarity in hashtag over time.



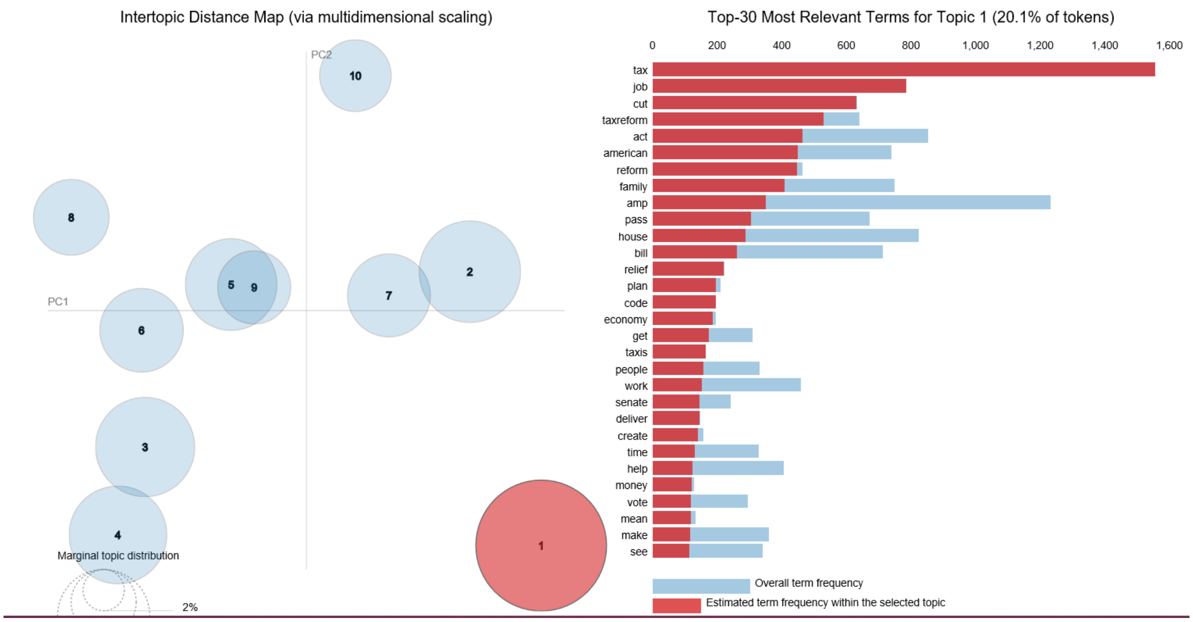
Notes: The figure plots the number of common hashtags used in the top 10, 20, 40, 50 and 100 hashtags used by Republicans and Democrats. The Before facet shows what the trend before the Nov 6th midterm election, whereas the After facet shows the trend after the Nov 6th midterm election.

Figure 6: Trend of inverse of standardized euclidean distance of hashtags interacted with sentiments



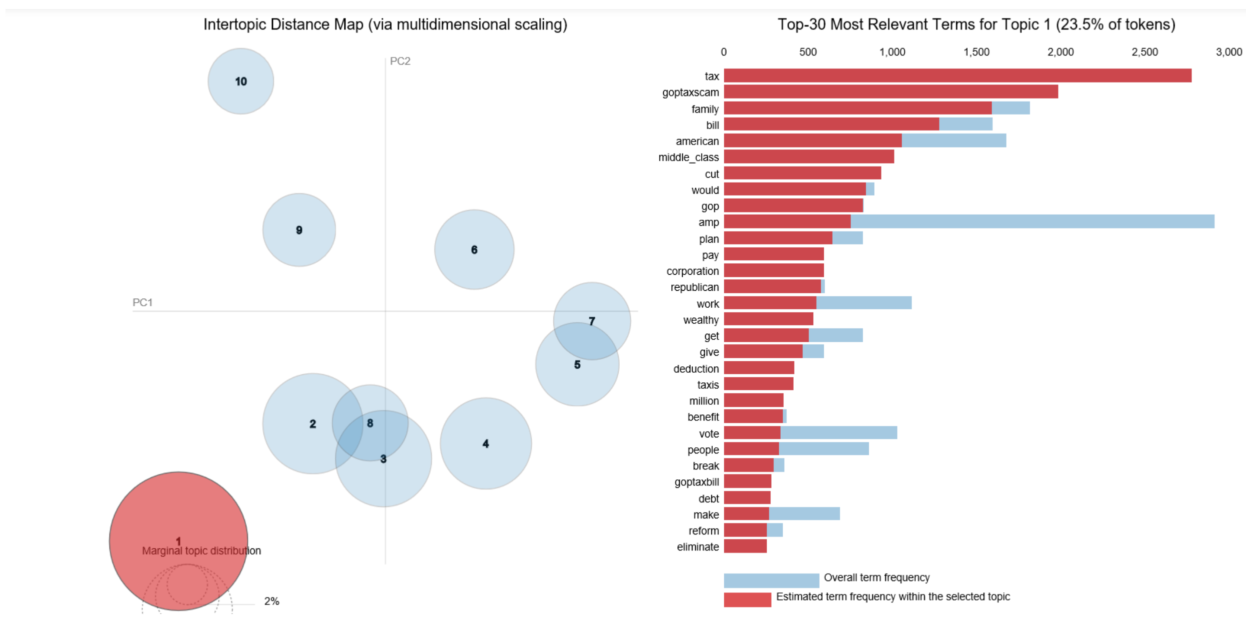
Notes: The figure inverse of the euclidean distance between negative tweets containing hashtags between Republicans and Democrats. A higher value implies low polarization whereas a higher value implied low polarization.

Figure 7: Token distribution for Topic 1 in Republican tweets



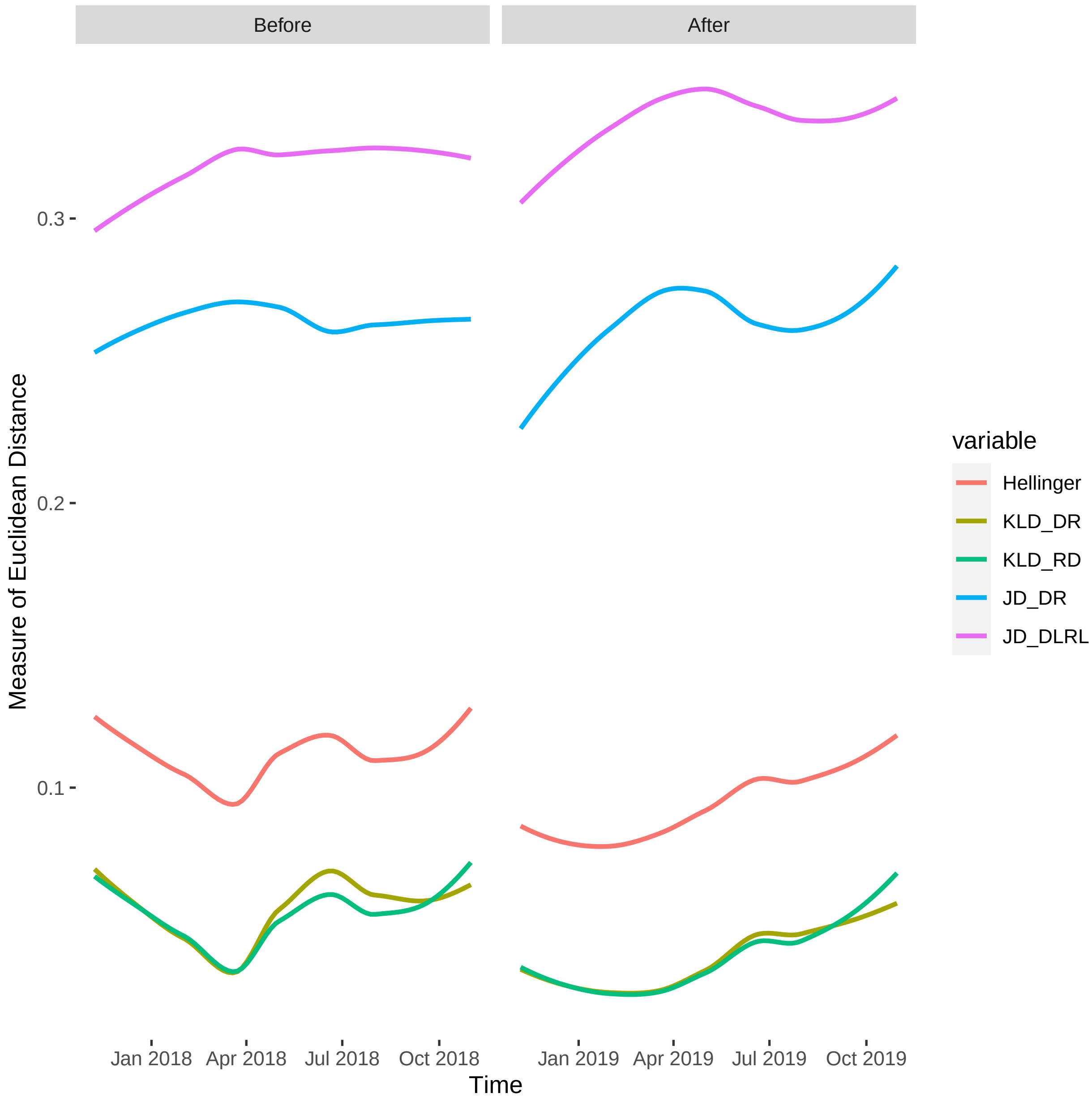
Notes: The top 30 most relevant words used in the first topic in the tweets by Republicans for the month of 7th November- 7th December, 2017 after fitting the LDA model.

Figure 8: Token distribution for Topic 1 in Democrat tweets



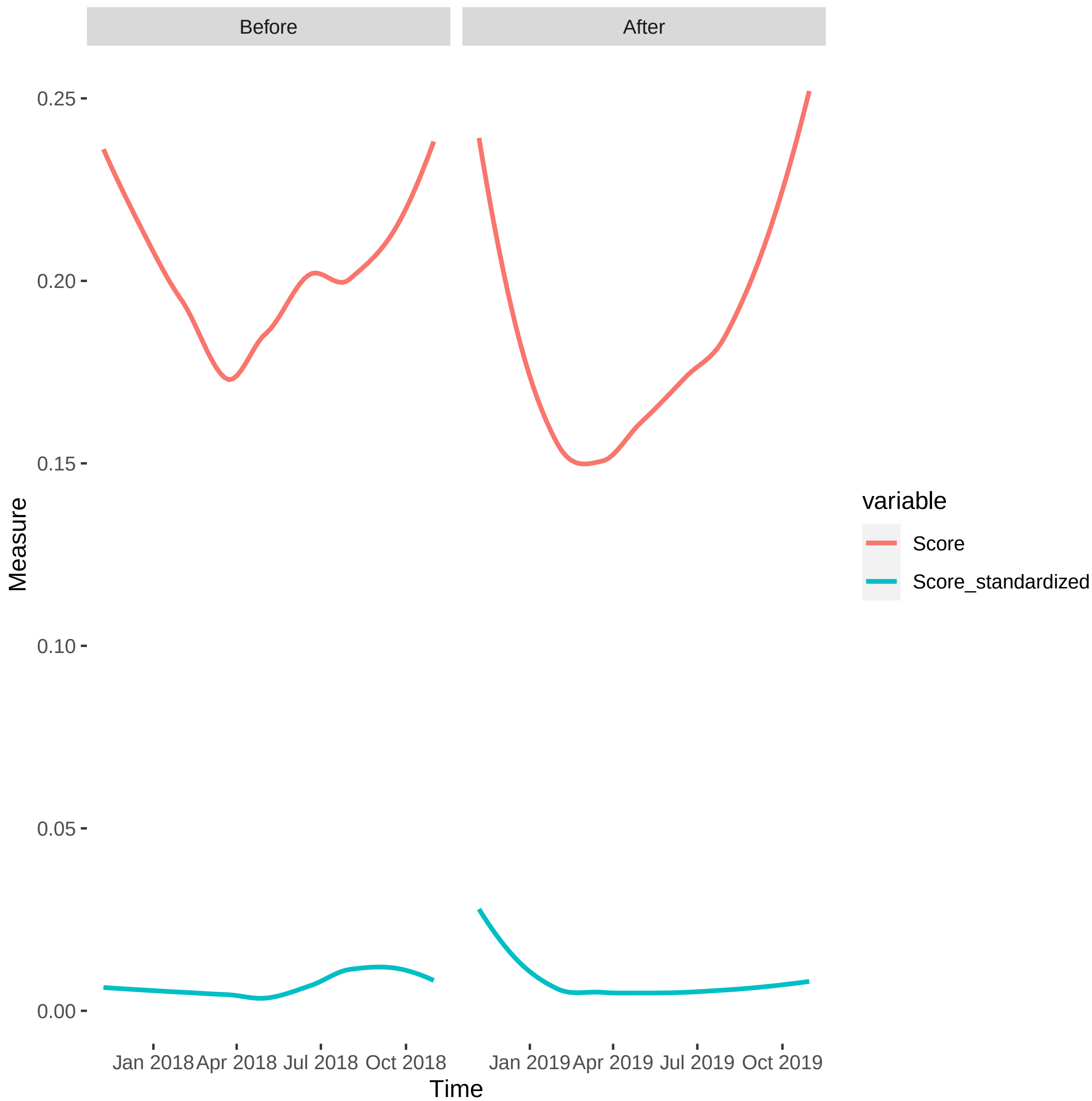
Notes: The top 30 most relevant words used in the first topic in the tweets by Democrats for the month of 7th November- 7th December, 2017 after fitting the LDA model.

Figure 9: Distance metrics over time between Democrat and Republicans



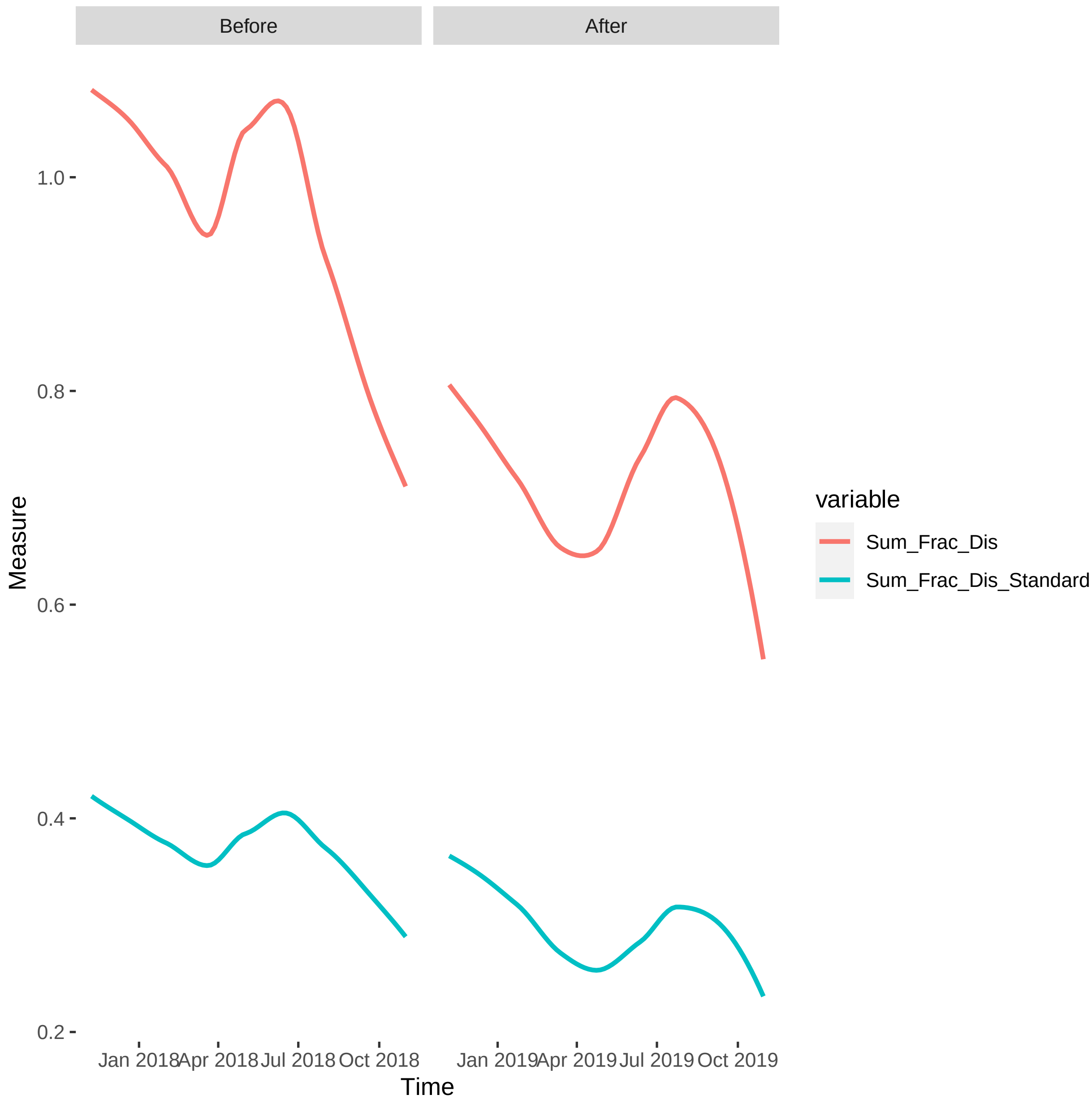
Notes: Euclidean distance between the probability distributions after applying the LDA model as measured by Hellinger, KLD_DR, KLD_RD. Distance between the words used and the list of words used as measured by JD_DR and JD_DLRL respectively.

Figure 10: Euclidean Distance between the dominant topics



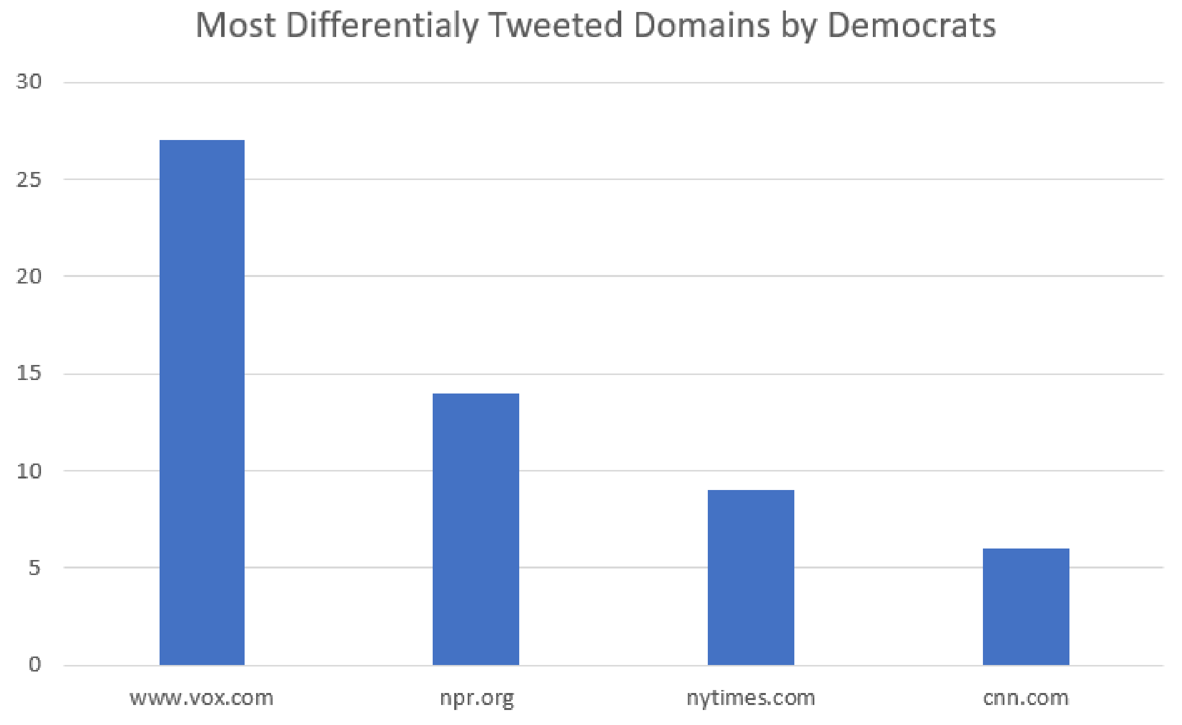
Notes: Euclidean distance between the vector of fractions of the tweets by Republicans and Democrats devoted to the dominant topic calculate at at a weekly basis. $Score$ calculates the euclidean distance whereas $Score_{standardized}$ is $Score$ normalized by the number of topics in that week.

Figure 11: Euclidean Distance for fraction of positive, negative and neutral sentiments in a topic between Democrats and Republicans



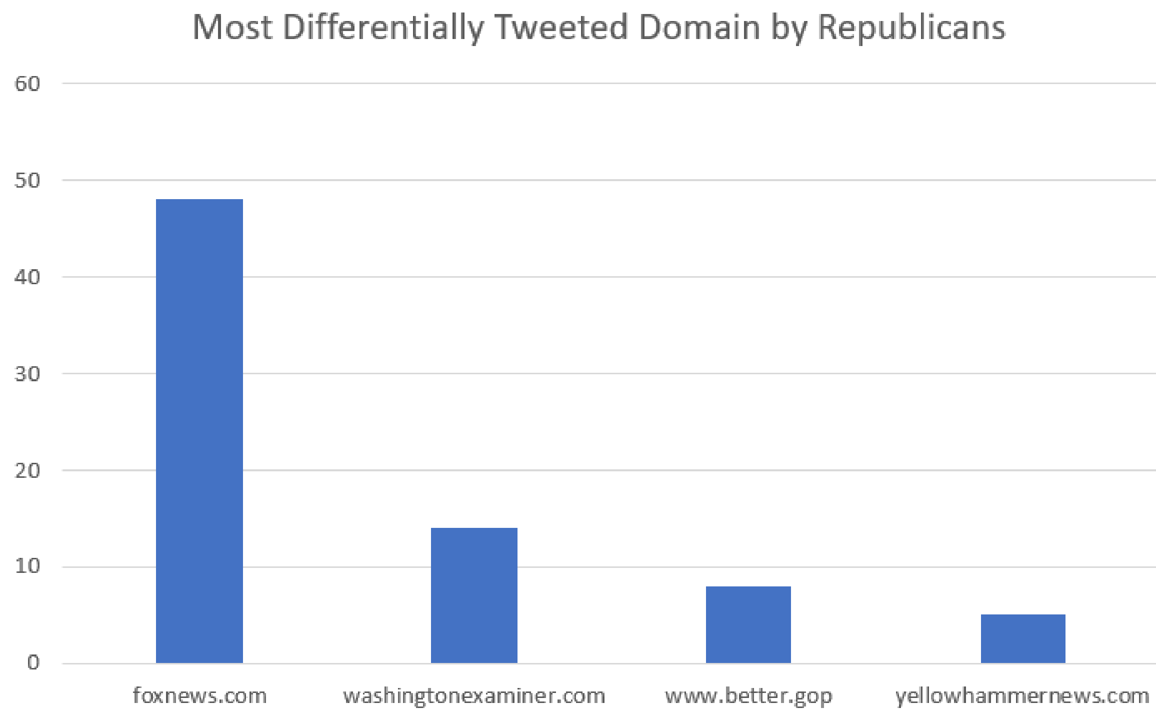
Notes: Euclidean distance between the vector of fractions of tweets by Democrats and Republicans used with a positive, negative and neutral sentiments. *Sum_Dis* is the euclidean distance whereas *Sum_Dis_Standard* is the euclidean distance normalized by the square root of the number of topics each week.

Figure 12: Domains which have tweeted relatively more by Democrats compared to Republicans



Notes: The top 5 domains that Democrats share more relative to Republicans as URLs. The number on the Y-axis shows the number of weeks that a particular domain has emerged the top domain.

Figure 13: Domains which have tweeted relatively more by Democrats comapred to Republicans



Notes: The top 5 domains that Democrats share more relative to Republicans as URLs. The number on the Y-axis shows the number of weeks that a particular domain has emerged the top domain.

Figure 14: Mean ideological score over time computed using Bayesian Ideal Point Estimation.

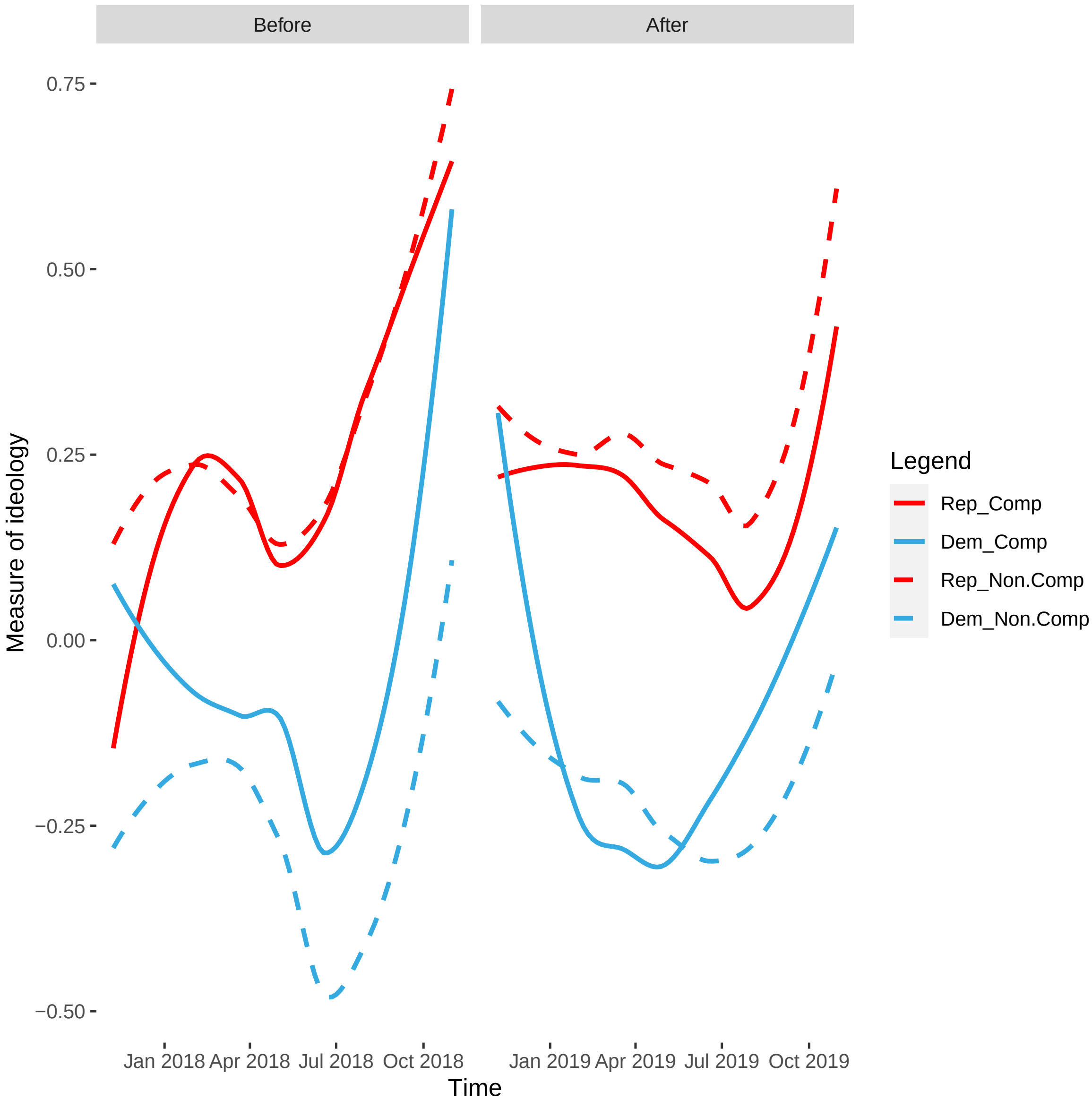


Figure 15: Mean ideological polarization over time computed using Bayesian Ideal Point Estimation.

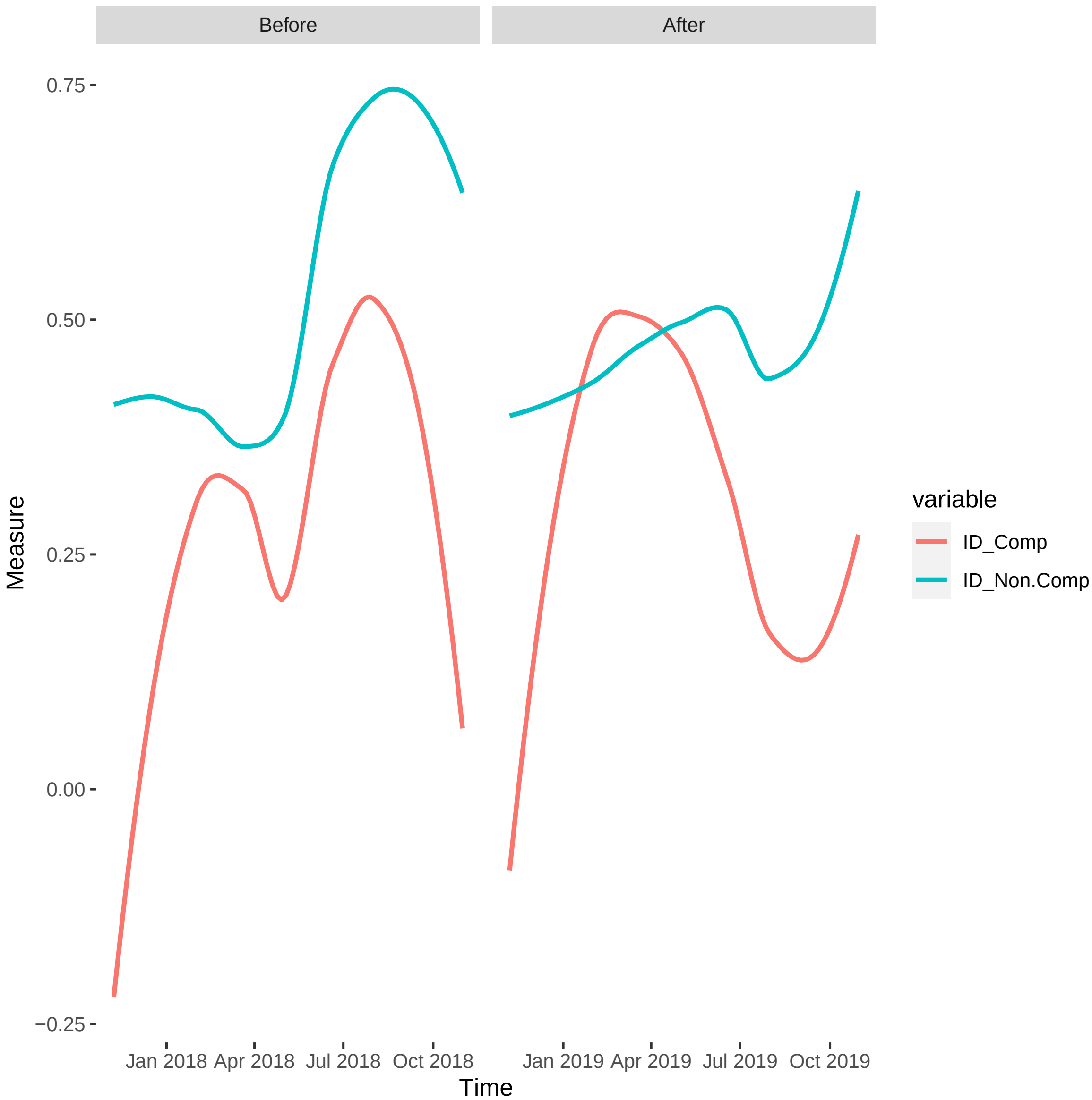
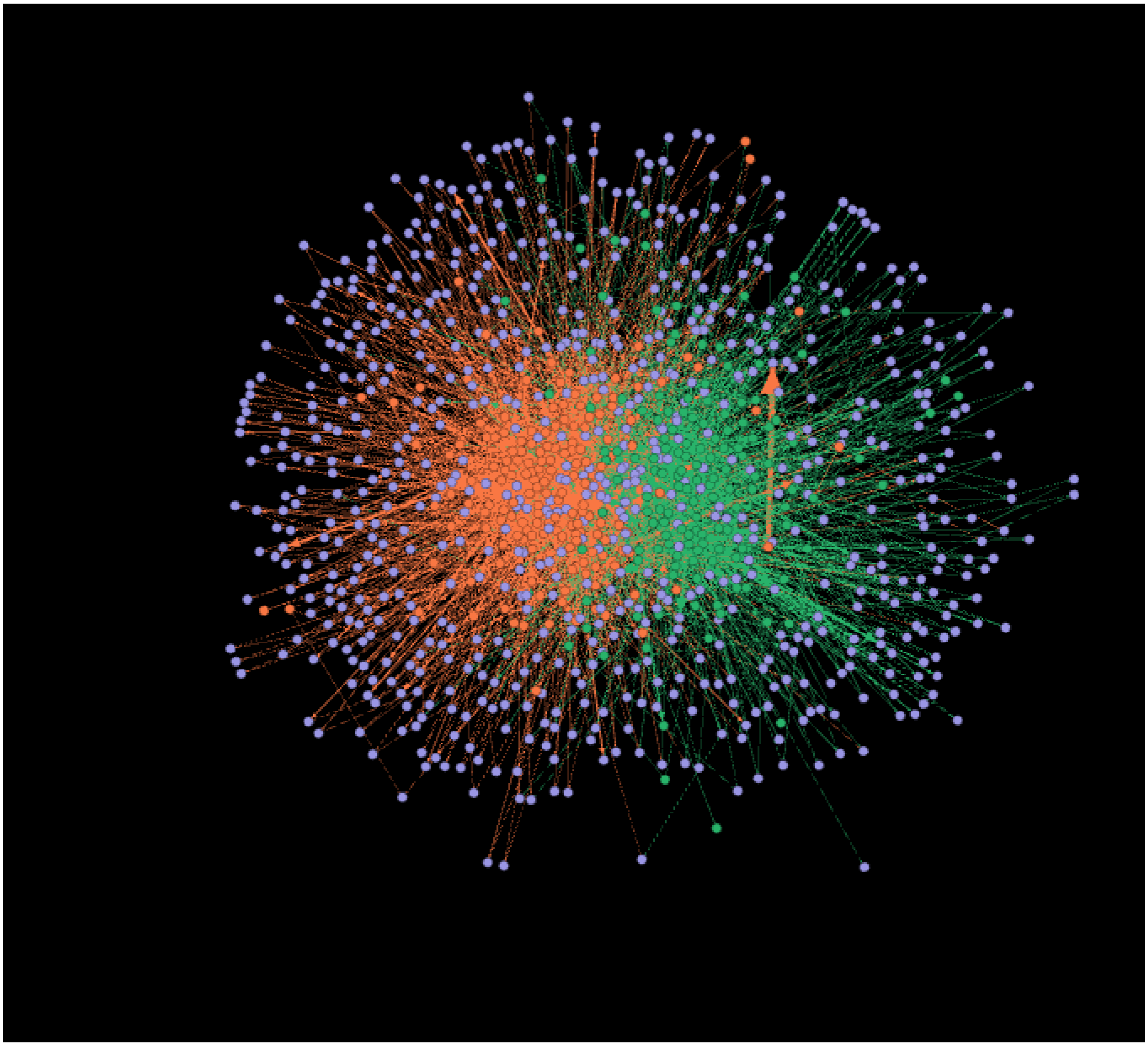
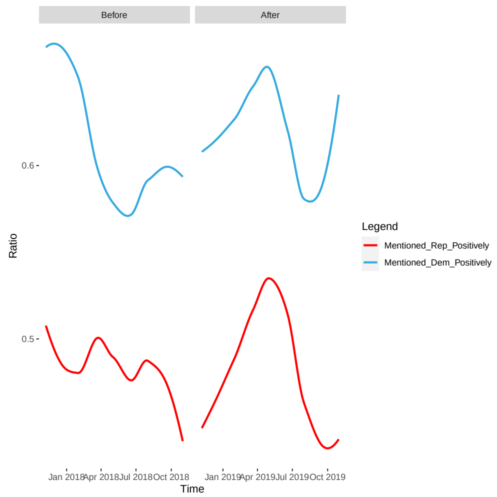


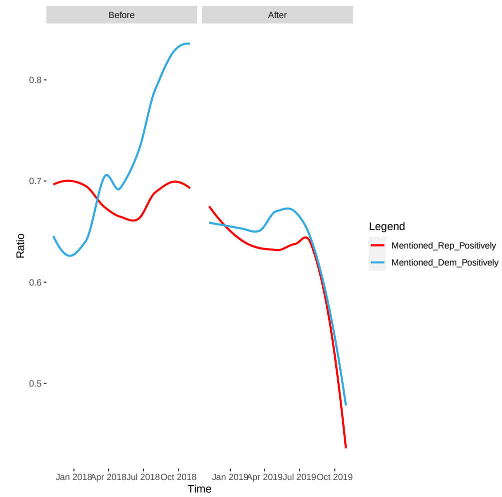
Figure 16: Polarization in the mention network



Notes: the green colored dots represents Republicans, orange represents Democrats, and mauve represents non politicians

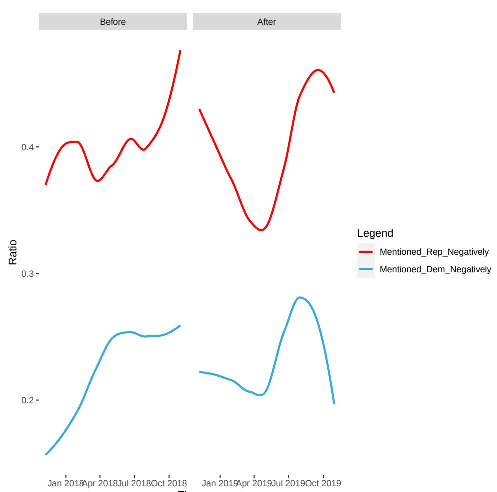


(a) Trend of positive mentions of Republicans by Democrats

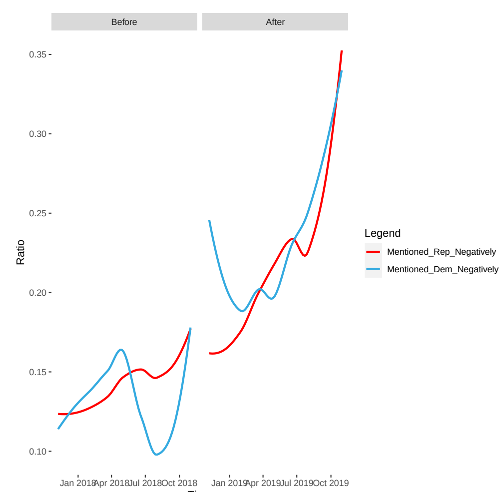


(b) Trend of positive mentions of Democrats by Republicans

Figure 17: Trend of positive mentions by parties over time

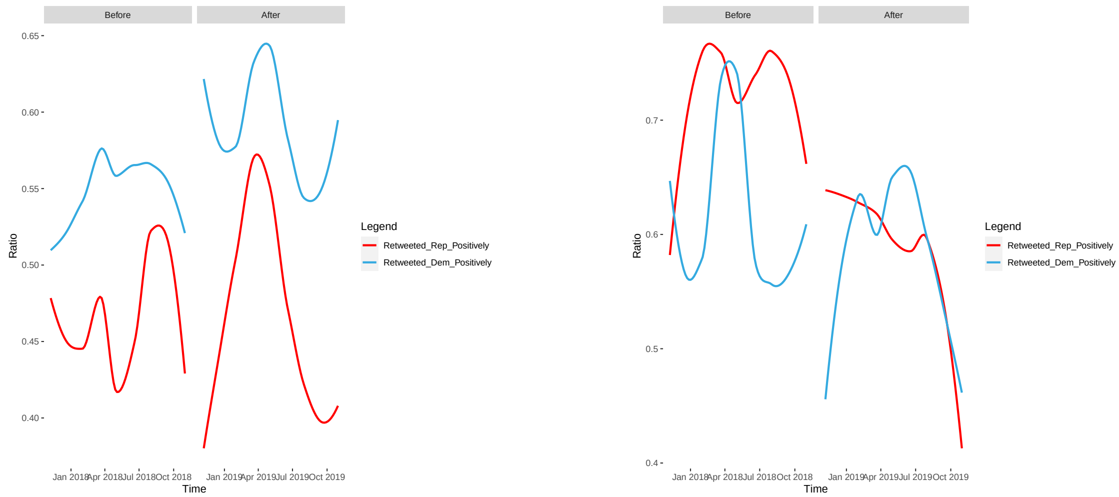


(a) Trend of negative mentions of Republicans by Democrats



(b) Trend of negative mentions of Democrats by Republicans

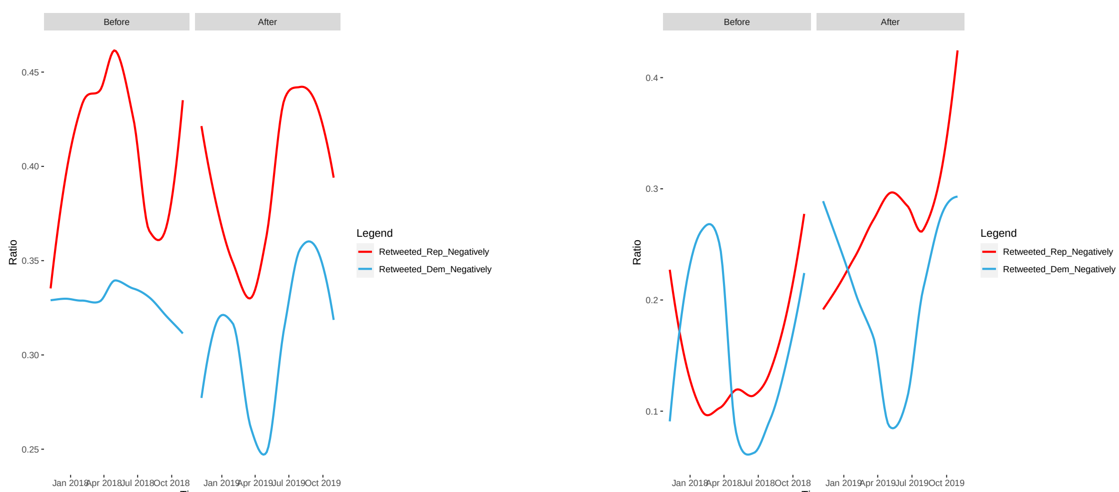
Figure 18: Trend of negative mentions by parties over time



(a) Trend of positive retweets of Republicans by Democrats

(b) Trend of positive retweets of Democrats by Republicans

Figure 19: Trend of positive retweets by parties over time



(a) Trend of negative retweets of Republicans by Democrats

(b) Trend of negative retweets of Democrats by Republicans

Figure 20: Trend of negative retweets by parties over time

8 Tables

Table 1: Top 20 hashtags used by the Democrats and Republicans

Panel A: One week before the election	
Democrats	Republicans
#forthepeople, #vote	#betteroffnow, #jobsreport
#getcovered, *#electionday*	#taxreform, #taxcutsandjobsact
#shirleychisholm, #protectourcare	#halloween, #hurricanemichael
#investigatezinke, #unboughtunbossed	#ms01, #maga
#govote, #latinaequalpayday	#ar3, #happyhalloween
#cultureofcorruption, #latinaequalpay	#veterans, #az05
#showupforshabbat, #midterms2018	#jobs, #ga10
#marianastrong, #openenrollment	#al03, #mobileoffice
#aca, #yutu	#nc06, *#electionday*
#pda40, #goptaxscam	#mi06, #az08

Panel B: One week after the election	
Democrats	Republicans
#veteransday, #getcovered	*#veteransday*, #campfire
#forthepeople, #woolseyfire	*#veterans*, #marinecorpsbirthday
#veterans, #protectmueller	#semperfi, *#veteransday2018*
#veteransday2018, #trump	#ar3, #findyourpark
#enoughisenough, #thousandoaks	#al03, #az05
#mueller, #daca	#ruralbizsummit, #betteroffnow
#counteveryvote, #hillfire	#floridarecount2018, #thankaveteran
#thxbirthcontrol, *#wwi*	#ms01, #ar4
#followthefacts, #yutu	#semperfidelis, *#wwi*
#gunviolence, #diwali	#nationaladoptionmonth, #nc06

Notes: **Panel A** shows the top 20 hashtags used by Republicans and Democrats one week before the election. The common hashtags are italicized and underlined. #electionday is the only common hashtag used by both Democrats and Republicans one week before the election. **Panel B** shows the top 20 hashtags used by Republicans and Democrats one week after the election. The common hashtags are #veteransday, #veterans, #veteransday2018 and #wwi.

Table 2: Regression Discontinuity Estimates for Hashtag Analysis

Panel A: RDD Estimates for Hashtag Similarity

	(1)	(2)	(3)	(4)	(5)
Degree	Hashtag_{10}	Hashtag_{20}	Hashtag_{40}	Hashtag_{50}	Hashtag_{100}
1	1.454	3.049***	3.389**	5.192***	12.048***
	(1.094)	(0.983)	(1.711)	(1.648)	(2.859)
2	3.586*	4.481***	6.278**	8.703***	17.272***
	(1.650)	(1.438)	(2.687)	(2.567)	(3.323)
3	3.296	4.330***	6.753**	9.189***	17.604***
	(1.723)	(1.617)	(3.190)	(2.963)	(4.249)
4	4.128*	4.163**	5.252	3.895	19.551***
	(1.851)	(1.774)	(3.838)	(3.820)	(4.862)
N	104	104	104	104	104

Panel B: RDD Estimates for $Inverse\_Score\_std$ estimates

	(1)	(2)	(3)	(4)	(5)
Degree	$\text{Inv\_Score\_std}_{10}$	$\text{Inv\_Score\_std}_{20}$	$\text{Inv\_Score\_std}_{40}$	$\text{Inv\_Score\_std}_{50}$	$\text{Inv\_Score\_std}_{100}$
1	3.050	6.967	0.074	-0.644	-3.344
	(12.147)	(8.164)	(5.578)	(6.207)	(3.808)
2	0.687	17.341	6.839	5.391	5.076
	(13.902)	(16.117)	(8.481)	(9.381)	(3.732)
3	10.732	19.138	3.968	6.733	11.191
	(24.855)	(18.158)	(9.968)	(10.767)	(9.210)
4	15.846	33.235	10.372	9.757	12.517
	(31.479)	(23.568)	(9.840)	(11.194)	(10.283)
N	104	104	104	104	104

[1] Standard error in parentheses. * denotes 10 percent significance, ** denotes 5 percent significance and *** denotes 1 percent significance.
[2] Hashtag similarity is defined as the number of common hashtags used by Democrats and Republicans conditional on the top hashtags used by them. $Inverse\_Score\_std$ is the inverse of distance between fraction of negative tweets for the Republicans and Democrats out of all the tweets that use a similar hashtag standardized for the number of topics. I perform a non-parametric RDD in Time using a data driven bandwidth selection method. The estimates are reported for 1, 2, 3 and 4 degrees of polynomials.

Table 3: Regression Discontinuity Estimates for Euclidean distance measures between topic distributions and word distributions

Degree	(1) Hellinger	(2) KLD$_{DR}$	(3) KLD$_{RD}$	(4) JD$_{DR}$	(5) JD$_{DLRL}$
1	-0.059	-0.060	-0.070	-0.018	-0.015*
	(0.051)	(0.053)	(0.062)	(0.015)	(0.008)
2	-0.070	-0.072	-0.085	-0.006	-0.012
	(0.062)	(0.067)	(0.077)	(0.028)	(0.011)
3	-0.095	-0.100	-0.108	-0.005	-0.004
	(0.065)	(0.068)	(0.080)	(0.035)	(0.018)
4	-0.081	-0.099	-0.105	-0.010	-0.007
	(0.086)	(0.085)	(0.118)	(0.041)	(0.018)
N	104	104	104	104	104

[1] Standard error in parentheses. * denotes 10 percent significance, ** denotes 5 percent significance and *** denotes 1 percent significance.
[2] $Hellinger$ Distance and the KLD_{DR} and KLD_{RD} measure the distance between the topic distributions after implementing the LDA model on the content of the tweets. The JD_{DR} and JD_{DLRL} measures the distance between the word distributions used by the Democrats and Republicans in their tweets. I perform a non-parametric RDD in Time using a data driven bandwidth selection method. The estimates are reported for 1, 2, 3 and 4 degrees of polynomials.

Table 4: Regression Discontinuity Estimates for Euclidean Distance and Euclidean distance interacted with sentiment for dominant topic

Panel A: RDD Estimates for Euclidean Distance		
	(1)	(2)
Degree	Score	Score_Std
1	-0.011	0.031
	(0.073)	(0.028)
2	-0.009	0.047
	(0.100)	(0.042)
3	-0.010	0.061
	(0.124)	(0.024)
4	0.042	0.075
	(0.170)	(0.078)
N	104	104
Panel B: RDD Estimates for sentiment augmented Euclidean distance		
	(1)	(2)
Degree	Sum_Frac_Dis	Sum_Frac_Dis_Std
1	0.023	0.005
	(0.212)	(0.080)
2	-0.205	-0.018
	(0.402)	(0.102)
3	-0.317	-0.040
	(0.620)	(0.158)
4	-0.269	-0.044
	(0.604)	(0.189)
N	104	104

[1] Standard error in parentheses. * denotes 10 percent significance, ** denotes 5 percent significance and *** denotes 1 percent significance.
[2] *Score* is defined as the euclidean distance between the vectors of fraction of each topic(dominant topic in each tweet) for the Democrats and Republican tweets for each week. $Score_std$ is standardized for the number of topics. Sum_Frac_Dis is the euclidean distance for each topic between Republicans and Democrats, and add the distances for all the topics. $Sum\_Frac\_Dis\_std$ is standardized for the number of topics. I perform a non-parametric RDD in Time using a data driven bandwidth selection method. The estimates are reported for 1, 2, 3 and 4 degrees of polynomials.

Table 5: Regression Discontinuity Estimates for Bayesian Ideal Point(BIP) Estimates for politicians and Ideological Difference in competitive and non-competitive districts

Panel A: RDD Estimates for BIP Estimates for politicians

	(1)	(2)	(3)	(4)
Degree	Rep_Comp	Dem_Comp	Rep_Non-Comp	Dem_Non-Comp
1	-0.308	-0.264	-0.428	-0.122
	(0.399)	(0.450)	(0.305)	(0.346)
2	-0.259	0.268	-0.172	0.232
	(0.474)	(0.920)	(0.524)	(0.709)
3	0.102	0.297	0.236	0.245
	(0.746)	(1.032)	(0.735)	(0.684)
4	0.104	-0.109	0.376	0.142
	(0.841)	(1.683)	(0.853)	(1.100)
N	104	104	104	104

Panel B: RDD Estimates for Ideological Difference

	(1)	(2)
Degree	IdeologicalDiff_Comp	IdeologicalDiff_Non-Comp
1	-0.281	-0.278
	(0.295)	(0.189)
2	-0.343	-0.107
	0.471	(0.227)
3	-0.336	-0.008
	(0.580)	(0.218)
4	0.772	0.148
	(0.820)	(0.296)
N	104	104

[1] Standard error in parentheses. * denotes 10 percent significance, ** denotes 5 percent significance and *** denotes 1 percent significance.
[2] Footnote 2: In **Panel A** *Rep_Comp* and *Dem_Comp* shows the mean ideological score for all Republicans and Democrats respectively in competitive districts.$Rep\_Non-Comp$ and $Dem\_Non-Comp$ shows the mean ideological score for all Republicans and Democrats respectively in non-competitive districts. In **Panel B** $IdeologicalDiff\_Comp$ and $IDeologicalDiff\_Non-Comp$ shows the difference between mean Republican and Democrat ideological scores in competitive and non-competitive districts respectively. I perform a non-parametric RDD in Time using a data driven bandwidth selection method. The estimates are reported for 1, 2, 3 and 4 degrees of polynomials.

Table 6: Regression Discontinuity Estimates for Mention Network Analysis

Panel A: RDD Estimates for Mentions by Democrats

	(1)	(2)	(3)	(4)
Degree	Dem_Positively	Rep_Positively	Dem_Negatively	Rep_Negatively
1	0.042	-0.012	-0.039	-0.043
	(0.063)	(0.029)	(0.022)	(0.038)
2	0.063	-0.004	-0.016	-0.056
	(0.069)	(0.036)	(0.042)	(0.043)
3	-0.105	-0.002	-0.008	-0.041
	(0.116)	(0.043)	(0.047)	(0.050)
4	-0.247	0.062	-0.024	0.044
	(0.179)	(0.068)	(0.053)	(0.068)
N	104	104	104	104

Panel B: RDD Estimates of mentions by Republicans

	(1)	(2)	(3)	(4)
Degree	Dem_Positively	Rep_Positively	Dem_Negatively	Rep_Negatively
1	-0.182	-0.012	0.004	-0.026
	(0.420)	(0.060)	(0.084)	(0.060)
2	-0.108	0.037	0.041***	-0.028
	(0.671)	(0.096)	(0.042)	(0.077)
3	0.046	-0.069	0.048***	-0.031
	(0.826)	(0.098)	(0.134)	(0.106)
4	1.147	-0.114	0.052***	-0.030
	(1.271)	(0.144)	(0.011)	(0.116)
N	104	104	104	104

[1] Standard error in parentheses. * denotes 10 percent significance, ** denotes 5 percent significance and *** denotes 1 percent significance.
[2] In **Panel A** *Dem_Positively* and *Dem_Negatively* are shares of how many times Democrats are mentioned positively and negatively relative to all mentions of Democrats by the Democrat politicians . *Rep_Positively* and *Dem_Negatively* are shares of how many times Republicans are mentioned positively and negatively relative to all mentions of Republicans by the Democrat politicians.In **Panel B** *Dem_Positively* and *Dem_Negatively* are shares of how many times Democrats are mentioned positively and negatively relative to all mentions of Democrats by the Republican politicians. *Rep_Positively* and *Dem_Negatively* are shares of how many times Republicans are mentioned positively and negatively relative to all mentions of Republicans by the Republican politicians. I perform a non-parametric RDD in Time using a data driven bandwidth selection method. The estimates are reported for 1, 2, 3 and 4 degrees of polynomials.

Table 7: Regression Discontinuity Estimates for Retweet Network Analysis

Panel A: RDD Estimates for Retweets by Democrats

	(1)	(2)	(3)	(4)
Degree	Dem_Positively	Rep_Positively	Dem_Negatively	Rep_Negatively
1	0.093	0.167	-0.146	-0.063
	(0.135)	(0.158)	(0.094)	(0.036)
2	0.125	0.216	-0.149	-0.094**
	(0.162)	(0.205)	(0.096)	(0.048)
3	0.166	0.297	-0.202	-0.108**
	(0.215)	(0.308)	(0.114)	(0.054)
4	0.171	0.421	-0.200	-0.122**
	(0.255)	(0.419)	(0.156)	(0.056)
N	104	104	104	104

Panel B: RDD Estimates for Retweets by Republicans

	(1)	(2)	(3)	(4)
Degree	Dem_Positively	Rep_Positively	Dem_Negatively	Rep_Negatively
1	0.031	-0.050	0.019	-0.090
	(0.088)	(0.077)	(0.029)	(0.080)
2	0.100	-0.096	0.013	-0.094
	(0.141)	(0.102)	(0.038)	(0.092)
3	0.155	-0.183	-0.019	-0.015
	(0.181)	(0.138)	(0.057)	(0.100)
4	0.227	-0.221	-0.111	0.007
	(0.249)	(0.146)	(0.090)	(0.110)
N	104	104	104	104

[1] Standard error in parentheses. * denotes 10 percent significance, ** denotes 5 percent significance and *** denotes 1 percent significance.
[2] In **Panel A** *Dem_Positively* and *Dem_Negatively* are shares of how many times Democrats are retweeted positively and negatively relative to all retweets of Democrats by the Democrat politicians . *Rep_Positively* and *Dem_Negatively* are shares of how many times Republicans are retweeted positively and negatively relative to all retweets of Republicans by the Democrat politicians.In **Panel B** *Dem_Positively* and *Dem_Negatively* are shares of how many times Democrats are retweeted positively and negatively relative to all retweets of Democrats by the Republican politicians. *Rep_Positively* and *Dem_Negatively* are shares of how many times Republicans are retweeted positively and negatively relative to all retweets of Republicans by the Republican politicians. I perform a non-parametric RDD in Time using a data driven bandwidth selection method. The estimates are reported for 1, 2, 3 and 4 degrees of polynomials.

Table 8: OLS Sub-sample Estimates for Hashtag Analysis

Panel A: Estimates for Hashtag Similarity

Metric	(1) Before	(2) After
$Hashtag_{10}$	-0.179	-0.179
	(0.1342)	(0.2712)
$Hashtag_{20}$	-0.333	0.0595
	(0.291)	(0.2576)
$Hashtag_{40}$	-0.536	-0.083
	(0.3323)	(0.381)
$Hashtag_{50}$	-0.8333**	0.0119
	(0.4123)	(0.4623)
$Hashtag_{100}$	-1.464**	-0.869
	(0.701)	(0.6209)
N	8	8

Panel B: Estimates of for $Inverse\_Score\_std$ **estimates**

Metric	(1) Before	(2) After
Inv_Score_10_std	-2.279	-6.285
	(1.543)	(4.426)
Inv_Score_20_std	-0.514	-2.979
	(0.8651)	(1.087)
Inv_Score_40_std	-0.368	-0.348
	(0.6268)	(0.3469)
Inv_Score_50_std	-0.262	-0.587
	(0.7059)	(0.3062)
Inv_Score_100_std	-0.288	-0.486
	(0.1866)	(0.1836)
N	8	8

[1] Standard error in parentheses. * denotes 10 percent significance, ** denotes 5 percent significance and *** denotes 1 percent significance.
[2] Hashtag similarity is defined as the number of common hashtags used by Democrats and Republicans conditional on the top hashtags used by them. $Inverse\_Score\_std$ is the inverse of distance between fraction of negative tweets for the Republicans and Democrats out of all the tweets that use a similar hashtag standardized for the number of topics. The Before column shows the estimates for 8 weeks before the election whereas the After column shows the estimates for 8 weeks after the election.

Table 9: OLS Sub-sample estimates for Euclidean distances for topic distributions and word distributions

Metric	(1) Before	(2) After
Hellinger	0.0055	-0.008
	(0.0052)	(0.0062)
KLD_DR	0.0064	-0.004
	(0.0057)	(0.0035)
KLD_RD	0.008	-0.004
	(0.0064)	(0.0035)
JD_DR	-0.006	0.0028
	(0.0017)	(0.0031)
JD_DLRL	-0.003	0.0008
	(0.0013)	(0.0028)
N	8	8

[1] Standard error in parentheses. * denotes 10 percent significance, ** denotes 5 percent significance and *** denotes 1 percent significance.
[2] *Hellinger* Distance and the KLD_{DR} and KLD_{RD} measure the distance between the topic distributions after implementing the LDA model on the content of the tweets. The JD_{DR} and JD_{DLRL} measures the distance between the word distributions used by the Democrats and Republicans in their tweets. The Before column shows the estimates for 8 weeks before the election whereas the After column shows the estimates for 8 weeks after the election.

Table 10: OLS Sub-sample Estimates for Euclidean Distance and Euclidean distance interacted with sentiment for dominant topic

Panel A: Estimates for Euclidean Distance		
	(1)	(2)
Metric	Before	After
Score	0.0046	-0.025
	(0.0101)	(0.0128)
Score$_{std}$	-4E-04	-0.006
	(0.0011)	(0.0037)
N	8	8

Panel B: Estimates for sentiment augmented Euclidean distance		
	(1)	(2)
Metric	Before	After
Sum_Frac_Dis	0.0111	0.0262
	(0.0557)	(0.0259)
Sum_Frac_Dis$_{std}$	-0.002	0.0123
	(0.014)	(0.0115)
N	8	8

[1] Standard error in parentheses. * denotes 10 percent significance, ** denotes 5 percent significance and *** denotes 1 percent significance.
[2]Score is defined as the euclidean distance between the vectors of fraction of each topic(dominant topic in each tweet) for the Democrats and Republican tweets for each week.*Score_std* is standardized for the number of topics.*Sum_Frac_Dis* is the euclidean distance for each topic between Republicans and Democrats, and add the distances for all the topics. *Sum_Frac_Dis_std* is standardized for the number of topics. The Before column shows the estimates for 8 weeks before the election whereas the After column shows the estimates for 8 weeks after the election.

Table 11: OLS Sub-sample Estimates for Bayesian Ideal Point(BIP) Estimates for politicians and Ideological Difference in competitive and non-competitive districts

Panel A: Estimates for BIP Estimates for politicians

	(1)	(2)
Metric	Before	After
Rep_Comp	-0.055	-0.021
	(0.0756)	(0.0732)
Dem_Comp	-0.008	-0.107
	(0.1056)	(0.076)
Rep_Non-Comp	-0.04	-0.065
	(0.0664)	(0.0813)
Dem_Non-Com	-0.039	-0.04
	(0.0811)	(0.0626)
N	8	8

Panel B: Estimates for Ideological Difference

	(1)	(2)
Metric	Before	After
ID_Comp	-0.047	0.0865
	(0.0363)	(0.0511)
ID_Non-Comp	-0.001	-0.025
	(0.0375)	(0.0313)
N	8	8

[1] Standard error in parentheses. * denotes 10 percent significance, ** denotes 5 percent significance and *** denotes 1 percent significance.
[2] In **Panel A** *Rep_Comp* and *Dem_Comp* shows the mean ideological score for all Republicans and Democrats respectively in competitive districts.$Rep\_Non-Comp$ and $Dem\_Non-Comp$ shows the mean ideological score for all Republicans and Democrats respectively in non-competitive districts. In **Panel B** $IdeologicalDiff\_Comp$ and $IdeologicalDiff\_Non-Comp$ shows the difference between mean Republican and Democrat ideological scores in competitive and non-competitive districts respectively. The Before column shows the estimates for 8 weeks before the election whereas the After column shows the estimates for 8 weeks after the election.

Table 12: OLS Sub-sample Estimates for Mention Network Analysis

Panel A: Estimates for Mentions by Democrats

Metric	(1) Before	(2) After
Dem_Positively	0.0054	-0.009
	(0.006)	(0.0116)
Rep_Positively	-0.003	0.0006
	(0.0117)	(0.0133)
Dem_Negatively	-0.006	0.0134
	(0.0051)	(0.0088)
Rep_Negatively	0.0134	-5E-04
	(0.0127)	(0.0134)
N	8	8

Panel B: Estimates of mentions by Republicans

Metric	(1) Before	(2) After
Dem_Positively	-1E-04	-0.0225
	(0.008)	(0.0245)
Rep_Positively	-0.009	-1E-04
	(0.0123)	(0.0062)
Dem_Negatively	0.0087	-0.036
	(0.0069)	(0.0198)
Rep_Negatively	0.0077	-0.012
	(0.0073)	(0.005)
N	8	8

[1] Standard error in parentheses. * denotes 10 percent significance, ** denotes 5 percent significance and *** denotes 1 percent significance.
[2] In **Panel A** *Dem_Positively* and *Dem_Negatively* are shares of how many times Democrats are mentioned positively and negatively relative to all mentions of Democrats by the Democrat politicians . *Rep_Positively* and *Dem_Negatively* are shares of how many times Republicans are mentioned positively and negatively relative to all mentions of Republicans by the Democrat politicians.In **Panel B** *Dem_Positively* and *Dem_Negatively* are shares of how many times Democrats are mentioned positively and negatively relative to all mentions of Democrats by the Republican politicians. *Rep_Positively* and *Dem_Negatively* are shares of how many times Republicans are mentioned positively and negatively relative to all mentions of Republicans by the Republican politicians. The Before column shows the estimates for 8 weeks before the election whereas the After column shows the estimates for 8 weeks after the election.

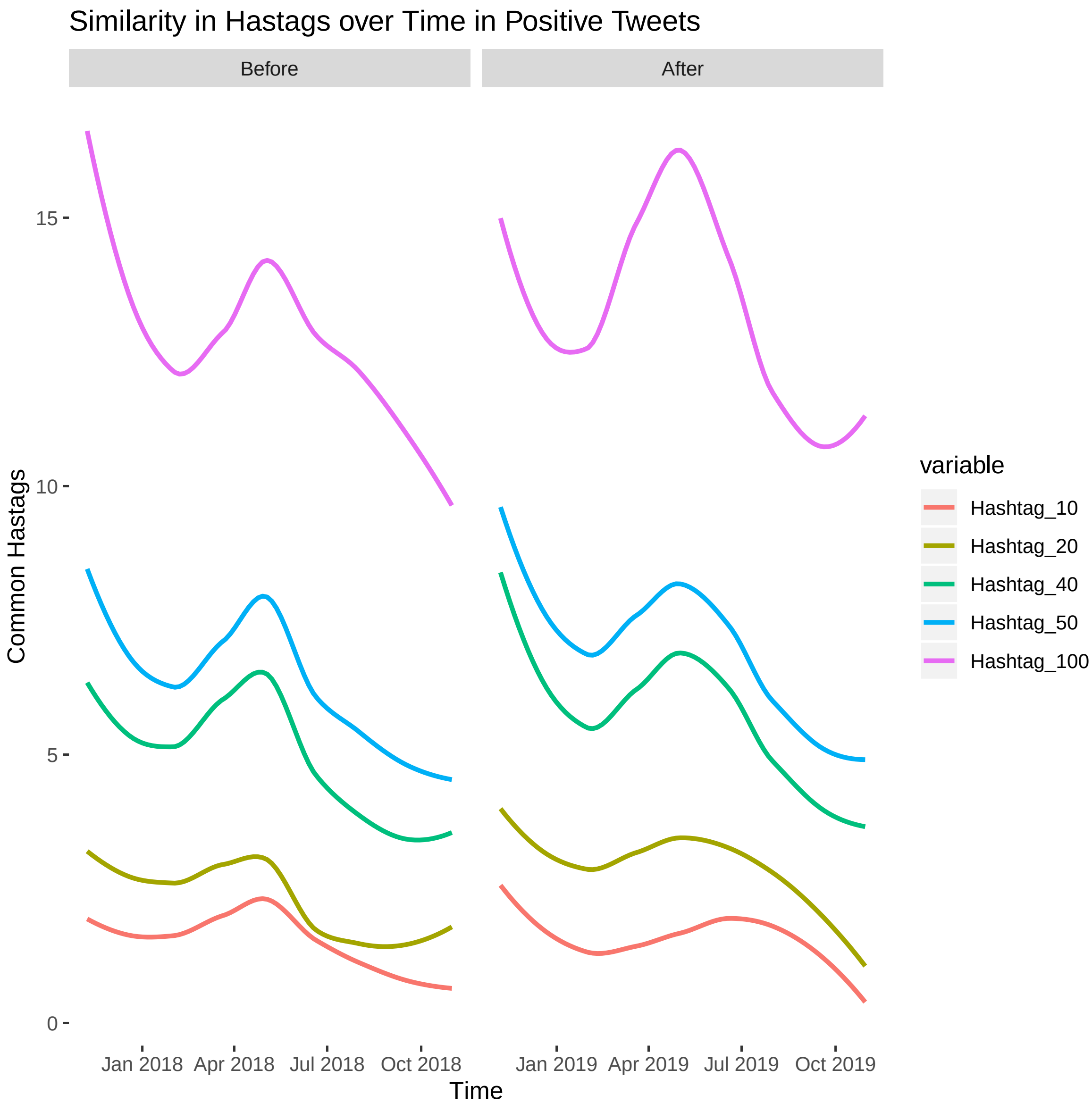
Table 13: OLS Sub-sample Estimates for Retweet Network Analysis

Panel A: Estimates for Retweets by Democrats		
	(1)	(2)
Metric	Before	After
Dem_Positively	-0.017	0.0371
	(0.019)	(0.0117)
Rep_Positively	-0.0016	-0.016
	(0.0392)	(0.0572)
Dem_Negatively	0.0041	0.0329
	(0.014)	(0.0113)
Rep_Negatively	-0.014	0.0198
	(0.0524)	(0.0566)
N	8	8
Panel B: Estimates for Retweets by Republicans		
	(1)	(2)
Metric	Before	After
Dem_Positively	0.08036	-.005952
	(0.0508)	(0.069501)
Rep_Positively	-0.02136	0.003568
	(0.01495)	(0.013303)
Dem_Negatively	0.006	-0.015
	(0.0431)	(0.0682)
Rep_Negatively	-0.001	0.0145
	(0.0191)	(0.017)
	8	8

[1] Standard error in parentheses. * denotes 10 percent significance, ** denotes 5 percent significance and *** denotes 1 percent significance.
[2] In **Panel A** *Dem_Positively* and *Dem_Negatively* are shares of how many times Democrats are retweeted positively and negatively relative to all retweets of Democrats by the Democrat politicians . *Rep_Positively* and *Dem_Negatively* are shares of how many times Republicans are retweeted positively and negatively relative to all retweets of Republicans by the Democrat politicians.In **Panel B** *Dem_Positively* and *Dem_Negatively* are shares of how many times Democrats are retweeted positively and negatively relative to all retweets of Democrats by the Republican politicians. *Rep_Positively* and *Dem_Negatively* are shares of how many times Republicans are retweeted positively and negatively relative to all retweets of Republicans by the Republican politicians. I perform a non-parametric RDD in Time using a data driven bandwidth selection method. The estimates are reported for 1, 2, 3 and 4 degrees of polynomials

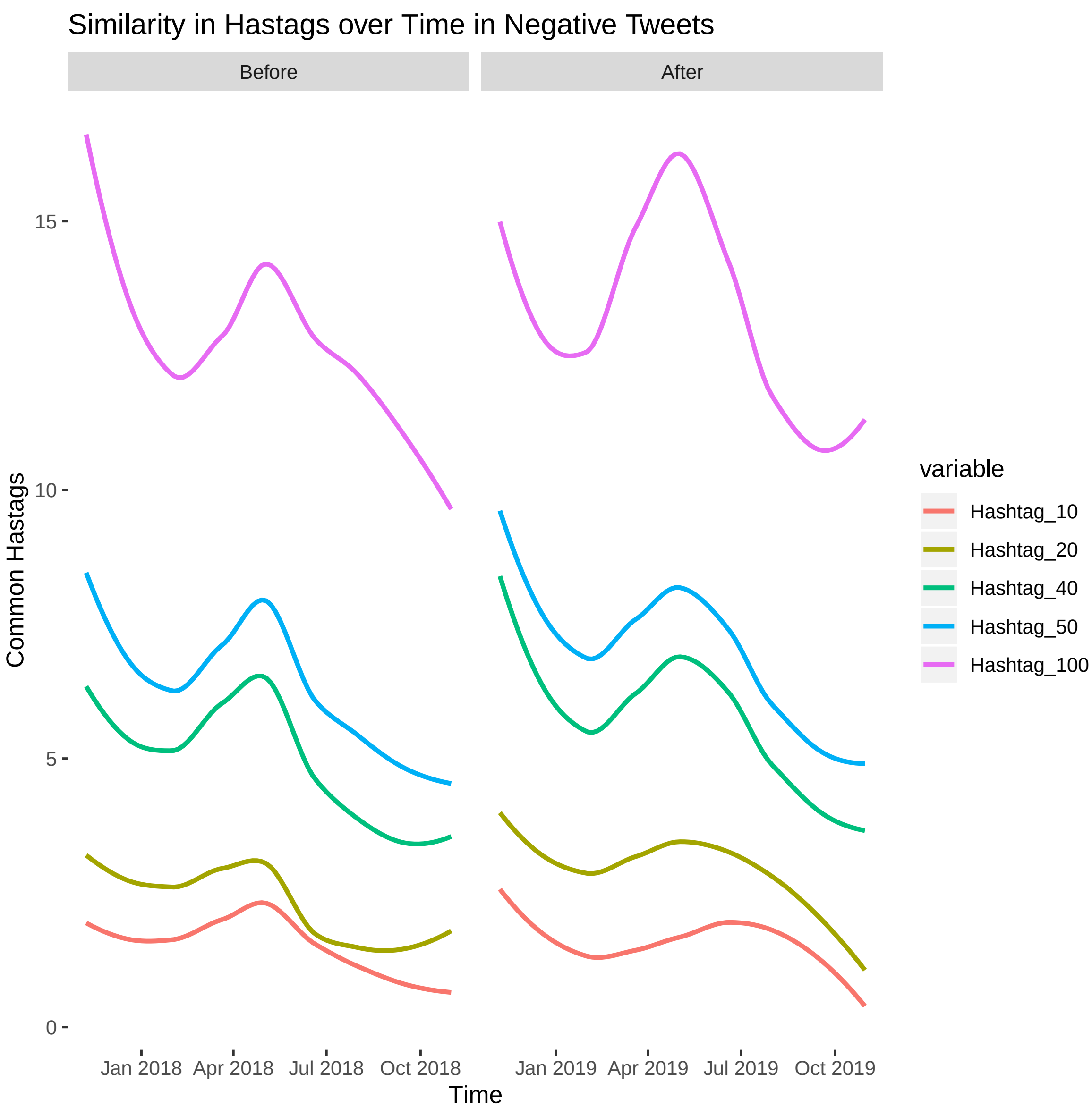
9 Appendix

Figure 21



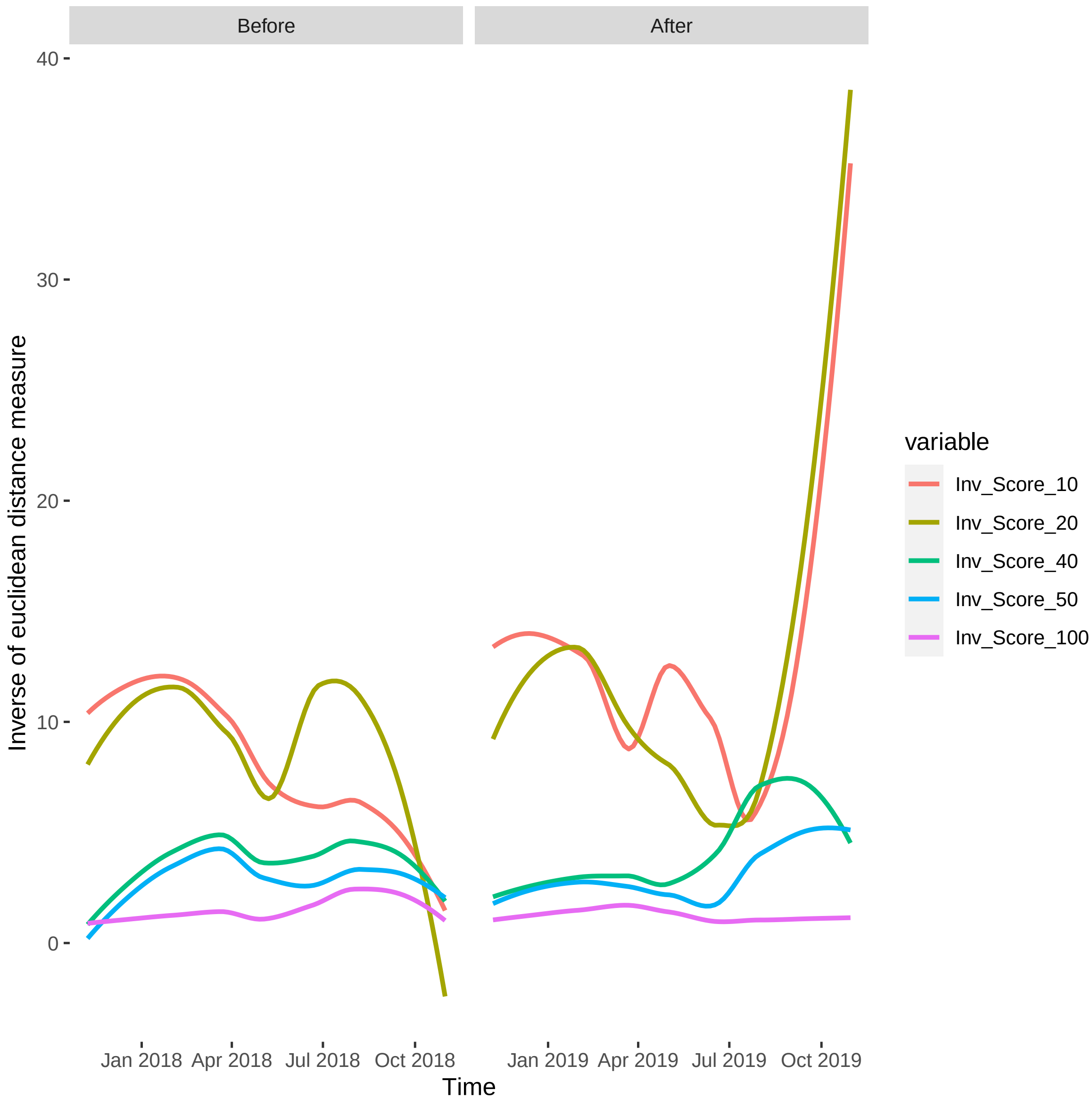
Notes: Similarity in hashtags were computed for only positive tweets, which were found by running the Sentiment Vader package on the tweets

Figure 22



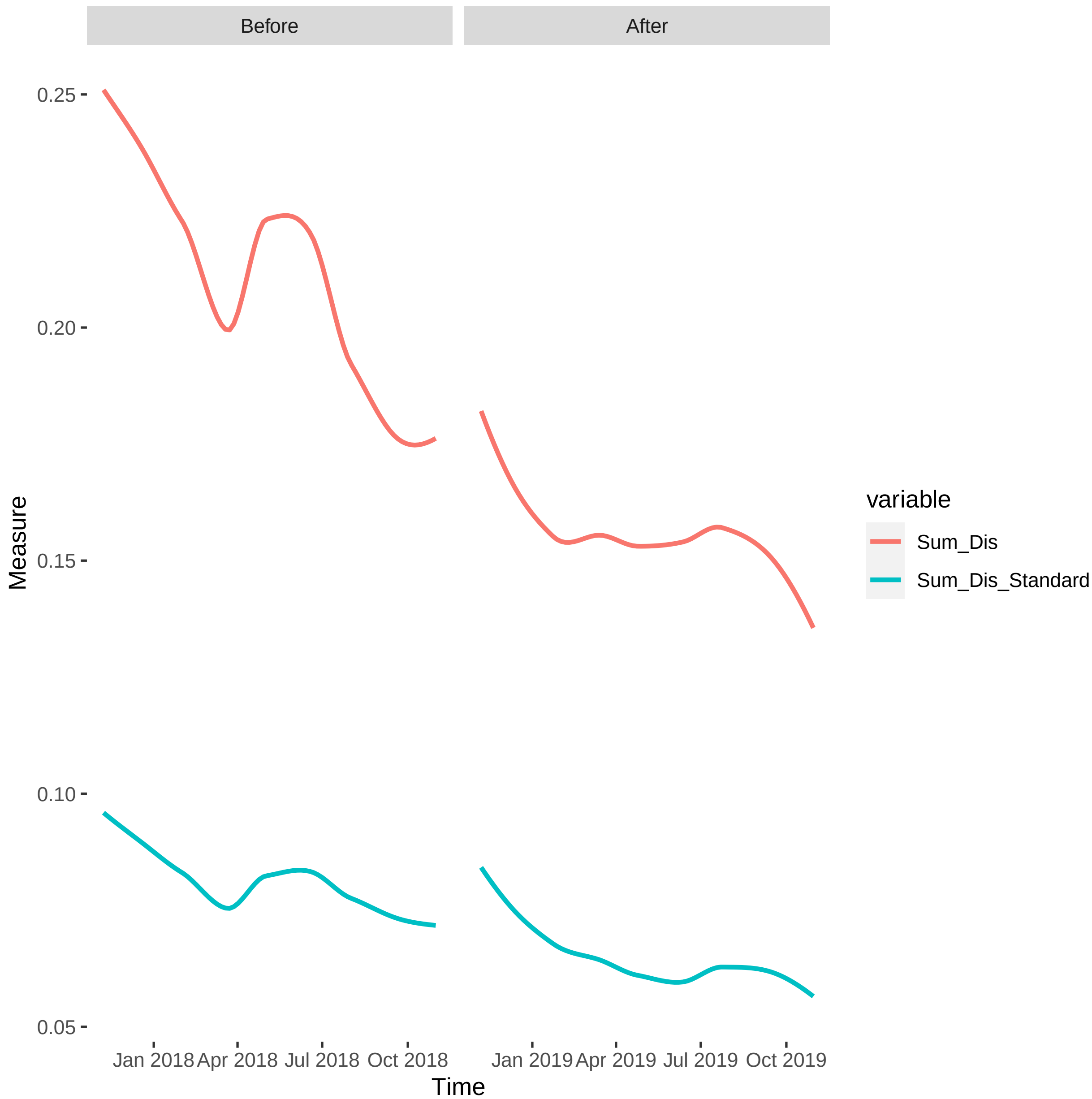
Notes: Similarity in hashtags were computed for only negative tweets, which were found by running the Sentiment Vader package on the tweets

Figure 23: Trend of inverse of euclidean distance of hashtags interacted with sentiments



Notes: Similarity in hashtags were computed for only negative tweets, which were found by running the Sentiment Vader package on the tweets

Figure 24: Euclidean Distance for intensity of positive, negative and neutral sentiments in a topic between Democrats and Republicans



Notes: Euclidean distance between the vector of fractions of tweets by Democrats and Republicans used with a positive, negative and neutral sentiments. *Sum_Dis* is the euclidean distance whereas *Sum_Dis_Standard* is the euclidean distance normalized by the square root of the number of topics each week.