The Impact of Twitter on Political Participation *

Tiziano Rotesi
University of Lausanne

Abstract

What is the effect of Twitter on political participation? I study how the spread of this social network has affected voting behavior and donations to politicians during the 2008, 2012, and 2016 US presidential elections. I construct a novel measure of Twitter penetration using a dataset containing around 6.8 million Twitter accounts matched with locations in the United States. To address endogeneity in the diffusion of Twitter, I exploit variation in the popularity of sport teams that have signed new players with Twitter accounts, thus making the social network more interesting for their fans. My instrumental variables estimates show significant effects of Twitter on political outcomes, which are differential across parties: the Democratic Party is penalized in terms of votes, while Republicans tend to receive more donations. I download and categorize around 150 million tweets written by users, complemented with survey data, to explore the underlying mechanisms. First, I show that Twitter reduces voters' information about politics and increases political polarization. Second, I show that the majority of users write about sports or entertainment and ignore politics for most of the year. Peaks in interest happen only during presidential debates, when both the quantity of partisan tweets and the average sentiment favor the Republican Party.

*Department of Economics, HEC HEC University of Lausanne. tiziano.rotesi@unil.ch.
I am grateful to Eliana La Ferrara and Guido Tabellini for their precious guidance on this paper. I also thank Alexsandros Cavgias, Gracia Cecere, Ruben Durante, Elena Esposito, Oliver Falck, Tommaso Giommoni, Pamela Giustelli, Massimo Morelli, Salvatore Nunnari, Santiago Perez-Vincent, Alessandro Saia, Mathias Thoenig, Diego Ubfal, as well as participants at the 2018 Economics of Digitization Conference at Telecom ParisTech, the 2018 EEA-ESEM at Cologne, the Transatlantic Doctoral Conference 2017 at the London Business School, and internal seminars at Bocconi University for helpful comments and suggestions.

1 Introduction

In the past decade, the Internet has undergone significant changes in the way in which individuals communicate and interact. With the rise of Internet penetration, the use of the Web has transitioned to a model in which users are the primary source of content and no longer play a passive role. This transition has led to the possibility for the platforms driving this change to influence not only the way in which individuals gather information, but also the way in which they are influenced by one another. Twitter[1] in particular has gained a relevant spot in the public debate so that it is now common to see news articles featuring the latest tweets from Congress representatives or users' reactions to them. Additionally, this platform has gained attention as one of the factors behind Barack Obama's success in 2008,[2] the protest that took place during the Arab Spring[3] or the events of January 6th 2021. While the role played by Twitter in shaping political outcomes has been extensively debated, little if any rigorous evidence exists to isolate its causal effect.

This paper presents causal evidence on the impact of Twitter on political participation in the US during the 2008, 2012, and 2016 presidential elections. The study investigates the effect of Twitter on electoral turnout, donations received by the Democratic and Republican parties, and the vote shares of these parties. Additionally, the paper employs machine learning techniques to analyze the characteristics of Twitter users and the content of their tweets in order to understand the mechanisms behind the results.

Ex ante it is difficult to predict the impact that Twitter had on politics, as several contrasting forces may be at play. First, this platform affects the amount of information available to users, though the direction is unclear. For example, Twitter could enlarge the set of entertainment opportunities already available and thus crowd out more informative media as online newspapers, in a way that is similar to what has been suggested for traditional media.[4] At the same time, through the network of contacts, users could discover pieces of information that they would have ignored otherwise, possibly becoming more knowledgeable about politics.[5] A second dimension that represents a strong novelty with respect to traditional media is the social interaction between users. Social media and Twitter in particular are characterized by their ability to foster interaction, making individuals part of a public debate that

[1]Twitter is one of the most popular microblogging platforms. It allows users to publish short public messages, called "tweets", that anyone can read, comment on, and share with others. More information about this social network is presented in Appendix A.

[2]Larcinese and Miner (2017) study how the internet affected the 2008 Presidential Elections give a description of how Obama's campaign made massive use of social media.

[3]Acemoglu et al.(2018) find that peaks in activity on Twitter could be used to predict protests in Tahrir Square.

[4]Gentzkow (2006) studies the effect of TV on political attitudes and suggests that the diffusion of TV may have crowded out other media like newspapers.

[5]See for example Fletcher and Nielsen (2017).

would have been hardly accessible otherwise. This could make users more politically engaged, especially in areas and at times characterized by greater animosity in the political debate. Yet, there is a concern that this interaction takes place predominantly among like-minded users, potentially leading to ideological segregation and therefore viewpoints that are harder to change.[6] Another concern is related with the presence of partisan propaganda. Social media could allow politically active participants, either independent or directly connected to political organizations, to exert influence on their contacts.[7]

To conduct my analysis I need to overcome a key limitation: Twitter does not provide any information on the number of accounts created *by region*. I thus develop a novel measure of Twitter penetration across Designated Market Areas (DMA) by matching accounts with counties, using location data provided by the users. In this way, I obtain a panel measure of the number of accounts created in each DMA from 2007 to 2016.

I then propose a novel identification strategy to study the causal effect of this social network on voting behavior. Endogeneity may arise both due to the presence of unobservable variables correlated with Twitter penetration and with local electoral conditions and due to reverse causality. For example, changes in local political discourse may lead individuals to use Twitter to express their opinions or gather information, which in turn may impact patterns of participation. Similarly, candidates may ask their supporters to join the platform to aid in their campaign, further complicating the relationship between Twitter and voting behavior.

My proposed instrumental variable strategy centers on the impact that celebrities have on their followers. Twitter features user-generated content, and the more compelling the users are, the more engaging the content becomes. For this reason, the presence of celebrities on Twitter increases the likelihood that their fans will join the platform to receive updates and messages from them.[8] In particular, I focus on players hired during the National Basketball Association (NBA) drafts between 2006 and 2016. Every year, the NBA draft is the event that closes the season. During the draft, teams pick new players that are willing to start playing in the league as professionals.

[6]Sunstein (2017) gives an extensive description of the risks connected to the creations of "echo chambers" online.

[7]See Tucker et al.(2018) for an overview of the main hypotheses that have been suggested in the literature.

[8]There is another way, more mechanical, in which celebrities could affect Twitter's popularity. When searching for a name of a person that happens to be on Twitter, among the first Google search results, there's usually a link to the Twitter profile. Therefore, people that could be looking for a particular name on Google or Bing, would become aware of the existence of Twitter. About this, the support page of Twitter says: *"Your Twitter profile shows up in Google searches because Twitter has a high Google search rank. Keep in mind that the words you write in your Twitter profile or public Tweets may be indexed by Google and other search engines, and cause your profile or Tweets to come up in a search for those terms."* Source: https://support.twitter.com/articles/15349.

This population of players has two important characteristics. First, the best players that participate in the draft receive a strong shock to their popularity, being the draft a very important event for the NBA league.[9] Second, the process that regulates the picks is based on teams records during the season and on a lottery, such that each player gets a destination that is quasi-random after controlling for these factors. This, combined with data on the location of each team's fan base allows me to compare the diffusion of Twitter between areas that were differently affected by the NBA draft.

The results of my instrumental variable estimates show that the effect of increasing the number of Twitter accounts on average political participation is weak: neither electoral turnout nor total campaign donations are affected. However, when analyzing the impact on the Democratic and Republican parties separately, interesting differences emerge. Twitter has a negative impact on the Democratic Party and a positive impact on the Republican Party. Specifically, we observe a decrease in the number of votes for the Democratic Party and an increase in the amount of donations received by the Republican Party.

The above estimates should be interpreted as the Local Average Treatment Effect (LATE) for the sub-population of individuals who open a Twitter account due to their interest in basketball. To gain further insight into the underlying mechanisms, it is necessary to study "compliers," or individuals who would not have opened a Twitter account in the absence of the intervention. To this end, I downloaded profile pictures for a random subsample of users that I had previously matched to a location. Using image recognition algorithms, I attached demographic characteristics - specifically age and gender - to these pictures. Results from approximately 1 million profile pictures show that in the population of compliers, users who are male and over 40 are overrepresented. Since these demographics are correlated with preferences for the Republican Party,[10] it is necessary to be careful in extrapolating the aforementioned estimates to the whole population.

Finally, I examine whether the observed pattern of results can be attributed to a divisive discourse that lacks substantive information. To accomplish this, I employ two additional data sets. The first is the Cooperative Congressional Election Study (CCES), which comprises surveys that gauge respondents' political knowledge and polarization levels. The second dataset encompasses around 150 million tweets generated by a representative sample of 200,000 users.

To study how Twitter influences interest in politics with the CCES survey, I follow Snyder and Strömberg (2010) and use knowledge of the name of

[9] The NBA draft is regularly watched by millions of viewers, see for example https://www.forbes.com/sites/maurybrown/2014/06/27/espn-sees-highest-rated-ever-tv-ratings-for-2014-nba-draft/

[10] See for example data from Pew Research Center: http://www.people-press.org/2018/03/20/1–in-party-affiliation-among-demographic-groups/

Senators and members of the House of Representatives as a proxy for political information. I find a negative effect on information, suggesting that Twitter has acted mostly as an additional source of entertainment. This fact is confirmed if we look at the tweets written on the platform. Using patterns in the co-occurrence of hashtags in the tweets, I assign categories to a sample of approximately 7 million tweets written after 2015. Only a minority of users write tweets about politics regularly, with the others sharing mostly comments about entertainment or sports. Politics becomes popular only at the time of the presidential debates.

I then analyze how Twitter impacted political polarization. Using the CCES data, I define two measures of political polarization and find that Twitter exacerbates political polarization. To code the political content of tweets, I examine hashtags that can be identified as partisan, as suggested by Bovet et al. (2018). I find that tweets with a Republican lean are more popular and receive a more positive sentiment on average. Overall, my findings indicate that Twitter did not influence political attitudes through the provision of more information to users, but rather by enabling the propagation of a partisan discourse that was favorable to the Republican party.

This paper contributes to the political economy literature that studies the relationship between media and political outcomes. Stromberg (2004), Gentzkow (2006), and Gentzkow, Shapiro, and Sinkinson (2011) estimate the effect of radio, TV, and newspapers on attitudes towards politics. Campante, Durante, and Sobbrio (2013), Falck, Gold, and Heblich (2014), Gavazza, Nardotto, and Valletti (2016), and Larcinese and Miner (2017) have studied the effect of broadband on voting behavior in Italy, Germany, England, and the US, respectively. In all cases except for the US, the authors find a negative effect of Internet availability on electoral turnout and no immediate impact on voting behavior. These papers suggest that the main mechanism is the quality of information provided by the media. By providing new and relevant information, newspapers and radio have a positive impact on participation. In contrast, both TV and the Internet were initially used primarily as sources of entertainment, leading to a reduction in the consumption of traditional media. In comparison to this literature, my paper focuses on social media, which can be seen as the second step in the evolution of the way the Internet is commonly used, with more emphasis on user-generated content. Additionally, my results suggest that social media may not have influenced political attitudes mainly by distracting voters, but rather by exposing them to a partisan debate that tends to favor one party, in line with the findings by Huszár et al. (2022). This also connects my work to the literature studying how new media are creating "echo chambers" where participants are only exposed to homogeneous opinions, increasing extremism and polarization (Gentzkow, 2016; Sunstein, 2017).[11] The literature discussing the political effects of social media is further

[11] Also related, but focused on political ads, Beknazar-Yuzbashev, and Stalinski (2022)

summarized in Zhuravskaya et al. (2020)[12] and Lorenz-Spreen et al. (2022).

The paper most closely related to this work is Fujiwara, Müller, and Schwarz (2021), which analyzes a similar set of outcomes from 2008 to 2020. Their identification strategy uses an instrumental variable approach based on the location of people who joined Twitter during the SXSW festival[13] in 2007. They find that Twitter reduced the vote shares of the Republican Party, interpreting this result as being driven by liberal content on the platform. The discrepancy in our results may be due to the fact that these effects are localized and our population of compliers differ. My instrumental variable strategy is based on accounts of basketball players, whereas theirs is based on attendees to the SXSW festival. For these reasons, in my case, the population of compliers seems to match demographics associated with the Republican Party, while the population of attendees to the SXSW festival may be disproportionately from urban and liberal areas, suggesting a bias towards the Democratic Party. Taken together, these papers could support the interpretation that social media make users more entrenched in their initial positions. Another closely related paper is Petrova, Sen, and Yildirim (2017), which shows that Twitter affected political competition by increasing campaign contributions for politicians active on the platform. The main difference between their paper and mine is probably that they focus on short-run outcomes for politicians entering the platform, using the exact timing of when politicians created new accounts on Twitter, while I study outcomes in a longer time horizon when the equilibrium effects of new politicians entering the platform have already played out.

Research outside the field of Economics has also extensively explored the impact of Twitter and other social media on participation. Even though this literature suffers from a lack of identification, it highlights the potential underlying mechanisms at work.[14] One area of research utilizes survey data to investigate how the use of social networking websites like Facebook or Twitter relates to acts of participation such as voting, with a general positive correlation.[15] Another strand of literature relies on data from the platforms. Barberá and Rivero (2015) use tweets to examine the ideological positions of users who wrote about politics, and find that individuals with extreme positions are disproportionately represented. Barberá (2015) measures the ideological position of millions of individuals and finds that users are usually embedded in ideo-

design a field experiment on Facebook to study the effect of political ads on behavior, finding insignificant effects of political ads on turnout.

[12]This review cites an early version of this paper (2018), with results covering only the 2008 and 2012 US presidential elections.

[13]The South by Southwest (SXSW) festival is an annual set of conferences and festivals centered around interactive media that take place in Austin, Texas.

[14]A notable exception is the work by Bond et al.(2012) who run an experiment on Facebook. They show the presence of peer effects in voting behavior among users using a particular message that appeared on the main page of the website.

[15]For a meta-analysis of this literature see Boulianne (2015).

logically diverse networks, suggesting that social media may mitigate political polarization. Compared to this literature, I contribute by studying the causal effect on political outcomes. My results can therefore shed light on the relative importance of the forces at play, in a relevant context such as the presidential elections.

2 Data

I perform the analysis at the Designated Market Area (DMA) level. DMAs are groups of counties defined by Nielsen on the basis of the television market in such a way that all counties that belong to the same DMA have a similar TV offering.[16] These regions are not the same as metropolitan areas, even though in some cases the differences are small. The reason for using this level of aggregation is that Google Trends data, which I use to measure popularity, are available at DMA level but not at the county level. Electoral data, demographic controls and the measure of Twitter penetration were collected at the county level and then aggregated at DMA level. The sample of counties that I use is such that each county belongs entirely to one DMA.

The sample includes observations for DMAs that had data in all periods, 2008, 2012 and 2016. In total the sample contains 207 DMA regions.[17]

2.1 Political and Census Data

I collected data on turnout and voting behavior for Presidential Election in 2008, 2012, and 2016 at the county level. The source is Dave Leip Atlas.[18] Data include information on the number of valid votes and votes received by the candidates. Controls were downloaded from Census and include age distribution across cohorts, income, race, gender and educational attainment. Table 2.1 includes summary statistics for the variables that are included in the analysis, once the data were aggregated at the DMA level.

Campaign contribution data were downloaded from Center for Responsive Politics (CRP).[19] This data is originally collected by the Federal Election Commission (FEC) and is complemented by CRP with additional information regarding the recipient. I use data on contributions to candidates and ignore contributions to PACs or other organizations. For each donation, the database contains information regarding the amount, the date, as well as name and location of the donor. Regarding the recipient, we know the name and the party

[16] From Nielsen website: "A DMA region is a group of counties that form an exclusive geographic area in which the home market television stations hold a dominance of total hours viewed."

[17] DMA regions that belong to Alaska were not considered.

[18] Link: https://uselectionatlas.org/

[19] https://www.opensecrets.org

the candidate belongs to. FEC reports donations from individuals that have donated at least 200$ in an election cycle. I follow Petrova et al.(2017) in reporting results for donations below 1000$, as small contributions could better represent supporters behavior. It is possible to notice that the total amount of donations below 1000$ was approximately equal to 472 million dollars in 2008 and reached 754 million dollars in 2016.

2.2 Voters' information and polarization

To measure voters' ideology and information, I downloaded data from the Cooperative Congressional Election Study (CCES).[20] CCES is a survey administered by YouGov every election year to over 50,000 responders. In particular, I make use of responses from the Common Content part, waves 2008, 2012, and 2016. This part contains answers regarding demographic characteristics, partisan identity and attitudes towards candidates. I use this survey to build a measure of voters' information and two measures of polarization.

To measure information, I use three questions in which respondents are asked whether they approve the way senators and house representatives are doing their job. In particular, I count the number of times (from 0 to 3) each respondent answers "Never Heard / Not Sure" to these three questions.

To measure information, I combine answers to questions in which respondents were asked whether they approved the way local politicians were doing their job. The same question[21] was asked about two state senators and one house representative, with names that depended on the respondent's electoral district. In particular, I count the number of times (from 0 to 3) each respondent answers "Never Heard / Not Sure" to these three questions.

I calculate two measures of polarization. "Partisan sorting" is related to the extent that self-reported partisan identity and self-reported ideology match. "Partisan polarization" captures a similar idea, as it is higher the stronger the ideological difference between republican and democrat respondents. Further details regarding the question used and the definitions of the two measures of polarization are presented in Appendix B.

2.3 Twitter Users

Twitter does not provide aggregated data on the number of active users and their geographic distribution. To build a measure of Twitter penetration across regions I relied on Twitter Search API[22] and downloaded information on ac-

[20] https://cces.gov.harvard.edu/

[21] The question was: *Do you approve of the way each is doing their job... [Name].* Possible answers ranged from *Strongly Disapprove* to *Strongly Approve*, with the additional option *Never Heard / Not Sure*

[22] API stands for Application Programming Interface. In this context, it can be considered as a set of tools that Twitter makes available to interact with their database.

Table 1: Control and Outcome Variables - Summary Statistics by DMA region

	2008		2012		2016	
	Mean	SD	Mean	SD	Mean	SD
Panel A: Control Variables						
Population	1,453	2,261	1,490	2,289	1,535	2,362
Male (share)	49.3	0.8	49.4	0.8	49.5	0.8
Age - under 18 (share)	24.2	2.6	23.6	2.5	22.9	2.5
Age - over 65 (share)	13.6	2.4	14.2	2.4	15.6	2.6
Race - White (share)	80.6	12.4	80.4	12.5	79.9	12.6
Race - Black (share)	10.5	11.5	10.7	11.6	10.8	11.5
Bachelor's degree or higher	23.1	5.8	24	6	25.6	6.3
Income lower 10k (share)	15.1	4.1	16.2	3.9	16.5	3.9
Income higher 200k (share)	2.5	1.5	2.9	1.7	3.6	2
Average Income	60.3	10.2	63.4	10.2	67.4	11.3
Internet Penetration	3.2	0.6	3.8	0.4	4.2	0.5
Panel B: Outcome Variables						
Turnout	58.5	7.5	55.3	8.2	54.2	8
% Dem	47	10.3	44.6	11.1	39.8	12.1
% Rep	51.3	10.4	53.4	11.2	55	12.2
Votes Dem	334.8	573.9	317.8	547.8	301.6	537.3
Votes Rep	288.4	353.4	293.4	350.8	295.0	340.2
Donations < 1k	2,283	5,560	2,470	5,348	3,645	8,350
Donations Dem < 1k	1,533	4,400	1,384	3,752	2,473	6,826
Donations Rep < 1k	749.4	1,266	1,086	1,754	1,172	1,788
N. DMA	207		207		207	

Note: Controls are provided at the DMA level. *Internet Penetration* refers to Residential Fixed High-Speed Connections per 1000 Households. Data were downloaded from Federal Communication Commission. Data are provided by county in a scale from 0 to 5 and were aggregated using population as weight. *Population* is expressed in thousands of inhabitants. *Average Income* is expressed in thousands of dollars. *Turnout* is given by the ratio between the number of votes and the voting age population. *% Dem* represents the share of votes received by the Democratic Party, while *% Rep* represents the share of votes received by the Republican Party. *Votes Dem* gives the number of votes received by the Democratic Party, expressed in thousands of votes. *Votes Rep* gives the number of votes received by the Republican Party, expressed in thousands of votes. *Donations < 1k* is the sum of all individual donations below 1000$ to politicians affiliated to the Democratic Party or to the Republican Party. *Donations Dem < 1k* and *Donations Rep < 1k* refer to donations received by members of the Democratic Party and the Republican Party, respectively. Variables that refer to donations are expressed in thousands of dollars.

counts. Over the span of a few months, I have made requests for approximately 310 million accounts.

Figure 5 shows the screenshot of an account page. On the left, below the profile picture, it is possible to read the username, description, location, and the date the account was created. In the center of the page, we can see the number of tweets (messages written by the user), and the number of other accounts that the user is following[23] (*Following*), the number of accounts that are following this user (*Followers*) and the number of messages that the user *liked*. In this case, we see that the username is *President Trump*, the location is *Washington, D.C.* and the account was created in January 2017. In total this user wrote 3,812 tweets and is receiving direct updates from 39 other accounts. There are 23.9 million users that are receiving every message written by *President Trump*.[24]

It is important to underline that, while the creation date is always provided by Twitter and cannot be modified by the user, the location field contains information that is self-reported. In order to obtain a reliable measure of the location and reduce issues while matching localities I only kept locations that were written as GPS coordinates or in the format <*City, State*>, that is the format suggested by Twitter on the basis of the IP address. Only a small fraction of users (less than 1%) uses GPS coordinates to specify their location. In Appendix C I provide more details regarding this process. The number of accounts that were matched at the county level was in total 6.8 million.

I then aggregated accounts at the DMA level. Figure 6 shows the kernel density of the number of accounts per 1000 inhabitants for the 207 DMA regions that are included in the sample. Twitter's popularity grew faster starting in early 2009. This fact is confirmed by Figure 7, which shows the number of tweets over time (data from Twitter). Finally, Figure 1 shows the distribution of accounts in 2016. Figures 8 and 9 describe the distribution of accounts in 2008 and 2012 respectively.

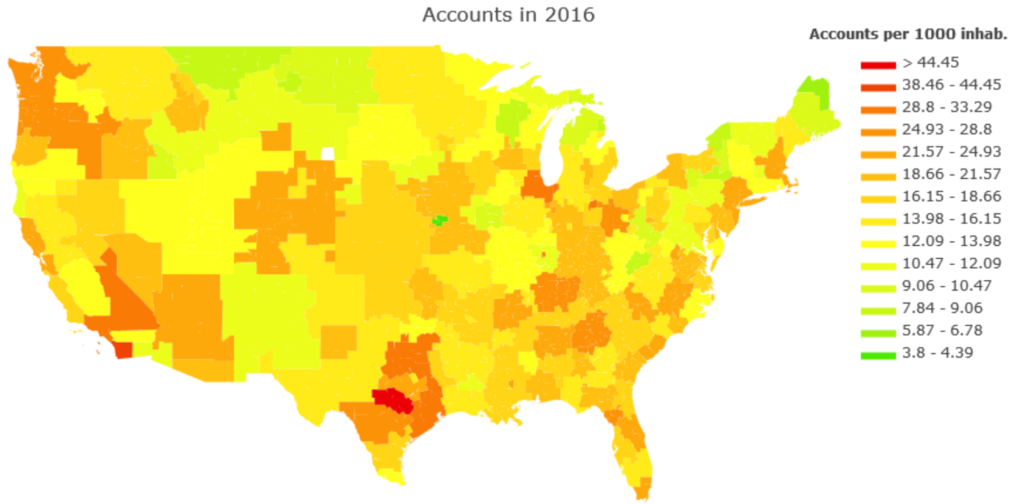
To gather more information about the user population, I downloaded the profile picture for a random subsample of 2 million accounts. Using machine learning algorithms, I selected images that contained at least one face and then analyzed the faces using facial recognition tools. This allowed me to obtain basic demographics for approximately 980,000 accounts. Table 2 contains summary statistics for the estimated share of male users and average age. Figure 10 shows the estimated age distribution across the three years.

Finally, I downloaded tweets for a subsample of approximately 200k users taken from the set of users I could match to a location. Using the Twitter API it is possible to download at most the last 3,200 tweets for each account, so

[23]On Twitter links are unidirectional. We say that user A is *following* user B when A is receiving all messages written by B. User B, instead, will not receive any update about A, unless he follows A back.

[24]The screenshot was taken on August 28, 2018.

Figure 1: Distribution of accounts per 1000 inhab. by DMA in 2016



that it is often not possible to retrieve tweets that are relatively old, especially for the most active users. I collected in total close to 155 million tweets, 24% of which are retweets. The use of hashtags is important, with 20% of tweets making use of at least one hashtag.

Table 2: Profile Pictures Demographics - Summary Statistics

	2008		2012		2016	
	Mean	SD	Mean	SD	Mean	SD
Male	57.5	8.8	46.3	3.1	45.9	2.8
Age	42.3	2.2	37.1	1.6	36.4	1.5
N. Images	334.5	527.6	3,751	4,453	4,677	5,537
N. DMA	207		207		207	

Note: Statistics are provided at the DMA level.

2.4 NBA players and teams' popularity

As it will become clear in Section 3, the instrument relies on tracking players from the National Basketball Association (NBA) league at the beginning of their career, the NBA draft. I collected a dataset of names of players from the NBA that were active in the period 2007-2016. I then identified those who had a Twitter account and the day they joined.

To measure how popular teams were over time and across regions I used Google Trends. As explained by Davidowitz and Varian (2015): *"Google Trends reports an index of search activity. The index measures the fraction of queries that include the term in question in the chosen geography at a particular time relative to the total number of queries at that time"*. The scores reported by Google Trends are normalized so that the maximum is always 100. Data were downloaded at the DMA level, by using teams' names as keywords. To exclude the possibility that these scores were directly affected by Twitter, I downloaded them for the period 2004-2008.[25] Figures 11 and 12 show two examples of such scores for Boston Celtics and Minnesota Timberwolves. We can notice that the highest scores are in the regions where the two teams are playing. At the same time, the scores do not decrease monotonically with distance, as the level of interest for the NBA is not uniformly distributed.

3 Empirical Specification

To investigate the effect that a stronger presence of Twitter had on participation, we need to relate the variation in Twitter penetration to changes in any of the outcome variables considered. The basic framework for our analysis is given by the following fixed effect model:

$$Y_{dt} = \beta_0 + \beta_1 Twitter_{dt} + X'_{dt}\beta_2 + \delta_t + \delta_d + \epsilon_{dt}$$

where t indexes years of election (2008, 2012, and 2016) and d indexes DMA regions. Outcome variables are presented in Table 2.1. The variable $Twitter_{dt}$ measures the number of accounts per 1000 inhabitants at time t in DMA region d. I control for the set of census variables described above. Finally, I include year fixed effects and DMA fixed effects.

A critical challenge in estimating the previous model is to address endogeneity concerns related to the presence of omitted variables and reverse causality. Changes in the political debate at the local level could for example drive users towards Twitter, to the extent that the platform allows them to express their opinion or gather information, while affecting patterns in participation. Similarly, strong candidates could ask their supporters to join the platform in order to help in the campaign.

To address this issue I implement an instrumental variable approach to exploit the fact that celebrities influenced Twitter's success by making the platform more interesting with their presence. In particular, the instrument is based on variation that comes from the NBA drafts.

[25] Google Trends data are not based on the full sample of past searches, but are instead based on a sample of Google search data (https://support.google.com/trends/answer/4365533?hl=en). This subsample is changed every 24 hours. In order to reduce noise I downloaded data from four different days, excluded areas with very low scores, and took the average.

3.1 The NBA Draft

Every year, after the end of the season, the NBA league organizes the Draft. During this event the teams pick players to add to their rosters, choosing from a population of players who wish to join the league. Therefore, the NBA Draft determines the allocation of new talents across teams. Given the characteristics of this sport, it is often the case that some of these new players can have a strong impact in the teams' performance, becoming idols for their newly acquired fan base. This makes the Draft a very popular event. For example, in 2015 ESPN counted 3.7 million viewers for the TV broadcast.

The process is organized as follows. Players that wish to participate to the Draft need to declare their eligibility no later than 60 days before the event. Players become eligible to participate one year after high school graduation if they are at least 19 years old. Approximately one month before the Draft, in May, the Draft Lottery takes place. This lottery determines the order that teams will follow when choosing the new players. The first three picks are allocated at random using a scheme that assigns higher chances to the teams that had a worse performance during the regular season (see Table 3). The other picks are assigned following again the reverse order of the regular season record. This system tries to balance between two forces. On the one hand, by assigning priority to teams with a weaker record, it brings balance in the league, as the best new players will go to these teams. On the other hand, the lottery is meant to reduce the incentive that teams have to worsen their record by losing matches in order to hire better players during the Draft[26].

The instrument I use exploits this mechanism. For every Draft between 2008 and 2016, I focus on the first 10 picks and check whether each player had a Twitter account at the time of the Draft. Moreover, I use the number of lottery tickets to control for the teams' records during the season. This allows me to take into consideration teams' choices that were possibly trying to obtain better chances of winning the lottery.

3.2 IV model

Equations 1 and 2 describe the IV model that I use to study how Twitter influenced political outcomes. Equation 2 gives the first stage regression I run. I instrument the number of accounts every 1000 inhabitants in a given DMA region d at time t using a variable that measures the degree of exposure to the Draft (Equation 3). With the same logic, I then control for the number of tickets each team had received for the Draft Lottery. The specification I

[26]Motivated by concerns that some teams had lost matches on purpose in order to obtain a higher number of lottery tickets, in 2017 the NBA league approved a new set of rules that regulate the Draft Lottery. These include in particular a new distribution of probabilities. A description of these rules can be found here.

Table 3: Draft Lottery - Number of Tickets

Ranking in the Regular Season	Number of Tickets	Ranking in the Regular Season	Number of Tickets
30	250	23	28
29	199	22	17
28	156	21	11
27	119	20	8
26	88	19	7
25	63	18	6
24	43	17	5

Note: Number of tickets received for the Draft Lottery, based of the regular season record. The 30 teams that belong to the NBA league are ranked based on their performance during the regular season. In total, 14 teams participate to the Draft Lottery. Source: NBA.com

consider is the following one:

$$Y_{dt} = \beta_0 + \beta_1 \mathit{Twitter}_{dt} + \beta_2 \log(\mathit{Tickets}_{dt}) + X'_{dt}\beta_3 + \delta_t + \delta_d + \epsilon_{dt} \tag{1}$$

$$\mathit{Twitter}_{dt} = \alpha_0 + \alpha_1 \log(\mathit{Draft}_{dt}) + \alpha_2 \log(\mathit{Tickets}_{dt}) + X'_{dt}\alpha_3 + \delta_t + \delta_d + \nu_{dt} \tag{2}$$

Where:

$$\mathit{Draft}_{dt} = \sum_c \mathit{IncomingTwitter}_{ct} \cdot \mathit{Popularity}_{cd} \tag{3}$$

Where c indexes teams, $t \in \{2008, 2012, 2016\}$ indexes election years and d indexes DMA regions.

$\mathit{Twitter}_{dt}$ indicates the number of accounts per 1000 inhabitants in region d at time t. $\mathit{IncomingTwitter}_{ct}$ measures the number of players that joined team c until time t during the NBA drafts and that had a Twitter account when the transfer was announced. I consider only players that were drafted as top 10 picks. $\mathit{Popularity}_{cd}$ refers to the measure of the popularity of team c in region d, calculated using Google Trends for the period 2004-2008. Finally, $\mathit{Tickets}_{ct}$ is the number of lottery tickets received by team c from 2008 until time t.

In words, the instrument captures the shock that is generated when a pick is realized. The player will receive a new wave of interest coming in particular from the fans of the team he will be playing for in the next season. In case the player has a Twitter account we expect part of this wave of interest to be transformed into new accounts on the social network, as some supporters will be interested in following the new member of their team.[27] Also, this effect would be magnified by network externalities, propagating among these fans' friends and relatives.[28]

The identifying assumption is that, conditional on observables, the popularity in region d of the team that has received a new player with a Twitter account is orthogonal to unobserved determinants of voting behavior in region d. Moreover, in order for the exclusion restriction to be valid we need to assume that drafted players, by moving to a new team, do not have a direct effect on political attitudes, for example by making their fans more aware of politics.

Table 10 contains results from the first stage regression described above. The F-stat of excluded instrument refers to the Kleinbergen-Paap F-statistic and is equal to 20.56 when I include controls while it is 31.99 without controls. The instrument appears therefore to be relevant. By looking at the first stage regression we can notice that the sign is as expected, with $Draft_{dt}$ having a positive effect on Twitter penetration. Back of the envelope calculations can be made to obtain a better intuition regarding the magnitude of the effect. The regression coefficient implies that a 1% increase in the *Draft* variable determines a 3.4/100 increase in the number of Twitter accounts over 1000 inhabitants. If we consider the total population, we get that this corresponds to 10,400 accounts. If we also consider that the number of drafted players with a Twitter account that appeared among the top 10 picks is 66, we see that an average player, representing 1.5% of the total, contributed to 15,600 accounts over the total 6.8 million.

4 Results

I analyze the effects of Twitter penetration on political outcomes using the instrumental variable strategy described above in equations 1 and 2. Since the instrument is defined at the DMA level, I aggregate outcome and control variables at this level. All regressions include DMA and Year fixed effects. Results are presented in Table 4. The upper part of the table shows the coefficients of Twitter penetration. The lower part instead reports results for the first-stage. In all regressions, Twitter penetration is standardized.

Tables 13 to 20 report results for the instrumental variables regressions

[27] An alternative explanation why the presence of drafted players should increase interest towards Twitter relies on the way Google search algorithm works. When searching for the name of players, in case they have a Twitter account, this account is shown on top of the results page. Fans who are searching for information on newly acquired talents would therefore become aware of the presence of the social network.

[28] In this sense the instrument described here relies on a similar mechanism as the one exploited in Enikolopov et al.(2018) for the Russian social network VK. In Enikolopov et al.(2018) the success of VK is instrumented using the location of VK's first members, at the time the social network was not allowing everyone to register.

Table 4: Instrumental variables estimates of Twitter on political outcomes

	Independent Variable: *Twitter*			
	(1)	(2)	(3)	(4)
Turnout	-0.709	-1.352	-1.734	-1.690
	(1.077)	(1.282)	(1.504)	(1.492)
% Dem	-2.157	-4.240*	-3.866*	-3.875*
	(1.931)	(2.497)	(2.184)	(2.153)
% Rep	1.271	3.154	3.033	3.114
	(1.794)	(2.273)	(2.049)	(2.024)
Donations < 1k	-126.1	126.3	-340.3	-305.9
	(1,669)	(1,961)	(2,161)	(2,130)
Donations Dem < 1k	-599.6	-276.1	-723.3	-688.4
	(578.6)	(653.9)	(714.9)	(704.6)
Donations Rep < 1k	473.5***	402.4***	383.0**	382.4**
	(132.6)	(142.6)	(170.3)	(166.9)
First Stage				
Draft	5.392***	3.744***	3.341***	3.397***
	(0.893)	(0.726)	(0.742)	(0.749)
F-value (instr)	31.99	26.63	20.30	20.56
Population, Male		Yes	Yes	Yes
Other Demographics			Yes	Yes
Internet				Yes
Number of DMA	207	207	207	207
Observations	621	621	621	621

Note: Effects on voting outcomes are calculated per 1,000 inhabitants. Effects on donations are expressed in thousands of dollars. The explanatory variable is Twitter penetration, that is defined as the number of accounts per 1,000 inhabitants and then standardized. All regressions include DMA and election year fixed effects. Controls are discussed in Section 2. The F-value refers to the Kleibergen-Paap F-statistic. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

summarized in Table 4, and offer more details concerning the other control variables. Tables 11 and 12 show OLS estimates for electoral outcomes and donations to politicians, respectively.

Two main facts emerge from the IV regression tables. Outcomes as voter turnout and the overall amount of donations are not significantly affected by the presence of Twitter. Nevertheless, I find effects that tend to favor the Republican Party, once the instrument is considered. I find that a 1 standard deviation increase in Twitter penetration reduced by 70,000 the number of votes received by the Democratic Party, while instead it increases by 382,400$ the amount of (small) donations received by the Republican Party. These effects are also sizable, as they matter respectively for 0.13 and 0.42 standard deviations of the two outcome variables.

4.1 Compliers

The estimates presented in Table 4, given the typology of analysis, should be interpreted as local effects. It is therefore necessary to study the characteristics of the population of those who are induced to open a Twitter account by the presence of NBA players. To study compliers I use the profile pictures data described above. In particular, I am interested in studying how the population of Twitter accounts is affected in terms of demographic characteristics by the instrument.

To do this, I consider the same regression model as Equation 2:

$$Y_{dt} = \alpha_0 + \alpha_1 \log(Draft_{dt}) + \alpha_2 \log(Tickets_{dt}) + X'_{dt}\alpha_3 + \delta_t + \delta_d + \nu_{dt}$$

The difference is that instead of the number of accounts per 1000 inhabitants, as Y_{dt} I consider the share of male users on the platform, and the average age. Regressions are weighted by the number of images matched to demographic characteristics for each DMA region.

Table 5 contains results for these regressions. Table 21 presents further details regarding control variables. From these regressions, it is possible to notice that compliers tend to be male and tend to be older than the average user. As Table 6 shows, these demographics correlate with stronger support for the Republican Party. Therefore we must be cautious in interpreting the effects described above as average treatment effects for the entire population.

5 Online Discussion, Information and Polarization

The results presented in the previous section can only be partially explained by the characteristics of the population of compliers. In particular, it is necessary to better understand why the effect on participation tends to be negative and why it seems relatively stronger for campaign donation, with respect to voting.

Table 5: Compliers' characteristics: gender, race and age

	Independent Variable: *Draft*			
	(1)	(2)	(3)	(4)
Share of Male users	5.674*	5.967*	5.378*	5.407*
	(3.065)	(3.055)	(3.062)	(3.065)
Average Age	2.682***	2.343***	1.345*	1.343*
	(0.774)	(0.768)	(0.700)	(0.700)
Population, Male		Yes	Yes	Yes
Other Demographics			Yes	Yes
Internet				Yes
Number of DMA	207	207	207	207
Observations	621	621	621	621

Note: Outcome variables are constructed using profile pictures from a sample of 980,000 Twitter accounts and are described in Section 2. All regressions include DMA and election year fixed effects. Controls are discussed in Section 2. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Partisan Identity by gender and age.

	2008			2012			2016		
	Dem	Rep	Ind	Dem	Rep	Ind	Dem	Rep	Ind
Male	43.9	40.3	12.8	46.1	37.5	14.8	42.9	38.4	15.9
Female	50.9	31.8	10.9	53.2	33.4	11.5	48.4	34	13.9
Age 18-24	55	24.7	12.4	55.2	29.6	13.5	48.2	29.1	16.8
Age 25-39	49.5	30.7	13.4	55.2	28.6	14.1	50.4	29.8	14.9
Age 40-64	45.6	38.9	11.8	48.2	36.6	13.6	43.8	38.5	15.6
Age 65+	42.1	47.1	8.9	42.7	45.6	9.9	41.9	44.5	12.3

Source: CCES survey. This table summarizes answers to the question "*Generally speaking, do you think yourself as a ... ?*". Answers are weighted according to survey weights.

To shed light on these questions I study the effect of Twitter on information and political polarization. In this section I report results from two additional analyses. First, using data from the CCES survey, I run an IV model that exploits the same identification strategy described previously. I find that Twitter had a negative effect on voters' knowledge about local politicians and that it has determined an increase in political polarization.

I then studied a dataset of tweets written during the last presidential campaign. Descriptive evidence shows that most people use Twitter to discuss about entertainment topics and pay attention to the elections for short lived periods, especially around the presidential debates. Moreover, an important fraction of the debate about politics contains clearly partisan views. The Republican Party had also a stronger presence on Twitter, both in terms of number of Tweets and in terms of sentiment. This difference was finally driven by the use of retweets, highlighting the importance of mechanisms that characterize social media.

5.1 Voters are less informed and more polarized

To study whether Twitter has influenced information and political polarization, I use data from the Cooperative Congressional Election Study (CCES). This allows me to build a measure of information regarding local politics and two measures of political polarization. As outcomes I use variables described in Section 2.2. The variable *No Infomation* counts the number of time each respondent has not been able to express an opinion regarding a state Senator or the incumbent House Representative. For each respondent, the variable takes values from 0 to 3, where 3 can be interpreted as a lower level of information. Variables *Partisan Sorting* and *Partisan Polarization* are instead constructed comparing respondents' ideology and partisan identity. High values in these two variables indicate a higher degree of homogeneity between ideology and partisan identity, with republican voters being more conservative and democratic voters being more liberal. I study the effect of Twitter penetration on these outcomes with an IV model that resembles the one used above:

$$Y_{i(dt)} = \beta_0 + \beta_1 Twitter_{dt} + \beta_2 \log(Tickets_{dt}) + X'_{idt}\beta_3 + \delta_t + \delta_d + \epsilon_i \quad (4)$$

$$Twitter_{dt} = \alpha_0 + \alpha_1 \log(Draft_{dt}) + \alpha_2 \log(Tickets_{dt}) + X'_{dt}\alpha_3 + \delta_t + \delta_d + \nu_{dt} \quad (5)$$

Where i indexes the respondent, t indexes time and d indexes DMA regions. As controls I include the same set of controls that were used previously and add dummies for income level (three categories), gender, educational attainment (two categories), race (four categories). I also control for the age of the respondent (both linear and quadratic). $Twitter_{dt}$ is instrumented using the same strategy described above. Variables $Draft_{dt}$ is defined as in Equation 3.

Importantly, CCES data do not include information regarding respondents' use of social media. I therefore use the same measure of Twitter penetration I was using before. This implies that all respondents from the same DMA region are assigned the same level of Twitter penetration. For this reason, standard errors are clustered at the DMA-year level.

Results are presented in table 7. Column (1) considers regressions for the full sample of respondents. Column (2) restricts the analysis to male respondents while column (3) uses only the subsample of people older than 40. These subsamples were chosen to match the analysis of compliers described before. IV regressions show that Twitter seems to have reduced information about politics, while at the same time increasing political polarization. For political polarization, effects are starker when considering the subsample of relatively older respondents, which is in line with what has been highlighted by Boxell et al.(2018), who show that older cohorts are the ones that have polarized the most during the last years.

Table 7: Instrumental variables estimates of the effect of Twitter on information and political polarization

	Explanatory variable: *Twitter*		
	(1)	(2)	(3)
No Information	0.166*	0.076	0.119
	(0.085)	(0.064)	(0.084)
Partisan Ideology	0.072	0.028	0.158**
	(0.0774)	(0.091)	(0.062)
Partisan Sorting	0.024	0.056**	0.042**
	(0.0242)	(0.027)	(0.018)
Subsample	All	Male	Age 40+
F-Stat	19.02	18.10	20.34

Outcome variables are described in Section 2.2. The explanatory variable is Twitter penetration, which is defined as the number of accounts per 1,000 inhabitants and then standardized. All regressions include DMA and election-year fixed effects. Controls are discussed in Section 2. The F-value refers to the Kleibergen-Paap F-statistic. Standard errors are clustered at the DMA, year level (two-way). *** p<0.01, ** p<0.05, * p<0.1

5.2 On Twitter, peaks of attention and a partisan debate

To corroborate the previous findings I downloaded tweets from a sample of approximately 200k users taken from the set of users I could match to a location. When downloading tweets, it is important to underline that the main limitation is that Twitter allows to download at most the last 3200 tweets, so that it is often not possible to retrieve tweets that are relatively old, especially for the most active users. I therefore decided to focus only on the last presidential elections and in particular on the year prior to the presidential elections.

The first step of this analysis has been the categorization of tweets into categories. To do that I followed a method suggested by Conover et al.(2011), which is based on hashtags. Starting from a set of hashtags that define various categories, it is possible to retrieve a wider set of tweets that contain hashtags that co-occur relatively often with the starting ones. In order to capture the most important hashtags in the period I cover, I first identified to 500 most popular hashtags in my sample. I then assigned a category to each of them, excluding those which did not clearly belong to any category[29]. This way I obtained a corpus of 1.1 million tweets divided into 10 categories.

The first two facts that emerge from these data are presented in Table 8 and Figure 2. It is possible to notice that the interest in politics is relatively low on average, especially compared to categories such as Sports or Entertainment. In particular, only 7.8% of used hashtags are on average related to Politics, during the year prior to the presidential elections. On the other hand, Sports, Entertainment, and Music attract almost 70% of the total flow of hashtags, suggesting that most users use Twitter as a source of entertainment. This is confirmed in Figure 13 which shows how most of the users write about politics very rarely. Nevertheless, interest in politics peaks during the electoral debates, when it becomes the category that receives the highest number of tweets. In Table 8 we see that at its peak, politics matters for almost 40% of the number of categorized tweets during the week. Another interesting element to notice is how important retweets are, especially for politics. Almost 48% of messages that include political hashtags are retweets, suggesting how important this element is for the debate on the platform. Finally, when looking at the sentiment, it is possible to see that politics is the category that has the lowest average sentiment. This is due to a higher share of negative messages regarding this topic, caused by a higher level of conflict, especially with respect to the other categories.

The majority of users use Twitter to discuss sports or other entertainment topics. It is only in a few moments during the electoral campaign that the majority of users become exposed to a debate around the elections and writes messages about that. Figures 14 and 15 show the most popular hashtags

[29]Typically, these include common words or locations.

Table 8: Topics on Twitter, statistics by week

Category	Avg share	Min share	Max share	Retweets (share)	Avg Sent.	Min Sent.	Max Sent.
Sport	33.0	19.2	51.9	41.4	0.18	0.15	0.26
Entertainment	25.1	12.5	35.7	24.5	0.18	0.1	0.24
Business	12.9	7.7	18.3	21.1	0.2	0.14	0.24
Politics	7.8	2.1	39.7	47.9	0.07	-0.1	0.21
Music	7.8	4.1	11.7	20.3	0.14	0.07	0.18
Science	2.2	1.1	3.1	33.2	0.23	0.1	0.34
Religion	2.5	1.4	4.7	29.6	0.19	0.14	0.27

Figure 2: Sports and Entertainment dominate the online discussion, Politics peaks just before the elections.

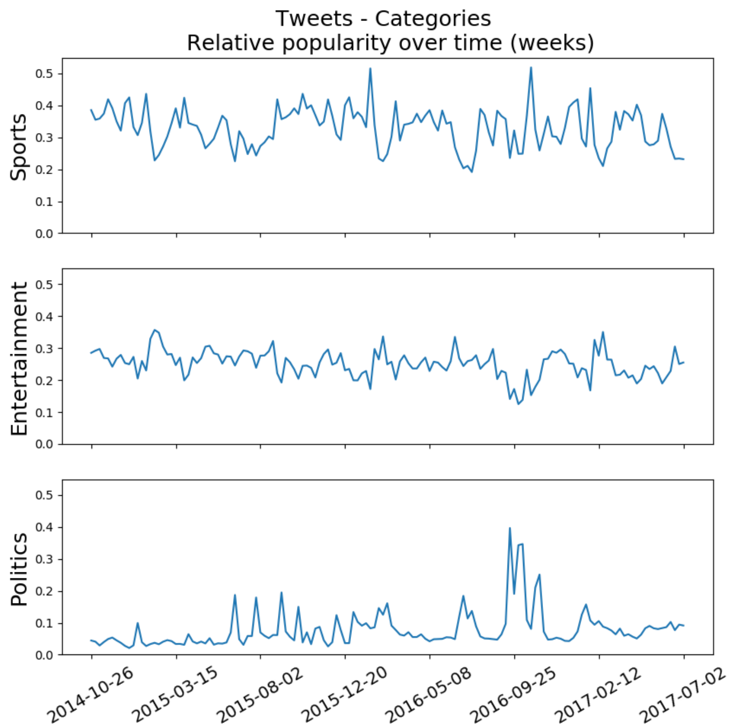
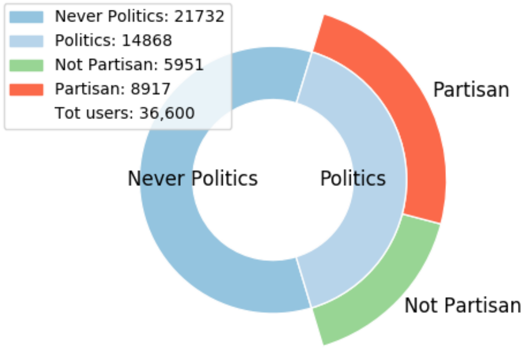


Figure 3: Users, attitudes towards politics during the last phase of the campaign



Note: This figure represents the number of users that wrote about politics or not, between July 1 and November 5, 2016. In Red, are the users who used at least once hashtags with a partisan leaning.

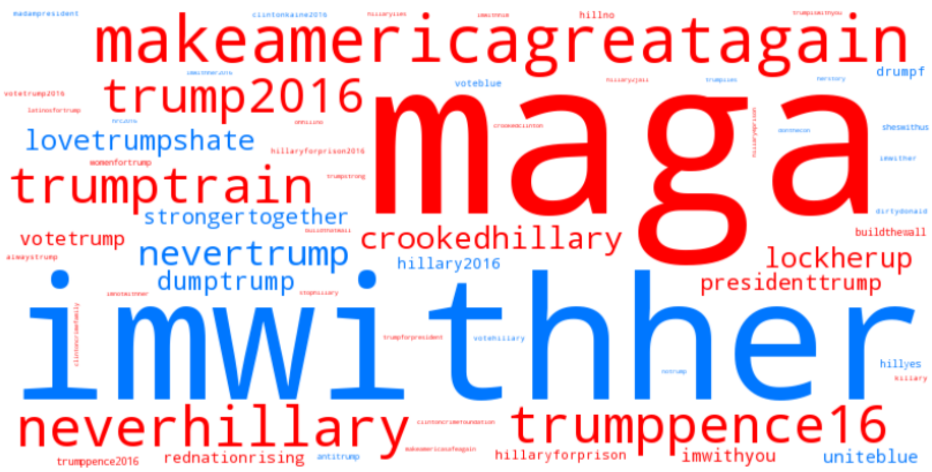
regarding politics, during the first phase of the electoral campaign and during the peak weeks respectively.

I then focused on the hashtags that belong to the politics category at the time of the spike in interest, to see whether there has been any difference between tweets written in support of the Democratic Party and the Republican Party. To do this I followed the categorization of hashtags suggested by Bovet et al.(2018) for the 2016 electoral campaign. Tweets that contain hashtags that have a clear partisan leaning are here considered.

Figure 3 focuses on users. It contains, information regarding the share of users who tweeted about politics or not and, in case they did, whether they used hashtags with a partisan connotation. It is possible to see how approximately 40% of the accounts used hashtags connected to politics during the last phase of the electoral campaign. Moreover, the majority of those who tweeted about politics used hashtags with a partisan leaning. This suggests that the wave of interest in politics exposed a large fraction of users to a partisan debate.

Figure 4 shows the most popular hashtags that belong to politics and have a clear partisan connotation. From this figure, it emerges how the support in favor of the Republican Party was stronger. Also, hashtags with an aggressive connotation were relatively more common, suggesting a difference in the rhetoric used by the two blocs. Table 9 offers more details. Partisan hashtags are here divided into four different groups: hashtags that support each of the two candidates and hashtags that attack each of the two candidates. Table 9 compares these groups of hashtags along three dimensions. The number of

Figure 4: Word cloud - hashtags leaning towards the Republican Party were more popular



Note: This figure represents the most popular hashtags in the "Politics" category when only considering hashtags with a partisan leaning. Size is proportional to the number of times each hashtag was used. In red (blue), hashtags whose leaning was in favor of the Republican (Democratic) Party.

Table 9: Tweets written between July 1 and November 5, by partisan leaning

	All			No Retweets Only		Retweets Only	
	Number of tweets	Avg. Sent.	Retweets (share)	Number of tweets	Avg. Sent.	Number of tweets	Avg. Sent.
All	16617	0.112	0.571	7056	0.084	9421	0.133
pro Clinton	4293	0.111	0.472	2261	0.109	2015	0.11
pro Trump	8935	0.164	0.663	2983	0.128	5899	0.182
anti Trump	1096	-0.039	0.339	708	-0.074	358	0.026
anti Clinton	2293	-0.016	0.51	1104	0.011	1149	-0.042

tweets that include any of those hashtags is shown together with the average sentiment and the share of retweets. I also split between original tweets and retweets. From these data, we can notice that on Twitter, if we consider the users in the sample I am using, there was a strong imbalance in favor of Donald Trump. This is true if we look at the total number of tweets written in support of the two candidates, but also if we consider the sentiment. Interestingly, this difference is driven mostly by retweets. Retweets supporting Donald Trump were almost three times as many as retweets supporting Hillary Clinton and were characterized by more positive sentiment. A similar pattern is present in the group of hashtags that were explicitly against the two candidates. The number of anti-Clinton retweets is more than three times higher than the number of anti-Trump hashtags, with a lower sentiment too.

6 Conclusions

To summarize, in the analysis presented above I studied the impact that Twitter had on political participation during the 2008, 2012, and 2016 US presidential elections. I first created a measure of Twitter penetration across regions and I proposed a novel identification strategy to deal with endogeneity in the diffusion of the platform. I found that Twitter penetration had a weak effect on turnout and on the total amount of donations received by candidates. When comparing the two parties I found that there was a negative effect against the Democratic Party, as they received a lower number of votes. On the contrary, candidates for the Republican Party received a higher amount of donations. Combining image recognition algorithms with users' profile pictures I then showed that the population of compliers tends to be male and older that the average user, so that the local average treatment effects pertain to a group that may not be representative of the whole population.

I then turned to outcomes related to information and political polarization and found that Twitter had a negative effect on the amount of information regarding (local) politicians and a positive effect on political polarization. Using tweets written during the presidential campaign I showed how the majority of users wrote mainly about sports and entertainment while turning to politics only at the peak of the electoral race. Tweets regarding politics were often partisan, with the Republican Party receiving more attention. Also, the role played by retweets was important, with a net advantage for Donald Trump's supporters.

My study speaks to the debate on the effect of social media on political outcomes, not only in the United States but also in European democracies. In particular, I find that Twitter had a twofold effect. On the one hand, the majority of content is about entertainment, with only a minority of users that are engaged in discussing politics. On the other hand, peaks in attention expose the average user to a partisan debate, which is not necessarily constructive, given the high share of negative messages. These dynamics could in turn negatively affect attitudes towards politics for the average user, thereby reducing information and turnout, while motivating a minority of active users to increase their effort, bringing more polarization and more donations.

The results that emerge from my analysis are only partially in line with the literature that studied the impact of the Internet or traditional media on politics. This difference is likely to be determined by the different natures of these media. Social media are indeed not only a source of entertainment and information. The content that is created and shared by users often contains opinions, suggests interpretations, and causes reactions by other users. Also, content that is more likely to be shared gets a disproportionate amount of coverage. This is likely to favor a particular kind of rhetoric, that does not need long messages to be appreciated by readers. A deeper understanding of the characteristics of the political debate on Twitter is nevertheless still needed in order to confirm these hypotheses.

References

Acemoglu, Daron, Tarek A. Hassan and Ahmed Tahoun. 2014. "The Power of the Street: Evidence from Egypt's Arab Spring" *The Review of Financial Studies* 2018, 31 (1)

Barberá, Pablo. 2015. "How Social Media Reduces Mass Political Polarization. Evidence from Germany, Spain, and the United States." Working Paper.

Barberá, Pablo and Gonzalo Rivero. 2015. "Understanding the political representativeness of Twitter users." *Social Science Computer Review* 2015, 33 (6), 712-729.

Bartik, Timothy. 1991. "Who Benefits from State and Local Economic Development Policies?" *W.E. Upjohn Institute.*

Beknazar-Yuzbashev, George, and Mateusz Stalinski. 2022. "Do social media ads matter for political behavior? A field experiment." Journal of Public Economics 214: 104735.

Boxell, Levi, Matthew Gentzkow, and Jesse M. Shapiro. 2017. "Greater Internet use is not associated with faster growth in political polarization among US demographic groups. *PNAS* 114 (40) 10612-10617.

Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. 2012. "A 61-million-person experiment in social influence and political mobilization" *Nature* 489(7415)

Boulianne, Shelley. 2015. "Social media use and participation: a meta-analysis of current research." *Information, Communication and Society* vol 18, Iss 5.2015.

Bovet, Alexandre, Flaviano Morone, and Hernán A. Makse 2018. "Validation of Twitter opinion trends with national polling aggregates: Hillary Clinton vs Donald Trump." *Scientific Reports* 8: 8673

Campante, Filipe R., Ruben Durante, and Francesco Sobbrio. 2013. "Politics 2.0: The Multifaceted Effect of Broadband Internet on Political Participation." National Bureau of Economic Research Working Paper 19029.

Conover, M. D., J. Ratkiewicz, M. Francisco, B. Gonc¸alves, A. Flammini, F. Menczer 2011. Political Polarization on Twitter. *Proceedings of the Fifth International Conference on Weblogs and Social Media.*

DellaVigna, Stefano, and Ethan Kaplan. 2007. "The Fox News Effect: Media Bias and Voting." *Quarterly Journal of Economics* 122 (3): 1187–1234.

Enikolopov, Ruben, Alexey Makarin, and Maria Petrova 2020. "Social Media and Protest Participation: Evidence from Russia." *Econometrica*, 88(4), 1479-1514.

Falck, Oliver, Robert Gold, and Stephan Heblich. 2014. "E-lections: Voting Behavior and the Internet." *American Economic Review* 104 (7):2238-2265.

Fletcher, Richard, and Rasmus Kleis Nielsen. 2018. "Are people incidentally exposed to news on social media? A comparative analysis" *New Media & Society* 20 (7)

Fujiwara, Thomas, Karsten Müller, and Carlo Schwarz. 2021. "The effect of social media on elections: Evidence from the United States." No. w28849. National Bureau of Economic Research.

Gavazza, Alessandro, Mattia Nardotto and Tommaso Valletti. 2019. "Internet and politics: Evidence from UK local elections and local government policies." *The Review of Economic Studies* 86.5: 2092-2135.

Gentzkow, Matthew. 2006. "Television and Voter Turnout." *Quarterly Journal of Economics* 121 (3): 931-72.

Gentzkow, Matthew. 2016. "Polarization in 2016." *Toulouse Network for Information Technology Whitepaper.*

Gentzkow, Matthew, Jesse M. Shapiro, and Michael Sinkinson. 2011. "The Effect of Newspaper Entry and Exit on Electoral Politics." *American Economic Review* 101 (7): 2980-3018.

Hall, Andrew, and James M. Snyder. 2015. "Information and Wasted Votes: A Study of U.S. Primary Elections." *Quarterly Journal of Political Science* 10 (4):433-459.

Hendel, Igal, Saul Lach, and Yossi Spiegel. 2016. "Consumers' Activism: The Cottage Cheese Boycott." Working paper.

Huszár, Ferenc, Sofia Ira Ktena, Conor O'Brien, Luca Belli, Andrew Schlaikjer, and Moritz Hardt. 2022. "Algorithmic amplification of politics on Twitter." Proceedings of the National Academy of Sciences 119.1: e2025334119.

Jurgens, David, Tyler Finnethy, James McCorriston, Yi Tian Xu, and Derek Ruths. 2015. "Geolocation Prediction in Twitter Using Social Networks: A Critical Analysis and Review of Current Practice." Proceedings of the 9th International AAAI Conference on Web and Social Media (ICWSM).

Larcinese, Valentino, and Luke Miner. 2017. The Political Impact of the Internet in US Presidential Elections. Working Paper.

Leip, David. *Dave Leip's Atlas of U.S. Presidential Elections.*

Lorenz-Spreen, Philipp, Lisa Oswald, Stephan Lewandowsky, and Ralph Hertwig. 2022. "A systematic review of worldwide causal and correlational evidence on digital media and democracy." *Nature human behaviour*: 1-28.

Margetts, Helen, Peter John, Scott Hale, and Taha Yasseri. 2016. "Political Turbulence: How Social Media Shape Collective Action." *Princeton University Press.*

Petrova, Maria, Ananya Sen, and Pinar Yildirim. 2017. "Social Media and Political Donations: New Technology and Incumbency Advantage in the United States." Working Paper.

Snyder, James M., Jr., and David Strömberg. 2010. "Press Coverage and Political Accountability." *Journal of Political Economy* 118 (2): 355-408.

Stephens-Davidowitz, Seth, and Hal Varian. 2015. A Hands-on Guide to Google Data. Working Paper.

Strömberg, David. 2004. "Radio's Impact on Public Spending." *Quarterly Journal of Economics* 119 (1): 189-221.

Sunstein Cass R. 2017 "#Republic. Divided Democracy in the Age of Social Media" Princeton University Press.

Tucker, Joshua A., Andrew Guess, Pablo Barberá, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal, and Brendan Nyhan 2018. "Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature".

Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov. 2020. "Political effects of the internet and social media." *Annual review of economics* 12: 415-438.

Tables

Table 10: First Stage Regression

	(1) Twitter	(2) Twitter	(3) Twitter	(4) Twitter
Draft	5.392***	3.744***	3.341***	3.397***
	(0.893)	(0.726)	(0.742)	(0.749)
Lottery	-1.696***	-1.526***	-1.292***	-1.341***
	(0.382)	(0.324)	(0.346)	(0.344)
Population		0.0134***	0.00961***	0.0104***
		(0.00205)	(0.00210)	(0.00220)
Male		-2.353***	-2.378**	-2.184**
		(0.762)	(0.985)	(0.958)
Age - under 18 (share)			-0.477	-0.437
			(0.500)	(0.498)
Age - over 65 (share)			-1.292***	-1.157***
			(0.409)	(0.412)
Race - White (share)			0.170	0.186
			(0.157)	(0.159)
Race - Black (share)			0.0973	0.120
			(0.422)	(0.414)
Bachelor's degree of higher			1.067***	1.003***
			(0.260)	(0.254)
Average Income			-0.561***	-0.545***
			(0.206)	(0.203)
Income higher 200k (share)			1.815**	1.775**
			(0.707)	(0.700)
Income lower 10k (share)			-0.253	-0.253
			(0.275)	(0.274)
Internet Penetration				-0.960*
				(0.574)
Observations	621	621	621	621
F-Stat	31.99	26.63	20.30	20.56

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 11: OLS estimates of Twitter on electoral outcomes

	(1) Turnout	(2) % Dem	(3) % Rep	(4) Votes Dem	(5) Votes Rep
Twitter	-0.179	1.236***	-0.899*	-11.28	-5.975
	(0.369)	(0.428)	(0.474)	(7.574)	(6.351)
Population	0.126	0.623***	-0.645**	14.24*	-6.786
	(0.183)	(0.193)	(0.251)	(7.323)	(7.178)
Male	-0.575	-2.368***	3.056***	7.069	-13.52*
	(0.603)	(0.650)	(0.688)	(7.916)	(6.889)
Age - under 18 (share)	0.672*	-2.626***	2.365***	14.81*	20.51***
	(0.381)	(0.384)	(0.442)	(7.550)	(6.715)
Age - over 65 (share)	0.464	-1.880***	1.597***	6.648	4.474
	(0.321)	(0.379)	(0.427)	(5.396)	(3.907)
Race - White (share)	0.241***	0.0852	-0.0751	4.109***	2.325**
	(0.0860)	(0.0798)	(0.0885)	(1.246)	(1.026)
Race - Black (share)	0.637*	-1.116***	1.405***	8.210	16.80***
	(0.339)	(0.326)	(0.345)	(7.083)	(5.620)
Bachelor's degree of higher	0.510**	-0.477**	0.461*	-5.926*	10.47***
	(0.247)	(0.231)	(0.236)	(3.225)	(2.842)
Average Income	-0.471	-3.906**	1.934	28.10	-30.81
	(1.681)	(1.949)	(2.054)	(22.78)	(21.03)
Income higher 200k (share)	-0.590	1.966***	-1.968***	-28.03**	-8.736
	(0.551)	(0.623)	(0.654)	(11.03)	(8.774)
Income lower 10k (share)	-0.184	-0.267	0.124	-0.905	-4.971*
	(0.171)	(0.231)	(0.246)	(2.652)	(2.600)
Internet Penetration	-0.136	0.631	-0.982**	-2.072	1.458
	(0.369)	(0.423)	(0.440)	(5.372)	(4.395)
Observations	621	621	621	621	621
R-squared	0.962	0.978	0.972	0.998	0.995
DMA FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Dep. Var. mean	55.98	43.82	53.26	318	292.3
Dep. Var. sd	8.107	11.56	11.37	552.5	347.6

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 12: OLS estimates of Twitter on donations

	(1) Donations < 1k	(2) Donations Dem < 1k	(3) Donations Rep < 1k
Twitter	-527.9**	-722.5***	194.6***
	(250.9)	(266.7)	(68.30)
Population	13.66***	11.44***	2.220***
	(3.443)	(3.728)	(0.535)
Male	686.6*	877.2**	-190.6**
	(371.1)	(398.2)	(75.95)
Age - under 18 (share)	-607.7**	-612.6**	4.897
	(254.7)	(253.8)	(48.95)
Age - over 65 (share)	189.5	227.7	-38.21
	(176.2)	(172.5)	(38.29)
Race - White (share)	-154.5***	-153.2***	-1.284
	(51.94)	(53.59)	(10.08)
Race - Black (share)	-395.7*	-443.8*	48.13
	(232.6)	(227.1)	(47.20)
Bachelor's degree of higher	2.025	-19.89	21.91
	(123.0)	(123.3)	(26.36)
Average Income	-45.16	-6.569	-38.59*
	(110.7)	(116.0)	(20.47)
Income higher 200k (share)	1,566***	1,419***	146.5
	(448.9)	(458.3)	(92.11)
Income lower 10k (share)	163.7	172.9	-9.246
	(119.9)	(124.8)	(21.35)
Internet Penetration	-257.7	-239.1	-18.56
	(227.0)	(235.6)	(46.48)
Observations	621	621	621
R-squared	0.977	0.960	0.981
DMA FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Dep. Var. mean	7200	1797	1003
Dep. Var. sd	17101	5179	1628

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 13: Turnout - IV Regressions

	(1) Turnout	(2) Turnout	(3) Turnout	(4) Turnout
Twitter	-0.709	-1.352	-1.734	-1.690
	(1.077)	(1.282)	(1.504)	(1.492)
Lottery	-0.0338	-0.169	-0.0904	-0.0992
	(0.234)	(0.225)	(0.241)	(0.242)
Population	0.00123	0.00111	0.00293	0.00313
	(0.00299)	(0.00292)	(0.00243)	(0.00251)
Male (share)		-1.286*	-1.228	-1.153
		(0.717)	(0.929)	(0.907)
Age - under 18 (share)			0.619	0.634
			(0.391)	(0.393)
Age - over 65 (share)			0.202	0.252
			(0.420)	(0.406)
Race - White (share)			0.268***	0.272***
			(0.0920)	(0.0925)
Race - Black (share)			0.650*	0.657*
			(0.351)	(0.350)
Bachelor's degree of higher			0.743**	0.717**
			(0.345)	(0.332)
Average Income			-0.178	-0.170
			(0.197)	(0.193)
Income higher 200k (share)			-0.154	-0.177
			(0.690)	(0.680)
Income lower 10k (share)			-0.262	-0.261
			(0.196)	(0.194)
Internet Penetration				-0.308
				(0.415)
Observations	621	621	621	621
R-squared	0.547	0.542	0.557	0.559
Number of DMA	207	207	207	207
DMA FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
F-Stat	31.99	26.63	20.30	20.56
Dep. Var. mean	55.98	55.98	55.98	55.98
Dep. Var. sd	8.107	8.107	8.107	8.107

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 14: Democratic Party, vote share - IV Regressions

	(1) Democratic Party %	(2) Democratic Party %	(3) Democratic Party %	(4) Democratic Party %
Twitter	-2.157	-4.240*	-3.866*	-3.875*
	(1.931)	(2.497)	(2.184)	(2.153)
Lottery	0.203	-0.236	-0.229	-0.228
	(0.328)	(0.286)	(0.259)	(0.261)
Population	0.0218***	0.0214***	0.0126***	0.0126***
	(0.00439)	(0.00462)	(0.00319)	(0.00334)
Male (share)		-4.169***	-4.225***	-4.241***
		(1.262)	(1.192)	(1.161)
Age - under 18 (share)			-2.732***	-2.735***
			(0.477)	(0.475)
Age - over 65 (share)			-2.596***	-2.606***
			(0.578)	(0.569)
Race - White (share)			0.199	0.198
			(0.127)	(0.128)
Race - Black (share)			-1.056**	-1.057**
			(0.431)	(0.431)
Bachelor's degree of higher			0.230	0.235
			(0.421)	(0.408)
Average Income			-0.803***	-0.805***
			(0.293)	(0.289)
Income higher 200k (share)			3.353***	3.357***
			(0.975)	(0.967)
Income lower 10k (share)			-0.511	-0.511
			(0.334)	(0.333)
Internet Penetration				0.0640
				(0.571)
Observations	621	621	621	621
R-squared	0.647	0.589	0.702	0.701
Number of DMA	207	207	207	207
DMA FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
F-Stat	31.99	26.63	20.30	20.56
Dep. Var. mean	43.82	43.82	43.82	43.82
Dep. Var. sd	11.56	11.56	11.56	11.56

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 15: Republican Party, vote share - IV Regressions

	(1) Republican Party %	(2) Republican Party %	(3) Republican Party %	(4) Republican Party %
Twitter	1.271	3.154	3.033	3.114
	(1.794)	(2.273)	(2.049)	(2.024)
Lottery	-0.397	-4.59e-05	-0.00884	-0.0249
	(0.321)	(0.278)	(0.249)	(0.252)
Population	-0.0209***	-0.0205***	-0.0119***	-0.0115***
	(0.00440)	(0.00452)	(0.00335)	(0.00343)
Male (share)		3.768***	4.236***	4.374***
		(1.169)	(1.139)	(1.120)
Age - under 18 (share)			2.390***	2.418***
			(0.485)	(0.480)
Age - over 65 (share)			2.095***	2.187***
			(0.575)	(0.568)
Race - White (share)			-0.186	-0.179
			(0.116)	(0.118)
Race - Black (share)			1.363***	1.375***
			(0.410)	(0.415)
Bachelor's degree of higher			-0.0766	-0.124
			(0.401)	(0.392)
Average Income			0.504*	0.519*
			(0.291)	(0.288)
Income higher 200k (share)			-3.004***	-3.045***
			(0.961)	(0.950)
Income lower 10k (share)			0.285	0.287
			(0.322)	(0.322)
Internet Penetration				-0.563
				(0.544)
Observations	621	621	621	621
R-squared	0.383	0.326	0.466	0.464
Number of DMA	207	207	207	207
DMA FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
F-Stat	31.99	26.63	20.30	20.56
Dep. Var. mean	53.26	53.26	53.26	53.26
Dep. Var. sd	11.37	11.37	11.37	11.37

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 16: Democratic Party, Number of votes - IV Regressions

	(1) Votes Dem	(2) Votes Dem	(3) Votes Dem	(4) Votes Dem
Twitter	-66.31***	-74.11***	-71.33***	-70.06***
	(18.21)	(20.60)	(22.85)	(22.52)
Lottery	-4.691	-6.334*	-4.554	-4.807
	(3.619)	(3.408)	(3.043)	(3.007)
Population	0.174**	0.172**	0.209***	0.215***
	(0.0844)	(0.0843)	(0.0765)	(0.0778)
Male (share)		-15.62	-18.28	-16.11
		(11.11)	(13.72)	(13.28)
Age - under 18 (share)			12.76	13.21*
			(7.801)	(7.857)
Age - over 65 (share)			-2.951	-1.504
			(7.019)	(6.875)
Race - White (share)			5.139***	5.253***
			(1.797)	(1.805)
Race - Black (share)			8.865	9.060
			(7.676)	(7.629)
Bachelor's degree of higher			2.749	1.997
			(4.928)	(4.680)
Average Income			-2.193	-1.957
			(3.411)	(3.311)
Income higher 200k (share)			-11.23	-11.88
			(14.37)	(14.07)
Income lower 10k (share)			-4.045	-4.008
			(3.712)	(3.674)
Internet Penetration				-8.881
				(7.654)
Observations	621	621	621	621
R-squared	0.071	0.028	0.102	0.113
Number of DMA	207	207	207	207
DMA FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
F-Stat	31.99	26.63	20.30	20.56
Dep. Var. mean	318	318	318	318
Dep. Var. sd	552.5	552.5	552.5	552.5

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 17: Republican Party, Number of votes - IV Regressions

	(1) Votes Rep	(2) Votes Rep	(3) Votes Rep	(4) Votes Rep
Twitter	21.83	18.61	12.39	11.92
	(17.98)	(20.61)	(20.59)	(20.32)
Lottery	-0.122	-0.800	-0.225	-0.130
	(2.373)	(2.168)	(2.516)	(2.502)
Population	-0.154*	-0.155*	-0.0883	-0.0905
	(0.0930)	(0.0924)	(0.0830)	(0.0837)
Male (share)		-6.442	-6.849	-7.659
		(9.324)	(9.363)	(9.301)
Age - under 18 (share)			20.91***	20.75***
			(6.689)	(6.685)
Age - over 65 (share)			7.645	7.104
			(5.602)	(5.537)
Race - White (share)			1.905*	1.862*
			(1.083)	(1.091)
Race - Black (share)			16.74***	16.66***
			(5.752)	(5.740)
Bachelor's degree of higher			7.578*	7.859*
			(4.194)	(4.134)
Average Income			-1.542	-1.631
			(2.747)	(2.710)
Income higher 200k (share)			-13.78	-13.54
			(9.596)	(9.513)
Income lower 10k (share)			-4.233	-4.247
			(2.749)	(2.745)
Internet Penetration				3.323
				(4.134)
Observations	621	621	621	621
R-squared	0.017	0.032	0.147	0.149
Number of DMA	207	207	207	207
DMA FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
F-Stat	31.99	26.63	20.30	20.56
Dep. Var. mean	292.3	292.3	292.3	292.3
Dep. Var. sd	347.6	347.6	347.6	347.6

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 18: Small Donation (less 1k $), total - IV Regressions

	(1) Donations < 1k	(2) Donations < 1k	(3) Donations < 1k	(4) Donations < 1k
Twitter	-126.1	126.3	-340.3	-305.9
	(573.0)	(646.5)	(698.9)	(687.5)
Lottery	-14.58	38.57	-39.82	-46.67
	(103.2)	(86.99)	(114.6)	(115.8)
Population	16.19***	16.24***	13.20***	13.36***
	(3.782)	(3.786)	(3.338)	(3.437)
Male		505.0	666.9	725.5*
		(413.3)	(425.5)	(439.6)
Age - under 18 (share)			-624.1**	-612.0**
			(247.6)	(246.8)
Age - over 65 (share)			187.2	226.4
			(235.8)	(235.2)
Race - White (share)			-166.6***	-163.5***
			(57.83)	(56.68)
Race - Black (share)			-399.1*	-393.8*
			(231.3)	(229.5)
Bachelor's degree of higher			-15.61	-35.96
			(171.7)	(171.6)
Average Income			-33.55	-27.17
			(132.8)	(132.6)
Income higher 200k (share)			1,527***	1,510***
			(476.1)	(475.3)
Income lower 10k (share)			165.5	166.4
			(123.4)	(122.9)
Internet Penetration				-240.4
				(223.2)
Observations	621	621	621	621
R-squared	0.512	0.509	0.597	0.598
Number of DMA	207	207	207	207
DMA FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
F-Stat	31.99	26.63	20.30	20.56
Dep. Var. mean	2799	2799	2799	2799
Dep. Var. sd	6581	6581	6581	6581

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 19: Small Donation (less 1k $), Democratic Party - IV Regressions

	(1) Donations Dem < 1k	(2) Donations Dem < 1k	(3) Donations Dem < 1k	(4) Donations Dem < 1k
Twitter	-599.6	-276.1	-723.3	-688.4
	(578.6)	(653.9)	(714.9)	(704.6)
Lottery	-51.48	16.65	-60.12	-67.08
	(109.5)	(95.86)	(125.7)	(126.9)
Population	14.13***	14.20***	11.21***	11.37***
	(3.941)	(3.943)	(3.559)	(3.668)
Male (share)		647.3	778.7*	838.2*
		(437.1)	(457.6)	(474.3)
Age - under 18 (share)			-635.1**	-622.9**
			(247.7)	(247.6)
Age - over 65 (share)			199.2	239.0
			(235.6)	(235.8)
Race - White (share)			-162.1***	-158.9***
			(58.90)	(57.66)
Race - Black (share)			-444.2**	-438.9**
			(226.1)	(223.6)
Bachelor's degree of higher			-12.50	-33.18
			(171.7)	(171.7)
Average Income			-10.30	-3.820
			(138.8)	(138.3)
Income higher 200k (share)			1,433***	1,415***
			(488.2)	(487.4)
Income lower 10k (share)			164.1	165.1
			(128.3)	(127.7)
Internet Penetration				-244.2
				(228.0)
Observations	621	621	621	621
R-squared	0.400	0.402	0.498	0.499
Number of DMA	207	207	207	207
DMA FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
F-Stat	31.99	26.63	20.30	20.56
Dep. Var. mean	1797	1797	1797	1797
Dep. Var. sd	5179	5179	5179	5179

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 20: Small Donation (less 1k $), Republican Party - IV Regressions

	(1) Donations Rep < 1k	(2) Donations Rep < 1k	(3) Donations Rep < 1k	(4) Donations Rep < 1k
Twitter	473.5***	402.4***	383.0**	382.4**
	(132.6)	(142.6)	(170.3)	(166.9)
Lottery	36.90	21.92	20.30	20.41
	(28.67)	(30.28)	(26.37)	(26.24)
Population	2.058***	2.044***	1.992***	1.989***
	(0.513)	(0.514)	(0.516)	(0.530)
Male (share)		-142.2	-111.7	-112.7
		(88.06)	(100.3)	(99.47)
Age - under 18 (share)			11.03	10.83
			(48.95)	(49.18)
Age - over 65 (share)			-12.00	-12.63
			(47.74)	(46.48)
Race - White (share)			-4.530	-4.579
			(10.86)	(11.02)
Race - Black (share)			45.11	45.02
			(47.84)	(47.89)
Bachelor's degree of higher			-3.107	-2.781
			(37.33)	(36.00)
Average Income			-23.25	-23.35
			(26.38)	(25.81)
Income higher 200k (share)			94.21	94.49
			(113.0)	(111.5)
Income lower 10k (share)			1.386	1.370
			(23.74)	(23.67)
Internet Penetration				3.852
				(54.08)
Observations	621	621	621	621
R-squared	0.559	0.576	0.584	0.584
Number of DMA	207	207	207	207
DMA FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
F-Stat	31.99	26.63	20.30	20.56
Dep. Var. mean	1003	1003	1003	1003
Dep. Var. sd	1628	1628	1628	1628

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 21: Images - Compliers

VARIABLES	(1) Male	(2) Black	(3) White	(4) Age
Draft	5.407*	-4.803	0.829	1.343*
	(3.065)	(3.669)	(3.141)	(0.700)
Lottery	-0.156	-1.679***	1.764***	-0.0184
	(0.391)	(0.633)	(0.562)	(0.132)
Population	-0.241*	-0.206	0.108	-0.0482
	(0.131)	(0.240)	(0.222)	(0.0587)
Male	0.248	-3.022**	3.597***	-0.498
	(0.969)	(1.468)	(1.366)	(0.336)
Age - under 18 (share)	-0.529	-3.094***	3.037***	0.196
	(0.575)	(1.059)	(0.973)	(0.172)
Age - over 65 (share)	-0.175	-2.638***	2.826***	-0.0728
	(0.465)	(0.606)	(0.579)	(0.150)
Race - White (share)	-0.165	-0.872***	0.500***	-0.0785
	(0.260)	(0.235)	(0.189)	(0.0596)
Race - Black (share)	0.0208	0.687	-0.723	-0.307**
	(0.459)	(0.814)	(0.769)	(0.145)
Bachelor's degree of higher	0.588*	0.193	0.00972	0.368***
	(0.348)	(0.581)	(0.530)	(0.120)
Average Income	-0.242	-6.581	4.475	-1.502*
	(2.550)	(4.039)	(3.599)	(0.854)
Income higher 200k (share)	0.164	0.115	0.484	0.748***
	(0.817)	(1.373)	(1.249)	(0.269)
Income lower 10k (share)	-0.0914	-1.283**	0.879*	0.0119
	(0.357)	(0.513)	(0.462)	(0.124)
Internet Penetration	-1.759***	-1.720*	1.625**	0.0963
	(0.486)	(0.877)	(0.821)	(0.198)
Observations	621	621	621	621
R-squared	0.883	0.784	0.818	0.948
Number of id_dma	207	207	207	207
DMA FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Dep. Var. mean	50.41	20.25	64.53	38.44
Dep. Var. sd	6.280	8.739	9.083	2.899

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 22: Tweets by partisan leaning

	All			No Retweets Only		Retweets Only	
	Number of tweets	Avg. Sent.	Retweets (share)	Number of tweets	Avg. Sentiment	Number of tweets	Avg. Sentiment
All	33286	0.115	0.58	14082	0.096	19085	0.130
pro Clinton	8809	0.109	0.50	4376	0.128	4416	0.092
pro Trump	18324	0.159	0.64	6496	0.124	11789	0.178
anti Clinton	3456	-0.011	0.53	1592	0.006	1823	-0.024
anti Trump	2697	-0.007	0.39	1618	-0.021	1057	0.016

Figures

Figure 5: Screenshot of a profile on Twitter.

Figure 6: Distribution of Twitter penetration

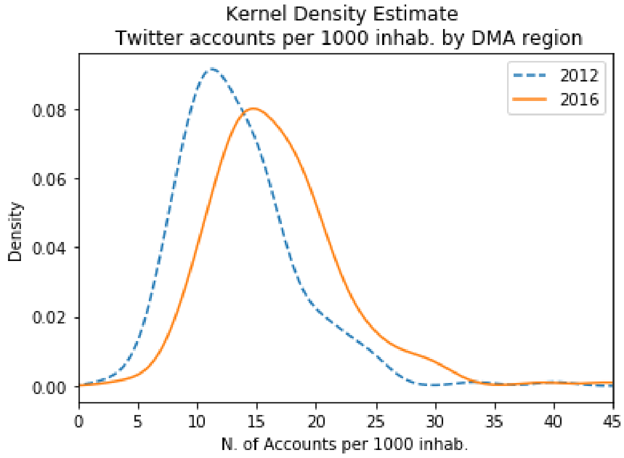


Figure 7: Tweets per day - Source: Twitter

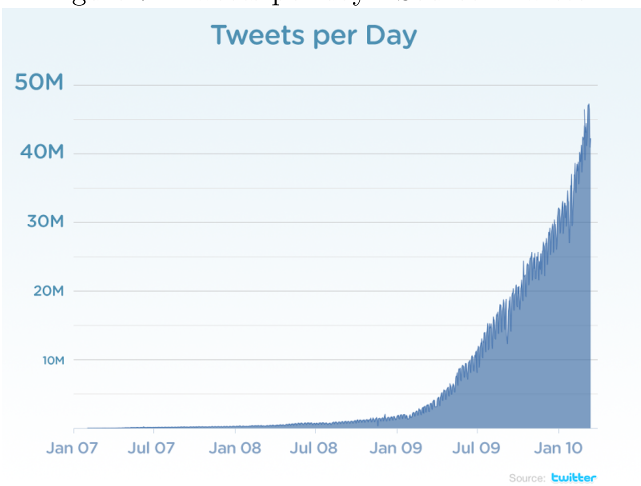


Figure 8: Distribution of accounts per 1000 inhab. by DMA in 2008

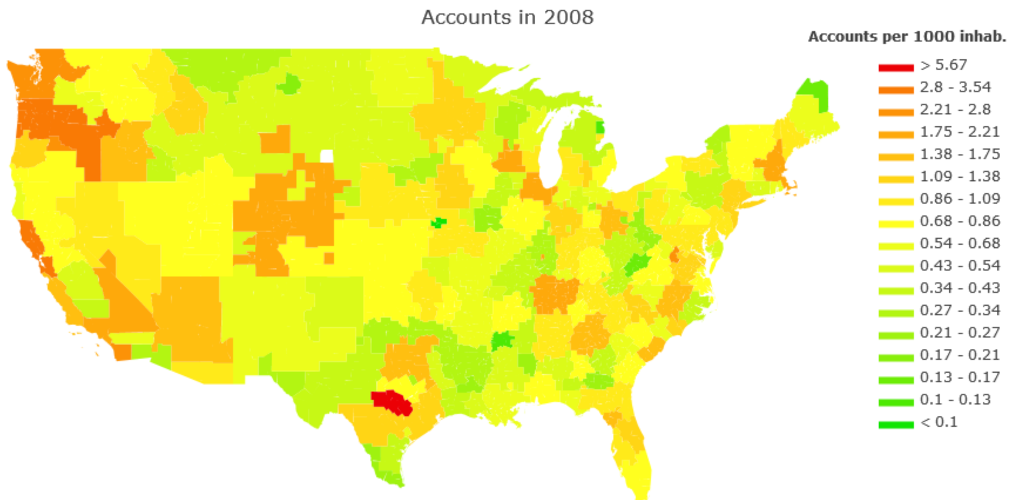


Figure 9: Distribution of accounts per 1000 inhab. by DMA in 2012

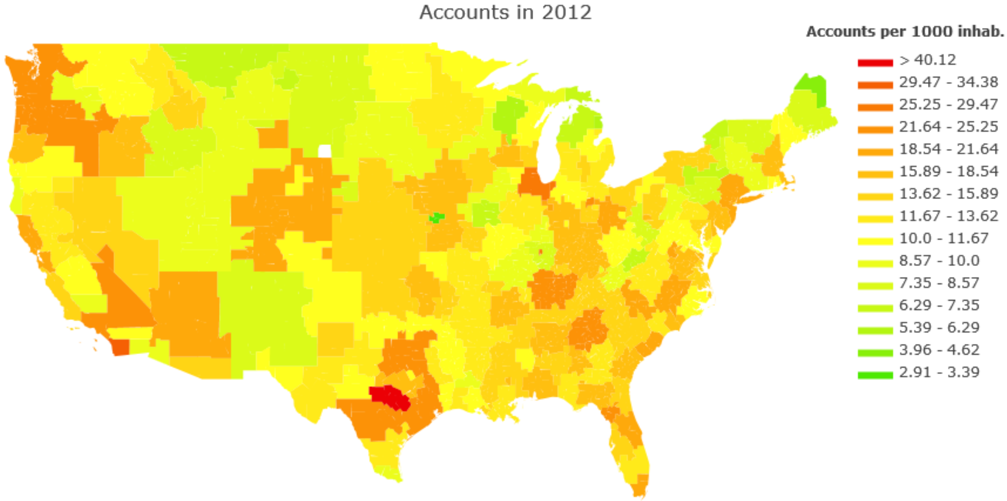


Figure 10: Age distribution - profile pictures

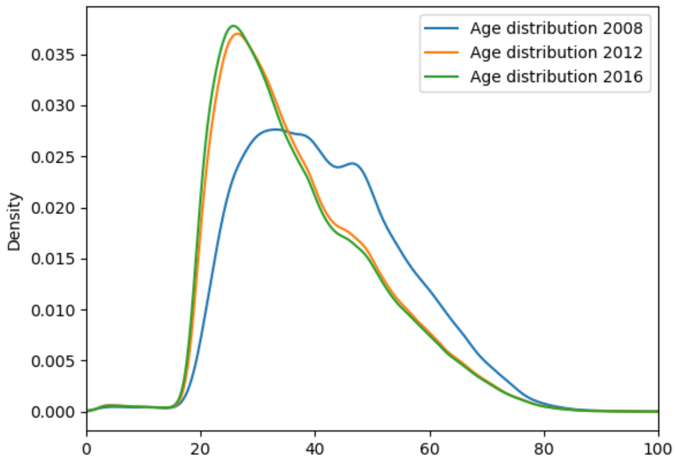


Figure 11: Distribution of Google Trends score by DMA - Boston Celtics

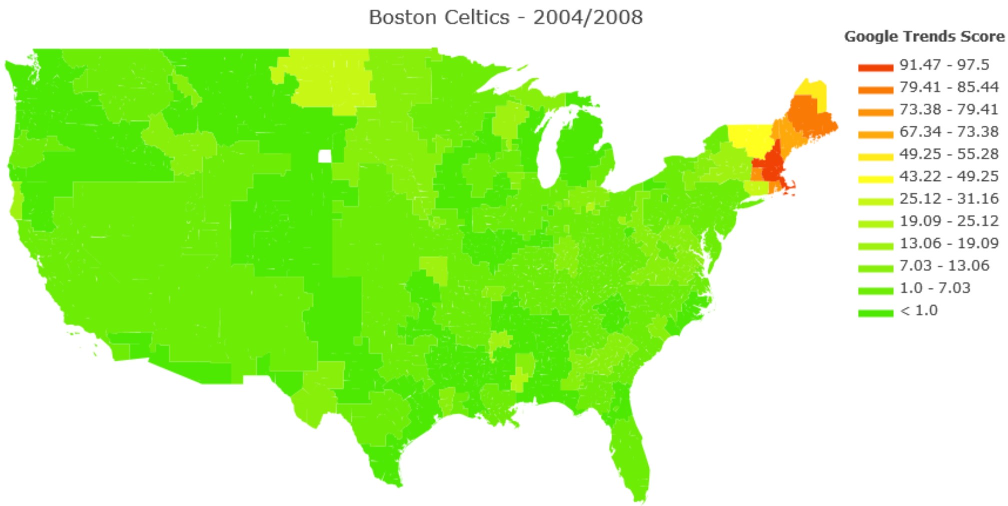


Figure 12: Distribution of Google Trends score by DMA - Minnesota Timberwolves

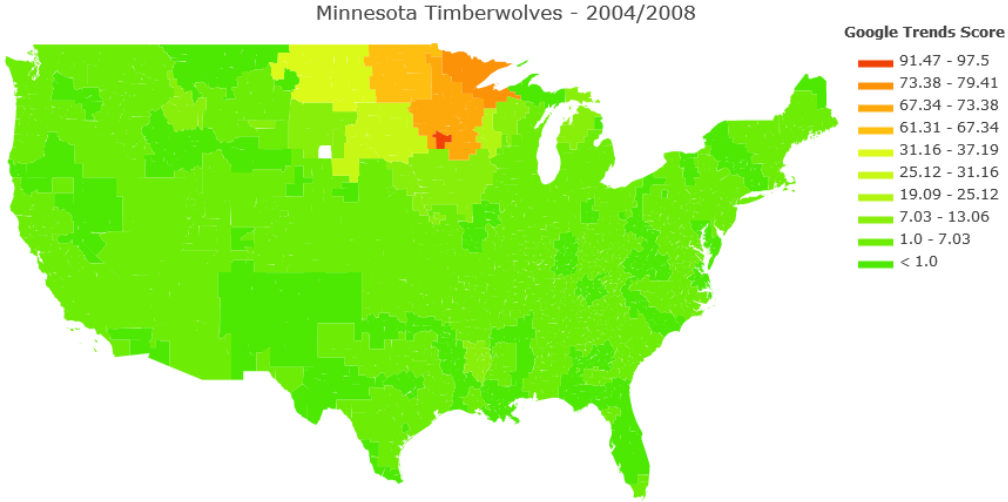


Figure 13: Most users never write about politics

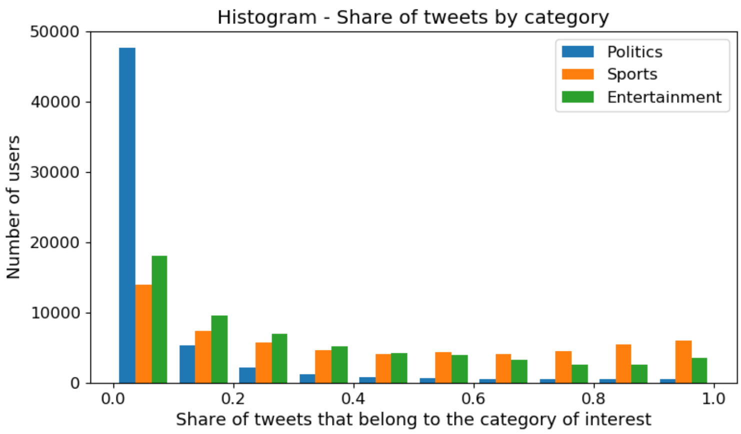
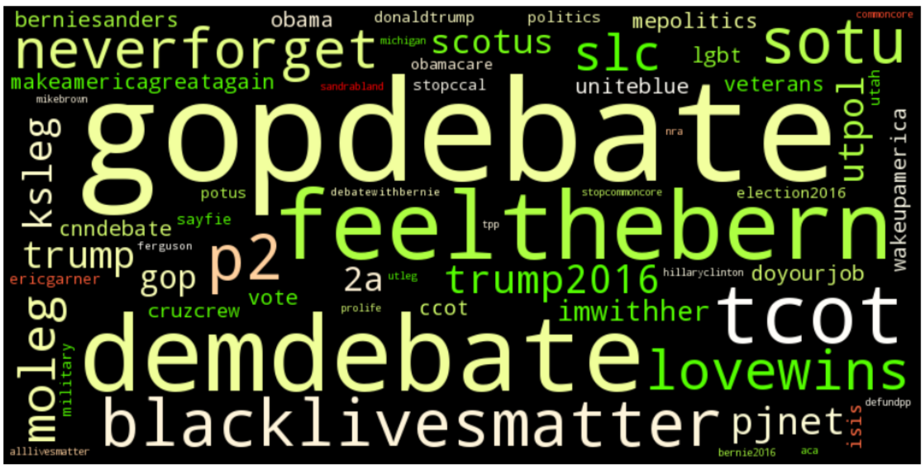
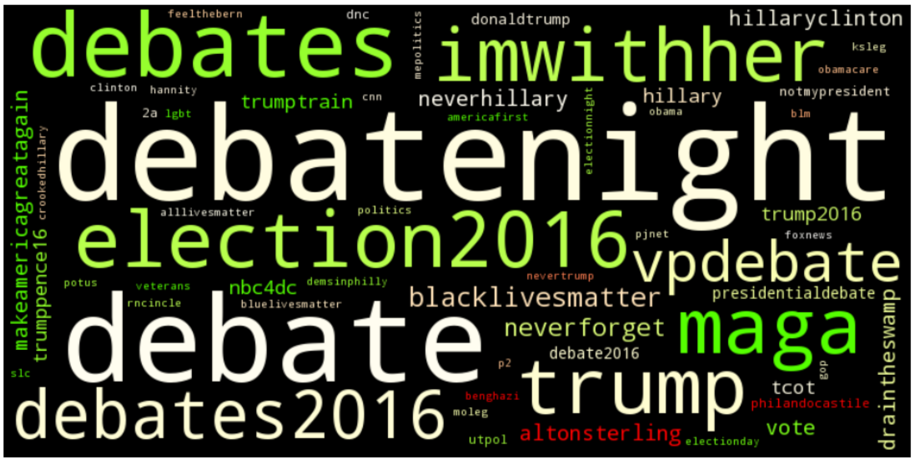


Figure 14: Word cloud - Most popular political hashtags during the first phase of 2016 electoral campaign



Note: This figure represents the most popular hashtags in the "Politics" category. Size is proportional to the number of times each hashtag was used. Colors refer to sentiment. Green words are associated with an average positive sentiment, while red words with a negative one.

Figure 15: Word cloud - Most popular political hashtags during the last phase of 2016 electoral campaign



Note: This figure represents the most popular hashtags in the "Politics" category. Size is proportional to the number of times each hashtag was used. Colors refer to sentiment. Green words are associated with an average positive sentiment, while red words with a negative one.

A Twitter

Twitter is a microblogging platform that allows users to publish short messages, tweets (max 140 characters), that are received by their followers. Tweets can then be shared with others or commented, possibly creating complex discussions involving a high number of participants. The website was launched in July 2006 and quickly became a mass phenomenon. In 2015 Twitter still ranked in the top 10 most popular websites[30], with approximately 66 million active users in the US and 320 worldwide. A survey made by PewResearch in 2014 shows that 21% of respondents were using Twitter. With respect to Facebook, the first social network in term of users, there are some relevant differences. In particular, from the beginning Twitter has appeared to be focused on the public sphere while the other was marketed as a tool to stay in touch with friends. This difference is evident under two aspects. First, Twitter accounts are public, while on Facebook there is a higher attention to privacy. Second, links on Twitter are unidirectional ("followers"), while on Facebook they are reciprocal ("friendship"). These differences are also reflected in the way users exploit the network. In particular 41% of users on Twitter say that reading comments by politicians, celebrities or athletes is a reason they use the website[31], share that is significantly higher than for Facebook. These facts motivate the identification strategy that I am using, as the presence of celebrities should affect users' interest in the platform and therefore Twitter penetration.

[30] Source: http://www.alexa.com/topsites

[31] For 11%, a major reason, 30% a minor reason. Source, Pew Research: http://www.pewinternet.org/2011/11/15/why-americans-use-social-media/

B Information and Polarization Measures

The variable *No Information* used in the text is constructed using answers to questions:

- *Do you approve of the way each is doing their job... [Incumbent Representative's Name]*

  Considered 1 if Answer equal to *"Never Heard / Not Sure"*

- *Do you approve of the way each is doing their job... [Incumbent Senator's Name]*

  Asked for two Senators, considered 1 if Answer equal to *"Never Heard / Not Sure"*

To calculate polarization scores I use answers to the following questions:

- **ideo5**: *In general, how would you describe your own political viewpoint?*

  The asnwers were rescaled in order to vary from -3 (Very Liberal) to +3 (Very Conservative).

- **pid7**: *Generally speaking, do you see yourself as a...?*

  The asnwers were rescaled in order to vary from -3 (Strong Democrat) to +3 (Strong Republican).

Following Boxell et al.(2018), Mason (2015), and Abramowitz and Saunders (2008) I used the following formulas.

Partisan Sorting:

$$\sum_{i \in S_t} [g(|pid7_i - ideo5_i| + 1)(|pid7_i| + 1)(|ideo5_i| + 1) - 7] \frac{1}{105}$$

where:

- S_t denotes the set of all respondents who did not answer *"Not Sure"* neither to *ideo5* not to *pid7*.
- $\gamma(x) = \max_{i \in \bigcup_t S_t}(|pid7_i - ideo5_i| + 1) + \min_{i \in \bigcup_t S_t}(|pid7_i - ideo5_i| + 1) - x$

Partisan Ideology:

$$\sum_{i \in R_t} ideo5_i - \sum_{i \in D_t} ideo5_i$$

where:

- $R_t := \{i : pid7_i > 1\}$
- $D_t := \{i : pid7_i < -1\}$

C Locations

When downloading information on accounts' locations, there are four cases that are typical. The user can:

- Specify a location using GPS.
- Indicate a location that corresponds to a clearly identifiable place.
- Indicate a location that does not match with any place.
- Decide not to provide any information.

I collected a random sample of user ids and matched locations to Counties in the US. Table 23 summarizes the results of this operation when considering a subsample of approximately 33 million accounts. Over the 33 million accounts in the subsample, 69% of them did not include any location. I could match in total 6 million accounts, of which 1.5 million at the county level. The remaining 4.5 million are either foreign users or accounts that I could only associate to a country or a state. Table 23 reports also the average number of tweets, the average number of followers and the average number of likes for these subsamples. We can notice that on average, the accounts that leave the location field empty appear to be less active that the others. Moreover, if we select only accounts with at least 100 tweets, two out of three of them are providing some location. Selection is therefore likely to operate towards the most active accounts.

Table 23: Collected accounts

	Accounts	%	Avg n. Tweets	Avg n. Followers	Avg n. Likes
Total	33,129,071	100	1,009	201	147
Empty location	22,956,140	69	392	65	63
Some location	10,172,931	31	2,400	506	338
Matched	6,056,009	18	1,693	516	266
Matched to county	1,523,558	4.6	1,737	658	364
Empty, 100+ tweets	2,182,070		4,084	622	640
Some, 100+ tweets	4,418,421		5,504	1,129	768