PhD THESIS DECLARATION
(use a pc to fill it out)

I, the undersigned

FAMILY NAME	Rotesi
NAME	Tiziano
Student ID no.	1759828

Thesis title:

Essays in Political Economy and Social Networks

PhD in	Economics and Finance
Cycle	29
Student's Advisor	Eliana La Ferrara
Calendar year of thesis defence	2019

DECLARE

under my responsibility:

1) that, according to Italian Republic Presidential Decree no. 445, 28th December 2000, mendacious declarations, falsifying records and the use of false records are punishable under the Italian penal code and related special laws. Should any of the above prove true, all benefits included in this declaration and those of the temporary "embargo" are automatically forfeited from the beginning;

2) that the University has the obligation, according to art. 6, par. 11, Ministerial Decree no. 224, 30th April 1999, to keep a copy of the thesis on deposit at the "Biblioteche Nazionali Centrali" (Italian National Libraries) in Rome and Florence, where consultation will be permitted, unless there is a temporary "embargo" protecting the rights of external bodies and the industrial/commercial exploitation of the thesis;

3) that the Bocconi Library will file the thesis in its "Archivio Istituzionale ad Accesso Aperto" (Institutional Registry) which permits online consultation of the complete text (except in cases of temporary "embargo");

4) that, in order to file the thesis at the Bocconi Library, the University requires that the thesis be submitted online by the student in unalterable format to Società NORMADEC (acting on behalf of the University), and that NORMADEC will indicate in each footnote the following information:

- PhD thesis (*write thesis title*) Essays on Applied Microeconomics;
- by (*Student's family name and name)* Tiziano Rotesi;

- defended at Università Commerciale "Luigi Bocconi" – Milano in the year (year of defence) 2019 ;
- the thesis is protected by the regulations governing copyright (Italian law no. 633, 22nd April 1941 and subsequent modifications). The exception is the right of Università Commerciale "Luigi Bocconi" to reproduce the same, quoting the source, for research and teaching purposes;

- <u>**only**</u> **when a separate "Embargo" Request has been undersigned**: the thesis is subject to "embargo" for (indicate duration of the "embargo") .... months;

5) that the copy of the thesis submitted online to Normadec is identical to the copies handed in/sent to the members of the Thesis Board and to any other paper or digital copy deposited at the University offices, and, as a consequence, the University is absolved from any responsibility regarding errors, inaccuracy or omissions in the contents of the thesis;

6) that the contents and organization of the thesis is an original work carried out by the undersigned and does not in any way compromise the rights of third parties (Italian law, no. 633, 22nd April 1941 and subsequent integrations and modifications), including those regarding security of personal details; therefore the University is in any case absolved from any responsibility whatsoever, civil, administrative or penal, and shall be exempt from any requests or claims from third parties;

7) that the thesis is not subject to "embargo", i.e. that it is not the result of work included in the regulations governing industrial property; it was not written as part of a project financed by public or private bodies with restrictions on the diffusion of the results; is not subject to patent or protection registrations.

Date 10/12/2018 Tiziano Rotesi

Abstract

In the first chapter I study the effect of Twitter on political participation. I address this question by studying how the spread of this social network affected voting behavior and donations to politicians during the last three US presidential elections. First, I develop a novel measure of Twitter penetration by using location data collected from users. To address endogeneity in the diffusion of Twitter across regions, I exploit variation in the popularity of sport teams that signed new players with Twitter accounts, making therefore the social network more interesting for their fans. Instrumental variables estimates do not show significant effects of Twitter on average participation. On the other hand, I find a differential effect across parties, with the Democratic Party being penalized. I provide two pieces of evidence on mechanisms. First, I show that Twitter reduces voters' information about politics and increases political polarization. Second, I show that on Twitter, peaks in interest happen towards politics only during presidential debates, when both the quantity of partisan tweets and the average sentiment favor the Republican Party.

The second chapter is coauthored with Eleonora Patacchini and Paolo Pin. Using an app for smartphones we run an experiment among high school students to study the pattern of aggregation of sparsely distributed information when competing agents are arranged in small networks and can share only non-verifiable pieces of information. Our first finding is that the level of cooperation is high, especially among students that belong to the same class. Nevertheless the level of centralization of the network significantly affects the final results, with the most central node benefiting in terms of payoffs. By adding a second node with a high centrality we see that the results change significantly, with more signals traveling through the links. We then turn to a parsimonious level-k approach to characterize players according to their behavior in the game. When estimating the model we see that data are consistent with a vast majority of players acting cooperatively, with the difference across networks driven mainly by a small share of strategic players.

Finally, the third chapter studies manipulation of reviews on Amazon.com. I identify the presence of fake reviews by combining two elements. First, I look at characteristics of the platform and in particular the way average ratings are rounded to the closest half star, thereby creating discontinuities in the effects that new positive reviews can have. Second, I exploit differences across business models by comparing traditional publishing companies with writers who self-publish their books. This allows me to frame the question as a difference-in-differences analysis. I find evidence that manipulation is at work.

Chapter 1

Do Social Media matter? The impact of Twitter on Political Participation

1.1 Introduction

During the last decade, Internet has changed the way we communicate and interact. Together with the strong increase in Internet penetration, the use of the Web has moved to a new model. Users are now the main source of content and no longer play a passive role. This introduces the possibility for the platforms that are driving this change to affect not only the way we collect information, but also the way we influence each other. Twitter[1] in particular has gained a relevant spot in the public debate so that it is now common to read in the news the last tweets by Donald Trump or users' reactions to them. Additionally, this platform has gained attention as one of the factors behind Barack Obama's success in 2008[2] or the protest that took place during the Arab Spring[3]. While the role played by Twitter in shaping

[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 and share with others.

[2] Larcinese and Miner (2017) study how internet affected 2008 Presidential Elections give a description of how Obama campaign made a 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.

political outcomes has been extensively debated, little if any rigorous evidence exists to isolate its causal effect.

This paper provides causal evidence on the effect of Twitter on political participation in the US, during the last three presidential elections. I study the impact of Twitter on electoral turnout, donations received by the Democratic and Republican parties, and vote shares of the two parties. I also try to shed light on the mechanisms underlying my results by applying machine learning tools to characterize Twitter users and analyze the content of their tweets.

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 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 would have been hardly accessible otherwise. This could make users more engaged politically, in particular in areas and in moments characterized by a stronger 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

[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).

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

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 the presence 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 the political debate at the local level could 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, candidates could ask their supporters to join the platform in order to help in the campaign.

My proposed instrumental variables strategy hinges on the influence that celebrities exert on their fans. Twitter offers content that is created by its users: the more interesting these users are, the more interesting the content. The fact that celebrities are on Twitter also makes their fans more willing to join the platform, in order to receive messages posted by these celebrities[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. This population of players has two important characteristics. First,

[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 the 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.

the best players that participate to 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.

My IV estimates show that the impact of increasing the number of Twitter accounts on average political participation is weak: neither electoral turnout nor total amount of campaign are affected. However, when one distinguishes between parties, interesting differences emerge. The impact of Twitter is negative for the Democratic Party and positive for the Republican Party. In particular, we see a reduction 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 people who open an account because of their interest in basketball. To learn more about the mechanism at work it is therefore necessary to study compliers. 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, gender and race - to these pictures. Results from approximately 1 million profile pictures show that in the population of compliers, users that are male and older than 40 tend to be overrepresented. There is no clear difference for what concerns race. Since these demographics correlate with preferences for the Republican Party[10], it is necessary to be careful in extrapolating the aforementioned estimates to the whole population.

Finally, I investigate whether this pattern of results can be explained as the consequence

[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/

of a partisan debate that does not contain relevant information. To do so, I employ two additional datasets. The first is the Cooperative Congressional Election Study (CCES), a survey containing questions that allow to measure respondents' level of information about politics and political polarization.

In particular, I follow Snyder and Strömberg (2010) and use knowledge of the name of 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 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.

To code the political content of tweets I look at hashtags that can be identified as partisan as suggested by Bovet et al (2018) and I see that Republican-leaning tweets are more popular and attract a more positive sentiment on average. I then define two measures of political polarization using the CCES data and I find that Twitter increased political polarization. Overall my results therefore suggest that Twitter did not affect political attitudes by improving information available to users, but rather by facilitating the spread of a partisan debate favoring the Republican Party.

This paper contributes to the political economy literature that studies the link between media and political outcomes. Stromberg (2004), Gentzkow (2006), and Gentzkow, Shapiro, and Sinkinson (2011) estimate the effect that Radio, TV, and Newspapers had on attitudes towards politics. Closer to the question addressed in this paper, the works by Campante, Durante, and Sobbrio (2013), Falck, Gold, and Heblich (2014), Gavazza, Nardotto, and Valletti (2016), and Larcinese and Miner (2017) study the effect of broadband on voting behavior, respectively in Italy, Germany, England, ans US. In all cases but 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 as main mechanism the quality of information offered by the media. By providing new and relevant information, newspapers and radio had a positive impact on participation. On the other hand, at least at first, both TV and Internet where used as a source of entertainment, reducing consumption of traditional media. With respect to this literature, my paper contributes by focusing on social media, that can be seen as characterizing the second step in the evolution of how Internet is commonly used, with more relevance given to user-generated content. Furthermore, my results point towards the possibility that social media did not influence political attitudes mainly by distracting voters, but rather by exposing them to a partisan debate that tends to favor one party. This 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).

Another literature has studied the role played by social networks in facilitating collective action, in particular protests participation and boycotts (e.g., Acemoglu, Hassan, and Tahoun (2017) for Egypt; Enikolopov, Makarin, and Petrova (2018) for Russia; and Hendel, Lach, and Spiegel (2016) for Israel). My research question differs in that I look at an outcome -voting- that is part of ordinary life in democracies and that for which coordination or collective action issues play a smaller role.

The closest related work is probably that of Petrova, Sen, and Yildirim (2017), who show that Twitter affected political competition by increasing campaign contributions for politicians active on the platform. They focus on short run outcomes, using the exact timing politicians created new accounts on Twitter. In addition to considering different outcomes (that include donations but also voters turnout and vote shares of the different parties), I exploit a different source of variation. My IV strategy is indeed based on accounts of basketball players. This has implications for the characteristics of compliers, but also advantages in terms of plausible exogeneity with respect to political incentives.

Outside the field of Economics, the impact of Twitter and other social media on participation has attracted a lot of interest by researchers. Even though this literature suffers from a lack of identification, it highlights the possible first order mechanisms at work[11]. A first strand of papers use data from surveys to study how the use of social networking websites like Facebook or Twitter correlates with acts of participation as voting. In general the correlation is positive[12]. Another strand of literature relies on data from the platforms. Barberá and Rivero (2015) use tweets to study ideological position of users that wrote about politics and find that individuals with extreme positions are overrepresented. Barberá (2015) measures ideological position of millions of individuals and finds that users are usually embedded in ideologically 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 as the presidential elections.

1.1.1 Background - 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[13], 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

[11]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 in the main page of the website.

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

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

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[14], 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.

1.2 Data

Before describing the data, it is necessary to specify that the analysis was carried at the Designated Market Area (DMA) level. DMAs are groups of counties defined by Nielsen on the basis of television market in such a way that all counties that belong to the same DMA have a similar TV offering[15]. These regions are not the same as metropolitan areas, even though in some cases the differences are small. In total there are 210 DMA regions. The reason why I use this level of aggregation is that Google Trends data, which are used 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

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

[15] 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."

2016. In total the sample contains 207 DMA regions[16].

Outcome variables and controls are standard to this literature and are described in subsection 1.2.1. Outcomes based on survey data are presented in subsection 1.2.2. Subsection 1.2.3 briefly illustrates the measure of Twitter penetration that I will then use as main explanatory variable. Finally, subsection 1.2.4 describes the data behind the instrument.

1.2.1 Political and Census Data

I collected data at the county level on turnout and voting behavior for Presidential Election in 2008, 2012, and 2016. The source is Dave Leip Atlas[17]. Data include information on the number of valid votes and votes received by the candidates. Table 1.1 includes summary statistics for the outcome variables that I consider, once the data were aggregated at the DMA level.

Controls were downloaded from Census and include age distribution across cohorts, income, race, gender and educational attainment. Table 1.2 includes summary statistics for the variables that are included in the analysis.

Campaign contribution data were downloaded from Center for Responsive Politics[18] (CRP). This data is originally available from the Federal Election Commission (FEC) and is enriched 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 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

[16] DMA regions that belong to Alaska were not considered.
[17] Link: https://uselectionatlas.org/
[18] https://www.opensecrets.org

below 1000$, as small contributions could better represent supporters behavior. Table 1.1 contains summary statistics for these variables. It is possible to notice that the total amount of donations below 1000$ was approximately equal to 472 million dollars in 2008 and reached 754 million dollars in 2016.

Table 1.1: Outcome Variables - Summary Statistics

	2008		2012		2016	
	Mean	SD	Mean	SD	Mean	SD
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. Obs.	207		207		207	

Note: *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.

1.2.2 Voters' information and polarization

To measure voters' ideology and information, I downloaded data from the Cooperative Congressional Election Study (CCES)[19]. CCES is a survey administered by YouGov that contains answers from 50,000+ responders every election year. 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. 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 the Appendix 1.A.

1.2.3 Twitter Users

Twitter does not provide aggregated data on the number of active users and their geographic distribution. In order to build a measure of Twitter penetration across regions I relied on Twitter Search API[20] and downloaded information on accounts. Over the span of a few months I have made requests for approximately 310 million accounts.

Figure 1.5 shows the screenshot of an account page. On the left, below the profile picture, it is possible to read 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), the

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

[20] 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 1.2: Control Variables - Summary Statistics by DMA region

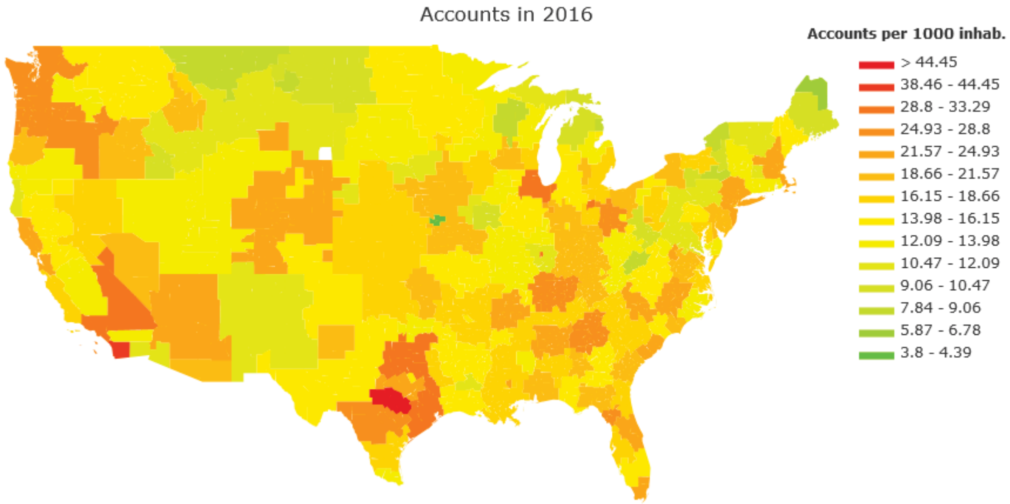
	2008		2012		2016	
	Mean	SD	Mean	SD	Mean	SD
Population	1,453	2,261	1,490	2,289	1,535	2,362
Male (share)	49.32	0.769	49.41	0.801	49.49	0.846
Age - under 18 (share)	24.16	2.571	23.60	2.467	22.92	2.462
Age - over 65 (share)	13.63	2.368	14.16	2.442	15.55	2.624
Race - White (share)	80.56	12.42	80.44	12.49	79.92	12.56
Race - Black (share)	10.51	11.50	10.70	11.57	10.81	11.53
Bachelor's degree or higher	23.10	5.785	23.98	6.023	25.55	6.265
Income lower 10k (share)	15.09	4.104	16.23	3.893	16.45	3.930
Income higher 200k (share)	2.527	1.452	2.932	1.653	3.626	2.033
Average Income	60.31	10.20	63.35	10.22	67.44	11.34
Internet Penetration	3.2	0.6	3.8	0.4	4.2	0.5
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.

number of other accounts that the user is following[21] (*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*.

[21] 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.

Figure 1.1: Distribution of accounts per 1000 inhab. by DMA in 2016



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 the location. In Appendix 1.B I provide more details regarding this process. The number of accounts that were matched at the county level were in total 6.8 million.

I then aggregated accounts at the DMA level. Figure 1.6 shows the kernel density of the number of accounts per 1000 inhabitants for the 207 DMA regions that are included in the sample. Twitter popularity grew faster starting from early 2009. This fact is confirmed by Figure 1.7, that shows the number of tweets over time (data from Twitter). Finally, Figure 1.1 shows the distribution of accounts in 2016. Figures 1.8 and 1.9 describe the distribution of accounts in 2008 and 2012 respectively.

To obtain more information regarding the population of users, I then downloaded the profile picture for a random subsample of 2 million accounts. Using machine learning algorithms, images were first selected in order to identify those which contained at least one face. Faces were then analyzed using facial recognition tools. This way I could obtain basic demographics for approximately 980,000 accounts. Table 1.3 contains summary statistics for the estimated share of male users, average age and share of black and white users. Figure 1.10 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 that is 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 1.3: 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
Black	12.2	5.2	21.9	10.1	21.9	9.7
White	73.5	7.2	63.8	9.5	63.3	9.2
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.

1.2.4 NBA players and teams' popularity

As it will become clear in Section 1.3, 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. In order to exclude the possibility that these scores were directly affected by Twitter, I downloaded them for the period 2004-2008[22]. Figures 1.11 and 1.12 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.

1.3 Empirical Specification

In order 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 1.1. The variable $Twitter_{dt}$ measures the number of accounts

[22]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.

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.

1.3.1 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 1.4). 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[23].

Table 1.4: Draft Lottery - Number of Tickets

Ranking in the Regular Season	Number of Tickets	Ranking in the Regular Season	Number of Tickets
30	250	29	199
28	156	27	119
26	88	25	63
24	43	23	28
22	17	21	11
20	8	19	7
18	6	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 http://www.nba.com/news/all-time-draft-lottery-probabilities/

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'

[23]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: http://www.nba.com/article/2017/09/28/nba-board-governors-approves-changes-draft-lottery-system .

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.

1.3.2 IV model

Equations 1.1 and 1.2 describe the IV model that I use to study how Twitter influenced political outcomes. Equation 1.2 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 1.3). With the same logic, I then control for the number of tickets each team had received for the Draft Lottery (Equation 1.4). The specification I consider is the following one:

$$Y_{dt} = \beta_0 + \beta_1 Twitter_{dt} + \beta_2 log(Tickets_{dt}) + X'_{dt}\beta_3 + \delta_t + \delta_d + \epsilon_{dt} \tag{1.1}$$

$$Twitter_{dt} = \alpha_0 + \alpha_1 log(Draft_{dt}) + \alpha_2 log(Tickets_{dt}) + X'_{dt}\alpha_3 + \delta_t + \delta_d + \nu_{dt} \tag{1.2}$$

Where:

$$Draft_{dt} = \sum_c Incoming\_Twitter_{ct} \cdot Popularity_{cd} \tag{1.3}$$

$$Lottery_{dt} = \sum_c Tickets_{ct} \cdot Popularity_{cd} \tag{1.4}$$

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

$Twitter_{dt}$ indicates the the number of accounts per 1000 inhabitants in region d at time t. $Incoming\_Twitter_{ct}$ 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. $Popularity_{cd}$ refers to the measure of popularity of team c in region d, calculated using Google Trends for the period 2004-2008.

Finally, $Tickets_{ct}$ 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 in new accounts on the social network, as some supporters will be interested in following the new member of their team[24]. Also, this effect would be magnified by network externalities, propagating among these fans' friends and relatives[25].

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 about politics.

Table 1.11 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 then total

[24]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 the newly acquired talents would therefore become aware of the presence of the social network.

[25]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.

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 6.8 million that make the total.

1.4 Results

I analyze the effects of Twitter penetration on political outcomes using the instrumental variable strategy described above in equations 1.1 and 1.2. 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 1.5. 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 1.14 to 1.21 report results for the instrumental variables regressions summarized in Table 1.5, and offer more details concerning the other control variables. Tables 1.12 and 1.13 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.

Table 1.5: 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)
Votes Dem	-66.31***	-74.11***	-71.33***	-70.06***
	(18.21)	(20.60)	(22.85)	(22.52)
Votes Rep	21.83	18.61	12.39	11.92
	(17.98)	(20.61)	(20.59)	(20.32)
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 1.2. The F-value refers to the Kleibergen-Paap F-statistic. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

1.4.1 Compliers

The estimates presented in Table 1.5, 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 1.2:

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

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, the share of white users, the share of black users and the average age. Regressions are weighted by the number of images matched to demographics characteristics for each DMA region.

Table 1.6 contains results for these regressions. Table 1.22 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 1.7 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.

1.5 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

Table 1.6: 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)
Share of Black users	-5.885	-7.974**	-4.830	-4.803
	(3.961)	(3.678)	(3.618)	(3.669)
Share of White users	1.741	4.284	0.855	0.829
	(3.368)	(3.099)	(3.117)	(3.141)
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 1.2. All regressions include DMA and election year fixed effects. Controls are discussed in Section 1.2. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.11

Table 1.7: 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.

campaign donation, with respect to voting.

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.

1.5.1 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 1.2.2. 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:

$$Y_{i(dt)} = \beta_0 + \beta_1 Twitter_{dt} + \beta_2 log(Tickets_{dt}) + X'_{idt}\beta_3 + \delta_t + \delta_d + \epsilon_i \tag{1.5}$$

$$Twitter_{dt} = \alpha_0 + \alpha_1 log(Draft_{dt}) + \alpha_2 log(Tickets_{dt}) + X'_{dt}\alpha_3 + \delta_t + \delta_d + \nu_{dt} \tag{1.6}$$

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 $Draft_{dt}$ and $Tickets_{dt}$ are defined as in Equations 1.3 and 1.4. 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 1.8. Column (1) considers regressions for the full sample of respondents. Column (2) restricts the analysis to male respondents while column (3) use only the subsample of people older than 40. This subsamples were chosen in order to match the analysis on 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.

1.5.2 On Twitter, peaks of attention and a partisan debate

To corroborate the previous findings I downloaded tweets from a sample of 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 is 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 in Conover et al (2011), that is based on hashtags. Starting from a set of hashtags that define various categories, it is possible to retrieve a wider set 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[26]. This way I obtained a corpus of 1.1

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

Table 1.8: 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 1.2.2. 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 1.2. 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

Table 1.9: 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

millions tweets divided into 10 categories.

The first two facts that emerge from these data are presented in Table 1.9 and Figure 1.2. It is possible to notice that the interest towards politics is relative 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 1.13 that shows how most of the users write about politics very rarely. Nevertheless, interest for politics peaks during the electoral debates, when it becomes the category that receives the highest number of tweets. In Table 1.9 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 which has the lowest average sentiment. This is due to a higher share of negative messages regarding this topic, cause by a higher level of conflict, especially with respect to the other categories.

Figure 1.2: Sports and Entertainment dominate the online discussion, Politics peaks just before the elections.

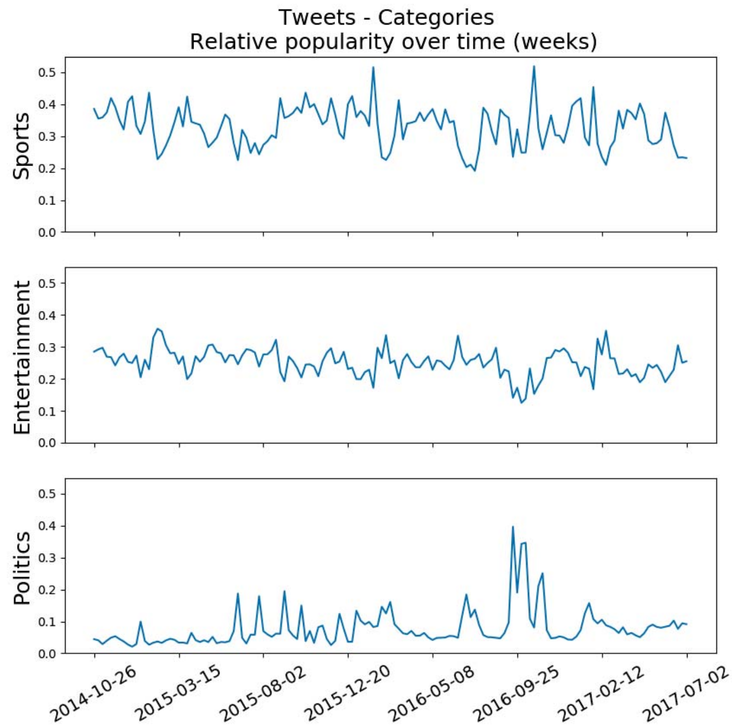


Table 1.10: Tweets written between July 1 and November 5, by partisan leaning

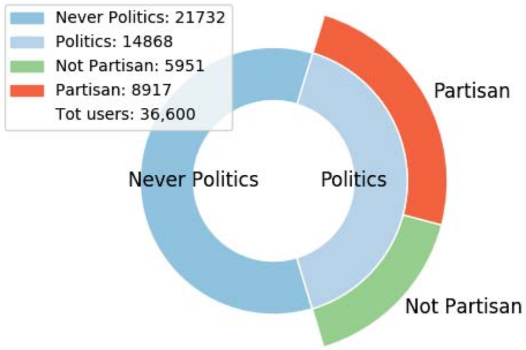
	All			No Retweets Only		Retweets Only	
	Number of tweets	Avg. Sent.	Retweets (share)	Number of tweets	Avg. Sent.	Number of tweets	Avg. Sent.
All	16617	0.112	0.571	7056	0.084	9421	0.133
pro Clinton	4293	0.111	0.472	2261	0.109	2015	0.11
pro Trump	8935	0.164	0.663	2983	0.128	5899	0.182
anti Trump	1096	-0.039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

The majority of users use Twitter to discuss about 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 write messages about that. Figures 1.14 and 1.15 show the most popular hashtags 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.

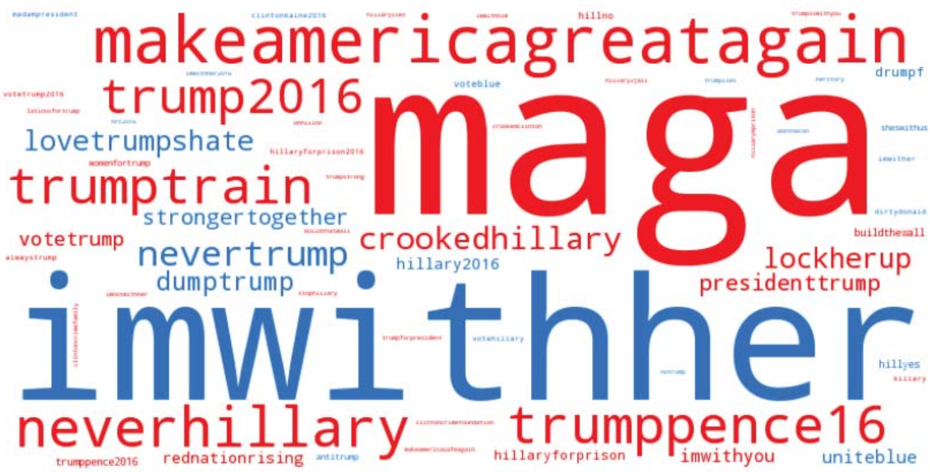
Figure 1.3 focuses on users. It contains, information regarding the share of users who tweeted about politics or not and, in case they did, whether they used hashtags with a partisan connotation. It is possible to see how approximately 40% of accounts used hashtags connected to politics during the last 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 towards politics exposed a large fraction of users to a partisan debate.

Figure 1.3: 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, the users who used at least once hashtags with a partisan leaning.

Figure 1.4: 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.

Figure 1.4 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 relative more common, suggesting a difference in the rhetoric used by the two blocs. Table 1.10 offers more details. Partisan hashtags are here dividend into four different groups: hashtags that support each of the two candidates and hashtags that attack each of the two candidates. Table 1.10 compares these groups of hashtags along three dimensions. The number of 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 wee 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 a 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 the number of anti Trump hashtags, with a lower sentiment too.

1.6 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 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's profile pictures I then showed that the population of compliers tend 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 for 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 about 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. This dynamics could in turn negatively affect attitudes towards politics for the average user, thereby reducing information and turnout, while motivate 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 Internet or traditional media on politics. This difference is likely to be determined by the different nature 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.*

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 Hernn 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. Goncalves, 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): 11871234.

Enikolopov, Ruben, Alexey Makarin and Maria Petrova. 2018. ”Social Media and Protest Participation: Evidence from Russia.” Available at SSRN https://ssrn.com/abstract=2696236

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)

Gavazza, Alessandro, Mattia Nardotto and Tommaso Valletti. 2016. "Internet and Politics: Evidence from U.K. Local Elections and Local Government Policies." Working Paper. AGGIORNA

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 AB, James M. Snyder J. 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.

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.*

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”

Additional Tables

Table 1.11: 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 1.12: 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 1.13: 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 1.14: 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 1.15: 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 1.16: 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 1.17: 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 1.18: 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 1.19: 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 1.20: 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 1.21: 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 1.22: 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 1.23: 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

Additional Figures

Figure 1.5: Screenshot of a profile on Twitter.

Figure 1.6: Distribution of Twitter penetration

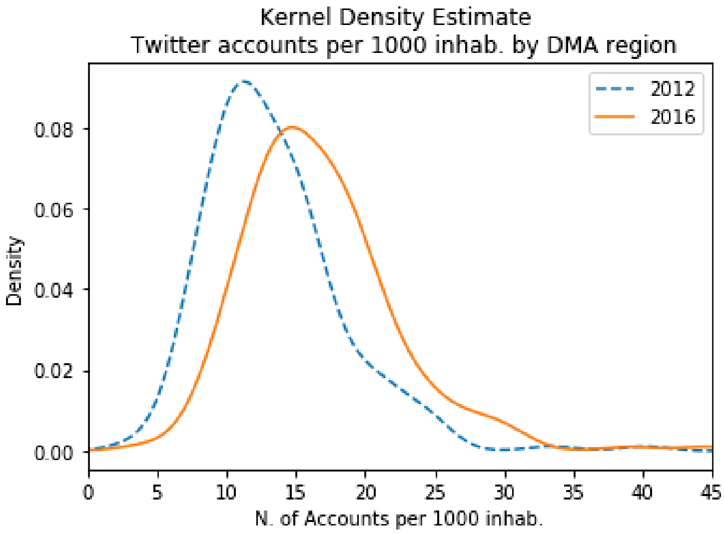


Figure 1.7: Tweets per day - Source: Twitter

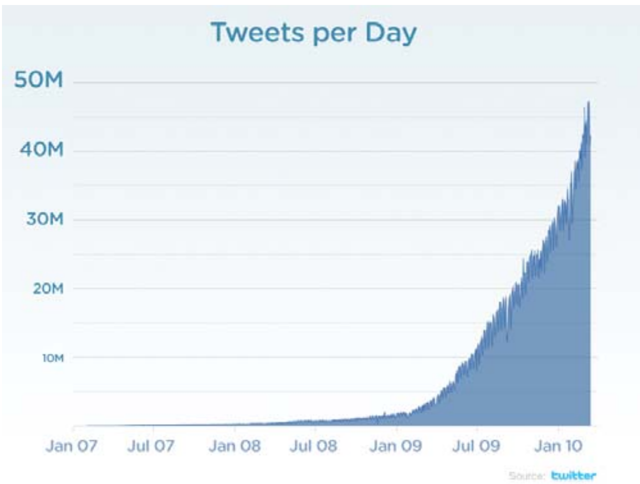


Figure 1.8: Distribution of accounts per 1000 inhab. by DMA in 2008

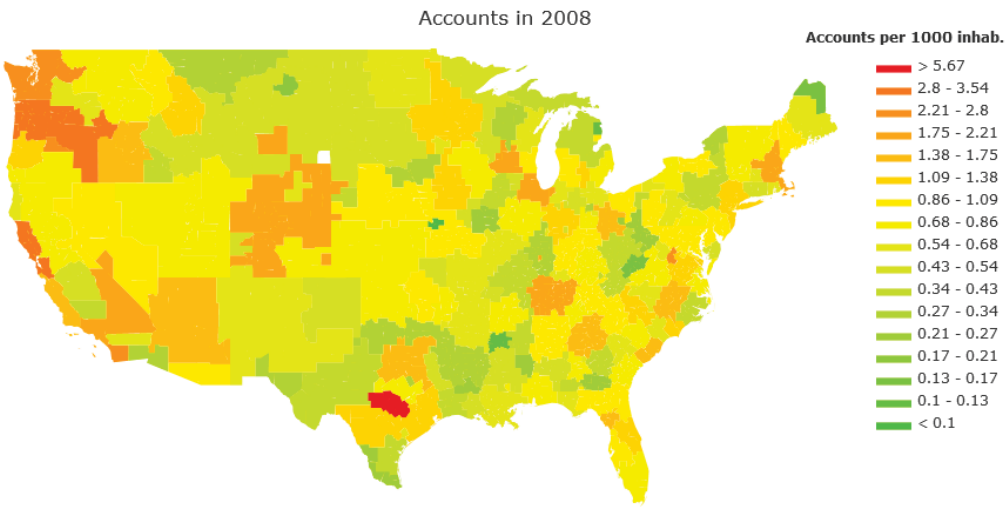


Figure 1.9: Distribution of accounts per 1000 inhab. by DMA in 2012

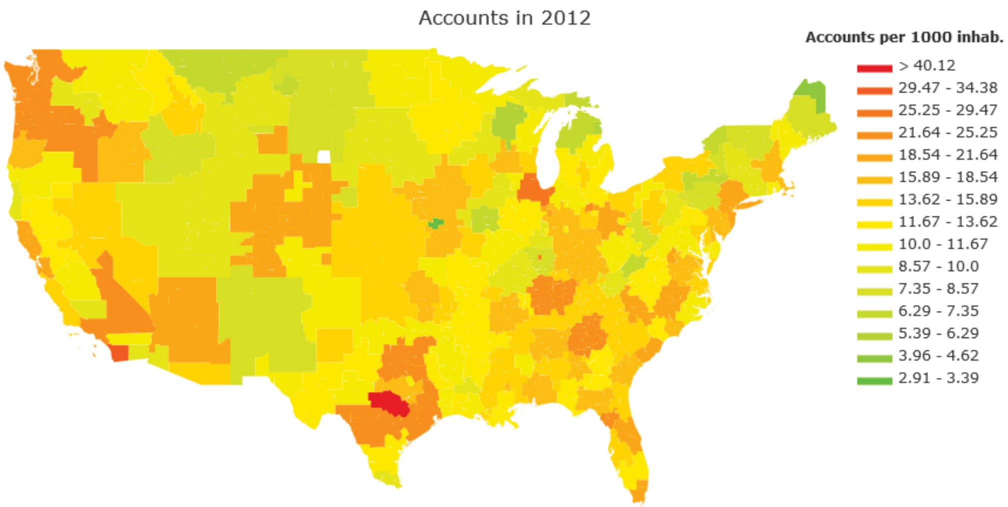


Figure 1.10: Age distribution - profile pictures

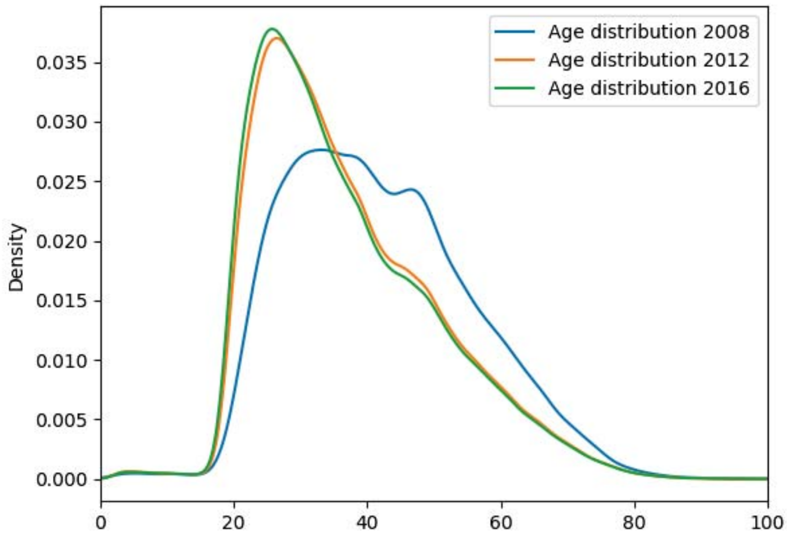


Figure 1.11: Distribution of Google Trends score by DMA - Boston Celtics

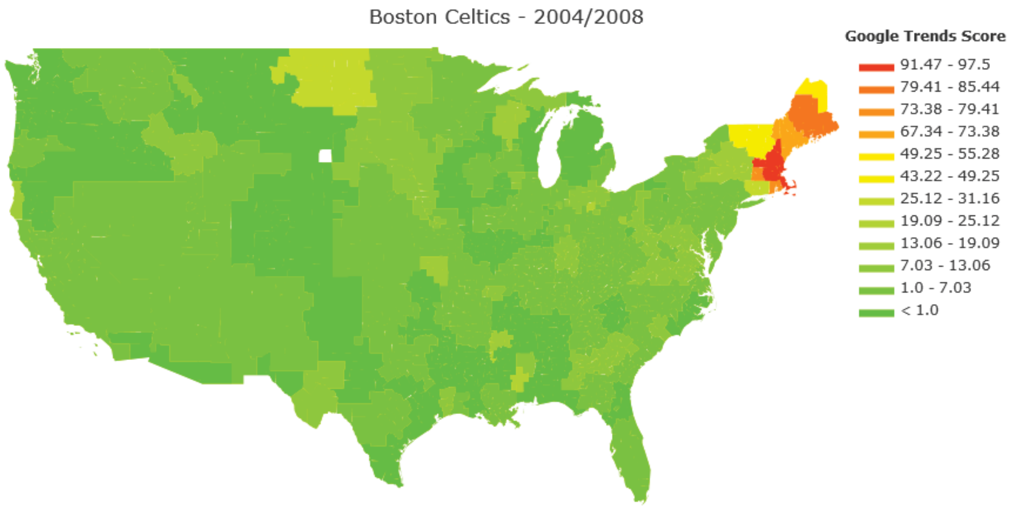


Figure 1.12: Distribution of Google Trends score by DMA - Minnesota Timberwolves

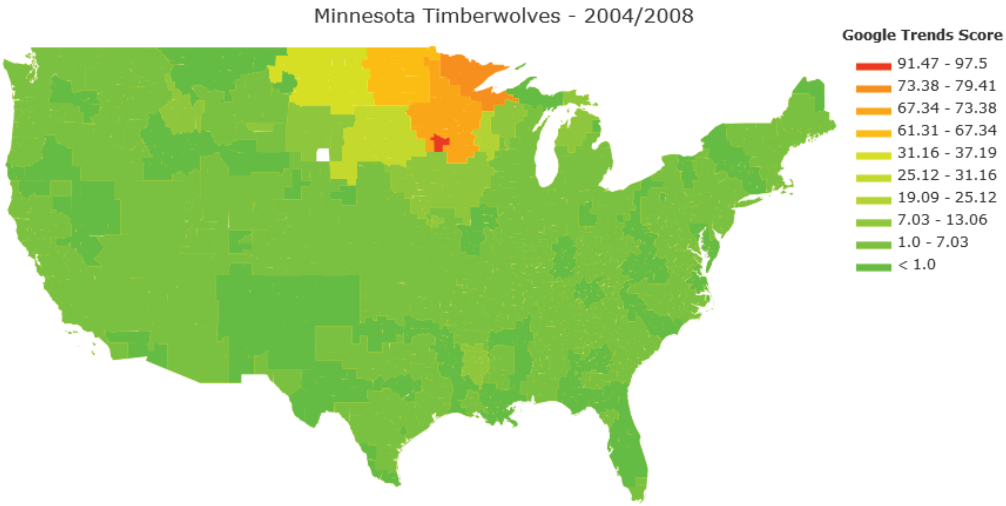


Figure 1.13: Most users never write about politics

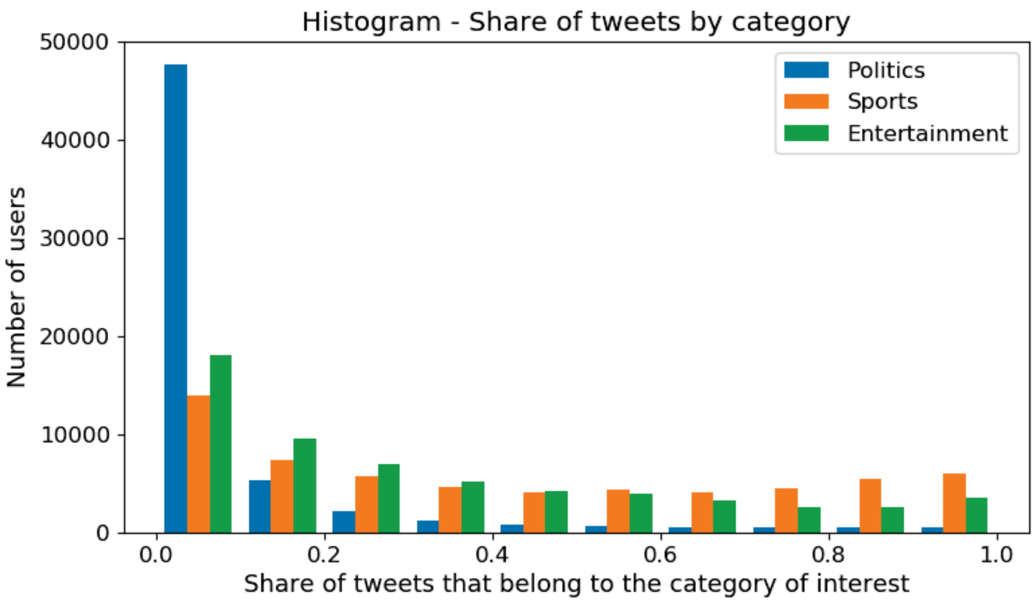
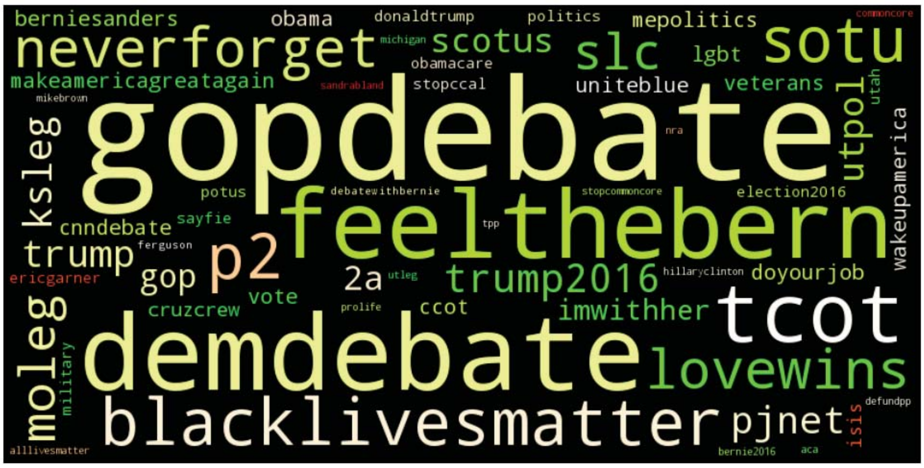
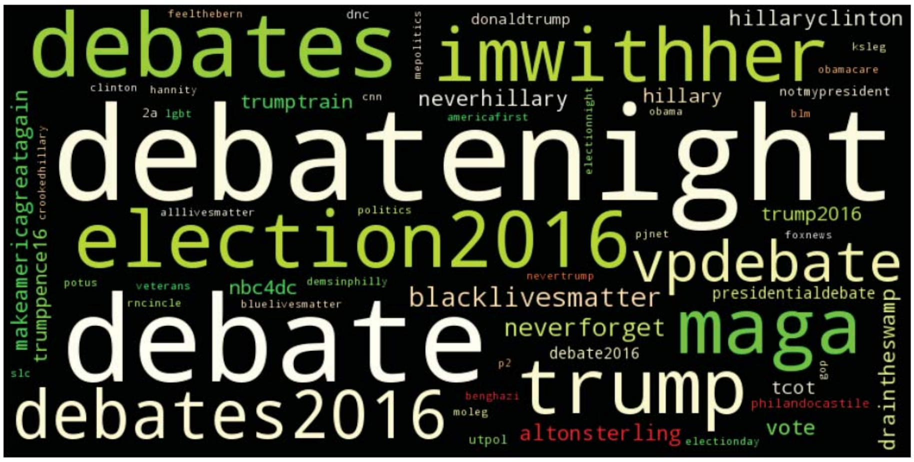


Figure 1.14: Word cloud - Most popular political hashtags during the first phase of 2016 electoral campaign



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 1.15: 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.

Appendix

1.A 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:

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

1.B 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 1.B.1 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 1.B.1 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 1.B.1: 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

Chapter 2

Information Transmission in a Social Network: A Field Experiment

with Eleonora Patacchini and Paolo Pin

2.1 Introduction

The way agents aggregate sparsely distributed information is a central topic in the economics literature. Examples of applications range from the theory of organizations, studying how to design the structure of communication inside a firm or a team in order to favor an efficient flow of knowledge, to the study of financial institutions, where strategic retention of information could prevent the correct functioning of the market. In many of these contexts there is an element of competition such that those who manage to have a better access to information get an advantage or, similarly, those who successfully hide pieces of knowledge to others are more likely to obtain a higher payoff. This paper studies this topic by focusing on a setting in which agents can exchange non verifiable pieces of information and are arranged in a network that defines their communication opportunities. In particular we are interested in studying how asymmetry among nodes alters the pattern of information diffusion. Is a less centralized network better at sustaining information exchange among agents?

To do this we use a field experiment with high school students in Italy. The experiment consisted in a game played using a smartphone app over the span of a week. Students were playing in groups of five members and were asked to guess five pieces of information that had previously been distributed among the five participants, one piece of information each. Communication was free in principle, but given the number of participants it was hard to identify in the school who was in possess of the information needed. The networks we used were then created by revealing players the identity of other members of the group. In particular, we would tell name, surname and class of some other player in the group. Peripheral players had little knowledge about the identity of the others, while central players received more information. We used three different networks in order to have variation in the structure of links among players. Crucially the scores were such that a bonus was given to the player who could better aggregate the available information. This introduced an element of competition as players, by truthfully sharing their information with others would reduce their chances of winning the bonus.

We believe this setup is a good description of what happens in many real life situations. Indeed, even inside an organization or a team, it is in principle possible to reach and exchange information with all the other active nodes. What we use as links can be interpreted as routes in which we expect communication to take place with higher probability. In the context of organizations this could be seen as the structure imposed by the management, while in teams it could be seen as frequent face-to-face interactions among team members.

The first finding is in line with the experimental literature on cheap talk games in which players were usually observed to exchange more information than what one would predict using theoretical models. We rarely observed the outcome that would constitute the only Nash equilibrium in a standard game with non-verifiable information, meaning the case in which people rely only on the hint they received and avoid exchanging pieces of information with the others. Instead the level of communication was high, especially among players that belonged to the same class.

Moreover, we find relevant effects of the network structure on the diffusion of information in the network. Central nodes managed to obtain on average higher payoffs, with a higher number of colors guessed correctly and a lower number of mistakes. In particular the network in which the asymmetry was stronger was associated with the lowest level of exchange of signals, but at the same time with the highest share of points going to the central nodes thanks to bonus points.

The presence of multiple nodes with a high centrality substantially reduced inequality in the final outcome and allowed the central nodes themselves to get access to more information. Interestingly, the strongest increase was observed for peripheral nodes who indirectly benefited from the change. With a lower level of asymmetry it was harder for central nodes to win alone, hiding information to the other nodes. For this reason bonus points were often divided among a larger audience.

In order to study the mechanism behind this result we turn to a parsimonious level-k model of strategic information transmission. We define a level-0 player as someone who is honest towards the others, passing correct information and who also believes what the others have told. Level-1 players would therefore react strategically by passing false information, but still believing what the other person is saying. Finally, level-2 players would pass false information and would not believe the others, a behavior that is in line with the cheap-talk literature. The model was then simulated in order to estimate structural parameters. We find that data are consistent with a high majority of players being cooperative and a small fraction (only 7%) acting strategically, passing therefore false pieces of information to their peers. 6These results highlight the role played by the network. Networks characterized by the presence of a single node with a high level of centrality offer better opportunities for strategic agents to manipulate information.

2.1.1 Literature Review

Classic studies of experiments regarding communication in networks are Bavelas (1950) and Leavitt (1951). More recently, the paper by Bonacich (1990) studies a context that is quite similar to ours, described by the author as a communication dilemma in which players face a tradeoff between sharing information with their links in order to accelerate the aggregation of signals at the group level and retaining pieces of information in order to increase the chances of winning a private reward. Hagenbach (2011) formalizes and generalizes the set-up described in the experiment in Bonacich (1990) and studies the ability of the group to solve communication dilemmas and how the speed at which aggregation takes place depends on the network structure. Mauleon et al. (2017) study a similar setting in which actions are taken sequentially, but the structure of the network is not exogenous. With respect to these papers we allow subjects to exchange non-verifiable information and this simplifies the set of possible strategies. At the same time we focus on different networks, as we are interested in understanding how the presence of multiple central nodes changes the overall outcome.

More recently, attention has been devoted to the study of how the network structure of communication allows players to coordinate on outcomes when multiple equilibria are possible (Choi and Lee (2014)) or to sustain collaborative norms (Gallo and Yan (2015)). Gallo and Yan (2015) in particular highlight how the introduction of asymmetric nodes in the network reduces the probability that players will avoid playing the Nash equilibrium and choose a collaborative norm, with higher average payoffs. This result is in line with what we find, with less communication when the structure of links is more centralized on one node. Also related to our results, Cassan (2007) study cooperation in networks by making agents play prisoner's dilemma games with their neighbors. The author finds that cooperation was hard to maintain. Networks in which neighbors shared the same links were able to sustain a slightly higher level of cooperation, suggesting that the network structure can play a role even when best-reply considerations would imply the absence of cooperation.

Related to our setting, there is a broad literature originated from Burt (1992) that studies the role played by structural holes in networks. This literature, using data from organizations, highlights how nodes who bridge structural holes can gain higher payoffs. In particular, by connecting parts of the network that contain different pieces of information agents can gain and advantage with respect to their peers. While the first contributions in this literature have considered static networks, more recently there has been attention to cases in which the network is endogenous (Goyal and Vega-Redondo (2007), Kleinberg et al (2008))[1]. With respect to this literature, while highlighting a similar mechanism, we are interested in comparing outcomes at the network level. Our results show how the presence of agents who bridge structural holes can have negative consequences for peripheral nodes as those who are central can easily alter or block the passage of information.

A paper that shares a similar structure with ours is the work by Mobius et al. (2015) that studies social learning using a field experiment involving college students. Authors, as in our case, find evidence of substantial social learning. Mason and Watts (2012) instead run web-based experiments to study how groups of agents who interact through a network can solve a complex problem in the presence of a tradeoff between exploitation of solutions already found by others and exploration of novel ones. As in our case, one of the main findings is that individuals that are more central in the network obtain higher payoffs than peripheral ones.

Among the theoretical contributions that study cheap-talk communication in networks when players have conflicting interests, the papers by Hagenbach and Koessler (2010) and Galeotti et al. (2013) focus on how the heterogeneity in preferences over outcomes affects the shape of the network, intended as the collection of truthful information exchanges among the members of the group. Forester (2018) extends Galeotti et al. (2013) to a dynamic

[1]Rand et al. (2011) describe an experiment that is connected to this. In particular, they consider how the possibility of breaking links in the network can help sustaining cooperation among agents. We abstract from this as in out case the network is fixed.

setting. Finally, our work is related to the literature that studies pre-play communication using level-k models of strategic reasoning. Ellingsen and Östling (2010) find for example that communication can facilitate coordination on Nash equilibrium outcomes.

2.2 Experimental Design and Data Description

The experiment was conducted in May 2017 with students from three high schools in Italy, in total were involved 645 students. Participation was voluntary and we relied on the help of teachers from each school to invite students to take part of the experiment. In order to register to play the game each student had to answer questions from an online survey and then download and install an app for her phone. The app contained the game and was made available both for Android and iPhone[2], so that we could allow the vast majority of interested students to play. The app contained a mock round with the explanation of the rules of the game and a set of questions made to make sure the participants had no doubt about the functioning of the game. The game took place over the span of a week, with three rounds of two days each. During each round every student was matched with four other colleagues to form a group of five. For this reason, every round, a total of 129 groups was active. During each round players could win from a minimum of 0 euro to a maximum of 15 euro. We then extracted randomly one round to be the one valid for the payments. Rewards were given using Amazon Gift Cards.

The game. The goal of the game was to guess the colors of the clothes of a character. There were 5 pieces of clothing: hat, shirt, gloves, trousers and shoes. Different groups had to guess colors for different characters, where each character was identified by a name. For example Figure 2.1 refers to a group that had to guess the colors of Andrea's clothes. Clothes,

[2]The app is called *VestiTito* and is still available for download on Play Store even though now the game is not active.

as shown in Figure 2.1 (a), appeared grey at the start of the round. The player could then change colors at any moment in time, with no cost, by choosing one out of nine possible alternatives (Figure 2.1 (b)). The only combination that was relevant for the payment was the last combination, the one that was left at the end of the round[3]. The player was allowed to leave clothes as grey and this would be counted as the player not choosing any alternative. The score was calculated according to the following rule: 10 points were assigned for each color that was correctly guessed, while 5 points were removed for each color that was guessed wrongly. Finally, 0 points were assigned for each piece of clothing left grey. On top of this, players could earn a bonus in case they were the ones that had the highest score among the members of the group of five. The bonus was 100 points, to be divided among the eligible players[4]. Each point was worth 10 cents, so that each player could win from 0 to 15 euro.

At the beginning of each round the player was given three pieces of information. First, the name of the character she had to guess. Second, the color of one of the clothes. Third, the identity of other members of the group she was assigned to. Figure 2.2 provides an example. Importantly, the hint regarding the color could be seen only once, hence the player had to memorize it[5]. Also, there was no overlap between the hints given to the players, so that information exchange was in principle possible and beneficial among any couple of players. Since all clothes were equivalent in term of points, the hint regarding the color was not introducing any asymmetry among the players. This was done only through the information regarding the other members of the group.

Players could be arranged in three different networks as shown in Figure 2.3. In these networks each node represents a players. A link connects two nodes if they were informed

[3]The round was set to end during the night, in order to reduce the probability that the students were together at the time of the collection of answers.

[4]For example, if only one player had managed to correctly guess 5 colors, the final score would have been 50 (because of 5 colors) plus 100 (for being the only one to get the highest score). Similarly, in case only two players got 40 points, with 40 being the highest score in the group, the bonus would have been 50 each, for a total of 90 points.

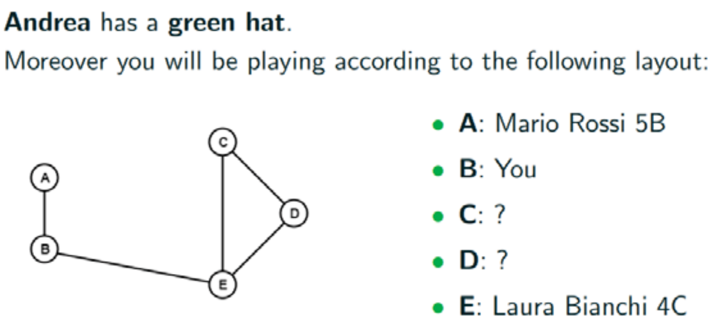
[5]In the app we also made sure it was not possible to take screenshots, that would have been an easy and fast way to memorize the hint and reproduce it when needed.

Figure 2.1: Screenshots from the app.

(a) Screenshot 1. (b) Screenshot 2. (c) Screenshot 3.

Note: The pages are in Italian as the game was done with students in Italy.

Figure 2.2: Hint received at the beginning of a round, example



Note: This figure represents and example of a hint received at the beginning of a round. We see that the player is told that she need to guess Andrea's colors (this element is unique for a group of 5 players). Moreover she is told that her position in the network is B, she therefore knows A's identity (Mario Rossi 5B) and E's identity (Laura Bianchi 4C). Finally, by looking at the network, it is possible to know that player E knows the identity of two other players.

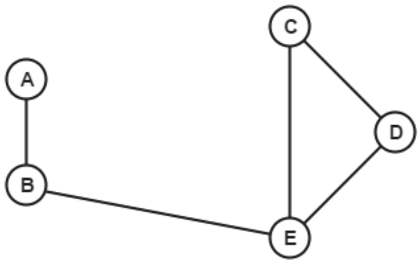
that they belonged to the same group. For example, Player A in Network 1 was only informed about the identity of Player B, while Player B knew about E also. It is important to underline that the app itself was not providing any chat to exchange messages with the other members of the group. All the players had to do was to talk with their colleagues. In a context of a high school, with hundreds of students, many of them not even aware of the existence of the game, knowing who to talk to can be considered as a relevant piece of information to hold. We therefore consider links are couples among which we expect communication to happen with high probability. Finally, each turn players could be assigned to any of the three networks and, given the network, to any of the five positions.

After receiving these pieces of information, players were given two days of time to choose a combination of colors to submit. At the end of the round, players were asked to answer some questions regarding the round they had just played. After that, the process was repeated, with new groups being generated. It is important to underline that groups were done in order to avoid two players to play together more than once. This feature of the randomization was known by the players and would prevent favors exchanges from one round to the other.

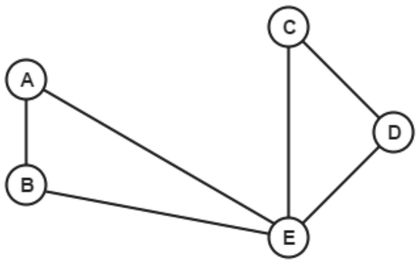
Data. During the registration process we asked students to answer a set of questions concerning personality. Moreover players could allow us to download their network of friendships on Facebook, that would be used as a proxy of friendship among students. Surprisingly, for one third of the students we could not collect this piece of information, either because the didn't allow us to have access to their account or because they had no account at the time of the experiment[6].

The app was designed to keep track of any change of color that was made by the players and the timing. This way we are able to reconstruct the history of changes made by the players. Moreover after the end of every round we asked a set of questions to the players.

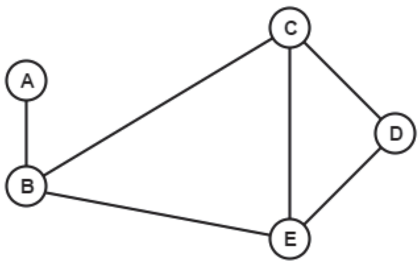
[6]We use Facebook Graph API. If the person accepts and decides to log into Facebook when playing the app, we receive the list of friends that have installed the app and logged into Facebook too.



(a) Network 1



(b) Network 2



(c) Network 3

Figure 2.3: Networks of contacts.

First, we were providing them a list of names and they had to choose the one of a person that was assigned to the group. This was done twice. Whenever possible we were using both one person for which they had received the hint and one for which they had not. The first was intended to be test to check whether the player had correctly understood the hint, while the other was intended to check the ability of the player to learn the names of the other participants, either by searching in the school or by asking other members of the group. Second, one question was about the score, as we were asking whether they thought they had achieved the highest score. Finally, in a second page, we were asking about the other members of the group. Each player was given the name of every other member of the group. For each name the player was asked to tell us whether they had exchanged information useful to the game and who had taken the initiative to contact the other.

Sample. In total 645 students registered and downloaded the app, this allowed us to create 129 groups of 5 players per round, for a total of 387. When the game began though, part of the registered players did not participate to the game. Additionally, we had to ignore 29 groups that included students from different schools. Table 2.1 reports the distribution of the number of active players per group, where we define as "active" a player that logged in during the round and checked the hint she was given. Importantly, when we generated groups of 5 for the second round we started by matching players that had not been active in the first round. This allowed us to obtain a higher share of groups with 5 active players in rounds 2 and 3. In what follows we report results only for groups with 5 active players, as it is unclear how to interpret incentives when one or more players are missing from the group.

Table 2.1: Number of active players per group.

	Five	Four	Three	Two	One	Zero
All	132	91	36	33	28	38
Network 1	29	31	13	10	11	12
Network 2	41	30	12	10	9	14
Network 3	52	30	11	13	8	12

2.3 Results

In this section we describe results at the group level that are suggestive of those that appear to be the mechanisms at work.

2.3.1 Information Exchange Across Classes and Within Classes

Players were arranged in groups so that links could be between students from the same class or from different classes. This seems a relevant dimension to consider if we want to study why players decided to share information. Table 2.1 highlights an interesting pattern in this dimension. The first column contains the probability that we observe info exchange in one link, meaning the probability that one node ended up guessing the right color for the clothes of the other node and vice-versa[7], the second column considers the probability of not having any exchange, while the third column considers asymmetric exchanges, that are when one player ends up knowing the other's hint, but not vice-versa. Asymmetric exchanges are particularly interesting because they would appear more frequently along links in which one

[7] For example, consider the link between node A and node B. Let us assume that A received a hint for the shoes and B a hint for the hat. Then we say that there has been information exchange if at the end of the round B chooses the right color for the shoes and A chooses the right color for the hat. This is a simplification since there are other reasons why both players could end up choosing the right hints. First of all they could be lucky and guess the right colors out of the nine alternatives. Second, other nodes could have told them which is the right color, therefore without having direct exchange of information.

Table 2.1: Info Exchange Within and Across classes

	Info Exchange	No Exchange	Asymmetric Ex.
Within	0.769	0.047	0.184
Across	0.358	0.431	0.211

Note: We use only the sample of groups in which all 5 players actively participated to the round by submitting a set of choices, so that we have 132 obs in total.

player manages to take advantage of the other one.

First of all we need to highlight that the probability of information exchange is high, especially if we consider as benchmark case what would happen in a standard model of information transmission when information is non-verifiable and players compete to get a higher score. This likelihood is particularly high for links within class and lower across classes. This can be rationalized in several ways, two examples being altruistic preferences or reputational concerns. Second, while Info Exchange and No Exchange show relevant differences across and within classes, the likelihood of Asymmetric Exchange is similar. If we think of the first two columns as being the result of an equilibrium behavior we see that a relevant fraction of players did not manage to optimally respond to their peers.

2.3.2 Diffusion

The most relevant dimension in which one should compare different networks is maybe the ability for the network to foster information aggregation. This could be measured in two ways. First, we could consider the total number of correct guesses as presented in Table 2.2. Second, we could instead focus on the maximum number of correct guesses among the members of the group as reported in Table 2.3. We believe that both outcomes are potentially interesting, with their relevance depending on the particular application that one

Table 2.2: Tot. number of colors guessed correctly in the group.

	Network 1 (16.46)	Network 2 (15.71)	Network 3 (18.46)
Network 1	-	-0.754 (1.29)	2* (1.1)
Network 2	-	-	2.75** (1.17)

Note: Each cell reports the difference between the total number of correct guesses obtained by players in a given group. We use only the sample of groups in which all 5 players actively participated to the round by submitting a set of choices, so that we have 29 obs for network 1, 41 obs for network 2, and 52 obs for network 3. In parentheses are reported standard errors for pairwise t-test.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

has in mind. Table 2.2 reports the average total number of correct guesses by networks type. First, we can notice once again that this number is higher than what we would expect in case agents were to rely only on their initial hints. We also notice that Network 3 is characterized by a higher degree of information sharing. Pairwise t-tests confirm that Network 3 does better that Network 1 and Network 2, with Network 2 being characterized by the lowest level of information sharing. The presence of a central node alone seems therefore to impact negatively on the total number of hints collected at the group level. Interestingly, we see that a high degree of asymmetry among nodes has negative effects also on the maximum number of correct guesses that are made by members of the group. From Table 2.3 we see indeed that this is the case, with Network 3 once again being characterized by higher values, especially when compared to Network 2.

Table 2.4 compares the average number of correct guesses made by players in a certain

Table 2.3: Max. number of colors guessed correctly in the group.

	Network 1 (4.13)	Network 2 (3.85)	Network 3 (4.3)
Network 1	-	-0.27 (.23)	0.18 (.19)
Network 2	-	-	0.45** (.2)

Note: Each cell reports the difference between the average of the maximum number of correct guesses obtained by players in a given group. We use only the sample of groups in which all 5 players actively participated to the round by submitting a set of choices, so that we have 29 obs for network 1, 41 obs for network 2, and 52 obs for network 3. In parentheses are reported standard errors for pair-wise t-test.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

position across the three different networks. There are three elements that we would like to stress. First, the see that in Network 3 the average for Position E is higher than in the other two networks. Even if the difference is not statistically different from zero, this suggests that having exclusive access to other nodes does not improve one's ability to retrieve information. Second, we see an effect even for Position A and Position D, that in some cases are easier to compare across networks as the set of neighbors is the same. In particular, the average score for Position A is significantly higher in Network 3 with respect to Network 1. For Position D there is a significant increase moving from Network 2 to Network 3. There are therefore second order effect, with nodes benefiting from the ability of their neighbors to have better access to information. Finally, Position B receives a clear advantage in Network 3 with respect to the other two, this is due to the increase in centrality that this position obtains.

Table 2.5 instead compares the number of correct guesses within the same network type, across positions in the network. Given the three structures we used, some comparisons are particularly interesting. First, we can notice that E does in general better than the other positions, especially in Network 2. Being in a central position therefore seem to be an important factor to obtain a high score. Second, in Network 3 differences across positions almost disappear, with the only exception of A having a lower score than B and E. This highlights how the presence of multiple central nodes allows for a better diffusion of information.

Table 2.6 instead reports for each node the probability of being in the winning group. It is possible to notice how node E, that is always the most central, manages to win with a higher probability than the other nodes, with this difference being stronger in network 2, when the informational advantage is higher.

Table 2.4: Test of pairwise equality of correct guesses.

Position A	Network 1 (2.82)	Network 2 (3)	Network 3 (3.42)
Network 1	-	0.18 (.36)	0.6* (.33)
Network 2	-	-	0.42 (.34)
Position B	Network 1 (3.1)	Network 2 (2.68)	Network 3 (3.79)
Network 1	-	-0.42 (.36)	0.69** (.29)
Network 2	-	-	1.1*** (.3)
Position C	Network 1 (3.41)	Network 2 (3.37)	Network 3 (3.64)
Network 1	-	-0.04 (.28)	0.22 (.26)
Network 2	-	-	0.27 (.27)
Position D	Network 1 (3.59)	Network 2 (3.1)	Network 3 (3.77)
Network 1	-	-0.49 (.29)	0.18 (.28)
Network 2	-	-	0.67** (.28)
Position E	Network 1 (3.54)	Network 2 (3.56)	Network 3 (3.85)
Network 1	-	0.02 (.3)	0.31 (.27)
Network 2	-	-	0.29 (.25)

Note: Each cell reports the difference between the avg. number of correct guesses obtained by players in a given position. We use only the sample of groups in which all 5 players actively participated to the round by submitting a set of choices, so that we have 29 obs for network 1, 41 obs for network 2, and 52 obs for network 3. In parentheses are reported standard errors for pairwise t-test.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.5: Test of pairwise equality of correct guesses.

Network 1	A (2.82)	B (3.1)	C (3.41)	D (3.59)	E (3.54)
A	-	0.28 (.19)	0.59** (.22)	0.77*** (.22)	0.72** (.29)
B	-	-	0.31 (.24)	0.49** (.23)	0.44* (.24)
C	-	-	-	0.18 (.21)	0.13 (.26)
D	-	-	-	-	-0.05 (.24)
Network 2	A (3)	B (2.68)	C (3.37)	D (3.09)	E (3.56)
A	-	-0.32* (.17)	0.37* (.18)	0.09 (.2)	0.56*** (.19)
B	-	-	0.68*** (.2)	0.41* (.24)	0.88*** (.21)
C	-	-	-	-0.27* (.16)	0.19 (.13)
D	-	-	-	-	0.46** (.19)
Network 3	A (3.42)	B (3.79)	C (3.64)	D (3.77)	E (3.84)
A	-	0.36** (.14)	0.21 (.19)	0.35 (.23)	0.42* (.22)
B	-	-	-0.15 (.17)	-0.02 (.19)	0.06 (.18)
C	-	-	-	0.13 (.19)	0.21 (.15)
D	-	-	-	-	0.08 (.19)

Note: Each cell reports the difference between the avg. number of correct guesses obtained by players in a given network. We use only the sample of groups in which all 5 players actively participated to the round by submitting a set of choices, so that we have 29 obs for network 1, 41 obs for network 2, and 52 obs for network 3. In parentheses are reported standard errors for pairwise t-test.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.6: Prob. winning by position in the network.

	A	B	C	D	E
Network 1	0.31	0.41	0.69	0.67	0.62
Network 2	0.46	0.32	0.60	0.56	0.78
Network 3	0.52	0.62	0.52	0.62	0.71

Note: Each cell reports the probability of winning by network and by position. We use only the sample of groups in which all 5 players actively participated to the round by submitting a set of choices, so that we have 29 obs for network 1, 41 obs for network 2, and 52 obs for network 3.

2.4 A Simple Strategic Model

We begin by considering a simple strategic model in which agents need to exchange non-verifiable pieces of information. Our experiment can be modeled as the following *Guess Colors Game*.

There are 5 players placed on a network, that can be either one of those in Figure 2.3. There is a set of *colors* that is $Col = \{1, 2, \ldots, 9\}$, and we call $Col_{\emptyset} = Col \cup \{\emptyset\}$. There are 5 *cells* to guess, so that there is a *true* message to guess $m_{true} \in Col^5$.

There is a message phase and a guessing phase.

In the message phase time is discrete. At time 0 each player receives as *hint* the projection of m_{true} on one of the 5 cells, and each player receives a different projection. At time 1 one link is picked at random (with uniform probabilities) and the two players can pass each others any vector $m \in Col_{\emptyset}^5$. The other 3 players are not aware of what link was picked, they only know that a link was picked that did not involve them.

After time t, with probability $1 - \delta$ the game stops and the guessing phase starts. With probability δ, instead, the message phase goes on to time $t+1$. A new link is picked at random with uniform probabilities (i.i.d. also across time), and the two involved players can again pass each others messages. Say that i and j are involved, we call $m^{t+1}_{i\to j}$ the message that i sends to j, and $m^{t+1}_{j\to i}$ the message that j sends to i.

In the guessing game, each player i plays one vector $m_i \in Col^5_\emptyset$. Payoffs are given in the following way:

- each player receives a *score* $s_i = \sum_{k=1}^{5} \sigma_i^k$, where

$$\sigma_i^k = \begin{cases} 10 & \text{if } m_i^k = m^k_{true} \\ 0 & \text{if } m_i^k = \emptyset \\ -5 & \text{otherwise} \end{cases} .$$

- call $S = \arg\max_{j\in\{1,\dots,5\}}\{s_j\}$ the set of players who get maximal score – a player i for which $i \in S$, receives a *premium* $v_i = 100/|S|$, a player j for which $j \notin S$, receives a *premium* of 0

- the payoff of player i is the sum of her score and her premium: $\pi_i = s_i + v_i$.

Now that we have fully described the game, we can derive a simple result, that relies on the implicit assumption that all players are selfish and rational.

Proposition: In every Nash equilibrium of the Guess Colors Game, in the guessing phase each player plays only the hint that she received at time 0 of the message phase. So, the payoff for each player is 30.

Proof: We start the proof by assuming that some players may use in the guessing phase some of the information that they receive in the message phase. If that was true, when i is picked to pass her message to j, she should consider if she wants to pass the information that she has, or some other pieces of information, or if she want to hide or falsify them. In general, if $t > 1$, i should also take into account that what she passes to j may be *double–checked* by j with other messages that she received.

Apart from all these conjectures, it is easy to see that a dominat strategy for i is to always say consistently the same falls information about her initial message. This strategy would make it impossible for any other player to check inconsistencies. This is actually a strictly dominant strategy if there is at least one other player that, with some positive probability, may believe that the information provided by i is true.

It requires a step of iterated dominance of strategies to understand that, once all players internalize that the initial pieces of information are falsified, they will not believe to any message that they receive.

Finally, at this stage of iterated dominance, since telling the false about own initial hint is only weakly dominat, we have also Nash equilibria in which some players may tell the thruth. However, they will do so with a probability that makes it dominated for the other players not to believe in their message. QED

The proof relies on the same reasoning that leads to *babbling models*. What it adds is only a simple consideration about our particular setup: the fact that the message phase has no pre–fixed finite time does not imply any form of Folk theorem, because there is no payoff feedback between the time steps of this phase. Suppose in fact that some form of collaboration was in place during the message phase between a coalition of players. The only *alarm* that could make them decide to stop collaborating and start *punishing* would be the detection of some inconsistency in the messages that they receive. However, since each player understand this and can falsify consistently her own piece of information, no detection is possible.

Table 2.1: Groups, by number of winners and number of hints guessed correctly.

	One win	Two win	Three win	Four win	Five win
One color	0	0	0	0	$1^{\leftarrow NASH}$
Two colors	0	$5^{\leftarrow Coalition}$	1	1	0
Three colors	$8^{\leftarrow cheat?}$	$10^{\leftarrow 1\ error?}$	$20^{\leftarrow Coalition}$	1	0
Four colors	$6^{\leftarrow cheat?}$	$7^{\leftarrow 2\ errors?}$	$3^{\leftarrow 1\ error?}$	$7^{\leftarrow Coalition}$	3
Five colors	$9^{\leftarrow cheat?}$	$14^{\leftarrow 3\ errors?}$	$11^{\leftarrow 2\ errors?}$	$14^{\leftarrow 1\ error?}$	$20^{\leftarrow Coalition}$

Note: Each cell reports the number of groups for which we observe a number of winners as indicated in the column that managed to collect a number of colors as indicated in the row. We use only the sample of groups in which all 5 players actively participated to the round by submitting a set of choices, so that we have 132 obs in total.

2.4.1 Winners

The first way to understand whether the model described above offers a compelling description of the dynamics that emerge from the experiment is to look at the distribution of the number of winners and the number of correct colors guessed by the winners. Table 2.1 contains the count for each combination. For example, the first entry in the first column tells that it was never the case that one player won alone with one correct guess only.

A striking fact that emerges from this table is that the outcome that would be predicted as the only Nash Equilibrium in a simple model of cheap talk is observed only once (upper right corner, highlighted in red). Indeed, with non-verifiable communication we would expect players to send misleading signals to the others, while not trusting what the others are saying. This should determine outcomes in which five players end up winning with only one correct guess (their original hint). As we see from the table, we observed five players winning with one color only once. Instead, we see that the most common outcomes were three winners guessing three colors and five winners guessing five colors, suggesting the presence of coalitions among players[8]. We see therefore that it was often the case that all players, or at least a group of them, managed to coordinate and truthfully share their hints. Interestingly we see that in many other cases we observe the winners collecting more hints (observations below the diagonal constitute 63% of all cases), with the extreme case of one individual gathering five hints (this happened 9 times). This suggests that information exchanges were often imperfect, with information flowing only one way. For this reason we move to a level-k model that allows us to consider different players and therefore explain the pattern described above.

2.5 A Level-k Model of Information Diffusion

In order to understand which is the mechanism behind this results we now move to a level-k model of strategic information diffusion in a network. Our purpose is twofold, first, we aim at understanding whether data can be explained by a simple model like the one presented in this section or instead we need to introduce more complex strategic reasoning by the players. This is particularly important since the app did not allow us to have full control over all variables of interest (for example we could not observe communication among players) and a model can be used to make educated guesses on what happened in the background. Second, estimates could be used to highlight which are the differences across networks that are more relevant.

The Model. We consider a population composed of a finite set of students, that are arranged in networks of five nodes each. In each network the set of nodes is N={A,B,C,D,E}. There are three possible combinations of links: L1={AB, BE, CD, CE, DE }, L2={AB, AE, BE, CD, CE, DE }, L3={AB, BC, BE, CD, CE, DE }. These three networks are repre-

[8] Of this coalitions we see that groups of three winners were often composed by nodes C, D, and E (75% of the times). Groups of five winners were observed relatively more often in Network 3 (9 times over the total of 20).

sented in Figure 2.3. There is a finite set of time periods t=1,2, ..., T in which agents can communicate and exchange information concerning the true state of the world.

Every period a link in the network is drawn at random as communication opportunity: the nodes involved meet and communicate according to their strategy. Communication is bilateral. In case the students belong to different classes they will meet with probability z, otherwise they will meet with probability 1. Each period the process continues with probability δ. When communicating, the agents will ignore new information that concerns signals they already have received and focus only on new information that the other party is bringing.

Agents can follow three different strategies, according to their level. A level-0 agent will behave cooperatively, passing always truthful information when communicating with another player and believing to what the other player says. A level-1 player will best respond by passing false information, but at the same time believing what the other player is saying. Finally, a level-2 player will pass only false information and won't update based on what the counterpart says. Given the context of the game we have that level-2 is the highest level of sophistication that one player can reach.

Estimation. This process can be easily simulated using as starting point the distribution of students across classes and network types that we used in the experiment. Simulations will depend on four parameters: z and δ are described above, while p and q will determine the share of type-1 and type-2 players respectively.

The model is then estimated using the method of simulated moments. Simulated data are used to generate moments that are compared with the moments we observe in the data. Parameters are chosen in order to minimize the quadratic distance between simulated and observed moments. We use as empirical moments the averages of correct guesses by network and by position that were discussed above, together with the matrix presented in Table 2.1,

Table 2.6, and Table 2.1.

Results. We estimate the model in three different versions and compare the results. In particular we run the same estimation first allowing all three types of players to be active, then we restrict to only type-0 and type-1. Finally we compare with a simple model of diffusion of information that emerges when only type-0 players are present.

Figure 3 to 6 report results from these simulations. In Figure 2.1 we see the distribution of our estimate for z across the three versions of the model described above. Similarly, Figure 2.2 does the same for δ. We can notice how the distributions overlap, so that the average values are ($z = 0.333, \delta = 0.937$) for the case of type-0 only, ($z = 0.345, \delta = 0.942$) for the case of type-0 and type-1, and ($z = 0.318, \delta = 0.947$) for the case in which we allow up to type-2 players. Figure 2.3 instead shows the comparison for the estimate of the probability of a player being type-0. When we only allow type-0 and type-1 players our estimate for the probability of a type-0 player is 0.942. When instead we allow for type-2 too, we estimate that a player is type-0 with probability 0.934. According to our estimates, results are therefore driven by a minority of strategic players who do not truthfully share information with their peers.

In order to compare the fit we show in Figure 2.4 the quadratic distance between simulated and empirical moments. It is possible to notice how this distance drops once we allow for type-1 players, but does not change significantly when we include type-2 players as well. Having players with a level higher than 0 indeed changes the simulated moment in a few striking ways, making them closer to the empirical moment. In particular with only level-0 players it would be impossible to observe cases in which one player wins alone with 3 or more hints. At the same time, groups of 5 players winning together should be observed more often. Also, the predicted share of asymmetric information exchanges would be much lower than the one observed one.

2.6 Conclusion

In this paper we studied how the structure of the communication network influences aggregation of information in a context in which agents compete and can only share non-verifiable messages. In particular, we asked whether a less centralized network structure can be better at fostering information diffusion.

To do this, we ran an experiment with high school students, using a smartphone app. During the game, students were arranged in groups of 5 players and could share hints regarding a combination of colors that they were asked to guess. An element of competition was in place, such that players were incentivized to try to be the best in their group because this would allow them to obtain a bonus payment.

The results discussed above show that the network structure has relevant effects on the overall pattern of aggregation of information. Players often managed to sustain symmetric exchange of information, but the node that enjoyed the most central position obtained higher payments on average. In particular, they could take advantage of their superior access to information by keeping the other players isolated. Moreover, by reducing asymmetry among nodes we could improve the performance of the group, both in terms of the average number of corrected guesses and the maximum number of guesses.

A level-k model of strategic information transmission fit the data well and points towards the fact that a small share of players managed to take advantage of the majority of collaborative players, increasing their own chance of winning the bonus score. In a context in which only a few players are not willing to cooperate and try to take advantage of the others, we see that the network structure has a strong impact on the overall outcome, precisely because a higher level of symmetry makes communication in the network more resilient to this type of players.

These results are therefore informative on which network structure an organization should prefer when there are concerns regarding the presence of strategic agents who may not fully cooperate with their peers. Future developments will go in two main directions. First, it would be interesting to use laboratory experiments. This would allow to observe the information exchanged by players, an element that would help in obtaining a better understanding of the strategies followed by the players. Finally, this would allow to compare the anonymous interaction in the lab with the face-to-face interaction that took place in the field experiment. Another direction to follow is to enrich the model, introducing a stronger interdependence between the position in the network and the strategy chosen during the game.

References

Bavelas, Alex. 1950. "Communication Patterns in Task-Oriented Groups" The Journal of the Acoustical Society of America 22,725.

Bonacich, Phillip. 1990. "Communication Dilemmas in Social Networks: An Experimental" American Sociological Review 55(3):448-459.

Burt, Ronald S. 1992. "Structural Holes" *Academic Press, New York.*

Cassar, Alessandra 2007. "Coordination and cooperation in local, random and small world networks: Experimental evidence. Games and Economic Behavior. 58 209-230.

Choi, Syngjoo and Jihong Lee. 2014. "Communication, coordination, and networks." Journal of the European Economic Association, 12(1):223-247.

Ellingsed, Tore and Robert Östling. 2010. "When Does Communication Improve Coordination?" American Economics Review, 100:1695-1724.

Galeotti, Andrea, Christian Ghiglino and Francesco Squintani. 2013. "Strategic Information transmission networks." Journal of Economic Theory 148(5):1751-1769.

Gallo, Edoardo, and Chang Yan. 2015. "Efficiency and equilibrium in network games: An experiment." Cambridge-INET Working Paper Series, 1503.

Goyal, Sanjeev, and Fernando Vega-Redondo 2007. "Structural holes in social networks." Journal of Economic Theory 137 460-492.

Foerster, Manuel 2018. "Dynamics of Strategic Information Transmission in Social Networks." Working Paper.

Hagenbach, Jeanne. 2011. "Centralizing information in networks." Games and Economic Behavior 72:149-162.

Hagenbach, Jeanne and Frédéric Koessler. 2010. "Strategic Communication Networks" The Review of Economic Studies 77:1072-1099.

Kleinberg, Jon, Siddharth Suri, Eva Tardos, and Tom Wexler 2008. "Strategic Network Formation with Structural Holes." Proc. 9th ACM Conference on Electronic Commerce.

Leavitt, Harold J.. 1951. "Some Effects of Certain Communication Patterns on Group

Performance" Journal of Abnormal and Social Psychology 46:38-50.

Mason, Winter, and Duncan J. Watts 2012. "Collaborative learning in networks." PNAS 109 (3) 764-769.

Mauleon, Ana, Simon Schopohl, and Vincent Vannetelbosch 2017. "Competition for leadership in teams." Working Paper.

Mobius, Markus, Tuan Phan and Adam Szeidl. 2015. "Treasure Hunt: Social Learning in the Field." Working Paper.

Rand, David G., Samuel Arbesman, and Nicholas A. Christakis 2011. "Dynamic social networks promote cooperation in experiments with humans." PNAS 108 (48) 19193-19198.

Additional Figures and Tables

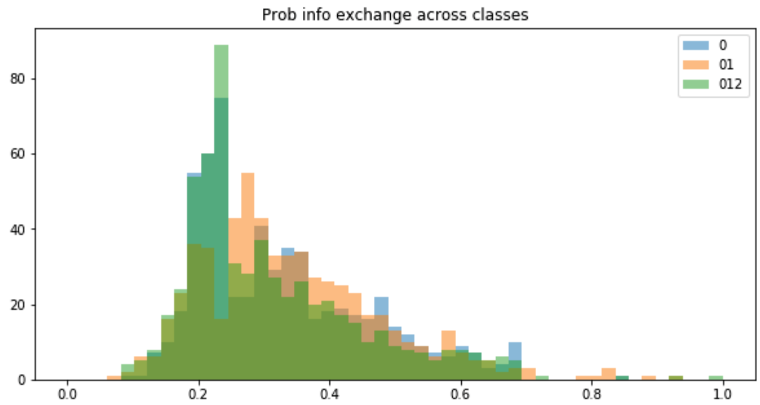


Figure 2.1: Estimates for z.

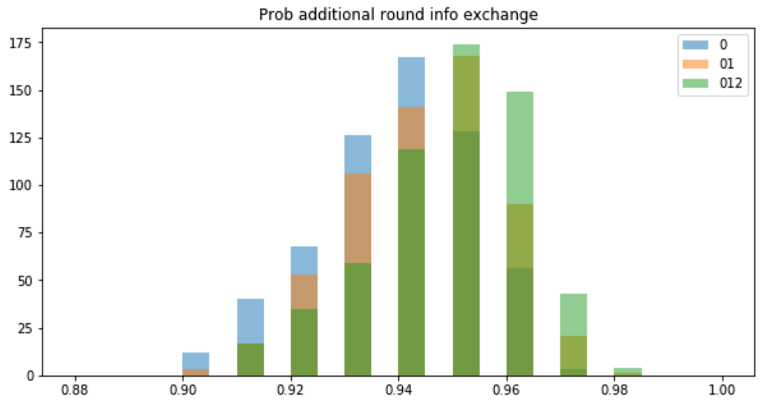


Figure 2.2: Estimates for δ.

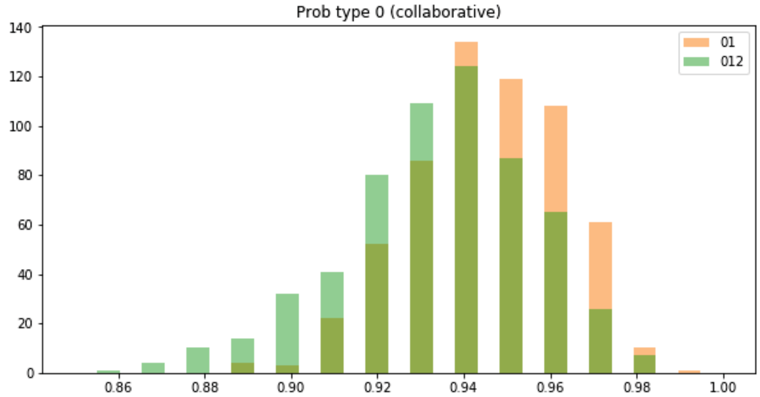


Figure 2.3: Prob. type-0 players.

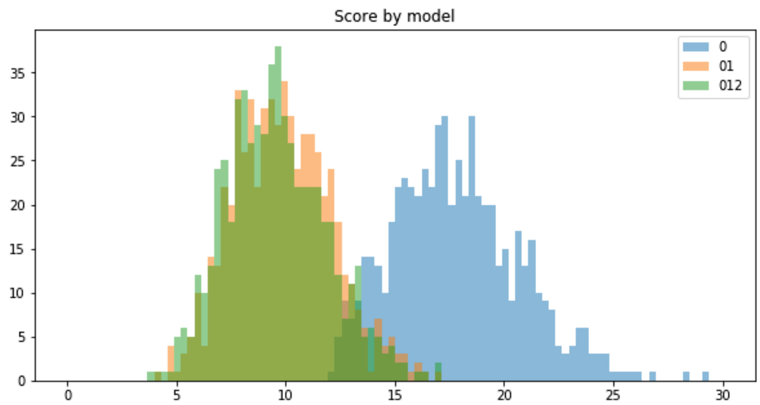


Figure 2.4: Quadratic distance from empirical moments.

Table 2.1: Prob. Info Exchange.

	AB	BE	CD	CE	DE
All	0.63	0.45	0.61	0.72	0.67
Network 1	0.65	0.41	0.63	0.77	0.63
Network 2	0.51	0.38	0.65	0.73	0.67
Network 3	0.71	0.52	0.57	0.67	0.72

Note: Each cell reports the probability of having an exchange of information between couples of nodes. We use only the sample of groups in which all 5 players actively participated to the round by submitting a set of choices, so that we have 29 obs for network 1, 41 obs for network 2, and 52 obs for network 3.

Chapter 3

Are Online Reviews Manipulated? Evidence from Amazon.com

3.1 Introduction

With the extraordinary growth of e-commerce that took place starting from the early 2000s, online reviews have acquired more and more importance as a source of information for consumers. Websites that allow users to write reviews have become popular and it is easy now to find reviews about any kind of product or service available on the market. The possibility given to anyone to share opinions concerning a good has raised concerns regarding the possible influence that biased individuals could exert in order to distort the market. This topic has attracted interest not only in the academic community. Industry leaders such as Amazon.com have taken a number of actions to address the problem of fake reviews. For example, suits were filed against owners of websites selling fake reviews[1] and new technologies have been developed to deal with this issue[2]. Still, identifying the presence of fake reviews is a hard task as biased reviewers write their reviews in such a way to mimic unbiased ones. The aim of this work is therefore to find conditions under which fake reviews are more likely to be written and show the effects on information available to consumers. To do that we exploit the fact that on Amazon.com average ratings are rounded to the closest half star when first

[1]http://www.forbes.com/sites/retailwire/2015/04/13/amazon-lawsuit-takes-on-fake-reviewers/
[2]http://www.theguardian.com/technology/2015/jun/22/amazon-ai-fake-reviews-star-ratings-astroturfing

showed to the consumers (see Figure 3.1 and Figure 3.2). As consumers are likely to pay attention to this piece of information, sellers may want to manipulate the rating in order to increase sales. In principle sellers would always benefit from an extra positive review. The average rating would increase together with the number of consumers highlighting positive features of the product. Some sellers (or other people interested in increasing the success of the good) would therefore always consider fake reviews as part of their promotional activity. Still, not all reviews have the same impact. Depending on the average rating prior to the review and the number of reviews, there could be the possibility of increasing the number of stars. Therefore the benefit is likely to depend on factors that vary over time. Our hypothesis will be that instead, costs associated to the submission of promotional reviews are constant over time and only depend on firm characteristics. We will see that this idea will be the basis of our empirical strategy.

In order to link to the previous literature I start by considering a regression discontinuity setting and see whether there is evidence for selection. In particular the treatment that will be considered is the increase in half-star determined by the crossing of .25 and .75 thresholds. For example, a book with average rating equal to 3.74 would get 3.5 stars while for another book with average rating 3.75 the number of stars would equal to 4. The first hypothesis that I will try test is therefore the presence of selection, that is, the tendency of average ratings to lie on the right of the cutoff level. I will also analyze characteristics of reviewers, comparing the left and right side of the cutoff.

As a second step I move from data at the item level (books) to data at the review level. This allows me to overcome some limitations of the previous approach. In particular I study which are the factors that explain higher grades. In the presence of unbiased reviewers we would expect the rating to only depend on factors such as quality, opinion of previous reviewers and idiosyncratic noise. On the contrary, I find that the possibility of increasing the number of stars explains some of the variation. In particular, I find that reviews which can have an impact on the number of stars tend to be associated with a higher grade. This is compatible with the hypothesis described above of sellers actively involved in trying to manipulate the average rating and therefore the number of stars associated to the item. This allows me to deal with the problem typically associated to this kind of analysis, that is the attempt of promotional reviewers to craft their reviews in such a way to appear unbiased. I will indeed avoid to focus on single characteristics of any particular review, but rather study the impact that the single review can have on the information presented to buyers.

Finally, I set up a difference in differences analysis to take advantage of heterogeneity among publishers. In particular, I distinguish between self-published books and books published by large companies that rely on a traditional business model. These two groups of publishers face different incentives. Big publishers rely on Amazon.com for a smaller fraction of their business and face higher reputation costs from engaging in promotional reviews. On the contrary, small independent publishers do not have well-known brands to defend, thus promotional reviews tend to be relatively less risky for them. I therefore compare reviews written for books sold by these two categories of publishers and compare ratings below and above the threshold that assigns an extra half star. I find that reviews for self-published books have higher ratings with respect to the other books, especially below the threshold. This confirms that smaller businesses are more likely to rely on promotional reviews.

This work proceeds as follows. In Section 3.1.1 I briefly discuss the related literature. Section 3.2 describes the data and presents summary statistics. In Section 3.3 the first part of the analysis is done, with data on books. Section 3.4 includes results for single reviews. Section 5 we conclude and discuss possible extensions.

3.1.1 Literature

The early literature on online reviews has focused on the effects that online reputation has on sales. Resnick and Zeckhauser (2002) and Resnick et al. (2006) are among the first contributions to show that on Ebay.com sellers characterized by higher ratings sell more units and are able to obtain higher prices. Babič et al (2015) contains a meta-analysis showing how electronic word of mouth has a positive effect on sales. The contributions by Chevalier and Mayzlin (2006), Luca (2011) and Anderson and Magruder (2012) are closer to my work. Chevalier and Mayzlin (2006) use data on books sales rankings from Amazon.com and BarnesandNobles.com and apply a difference-in-differences analysis to compare the evolution of sales in the two websites when ratings are different. They find that positive reviews are related to an increase in relative sales. Luca (2011) and Anderson and Magruder (2012) focus instead on restaurants reviews on Yelp.com. This website displays average rating in a way that is similar to what is done by Amazon.com: ratings are rounded to the closest half star and the star rating is shown in the main search pages. This fact allows the authors of both papers to use regression discontinuity design to measure the impact that an extra half-star has on sales. In Luca (2011) the author measures sales using data from the Washington State Department of Revenues, instead Anderson and Magruder (2012) rely on a database of restaurant reservation availability. In both papers the authors find a sizable effect of star rating on sales. In particular, Luca (2011) finds that a one-star increase in Yelp rating is associated with a 5-9 percent increase in revenues, while Anderson and Magruder (2012) suggest that an extra half-star leads on average to a gain of approximately 800$ per week in pre-tax profits. These works therefore motivate my approach, as they show that there are strong potential gains from a extra half-star rating in revenues[3].

Positive reviews seems therefore to be associated with better sales. Hence, there are incentives for the sellers to manipulate these ratings. For this reason, recent contributions have moved in the direction of studying fake reviews. Mayzlin et al. (2014) study promotional reviews for hotels. They compare two platforms that follow different criteria in determining

[3] While showing a strong effect on revenues, these papers bring evidence that there is no selection around the cutoff in the type of restaurant. This is needed to justify the use of a regression discontinuity design. In this paper instead I provide evidence that on Amazon.com manipulation is at work.

who is allowed to write reviews. On TripAdvisor.com anyone can submit a review, while on Expedia.com only those who have actually used the website to pay for a stay are allowed to. This generates a difference in the costs associated to the submission of a promotional review between the two platforms. The authors exploit this fact in conjunction to the hypothesis that independent hotel with single unit owners would benefit more from positive reviews than chain hotels that rely on better known brands. In this context, a difference in differences approach allow them to establish the presence of reviews manipulation done in particular by independent owners. Fake reviews are used to increase their average rating or to damage reputation of their closest competitors. Another relevant contribution in this literature is Luca and Zervas (2015). In this paper the authors study incentives to commit review fraud on Yelp.com. By looking at reviews that were filtered by Yelp.com (because considered fake) they find that fake reviews are more likely to appear when the number of reviews is low or after bad reviews. Moreover, big chains are less likely to manipulate reviews and increases in competition lead to higher number of fake reviews. In my analysis I study a different platform, Amazon.com, with a novel identification strategy that combines the difference-in-differences approach used by Mayzlin et al. (2014) with the regression discontinuity design employed by Luca (2011) and Anderson and Magruder (2012). My findings therefore extend this literature to a new platform (Amazon.com) and show how sellers are exploiting features of the design at their own advantage.

There are several reasons why one could expect fake reviews to have detrimental effects. First, by providing misleading information about products, it would lead consumers to choices that do not reflect their preferences (Mayzlin (2006), Dellarocas (2006)). This in turn, as pointed out by Nasko and Tadelis (2015) would damage also the platform. Bad sellers have indeed a negative externality as they reduce consumers' trust on the platform. Moreover, with a mechanism similar as the one highlighted in Salganik et al (2006) for music, fake reviews could influence the market also by making consumers pay attention on fewer products (those which receive good fake reviews), creating therefore inequality among sellers that is not justified by differences in quality.

3.2 Data

I use data on reviews from Amazon.com. The dataset I use has been collected by Julian McAuley[4] and has been used by McAuley and coauthors for two publications: McAuley, Targett, Shi and van den Hengel (2015) and McAuley, Pandey and Leskovec (2015).

The dataset contains information on 83.06 million reviews written on Amazon.com between May 1996 and July 2014[5]. For each review we know the item it refers to, who wrote it, whether it was voted as helpful by other users, the text, the day it was written and finally the rating. The dataset contains also products metadata. In this case I have information on the title, price, related items (”also bought”, ”also viewed”), the sales rank and the product category.

For the analysis presented here I only used metadata to select a random sample of 476,342 books with more than 20 reviews. For each item I collected the reviews, ordered them in time and calculated other variables such as the average before the review and the variance of ratings. Variables that refer to single reviewers are built in a similar manner, but in this case we take advantage of the availability of data and use the whole sample of reviews. This allows me to map the complete activity of each reviewer over time. Table 3.1 includes summary statistics for the three levels of observation: review, reviewer, product.

One feature of this dataset that it is important to underline is the fact that it has been collected in 2014. This could offer an advantage with respect to data downloaded now because Amazon.com seems to have changed strategy with respect to fake reviews. It was already filtering promotional reviews, when detected, but from June 2015 the effort in this direction

[4]link: http://jmcauley.ucsd.edu/data/amazon/3

[5]Amazon has not made significant changes to the platform in this period for what concerns reviews.

Table 3.1: Books with more than 20 reviews - Summary Statistics

	Mean	SD	Min	Max
N. of Reviews - Book	9.49	54	1	8280
Avg. Rating - Book	4.29	0.895	1	5
N. of Reviews - Reviewer	11.06	52.22	1	44557
Avg. Rating - Reviewer	4.27	0.85	1	5
Tot. N. Reviews	4,522,001			
Tot. N. Reviewers	2,598,773			
Tot. N. Books	476,342			

Note: This table shows summary statistics for reviews written about 476,342 books, randomly selected from the whole sample of books with more than 20 reviews. Data on reviewers refer to users who have written at least one review about one of the books considered, but are then calculated using the whole sample of reviews.

has increased (see the link mentioned in the introduction). In Section 5 we discuss how to improve on these data, as some pieces of information presented in the website have not been downloaded or tracked over time.

I then enriched these data with information concerning publishers. The first step has been to use Amazon.com API to collect additional metadata on the books included in the analysis. For the books included in the analysis, I therefore collected the names of the publishers and ran an online search for each of them. I was therefore able to distinguish whether each book was published by well-established companies (e.g. Hachette, MacMillan, Penguin) or was instead self-published by the author. In total, I found 459 imprints that I could link to publishing companies and approximately 19,200 other names that can be identified as connected to self-published books.

3.3 Book level analysis

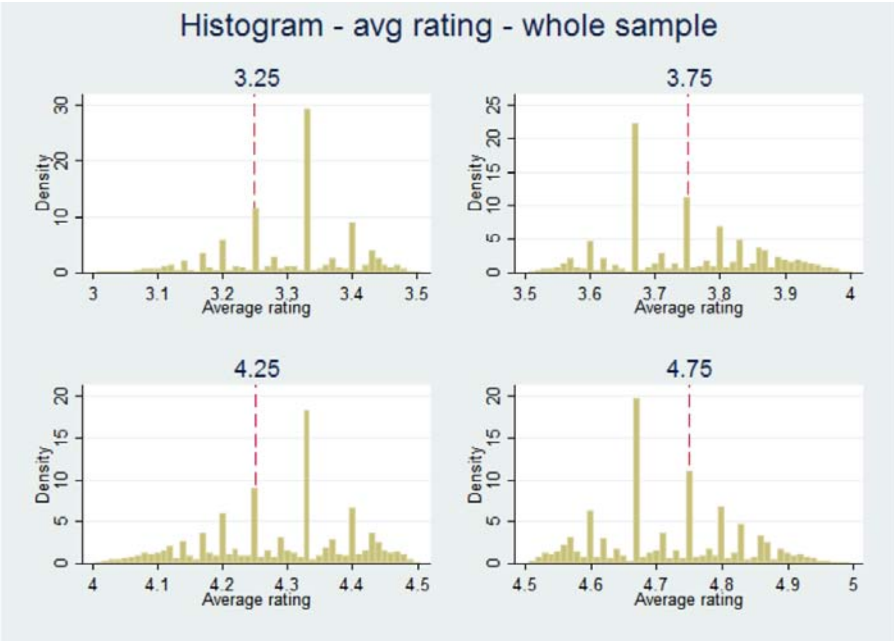
As a first step I study selection in a context that is comparable with what has been done in Luca (2011) and Anderson and Magruder(2012). I consider a regression discontinuity design in which however the question is not the effect of stars on sales (as it was in the afore mentioned papers) but whether individuals self-select into the treatment. With treatment in this case I mean having an extra half star, while the exogenous rule that assigns the treatment is the criterion used to round the ratings. The variable that determines the treatment is therefore the average rating. The average rating depends on reviews written by consumers and possibly on fake reviews. Unbiased reviews can be in principle considered to be a random process that mainly depends on the underlying quality of the product. If this is the case, unbiased reviews should generate a distribution of average ratings that is continuous around the .25 and .75 cutoffs[6]. Similarly, if all sellers manipulate the average rating with the same ability, the final distribution should still be continuous. Instead, if we observe discontinuities around the cutoffs, this could be interpreted as evidence in favor of manipulation by a group of sellers. We will be looking for two types of discontinuities. First we will apply McCrary test to see whether the density of observations jumps so that fewer books lie on the left of the threshold. Second, we will consider jumps in other variables, like characteristics of reviewers, that could suggest the presence of manipulation.

3.3.1 McCrary Test for manipulation of average ratings

We want to test whether sellers are able to manipulate their average grade in order to gain extra stars. If fake reviews are written when the rating is just below the threshold in order to cross it, then we should observe more books with average rating just above the cutoff. To

[6] The .25 threshold would bring an extra half star so that the rounded score would be either half, one and a half, tow and a half, three and a half or four and a half stars. Similarly, the .75 threshold would imply a rounded score of one, two, three, four or five stars.

Figure 3.1: Density of observations around cutoff levels



understand whether this is the case it is useful to start from the observation of histograms. Figure 3.3, Figure 3.4 display histograms for density of observations around the cutoffs. In both cases the variable I am considering takes value in the interval (-0.25, 0.25) that is the distance between the average rating from the closest cutoff. For example for two books with average rating 3.15 and 4.65 the variable would take value -0.1 while for a book with average rating 4.4 it would take value 0.15. The difference between the two figures is the fact that in Figure 3.4 I focus on books that have more than 20 reviews. The two histograms share the characteristics of being almost symmetric around 0. It is hard to identify jumps apart those that are determined by the fact that these ratings are the average of integer numbers that go from 1 to 5 (for example .25, .75, .33, .66). Figure 3.5 and Figure 3.1 present histograms for the average rating around cutoffs 3.25, 3.75, 4.25 and 4.75. Again, it is hard to spot jumps in densities.

In order to formally test the hypothesis described above I implement McCrary test as described in McCrary (2008). Results are presented in Table 3.1 and Figures 3.1, 3.6 and 3.7.

I find that peaks in densities at the cutoff level play a crucial role. For samples that include books with fewer reviews the test rejects the null hypothesis of the absence of discontinuity (columns 1 and 2). However this discontinuity could be determined by the density in .25 and .75 that can be explained with the fact that ratings are the average of integer numbers. It is indeed necessary to underline that for the subsample of books with more than 100 reviews the test does not identify any jump in densities (column 3).

Table 3.1: McCrary test

	(1) More than 20 rev	(2) More than 20 rev	(3) More than 100 rev
Est. Discontinuity	0.28***	0.171***	0
	(.035)	(.028)	(.08)
Bandwidth	0.1	0.15	0.114
Obs.	35,854	35,854	5,500

Note: This table shows MaCrary test results for discontinuity around the thresholds .25 and .75. Columns (1) and (2) refer to two different bandwidths. Column (3) refers to books with more than 100 reviews.

3.3.2 Regression Discontinuity

The previous analysis suggests, in line with the literature, that manipulation is more likely to occur when the number of reviews is lower. Here I implement a regression discontinuity analysis on reviewers and review characteristics to better understand differences between the two subsamples. I proceed as follows. For each book in the sample I take the average of reviewers' characteristics described in Table 3.2 and Table 3.3. For each book I calculate the average length of reviews, the average number of reviews written by reviewers etc. I then

Table 3.2: Reviews - variables - Summary Statistics

Variable	Obs	Mean	SD	Min	Max
Number of Reviews	4,522,001	157.9	511.5	0	8280
Review Length	4,522,001	641.9	890.9	0	32,729
Ratings before the review - Average	4,045,659	4.34	0.59	1	5
Ratings before the review - Variance	4,045,659	0.82	0.75	0	4
Last four ratings - Variance	3,340,851	0.8	0.88	0	3.84

Note: *Number of Reviews* describes the number of reviews before the review considered; *Review Length* expresses the length of the review (number of characters); *Ratings before the review - Average* is the average rating before the review; *Ratings before the review - Variance* is the variance of ratings before the review; *Last four ratings - Variance* is the variance of the previous 4 ratings before the review.

run the following regression:

$$Y_j = \alpha + \beta I(R_j > \hat{R}) + \gamma R_j \tag{3.1}$$

Where R_j measures the distance from the cutoff and $I(R_j > \hat{R})$ is a dummy variable that takes values 1 when the rating is above the cutoff. Table 3.4 reports regression discontinuities estimates for reviewers' and reviews' characteristics. Figure 3.8 report scatter plots around the threshold. It is important to underline that although the coefficients are significantly different from zero, they can not explain a significant fraction of the variance. Results show that books that lie above the threshold have reviews written by users who are more active (more reviews written, longer text) and assign a slightly lower rating to the other products reviewed.

Table 3.3: Reviewers - variables - Summary Statistics

Variable	Obs	Mean	SD	Min	Max
Number of reviews, same ID	2,598,773	11.1	52.2	1	44557
Average Rating, same ID	2,598,773	4.27	0.85	1	5
Ratings Variance, same ID	2,598,773	0.69	0.95	0	4
Time first-last review	2,598,773	2.22	3.13	0	18.2

Note: *Number of reviews, same ID* describes the number of reviews from the same reviewer; *Average Rating, same ID* expresses the average rating from the same reviewer; *Ratings Variance, same ID* is the variance of ratings from the reviewer; *Time first-last review* is the distance measured in years from the first to the last review.

3.4 Review Level Analysis

In order to gain better insights regarding the presence of review manipulation I now use data on single reviews. For each book I have information concerning the whole history of grades so that I can reconstruct the evolution in time of the statistics that are displayed to the user. When submitting a review, an unbiased user should base her decision on the perceived quality of the item. It is also possible that the decision of writing a review and the rating depend on the statistics presented in the website. A higher average, a higher number of stars or a higher number of previous reviews are, among others, elements that can potentially influence one person's opinion regarding a product. I therefore expect the grade to depend on some unobservable quantities like the true quality of the book and on the statistics that the user finds available on Amazon.com. Instead, consider the case of a seller interested in manipulating the rating in order to sell more copies. When writing or buying fake reviews the seller incurs a monetary cost but also risks to be identified by customers or by Amazon as a dishonest seller. Let us assume that this cost does not vary over time. In particular, similarly as in Mayzlin et al. (2014) different sellers can be characterized by different costs, but these costs are constant. If we instead consider the benefits from writing promotional

Table 3.4: Regression Discontinuity Estimates

Variable	Coef.	SD	z	P>z
Average Rating, same ID	-0.034	0.006	-5.25	0.000
Ratings Variance, same ID	-0.012	0.005	-2.43	0.015
Number of reviews, same ID	72.8	14.3	5.09	0.000
Time first-last review	0.29	0.03	8.48	0.000
Review Length	23.37	6.96	3.35	0.001

Note: This table reports regression discontinuity estimates for five different outcomes around the .25 and .75 thresholds described above. *Number of reviews, same ID* describes the number of reviews from the same reviewer; *Average Rating, same ID* expresses the average rating from the same reviewer; *Ratings Variance, same ID* is the variance of ratings from the reviewer; *Time first-last review* is the distance measured in years from the first to the last review; *Review Length* expresses the length of the review (number of characters).

reviews, it is clear that these benefits are a function of the history of grades that evolves over time. Over time, new reviews change the average rating and, with that, increase or decrease the influence that a new review can have on the statistics that buyers are likely to consider when making their decisions. The clearest example is the case of stars. Depending on the average at the moment of writing, a high grade can increase the number of stars that will be assigned to the product. My empirical strategy goes in this direction. I want to see whether the possibility of increasing the number of stars is a factor that helps explaining the rating assigned in a review. First, for a seller it is easy to understand the extent that her review can alter the rounded rating. Second, this information is in principle available also to any unbiased reviewer, but in this case it is hard to imagine why they should find it relevant. Therefore I test the hypothesis that the seller, while keeping track of the evolution of the rating and the number of reviews, is ready to take advantage of those cases in which the number of stars can be easily manipulated. It is clear that I am considering only one of the strategies that sellers may be implementing in order to modify buyers' perception of their products. Still, I take advantage of the fact that in this context the ability to manipulate information varies over time and depends on factors that are not perfectly controlled by the seller, like in particular the grades of unbiased users. I therefore propose a regression model in which the overall grade is regressed on a set of observables at the time of the review and on a set of reviewer characteristics. Among the observables, I am particularly interested on a variable that measures the ability of the review to alter the number of stars that are displayed in the product page. The estimating equation will be as follows:

$$r_i = \alpha + \beta Gain_i + \gamma B1_i + \delta B2_j + \theta B3_r \tag{3.2}$$

Where r_i is the rating attributed in review i by reviewer r for the product j. The variable Gain_i describes the number of stars that can be gained in case the rating was equal to 5. This variable depends on the average at the time of the review and on the number of reviews. The closer (to the left) the average is from the cutoff, the higher Gain_i will be. Similarly, a lower number of reviews will typically be associated with higher levels of Gain_i. $B1_i$ contains controls that pertain to the single review, like the average grade and number of reviews before that review. $B2_j$ will include characteristics of the book as the total number of reviews. Ideally I would include a measure of quality, but this is not observable. Finally, $B3_r$ refers to controls about the reviewers' characteristics like the total number of reviews written over time, the average rating assigned and the variance. I am interested in the correlation between the rating and the number of stars that can be gained. With this model we want to capture the kind of opportunistic calculations that the seller may be doing.

3.4.1 Results

Table 3.1 presents the results of the estimation of the equation described above. Columns (1) and (2) refer to the same model, but consider different samples. In column (3) I included two more variables. The first specification is done using the whole sample of reviews that were preceded by at least another review (I cannot calculate the variable average before for the first review). Controls include the observables that are likely to influence reviewers' opinion regarding the quality of the book. As described above, I am interested in the coefficient of the variable called Gain_i. When I include the whole sample, results show a non negligible and statistically significant effect of the variable on the overall rating. In particular, 1 standard deviation change in Gain_i explains approximately 1.5% variation in the overall rating. This effect is again positive, but smaller, when I only consider books with more that 50 reviews. This result is in line with the findings by Mayzlin et al (2014): books with a higher number of reviews are likely to be have been published by bigger firms that could be less prone to manipulate reviews. In the third specification I include two more explanatory variables. I use the average rating after all recorder reviews as a proxy for quality and we exploit the variable *Variance last 4* to look for evidence in favor or coordination among reviewers. The variable *Variance last 4* indeed measures the variance of rating of the four reviews that preceded the one we consider. Importantly, this variable forces me to consider only reviews that are at least fifth in the sequence of all reviews for a book. There are three interesting elements to underline about Column(3). The coefficient for Gain_i signals now the presence of weak correlation. The coefficient for average before switches sign, likely as an effect of avg end. Finally, the coefficient for *Variance last 4* is negative. This last point suggests the possibility that sellers tend to submit more reviews in a short time period.

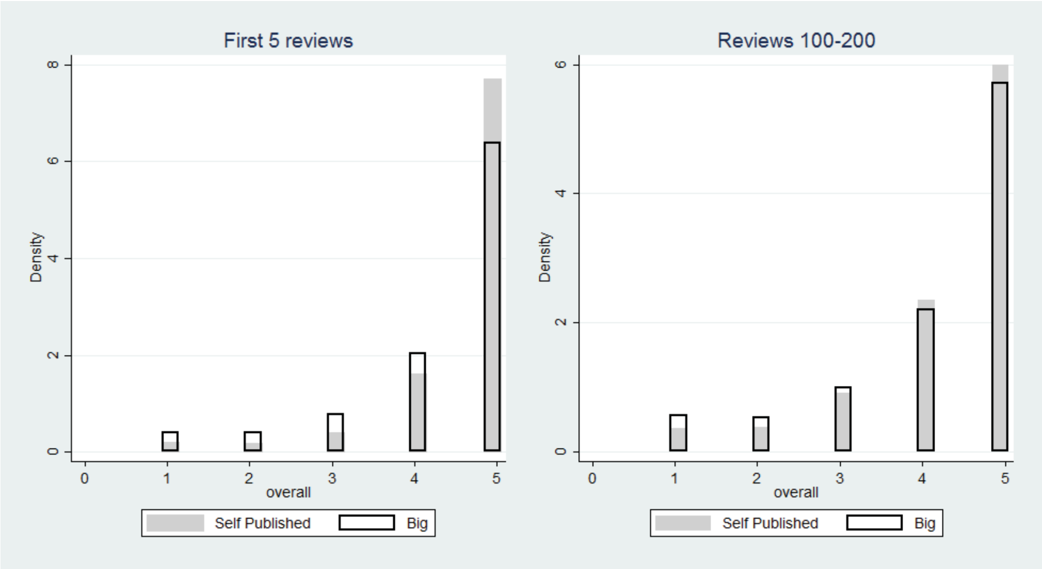
3.5 Publisher type and promotional reviews

This section builds upon the previous analysis adding one element. I combine the regression discontinuity design described above with differences across publishers, to provide further evidence that manipulation is at work.

The main idea is to introduce in the analysis the publisher type. This because publishers of different sizes face different incentives. As mentioned in Section 3.2, there is a vast number of publishers that sell their books on Amazon.com. A stark distinction can be made regarding the size. On the one hand, we have big firms, which rely on a traditional business model. For them Amazon.com is one of the many channels they use to distribute their books. They also rely on advertising and they give high value to their reputation, as customers recognize their brand. On the other hand, Amazon.com allows very small businesses to operate in this industry too. People can easily self-publish their books and sell them on the platform. Amazon.com would then matter for all their revenues from this activity. As these subjects are not known to the public, reputation costs are low. Even in case users realized about their misbehavior, they could easily create a new account in case they wanted to publish a new book. A difference between these two groups emerges directly from the comparison of reviews. Figure 3.1 shows the histogram of reviews received by self-published books and compares it with the histogram related to big publishers. It emerges that self-published books have better reviews, with a difference that is stronger for the first ones. This could be due to promotional reviews but for the moment we cannot exclude other hypotheses, for example the presence of niche readers that are particularly passionate and interested in self-published books.

The discussion above suggest the use of a difference-in-differences set up. One dimension is given by the average rating before the review. If the average rating is just below the threshold for the extra half star, we have that the gain from an additional review is particularly high. The second dimension is given by the type of publisher, as this correlates with differences in

Figure 3.1: Self-published books receive better reviews at first



Note: This figure represents histograms for reviews received by self-published books and book published by big companies. The histogram of the left refers to the first 5 reviews received by books that belong to the two categories. The histogram on the right refers to the reviews from the 100th to the 200th.

costs associated with the fake review.

Notice how this setup shares some similarities with the one used in Mayzlin et al (2014). While studying hotels, they compared independent businesses with branded chain hotels, as these two groups differ in the net gains from promotional reviews. By comparing independent publishers with bigger companies I exploit a similar mechanism. I then combine this consideration with the difference in potential gains that is given by the use of half stars[7].

In the presence of promotional reviews we would therefore not only expect reviews to be better just below the threshold. We would also expect this difference to be starker for self-published books. This is therefore in line with a difference-in-differences approach. The identifying assumption is that sincere reviewers do not change their behavior when reviewing self-published books and other books if the average is below the threshold[8]. In other words

[7]Mayzlin et al (2014) compare instead reviews across two different platforms, Expedia and Tripadvisor. They exploit the fact that costs of writing reviews differ across these two platforms.

[8]This is a way to reformulate the common trend assumption in this context.

I need to assume that sincere reviewers do not directly try to increase the star rating when they see a self-published book.

The previous discussion can be condensed in the following linear probability model.

$$\mathit{Above\ the\ mean}_i = \beta_1 + \beta_2 \mathit{Close}_i + \beta_3 \mathit{Self\ Published}_i + \beta_4 \mathit{Close}_i \cdot \mathit{Self\ Published}_i + X_i\beta_5 + \epsilon_i \tag{3.3}$$

Where *Above the mean* is a dummy variable that takes value 1 if the review has a rating that increases the average. *Close* is a dummy variable that takes value 1 if the average rating before the review is just below the threshold for the half star and value 0 if it just above. *Self Published* is a dummy variable that identifies self-published books. I include controls on the number of reviews, the average rating before the review and other book characteristics.

Table 3.1 reports results for the main specification run with three different subsamples of reviews. Table 3.2 considers the use of wider intervals around the threshold. In particular, the difference-in-differences term of interest is β_4. This coefficient is positive and significant, in line with the previous hypothesis. I therefore find that reviews written for self-published book when the average rating is just below the threshold have a high probability of being above the average rating. This is result is robust if I consider different bandwidths and different samples of reviews.

3.6 Conclusions

Overall, the picture that emerges from the analysis described above suggests that on Amazon.com a fraction of sellers are using fake reviews as a way to manipulate information available to buyers. In particular I showed that this activity may be aimed at increasing the number of stars associated to the products they sell. To show this I used data on book reviews. First, I aggregated reviews by book and ran tests aimed at identifying the presence

Table 3.1: Difference in Differences Estimates

	(1) Above the mean	(2) Above the mean	(3) Above the mean
Close	0.149 (111.16)***	0.167 (109.28)***	0.210 (106.48)***
Self Published	-0.0911 (-65.21)***	-0.0805 (-55.66)***	-0.0656 (-43.17)***
Close · Self Published	0.126 (56.21)***	0.126 (50.92)***	0.128 (41.24)***
Average Rating before	-0.174 (-180.73)***	-0.195 (-187.64)***	-0.246 (-207.43)***
Number of reviews	0.00125 (107.72)***	0.00249 (102.62)***	0.00426 (64.18)***
Observations	887338	740764	536731

t statistics in parentheses
(1) First 200 reviews
(2) First 100 reviews
(3) First 40 reviews
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

of selection around the threshold that assigns ax extra half star. These tests are McCrary test on the density of observations around cutoff levels and regression discontinuity estimates of jumps in average books characteristics around these thresholds. In both cases I found evidence that point towards manipulation. I then moved to consider data on single reviews. In this case I exploited variation in the ability to manipulate the number of stars and measured the response of ratings. I found a positive correlation between the ability to manipulate the number of stars and the assigned rating. I then provided additional evidence by considering a difference-in-differences model in which I compared reviews written for self-published books and reviews written for books published by bigger firms, as big publishers have a higher cost from engaging in the use of promotional reviews coming from reputation costs. I found that below the threshold, average rating are higher for self-published books, in line with the hypothesis that these group of publishers is more likely to use fake reviews.

I therefore contribute to the literature that studies manipulation of information in online platforms. I do this by considering the case of Amazon.com and employing a novel identification strategy that combines two strands of contributions. On the one hand, I exploit characteristics of the platform, such as the use of half stars to round average ratings. On the other hand I exploit differences across business models by comparing traditional publishing companies with self-published books.

This project has implications for the design of review systems in online platforms. The use of a system that is transparent as the one used by Amazon.com facilitates the action of agents who are interested in manipulating their reputation online. By investing in more complex reputation systems, online platforms would therefore reduce incentives for misbehavior.

The lack of data on sales prevented me from studying how consumers react to the presence of manipulation. In particular, it is important to understand whether consumers are able to detect behaviors put in place by firms that are trying to manipulate information available on the platform. This would have important policy implications.

References

Anderson M., and J. Magruder 2012. "Learning from the Crowd: Regression Discontinuity Estimates of the Effects of an Online Review Database" The Economic Journal.

Babić Ana, Francesca Sotgiu, Kristine de Valck, Tammo H.A. Bijmolt 2016. "The Effect of Electronic Word of Mouth on Sales: A Meta-Analytic Review of Platform, Product, and Metric Factors" Journal of Marketing Research, Volume 53, Issue 3.

Chevalier J., and D. Mayzlin 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews" Journal of Marketing Research 43(3):345-354.

Chintagunta P., S. Gopinath, and S. Venkataraman 2010. "The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout, and Aggregation Across Local Markets" Marketing Science 29(5):783-961.

Imbens G., and K. Kalyanaraman 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator" The Review of Economic Studies 79(3):933959.

Luca M. 2011. "Reviews, Reputation, and Revenue: The Case of Yelp.com" Working Paper.

Luca M., and G. Zervas 2015. "Fake It Till You Make It: Reputation, Competition, and Yelp Review Fraud" Management Science 62(12):3393-3672.

Mayzlin, Dina 2006. "Promotional chat on the Internet" Marketing Science 25(2):155-163.

Mayzlin, Dina, Yaniv Dover and Judith Chevalier 2014. "Promotional Reviews: An Empirical Investigation of Online Review Manipulation" The American Economic Review 104(8):2421-2455.

McAuley J., R. Pandey, J. Leskovec 2015. "Inferring networks of substitutable and complementary products." Knowledge Discovery and Data Mining (KDD).

McAuley J., C. Targett, J. Shi, A. van den Hengel 2015. "Image-based rec- ommendations on styles and substitutes." SIGIR.

McCrary J. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test" Journal of Econometrics 142(2):698-714.

Nasko, Chris, and Steven Tadelis 2015. "The Limits of Reputation in Platform Markets: an Empirical Analysis and Field Experiment." NBER Working Paper No. 20830.

Resnick, P., and R. Zeckhauser 2002. "Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System." Advances in Applied Microeconomics.

Resnick P., R. Zeckhauser, J. Swanson, and K. Lockwood 2006. "The value of reputation on eBay: A controlled experiment" Experimental Economics 9(2):79101.

Salganik, Matthew J., Peter Sheridan Dodds, and Duncan J. Watts 2006. "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market" Science Vol. 311, Issue 5762, pp. 854-856.

Additional Figures and Tables

Figure 3.1: Screenshot from Amazon.com - Bestsellers

Figure 3.2: Screenshot from Amazon.com - Review Page

Figure 3.3: Density of observations around cutoff levels

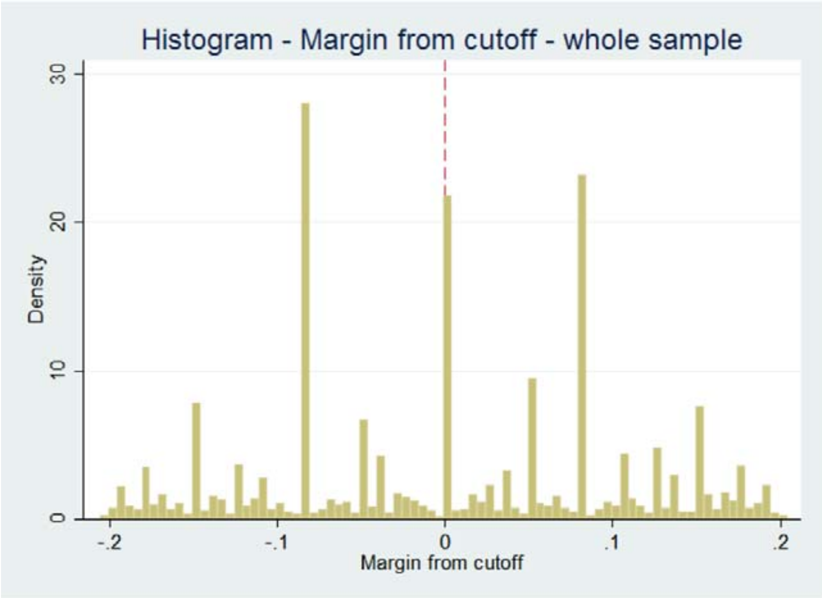


Figure 3.4: Density of observations around cutoff levels

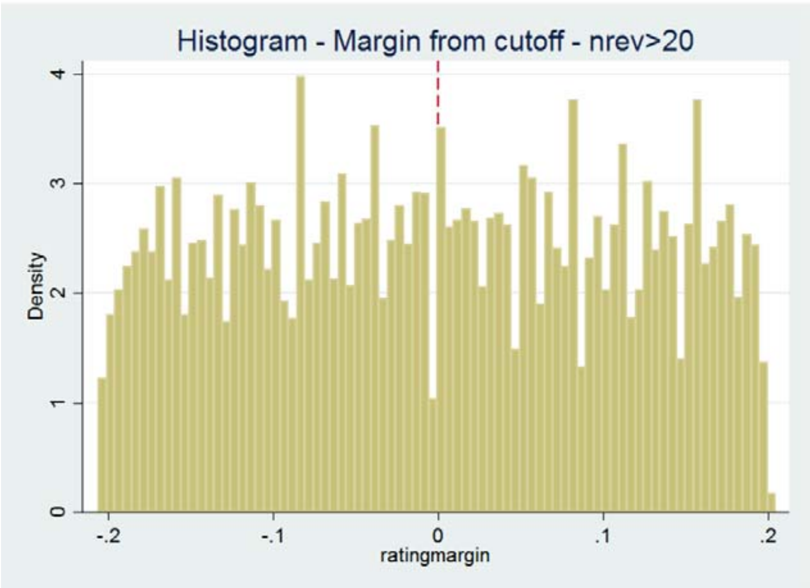


Figure 3.5: Density of observations around cutoff levels

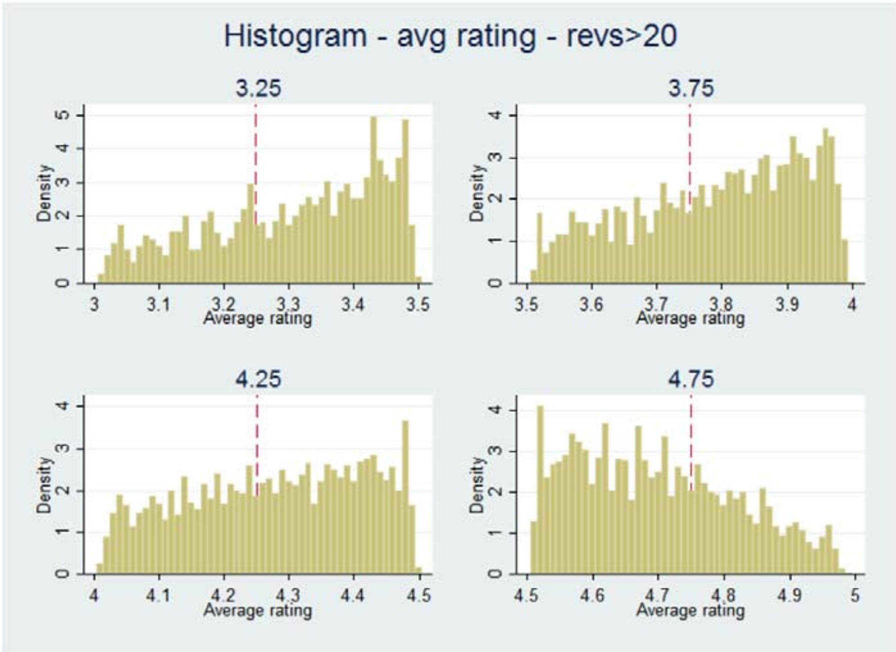


Figure 3.6: McCrary Test, first specification

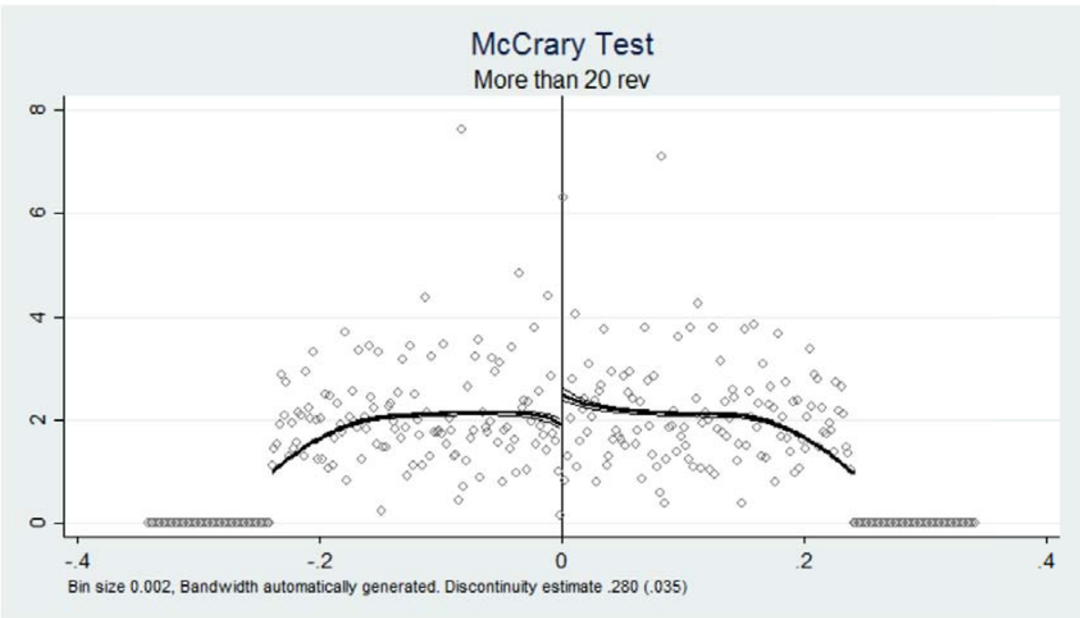


Figure 3.7: McCrary Test, second specification

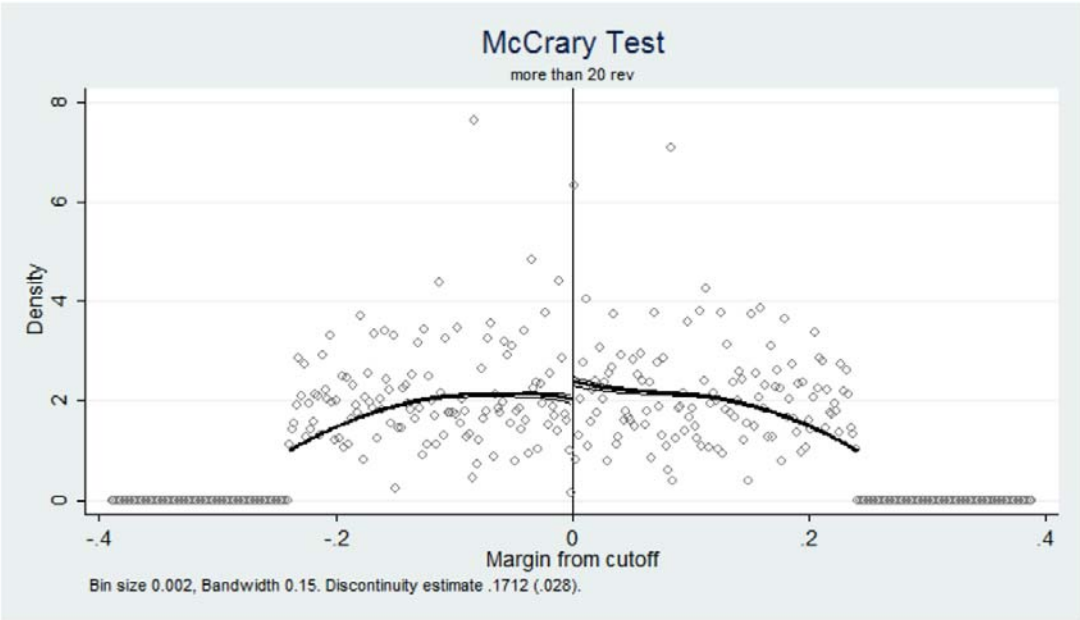


Figure 3.8: McCrary Test, third specification

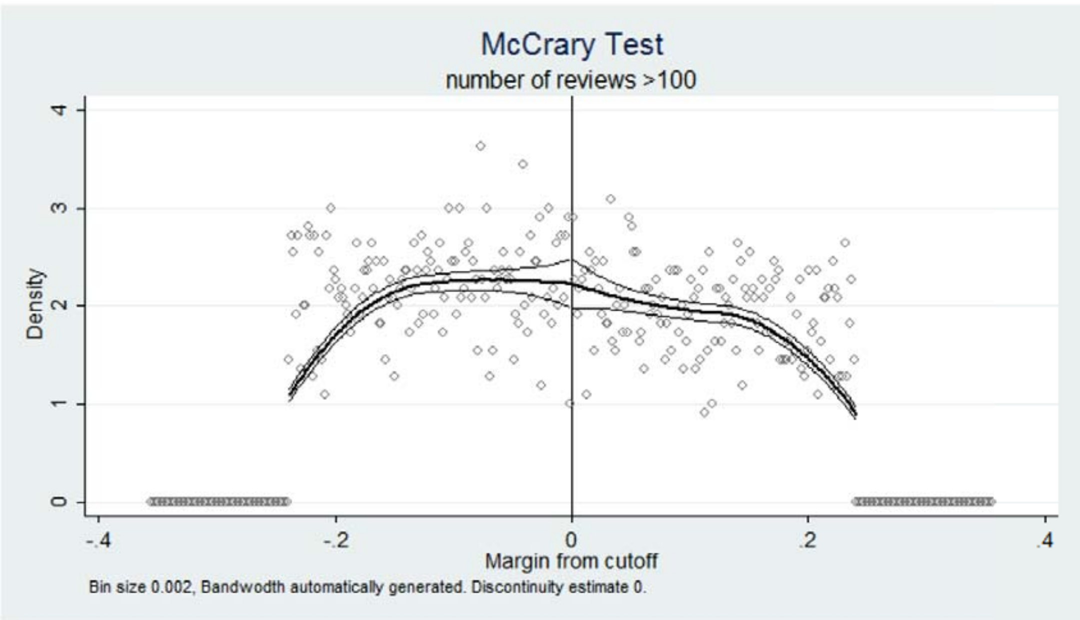


Figure 3.9: Scatter plot around cutoff values

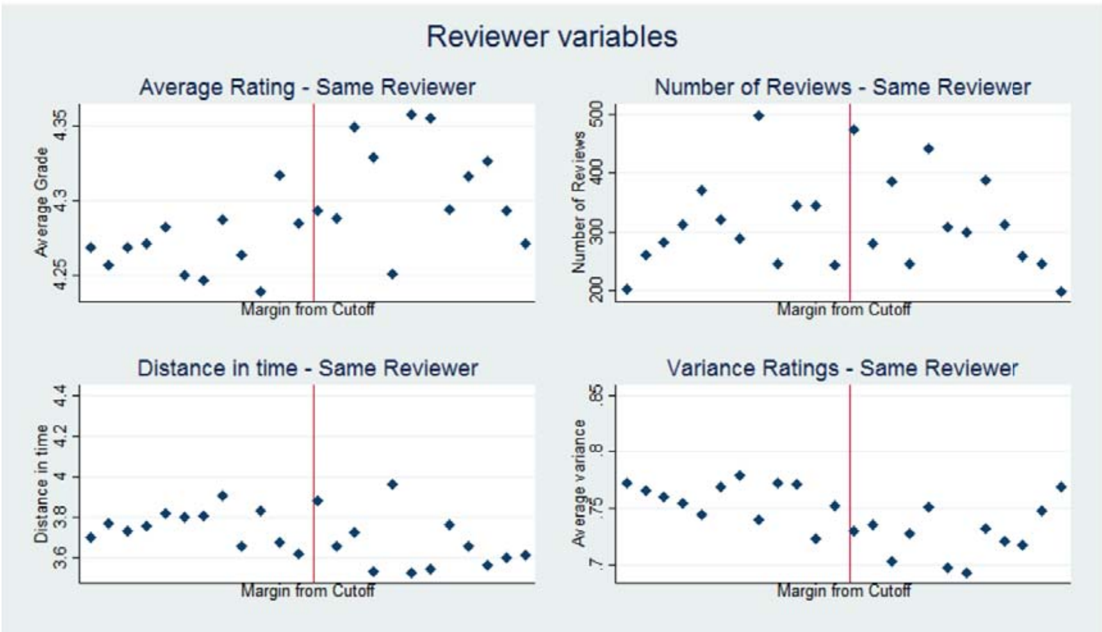


Table 3.1: Regression Discontinuity Estimates

Variable	(1) Whole Sample	(2) More than 50 rev	(3) More than 50 rev
Gain	0.282***	0.180***	0.101***
	(.008)	(.009)	(.007)
Number of Reviews	2.22e-06	3.15e-06	-1.57e-05**
	(5.74e-06)	(6.12e-06)	(7.66e-06)
Average Rating Before	0.286***	0.474***	-0.162***
	(.007)	(.012)	(.009)
Variance Rating Before	-0.007***	0.01**	0.23***
	(.002)	(.005)	(.003)
Number of Stars Before	0.08***	0.014*	0.04***
	(.007)	(.008)	(.005)
Number of Reviews Final	2.69e-05***	1.17 e-05**	9.25e-06
	(5.31e-06)	(4.59e-06)	(7.22e-06)
Number of Reviews, same ID	1.14e-06***	4.52e-06***	4.71e-06***
	(1.28e-07)	(3.63e-07)	(2.21e-07)
Average Rating, same ID	0.949***	0.929***	0.836***
	(.002)	(.003)	(.002)
Variance Rating, same ID	0.0650***	0.0664***	0.0596***
	(.003)	(.005)	(.003)
Time first-last review	-0.005***	-0.006***	-0.003***
	(.0002)	(.0004)	(.0002)
Average Rating Final			0.751***
			(0.007)
Variance last 4			-0.31***
			(.0011)
Observations	4,045,659	2,138,477	3,340,851
R-squared	0.31	0.309	0.353
σ_{Gain}	0.213	0.094	0.114

Note:*Number of Reviews* describes the number of reviews from the same reviewer; *Average Rating, same ID* expresses average rating from the same reviewer; *Variance Rating, same ID* is the variance of ratings from the reviewer; *Time first-last review* is the distance measured in years from the first to the last review.

Table 3.2: Regression Discontinuity Estimates

	(1) Above the mean	(2) Above the mean	(3) Above the mean
Close 02	0.149		
	(111.16)***		
Close 05		0.0973	
		(106.60)***	
Close 10			0.0596
			(92.52)***
Self Published	-0.0911	-0.0870	-0.0641
	(-65.21)***	(-75.15)***	(-72.91)***
Closed · Self Published	0.126	0.0978	0.0626
	(56.21)***	(61.30)***	(54.37)***
Average Rating before	-0.174	-0.0874	-0.0137
	(-180.73)***	(-119.60)***	(-24.93)***
Number of reviews	0.00125	0.00102	0.000659
	(107.72)***	(131.58)***	(120.09)***
Observations	887338	1712930	3342857

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Acknowledgments

I thank my advisor Eliana La Ferrara for her support, her comments, and her kindness, which has been a great source of inspiration. I'm also grateful to Paolo Pin with whom I have the pleasure to collaborate. I also thank Guido Tabellini, Fernando Vega-Redondo, Diego Ubfal, Pamela Giustinelli, Nenad Kos, Salvatore Nunnari, and Nicola Pavoni for the help they gave me during these years. I want to mention also Angela Baldassarre and Silvia Acquati.

I'm grateful to my fellow PhD students, Alex, Carrie, Francesco, Katya, Marco, Santiago, Stefano and Tommaso, for their support and for their friendship.

Finally I want to thank my family, my parents, my brother, and my grandmother, for always being there for me.