# Birds of the Same Feather Tweet Together. Bayesian Ideal Point Estimation Using Twitter Data

**Pablo Barberá**
pablo.barbera@nyu.edu

**June 2013**

**Abstract**

Parties, candidates, and voters are becoming increasingly engaged in political conversations through the micro-blogging platform Twitter. In this paper I show that the structure of the social networks in which they are embedded has the potential to become a source of information about policy positions. Under the assumption that social networks are homophilic (McPherson et al., 2001), this is, the propensity of users to cluster along partisan lines, I develop a Bayesian Spatial Following model that scales Twitter users along a common ideological dimension based on who they follow. I apply this network-based method to estimate ideal points for Twitter users in the US, the UK, Spain, Italy, and the Netherlands. The resulting positions of the party accounts on Twitter are highly correlated with offline measures based on their voting records and their manifestos. Similarly, this method is able to successfully classify individuals who state their political orientation publicly, and a sample of users from the state of Ohio whose Twitter accounts are matched with their voter registration history. To illustrate the potential contribution of these estimates, I examine the extent to which online behavior is polarized along ideological lines. Using the 2012 US presidential election campaign as a case study, I find that public exchanges on Twitter take place predominantly among users with similar viewpoints.

# 1 Introduction

The micro-blogging service Twitter has become one of the most important communication arenas in daily politics. Despite being initially conceived as a website to share personal status updates, it has now become a massive phenomenon, with 200 million montly active users worldwide[1], including 15% of all online Americans[2]. All these users engage in a permanent interaction, 140 characters at a time, exchanging opinions and debating about news events in real-time. The content and structure of this conversation, easily accessible through the Twitter API, represents a unique opportunity for researchers interested in the study of elections and public opinion.

One distinct characteristic of this online social network is the presence of not only ordinary citizens, but also public officials, political parties, and candidates. A vast amount of information is publicly available about each of them: the content of their messages, who they decide to follow, and how they interact with other users. The purpose of this paper is to explore to what extent this new source of data can be used to estimate reliable policy positions for both types of actors.

Measuring parties' and voters' policy positions is a relevant, yet complex, scientific endeavor. Studies of government formation and stability (Strom, 1990; Laver and Shepsle, 1996; King et al., 1990), political competition (Inglehart, 1990; Franklin, 2004; Dow, 2001; Adams et al., 2005), policy outcomes (Blais et al., 1993; Kedar, 2005), and institutional reform (Sartori, 1994; Reynolds, 2002) require systematic information on the placement of key political actors on the relevant policy dimensions. Empirical tests of spatial voting models (Downs, 1957; Stokes, 1963; Lau and Redlawsk, 1997; Jessee, 2009) also rely on measures of citizens' positions on these dimensions. This type of information is necessary not only to test whether they are accurate predictors of vote choice, but also in order to advance our knowledge of electoral behavior and public opinion.

A particularly promising aspect of Twitter is that ordinary users and politicians interact within the same symbolic framework. They use the same type of language, in messages of identical length, which feature the same external references and very frequently even the same content – as a result of the use of hashtags and "re-tweets" –, and most importantly, they are embedded in a common social network. This opens the possibility of estimating ideological positions of both political types of actors on the same scale, which could help overcome a significant limitation of the existing methods: the lack of a common ideological scale for legislators and voters (Shor et al., 2010).

The method I propose relies on the characteristics of the social ties that Twitter users develop with each other and, in particular, with the political actors (politicians,

[1]Source: Twitter's Official Twitter Account, December 18, 2012. [link]
[2]Source: The Pew Research Center's Internet & American Life Project, February 2012. [link]

think tanks, news outlets...) they decide to follow. I argue that valid policy positions for ordinary users *and* political actors can be inferred from the structure of the following links across these two sets of Twitter users. Following decisions are considered costly signals that provide information about Twitter users' perceptions of both their ideological location and that of the political accounts.

My argument hinges on the assumption that Twitter users prefer to follow other accounts whose ideology is similar to theirs. This is not a strong assumption for two reasons. First, it is a well-established finding that social networks are homophilic (McPherson et al., 2001) and thus individuals tend to relate and interact more often with those of similar traits. Second, given that Twitter is also a news media (Kwak et al., 2010), this pattern is reinforced by "selective exposure" (Bryant and Miron, 2004) to sources of information biased in the same direction as each user. Drawing an analogy with offline behavior, this argument would be equivalent to using the choice of sources of political information voters make as a proxy for their political preference.

More specifically, I implement a spatial following model that considers ideology as a latent variable, whose value can be inferred by examining which political actors each user is following. Whether a user $i$ decides to follow a political account $j$ is modeled as a function of the Euclidean distance between them on the latent ideological dimension, as well as two other parameters measuring the popularity of user $j$ and how interested in politics user $i$ is. Unlike similar studies (Conover et al., 2010; King et al., 2011; Boutet et al., 2012), the method I propose allows us to estimate policy positions, with standard errors, on a multidimensional scale, for all types of Twitter users, and across different countries.

To illustrate the method, I generate ideal point estimates for a large sample of active users in the US and the four European countries with the highest number of Twitter accounts – the United Kingdom, Spain, Italy, and the Netherlands. In all five countries, parties and individual candidates have a very visible presence in Twitter, and engage in frequent conversations between each other and with ordinary citizens, which is a clear sign of the increasing use of this social network as a tool for political debate. At the same time, the variation in the size of their party systems allows me to examine whether the positions of the different parties on the resulting scale are congruent with their location on the left-right axis.

My results show that this method generates valid ideology estimates for both politicians and citizens. In the US, the resulting ideal point estimates for the members of the House and Senate are highly correlated with measures that rely on their roll-call votes (Poole and Rosenthal, 2007). Similarly, most individuals who self-identify as "liberal", "moderate', and "conservative" on their Twitter profiles are successfully classified on the left, center, and right of the resulting ideological scale. Average ideal points by state are also highly correlated with ideology as measured by surveys (Lax and Phillips, 2012). In order to further validate the method, I also match a sample of Twitter accounts from the state of Ohio with their voter registration records, based on their full name and county, finding that Twitter-based ideal points are good predictors of party registration. Finally, in the UK, Spain, Italy, and the Netherlands, the method I propose is able to cluster members of the same political party on similar locations of the latent dimension, and their positions are congruent with other measures based on manifestos or surveys (Bakker et al., 2012).

To illustrate a potential use of these estimates, I provide an application where the method I propose can make a substantive contribution. I examine the extent to which online behavior during the 2012 US presidential election campaign is clustered along ideological lines – finding support for the so-called "echo-chamber" theory – and high levels of political polarization at the mass level on Twitter.

The rest of this article proceeds as follows. Section 2 examines the existing literature on the use of Twitter data in the Social Sciences and discusses the opportunities and challenges in this field. Section 3 presents the ideal point estimation method I propose in the context of the different alternatives available. Section 4 describes the data, together with some basic summary statistics. Results of my analysis are shown in section 5, and an application that uses the resulting measures is presented in section 6. The article concludes in section 7 with a summary of my main findings and a list of possible paths for future research.

## 2 Background. Wading into the (Political) Tweet Stream.

The increase in the use of social media has led many social scientists to examine whether specific patterns in the stream of tweets might be able to predict real-world outcomes, such as movies' box-office revenue (Asur and Huberman, 2010), prevalence of the flue Lampos et al. (2010), temporal patterns of happiness Golder and Macy (2011); Dodds et al. (2011), and even the epicenter of earthquakes (Sakaki et al., 2010).

Given the accuracy of these predictions, and the consolidation of Twitter as a source of political information, a battlefield for campaigning, and a public forum of political expression, some researchers have argued that "tweets" validly mirror offline public opinion and can even predict elections (O'Connor et al., 2010; Tumasjan et al., 2010; Sang and Bos, 2012; Congosto et al., 2011). Critics (Metaxas et al., 2011; Gayo-Avello, 2012) have respondent that the "predictive power of Twitter regarding elections has been greatly exaggerated": most of these electoral predictions do not perform better than mere chance. In their view, an accurate prediction can only come through "correctly identifying likely voters and getting an un-biased representative sample of them". The average internet user is younger, more interested in politics, and comes from a higher socioeconomic background than the average citizen, which raises concerns about external validity (Mislove et al., 2011; Gong, 2011). It is therefore necessary to obtain more background information about each individual user, so that it is possible to calibrate any analysis that relies on social media data. This paper represents a step forward in that direction.

Extracting socioeconomic information in Twitter is a difficult task, because users are not even asked to provide their age or gender, as it is the case in other social networks. However, developing techniques to estimate social media users' individual attributes serves three important purposes. First, this type of information improves our understanding of the profile of who participates in online social networks and how representative of the entire population is a random sample of social media users. Second, individual-level data can be very valuable in the process of generating reliable public opinion estimates. For example, if we are interested in studying the Republican primary election, it would allow us to sample only supporters of this party and avoid simplifying assumptions (see King et al., 2012). Future studies that aim at capturing public opinion trends using language processing techniques could use these data to stratify and weigh their estimates.

Most importantly, ideology and other personal traits could be particularly useful as a covariate in future studies about online and offline political behavior. For instance, given the use of Twitter as a coordination mechanism in an era of new types of social protest, this variable might be useful to study how political action spreads across partisan networks. It could also prove to be useful in the study of party competition and electoral behavior, for it might provide ideal point estimates for legislators and ordinary citizens within different regions or states, which would improve the existing empirical tests of spatial voting models. These reasons justify the relevance of the method I present in this paper.

# 3 Ideal Point Estimation Using Twitter Data

## 3.1 Previous Studies

There is a limited but increasing literature on the measurement of users' attributes in social media, particularly in the field of computer science. Despite ideology[3] being one of the key predictors of political behavior online, their measurement through social media

[3]Ideology is defined here as the main policy dimension that articulates political competition: "a line whose left end is understood to reflect an extremely liberal position and whose right end corresponds to extreme conservatism" (Bafumi et al., 2005, p.171). Each individual's ideal point or policy preference corresponds to their position on this scale.

data has only been examined in a handful of studies.

These studies have relied on three different sources of information to infer Twitter users' ideology. First, Conover et al. (2010) focus on the structure of the conversation on Twitter: who replies to whom, and who retweets whose messages. Using a community detection algorithm, they find two segregated political communities in the US, which they identify as democrats and republicans. Second, Boutet et al. (2012) argue that the number of tweets referring to a British political party sent by each user before the 2010 elections are a good predictor of his/her party identification. However, Pennacchiotti and Popescu (2011) and Al Zamal et al. (2012) have found that the inference accuracy of these two sources of information is outperformed by a machine learning algorithm based on a user's social network properties. In particular, their results show that the network of friends (who each individual follows on Twitter) allows to infer political orientation even in the absence of any information about the user. Similarly, the only (to my knowledge) political science study that aims at measuring ideology (King et al., 2011) uses this type of information. These authors apply a data-reduction technique to the complete network of followers of the U.S. Congress, and find that their estimates of the ideology of its members are highly correlated with estimates based on roll-call votes.

From a theoretical perspective, the use of network properties to measure ideology has several advantages in comparison to the alternatives. Text-based measures need to solve the potentially severe problem with disambiguation caused by contractions designed to fit the 140-character limit, and are vulnerable to the phenomenon of 'content injection'. As Conover et al. (2010) show, hashtags are often used incorrectly for political reasons: "politically-motivated individuals often annotate content with hashtags whose primary audience would not likely choose to see such information ahead of time". This reduces the efficiency of this measure and results in bias if content injection is more frequent among one side of the political spectrum. Similarly, conversation analysis is sensitive to two common situations: the use of 'retweets' for ironic purposes, and '@-replies' whose purpose is to criticize or debate with another user. As a result, it is hard to characterize the emerging communities, and whether this divide overlaps with the ideological composition of the electorate, or even if it is stable over time.

In conclusion, a critical reading of the literature suggest the need to develop new, network-based measures of political orientation. It is also necessary to improve the existing statistical methods that have been applied. Pennacchiotti and Popescu (2011) and Al Zamal et al. (2012) focus only on classifying users, but most Political Science applications require a continuous measure of ideology, for it is considered a latent variable that is scaled on a single dimension. In order to draw correct inferences, it is also important to indicate the uncertainty of the estimates. King et al. (2011), for example, do not provide standard errors for their measures of members of Congress' ideology.

Without these, it is not possible to make inferences about their rank-ordering. Similarly, these authors do not explore the possibility of placing ordinary citizens and legislators on a common scale. The main contribution of this paper is thus to implement a method to provide reliable and valid estimates (and standard errors) of Twitter users' ideology on a continuous scale.

### 3.2 Implementing a Bayesian Spatial Following Model of Ideology

The purpose of this paper is to demonstrate that valid ideal point estimates of individual Twitter users *and* political actors with a Twitter account can be derived from the structure of the 'following' links across these two sets of users. In order to do so, I develop a Bayesian spatial model of Twitter users' following behavior. Ideology is defined as a position in a latent ideological space (Poole and Rosenthal, 1997, 2007), and individual estimates are derived on the basis of the observed 'following' decisions, under the assumption that Twitter users are instrumentally rational.

The key assumption of this model is that Twitter users prefer to follow politicians whose position on the latent ideological dimension are similar to theirs. There exists broad theoretical and empirical support for this notion. Firstly, the vast body of research about homophily in personal interactions can easily be extended to online social networks such as Twitter. As McPherson et al. (2001) theorize, individuals tend to be embedded in homogenous networks with regard to many sociodemographic and behavioral traits, because of a dual process of creation and dissolution of social ties due to shared geographical, organizational, and symbolic spaces. A very small individual preference for interactions with similar people is required for segregated communities to emerge at the aggregate level (Schelling, 1978). Different studies have provided solid evidence of the existence of these patterns in Twitter (Gayo-Avello, 2010; Wu et al., 2011; Conover et al., 2012).

However, Twitter is not only an online social network – it is also a news media (Kwak et al., 2010). From this perspective, we can also rely on the existing literature on the selective exposure theory (Lazarsfeld et al., 1944; Bryant and Miron, 2004) to argue that Twitter users exhibit a preference for opinion-reinforcing political information and that they systematically avoid opinion challenges. Given the fast-paced nature of this social network and individuals' finite ability to process incoming information (Oken Hodas and Lerman, 2012), we should expect Twitter users to maximize the value of their online experience by choosing to follow political actors who can provide information that can be of higher value to them (Chan and Suen, 2008).

The theoretical model I employ relies on similar assumptions as the Euclidean spatial voting model (Enelow and Hinich, 1984). Suppose that each Twitter user $i \in \{1, \ldots, n\}$

is presented with a choice between following or not following another target user $j \in \{1, \ldots, m\}$, where $j$ is a political actor who has a Twitter account[4]. Let $y_{ij} = 1$ if user $i$ decides to follow user $j$, and $y_{ij} = 0$ otherwise. For the reasons explained above, I expect this decision to be a function of the squared Euclidean distance in the latent ideological dimension[5] between user $i$ and $j$: $-\gamma||\theta_i - \phi_j||^2$, where $\theta_i \in \mathbb{R}$ is the ideal point of Twitter user $i$, $\phi_j \in \mathbb{R}$ is the ideal point of Twitter user $j$, and $\gamma$ is a normalizing constant.

To this basic setup, I add two extra parameters, $\alpha_j$ and $\beta_i$. The former measures the popularity or "indegree" of user $j$. This parameter accounts for the fact that some political accounts are more likely to be followed, because of the role of the politicians behind it (for example, we would expect the probability of following `@BarackObama` to be higher than the equivalent for a member of Congress, all else equal) or other reasons (politicians who 'tweet' more often are more likely to be highly visible and therefore also to have more followers, all else equal). The latter measures the level of political interest or "outdegree" of users $i$. Similarly, this parameter accounts for the differences in the number of political accounts each user $i$ decides to follow, which could be due to the overall number of Twitter users he follows, or how interested in politics he/she is.

The probability that user $i$ follows a political account $j$ is thus formulated as a logit model:

$$P(y_{ij} = 1|\alpha_j, \beta_i, \gamma, \theta_i, \phi_j) = \text{logit}^{-1}\left(\alpha_j + \beta_i - \gamma||\theta_i - \phi_j||^2\right) \quad (1)$$

To illustrate the intuition behind this statistical model, Figure 1 shows the predicted probability[6] that a user $i$ follows Barack Obama or Mitt Romney, at different values of $\theta_i$, and holding all other parameters at their means. Liberal Twitter users are more likely to follow Barack Obama, and this probability is maximized when their ideology equals the estimated ideology for Barack Obama ($\theta_i = \phi_{j1}$)[7]. The same logic applies to the estimated probability of following Mitt Romney, but in this case the predicted probability when $\theta_i = \phi_{j2}$ is lower because the popularity parameter for Romney's Twitter account

[4] If we considered not only politicians, but the entire Twitter network, then $n = m$. In that case, the model would still yield valid estimates, but the estimation would be computationally intractable and inefficient and, as I argue below, the resulting latent dimension might not be ideology. In this paper I show that it is possible to obtain valid ideal point estimates choosing a small $m$ whose characteristics make 'following' decisions informative about the ideology of users $i$ and $j$.

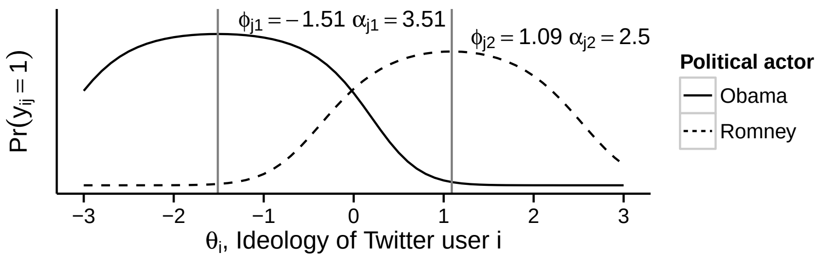
[5] I assume that ideology is unidimensional, which is a fairly standard assumption in the literature (e.g., see Poole and Rosenthal, 1997, 2007). However, the model I estimate could be easily generalized to multiple dimensions.

[6] See Sections 4 and 5 for details on the Twitter data that was used to estimate these parameters.

[7] Note that, unlike in the standard item-response theory models, the probability of a positive outcome is not monotonically increasing or decreasing in ideology. On the contrary, it is decreasing as the distance between users $i$ and $j$ increases. Continuing with the example, this model is consistent with the intuition that extremely liberal individuals are less likely to follow Barack Obama because they do not view him as "liberal enough".

is smaller ($\alpha_{j1} = 3.51$ vs $\alpha_{j2} = 2.5$). To understand the fit of the model, note that of the 301,537 Twitter users included in the US sample, 46% follow Barack Obama, 20% follow Mitt Romney, and 17% follow both, which roughly matches the areas under the curves in this Figure.

Figure 1: Estimated probability that a given Twitter user $i$ follows Barack Obama ($j_1$) or Mitt Romney ($j_2$), as a function of the user's ideal point




Estimation and inference for this type of model is not trivial. Maximum-likelihood estimation methods are usually intractable given the large number of parameters involved. For this reason, I implement a Bayesian method, where the posterior density of the five sets of parameters is explored via Markov Chain Monte Carlo methods and, more specifically, a Hamiltonian Monte Carlo algorithm. A detailed explanation of the model and the estimation algorithm is provided in section B.1 of the Appendix.

This approach presents three important advantages. Firstly, it allows the researcher to incorporate prior information into the estimation, based on observed data or relevant new information, but it is also possible to choose priors that reflect complete ignorance. Secondly, it provides proper estimates of the stochastic error at a low computational cost. And third, samples from the posterior distribution of the ideology parameters can be easily combined with simulated values of covariates to propagate uncertainty and obtain more accurate standard errors.

An important challenge regarding the implementation of method I propose is the choice of $m$, this is, those Twitter users with such "discriminatory" predictive power that the decision to follow them or not can provide information about an individual's ideology. Following Conover et al. (2010), we could analyze the entire networks of friendships in Twitter and let the different clusters emerge naturally, this is, without pre-imposing any structure or any reference point. However, this decision can violate the independence assumptions, as "homophilic" networks emerge based not only on political traits, but also as a result of similarities in other dimensions. Instead, the approach I suggest is to select a limited number of target users that includes politicians, think tanks, news outlets with a clear ideological profile, etc. Considering only those accounts that can be more informative about individuals' ideal points will ensure that the estimation is efficient, and that the latent dimension in which we are locating the ideal points is political ideology[8].

# 4 Data

The estimation method I propose in this paper can be applied to any country where a high number of citizens are discussing politics on Twitter[9]. However, in order to test the validity of the estimated parameters, I will focus on four countries where high-quality ideology measures are available for at least a subset of all Twitter users: the US, the UK, Spain, Italy, and the Netherlands. Furthermore, the increasing complexity of the party system in each of these countries will show how the method performs where more than two parties are present on Twitter.

For each of these countries, I identified a list of political representatives in national-level institutions, parties, and individuals with a highly political profile who are active on Twitter[10]. This represents a total of $m = 548$ target users in the US, $m = 244$ in the UK, $m = 298$ in Spain, $m = 215$ in Italy, and $m = 118$ in the Netherlands.

Next, using the Twitter REST API, I obtained the entire list of followers (as of November 4th, 2012) for all $m$ users in each country, resulting in a entire universe of Twitter users following at least one politician of $n = 32{,}919{,}418$ in the US, $n = 2{,}647{,}413$ in the UK, $n = 1{,}059{,}890$ in Spain, $n = 1,119,763$ in Italy, and $n = 856{,}201$ in the Netherlands. However, an extremely high proportion of these users are either inactive, spam bots or reside in different countries[11]. To overcome this problem, I extracted the available personal attributes from each user's profile, and discarded from the sample those who 1) have sent less than 100 tweets, 2) have not sent one tweet in the past six

[8] In the examples I show in this paper, accurate ideal point estimates can be obtained once $m > 200$ if the sample of political Twitter accounts in $m$ is informative about ideology. The resulting ideal points remain essentially constant past that sample size, with a marginal increase in accuracy, particularly at the extremes of the ideological dimension.

[9] Estimating ideal points using data from different countries simultaneously is more complex, given the high intra-country locality effect (Gonzalez et al., 2011).

[10] These lists combine information from different sources. In the US, I have used the NY Times Congress API, the Sunlight Labs Congress API and the GovTwit directory. In the UK, I have used the Twitter lists compiled by Tweetminster. In Spain, I have used the Spanish Congress Widget developed by Antonio Gutierrez-Rubi, and the website politweets.es. In Italy, I used a list of Twitter users collected by Cristian Vaccari and Augusto Valeriani, to whom I express my gratitude. In the Netherlands, I have used the data set of politiekentwitter.nl. I considered only political Twitter users with more than 5,000 (US) or 2,000 (UK, Spain, Italy, Netherlands) followers. A complete list of the criteria I followed when collecting these lists is available upon request.

[11] For example, in my analysis I found that only around 56% of Barack Obama's 23 million followers as of December 2012 are located in the United States.

months, 3) have less than 25 followers, 4) are located outside the borders of the country of interest, and 5) follow less than three political Twitter accounts. The final sample size is $n$ = 473,640 users in the US[12], $n$ = 135,015 in the UK, $n$ = 123,846 in Spain, $n = 150,143$, and $n$ = 96,624 in the Netherlands.

Of course, this is a highly self-selected sample. Twitter users are not a representative sample of the population: they tend to be younger and to have a higher income level than the average citizen, and their educational background and racial composition is different than of the entire country (Mislove et al., 2011; Parmelee and Bichard, 2011). In the context of this paper, the inferences I make based on my sample won't even be valid for the entire universe of Twitter users, since I am only selecting those who follow a certain number of political accounts. However, this should not affect the inference of *politicians'* ideal points, because these users can indeed be considered as "authoritative" when it comes to politics. Precisely because they are more likely to be knowledgeable and interested in politics than the average citizen, examining their online behavior can be highly informative about policy positions. This procedure is somehow analogous to an expert survey with many respondents where each of them provides a small amount of information that, when aggregated, results in highly accurate policy estimates.

The sample selection process requires identifying the specific country from where each user tweets. This information was extracted from the "location" field in the user profile, and structured using the Yahoo geolocation API[13]. Interestingly, the geographical distribution of Twitter users in the US resembles that of the general population. Figure 2 shows the percentage of Twitter accounts located in each state. The correlation with the distribution of population according to the 2012 U.S. census estimates is $\rho = 0.975$.

In the case of the US, I also imputed the gender of all Twitter users in my sample. This variable was estimated using the first name of each Twitter user (when it was provided), and applying a probabilistic model that relies on the list of most common first names by gender in anonymized databases, available in the RandomNames `R` package (Betebenner, 2012). The table below provides descriptive statistics for this variable. A more detailed explanation of the imputation method I use can be found in Appendix C.

The substantive application I present in section 6 analyzes the structure and con-

[12]The actual sample size in the US is 301,537 users, which is the number of Twitter accounts who tweeted at least three times mentioning 'Obama' or 'Romney' and can thus be included in the analysis in section 6.

[13] Note that location information is only available for the subset of users who decide to provide it on their profiles, which is around 70% (Hale et al., 2012). Even if they decide to provide such information, it is highly unstructured and sometimes states only the country but not the region or city, or it refers to imaginary places (Hecht et al., 2011). However, in combination with the information about the time zone in each user's profile, it is sufficient to identify the country of residence in 90% of the cases. This proportion is lower when we consider more specific geographical levels, such as state in the US (71%), country in the UK (67%), or province in Spain (61%) and the Netherlands (64%).

Figure 2: Distribution of Twitter Users in the Continental US, by State

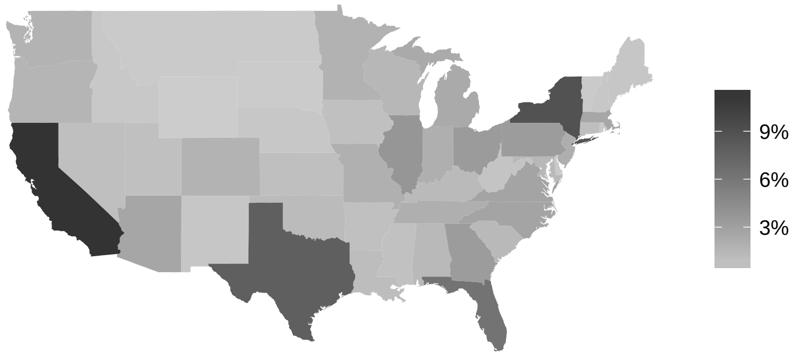



Table 1: Distribution of Twitter Users, by Gender

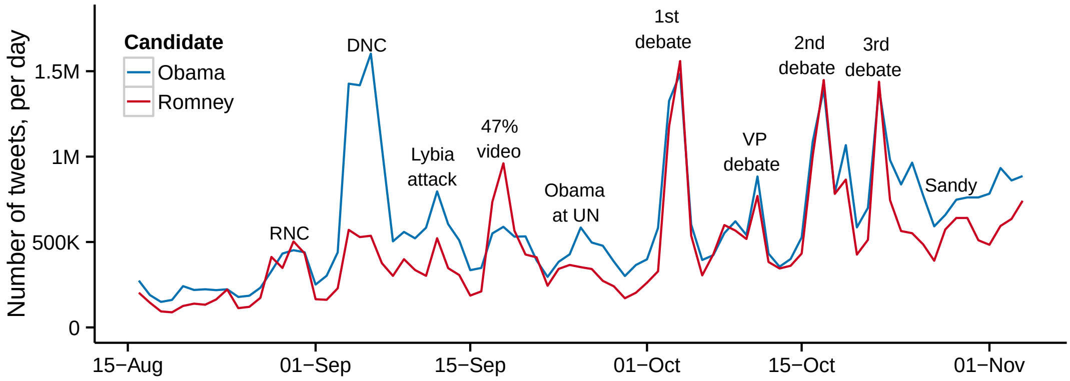
| | Male | Female | Unknown | Total |
|---|---|---|---|---|
| Number of Twitter Users | 151,497<br>(50.2%) | 105,811<br>(35.1%) | 44,229<br>(14.7%) | 301,537<br>(100%) |

tent of the political conversation in Twitter during the 2012 US Presidential election campaign. The data I use consists of all public tweets that mentioned 'Obama' or 'Romney' from August 15th to November 4th. These messages total over 75 million tweets (15 million of which were published by the 301,537 users in my sample – an average of 50 tweets per person) and were collected using the Twitter Streaming API and the `streamR` package for `R` (Barberá, 2013). Figure 3 plots the evolution in the daily number of tweets sent over the course of the electoral campaign. As expected, this metric peaks during significant political events, such as the party conventions or the three presidential debates.

Finally, in order to improve the validation of the ideal points I estimate in the US, a sample of Twiter users from the state of Ohio were matched with their voting registration records. This choice was based on it being considered one of the prime 'battleground' states, and also for reasons of data availability. (The entire voter file is available online at the Ohio Secretary of State website.) A total of 2,462 Twitter users were matched with their individual records, based on perfect matches of their reported full name and county of residence, which represents over 12% of the 20,153 Twitter users from Ohio in the full sample[14]. Again, this subset cannot be considered representative for any

[14]This proportion is comparatively not too small, particularly if we consider that most Twitter users

Figure 3: Evolution of mentions to Obama and Romney in Twitter




population of interest, and will only be used to examine whether the resulting ideology estimates are good predictors of the party under which they are registered.

# 5 Results

In this section I provide a summary of the ideology estimates for the five countries included in my study, and the oversampled set of Twitter users in the state of Ohio. To validate the method, I will use different sources of external information to assess whether this procedure is able to correctly classify and scale Twitter users on the left or right side of the ideological dimension.

## 5.1 United States

The first set of results I focus on are those from the United States. Figure 4 compares $\phi_j$, the ideal point estimates, of 231 members of the 112th U.S. Congress[15] based on their Twitter network of followers ($y$ axis) with their DW-NOMINATE scores[16], based on their roll-call voting records (Poole and Rosenthal, 2007), on the $x$ axis. Each letter correspond to a different member of congress, where $D$ stands for democrats and $R$

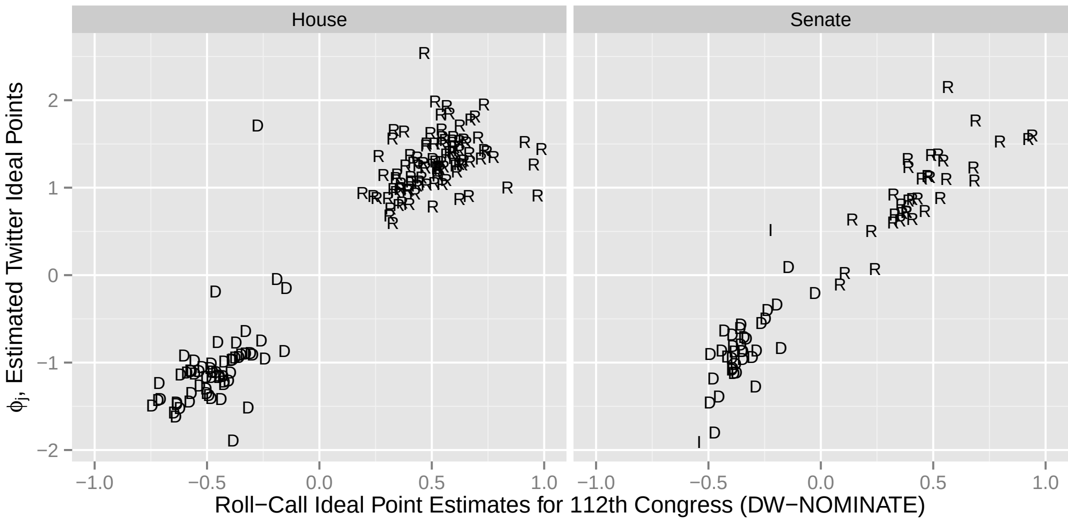
do not provide their real full name. The other existing study that performed a similar analysis (Bond et al., 2012) was only able to successfully match 1 in 3 Facebook users to voter records.

[15]Only members of congress whose Twitter acounts have more than 5,000 followers are included in the sample.

[16]Source: voteview.com

stands for republicans, and the two panels split the sample according to the chamber of Congress to which they were elected.

Figure 4: Comparing Ideal Points Based on Roll-Call Records and Based on Twitter Network of Followers in the U.S. Congress




As we can see, the estimated ideal points are clustered in two different groups, which align almost perfectly with party membership. The correlation between Twitter- and roll-call-based ideal points is $\rho = .941$ in the House and $\rho = .954$ in the Senate. Furthermore, if we examine the most extreme legislators, we find that their Twitter-based estimates also position them among those with the highest and lowest values on the ideological scale. Within-party correlations are also relatively high: $\rho = .546$ for republicans, $\rho = .610$ for democrats[17].

Ideal points for a wider set of political actors are plotted in Figure 5. (See Appendix A for an expanded version of this plot.) As it was the case with members of congress, the resulting estimates show a clear division across members of each party, and the ideal points for all non-partisan actors have face validity. Note also their positions within each cluster are also what we would expect based on anecdotal evidence. For example, Schwarzenegger and Jon Huntsman appear among the most liberal Twitter accounts in the Republican Party, while Rush Limbaugh and Glenn Beck are in the group of most

[17]These results are essentially identical if compare my Twitter-based estimates with other ideal points based on voting records, such as those estimated by Jackman (2012) using an item-response theory scaling method (Clinton et al., 2004): $\rho = .966$ in the House, $\rho = .950$ in the Senate, $\rho = .538$ for republicans, and $\rho = .749$ for democrats.

conservative nonpartisan Twitter accounts. On the left side of the ideological dimension we find Keith Olbermann, Michael Moore, Rachel Maddow or the HRC as the most liberal Twitter acounts.

Figure 5: Estimated Ideal Points for Key Political Actors

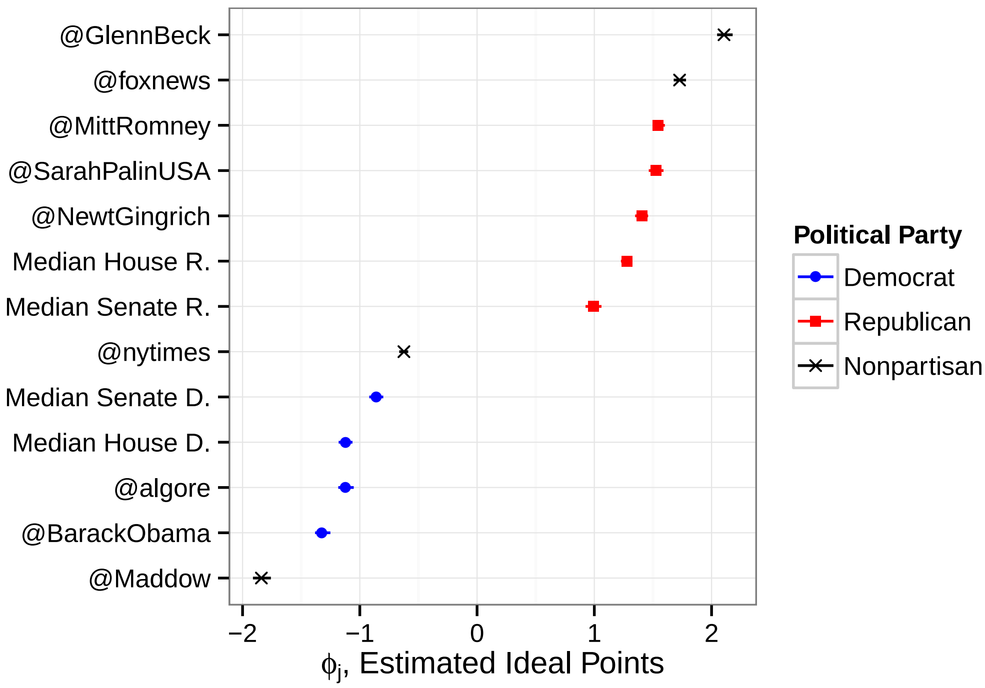



Figure 6 compares the distribution of ideological ideal points for the two types of Twitter users in the sample – political actors and ordinary citizens. The pattern that emerges is an almost exact replication of the standard result in the literature (see for example Figure 5 in Bafumi and Herron, 2010). Both distributions are bimodal, liberal citizens represent a majority of the population, and political actors are more polarized than mass voters.

Now I turn to assess whether the estimated ideal points for ordinary citizens are also valid. In Figure 7 I plot the distribution of the ideology estimates for different groups of individuals. Here I exploit the fact that many Twitter users define themselves politically in their profiles[18]. Using this information, I extracted five subsets of accounts, according

[18]Three different examples of profiles that can be used to identify ideology would be: "Student of History and Politics. Christian and **Conservative**. Southern and Saved [...]", "Idaho native. Oregon **democrat**. Fly Fisherwoman. Political Nerd [...]", and "reader, citizen patriot, concerned, recently

Figure 6: Distribution of Political Actors and Ordinary Twitter Users' Ideal Points

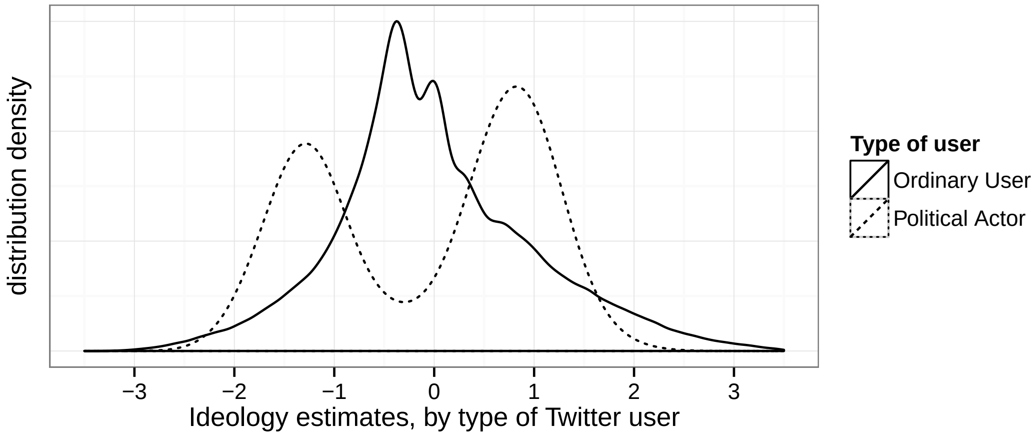
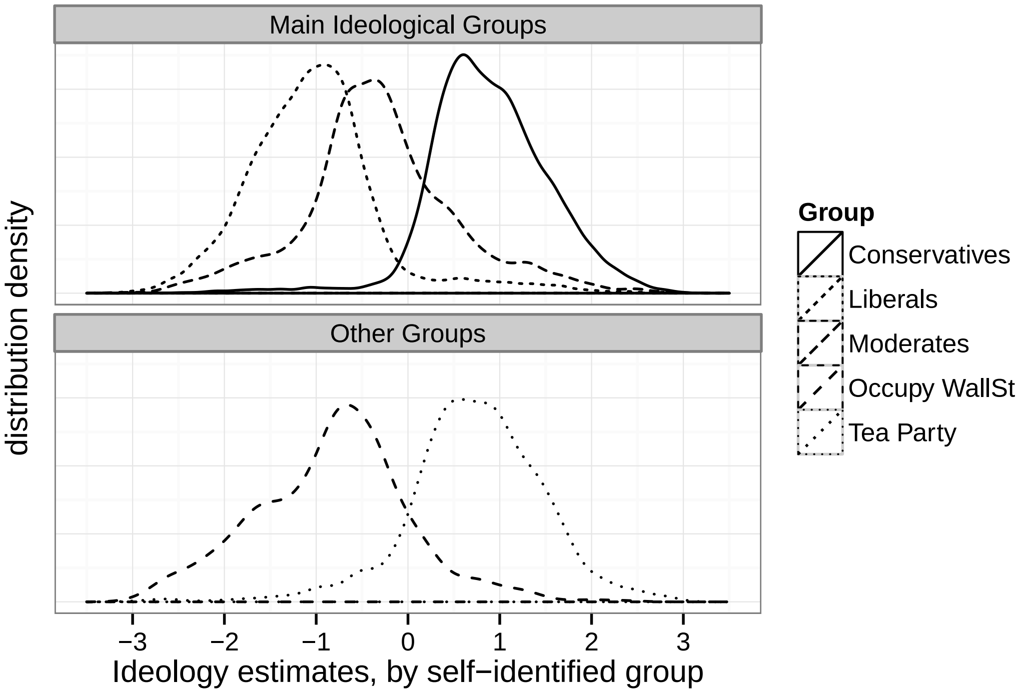



Figure 7: Distribution of Ideal Point Estimates, by Self-Identified Political Group



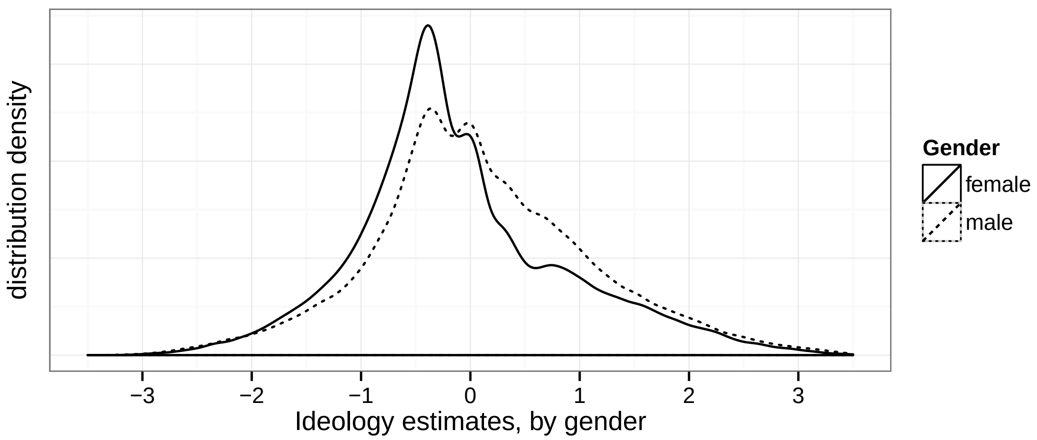

changed registration to **Independent**".

to whether they mention specific keywords on their profiles: conservatives ("conservative", "GOP", "republican"), independents ("independent", "moderate"), liberals ("liberal", "progressive", "democrat"), Tea Party members ("tea party", "constitution"), and Occupy Wall Street members ("occupy", "ows").

The distribution of ideal points for each group closely resembles what one would expect: conservatives are located to the right of independents, and independents are located to the right of liberals, with some overlap. Similarly, supporters of the Occupy Wall Street movement tend to be more liberal than Tea Party members, although they do not appear to be different than the median conservative or liberal Twitter user, which suggests that this scale could be capturing the intensity of party support rather than strictly ideology. While classification is not completely perfect, this plot shows that the estimation is able to distinguish and scale Twitter accounts according to the policy position of who updates them.

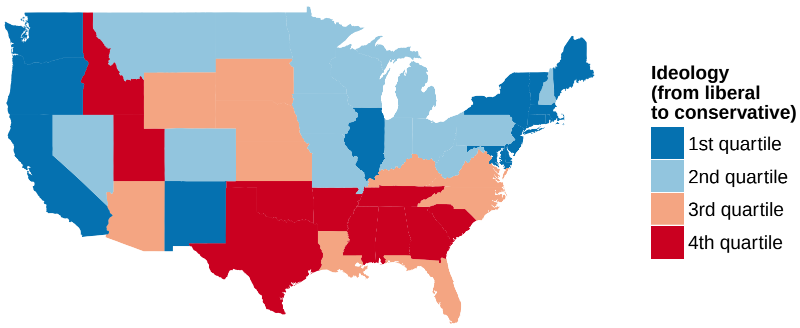
Figure 8 compares the distribution of ideal points by gender in my sample of Twitter users, showing that women tend to be slightly more liberal than men. This result is consistent with what can be found in political surveys. For example, the average ideological placement (in a scale from 1, extremely liberal, to 7, extremely conservative) in the 2008 American National Election Survey was 4.05 for women and 4.24 for men.

Figure 8: Distribution of Ideal Point Estimates, by Gender




As an additional validation test, in Figure 9 I show the ideology of the median Twitter user in each state, where the shade of the color indicates the quartile of the distribution. Despite Twitter users being a highly self-selected sample of the population, this figure nonetheless presents a close resemblance to ideology estimates based on surveys. As I show in Figure 10, Twitter-based ideal point estimates by state are highly correlated ($\rho = .880$) with the proportion of citizens in each state that hold liberal opinions across different issues, as estimated by Lax and Phillips (2012) combining surveys and socioeconomic indicators[19]. Ideology by state is also a good predictor of the proportion of the two-party vote that went for Obama in 2012, as shown in the right panel of Figure 10, but the correlation coefficient is smaller ($\rho = -.792$), which suggests that the meaning of the emerging dimension in my estimation is closer to ideology than to partisanship.

Figure 9: Ideal Point of the Average Twitter User in the Continental US, by State




Finally, in order to show that the relationship between estimated ideology and vote also holds at the individual level, in Figure 11 I show the distribution of ideal points for a sample of Twitter users who "self-reported" their vote for Obama ($N = 2539$) or Romney ($N = 1601$) during Election day. To construct this dataset, I captured all tweets mentioning the word "vote" and either "obama" or "romney" and then applied a simple classification scheme to select only tweets where it was openly stated that the user had casted a vote for one of the two candidates[20]. As expected, ideology is highly associated with vote orientation.

[19] This correlation is slightly weaker ($\rho = .832$) if instead we use Gallup's 2012 "State of the States" estimates of the conservative advantage by state – measured as the percentage of conservative citizens minus the percentage of liberal citizens in each state, which were based based on a survey conducted in January 2012 with a random sample of 211,972 adults

[20] For example, in the case of Obama I selected those tweets that mentioned "I just voted for (president, pres, Barack) Obama", "I am voting for Obama", "my vote goes to obama", "proud to vote Obama", and different variations of this pattern, while excluding those that mentioned "didn't vote for Obama", "never vote for Obama", etc. 25 sample tweets and their imputed votes are provided in Appendix D

Figure 10: Twitter-Based Ideal Points, by State

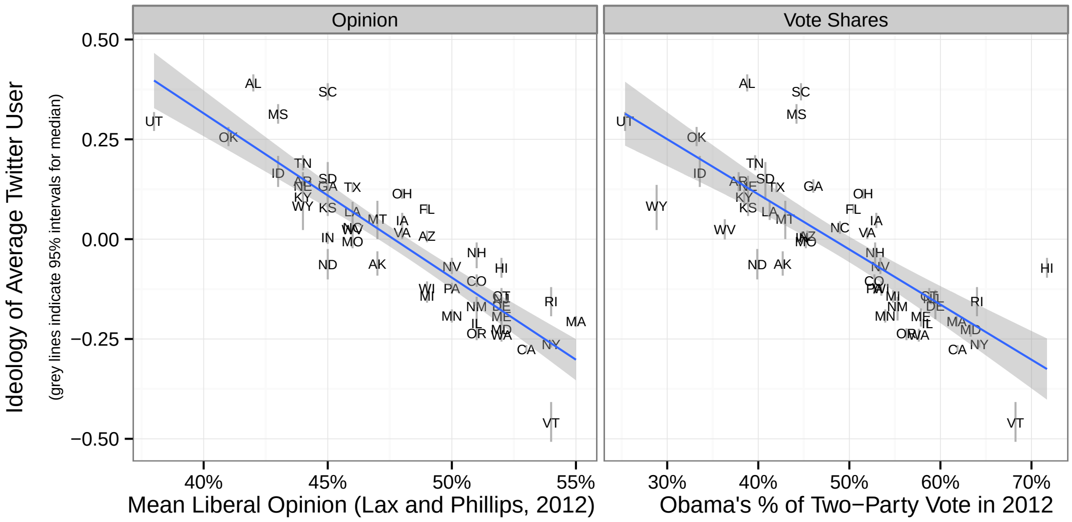
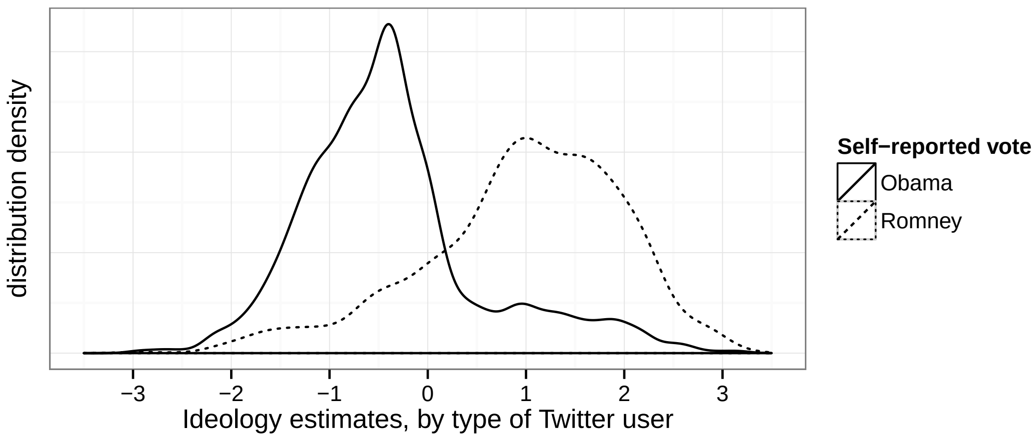



Figure 11: Distribution of Users' Ideal Points, by Self-Reported Votes

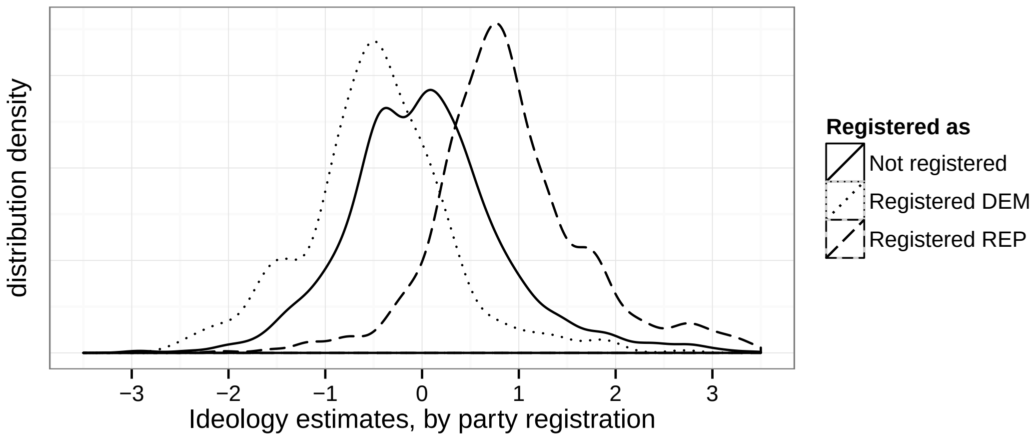



## 5.2 Ohio

To further validate the ideal point estimates I introduced in the previous section, now I turn to examine the results from the sample of 2,360 Twitter users from Ohio whose names were matched with the voter file.

In Figure 12, I plot the distribution of the ideology estimates across three different groups of voters, based in their most recent party registration in the period 2008–2012: not registered, registered as democratic, registered as republican. (Note that this variable is available because being registered with a party is a necessary condition in order to vote in the primary elections in Ohio.)

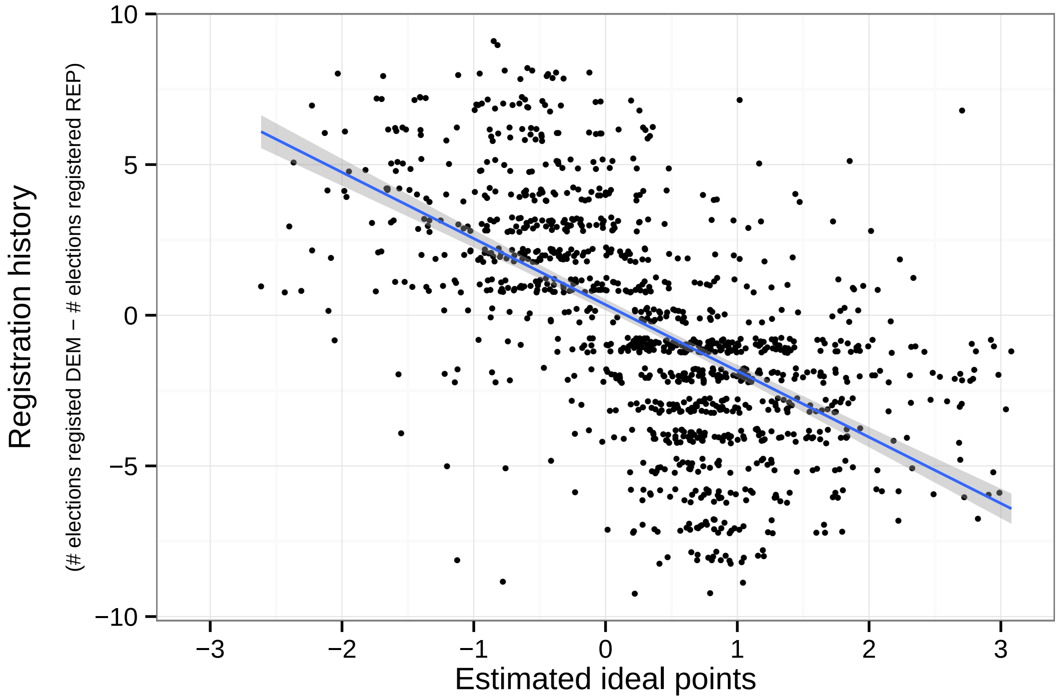
Figure 12: Distribution of Ideal Point Estimates, by Party of Registration




The evidence I present provides additional support for the external validity of my method: the average registered republican is located to the right of the average democrat, and this difference is large and statistically significant. If we consider the distribution of ideal points across these two groups, we see that most individuals are correctly classified to the left or right based on their party registration, with some overlap, particularly in the case of liberal voters.

Since each voter's registration history is available since 2000, we can examine if, as expected, the most conservative (liberal) voters in Ohio tend to consistently register as Republican (Democrat) in the primary elections. Figure 13 shows that this is indeed the case. The vertical axis shows the number of primary elections each voter was registered as Democrat minus the number of primary elections registered as Republican, with some jittering to facilitate the interpretation of this result. This plot shows that ideology is a very powerful predictor of each voter's registration history (the $R^2$ of a regression of registration history on ideology is 0.24).

Figure 13: Ideal Point Estimates and Party Registration History



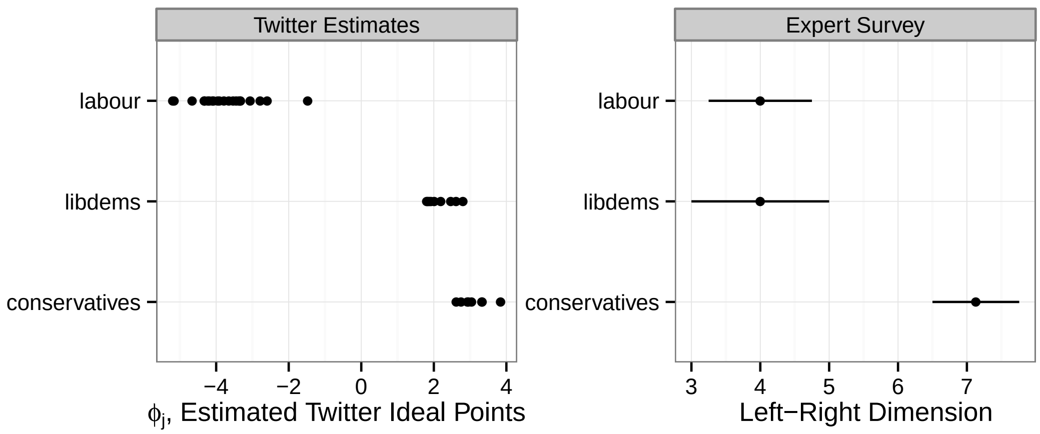

## 5.3 UK, Spain, Italy, and Netherlands

The next four figures refer to the other countries I included in the study, and display the ideal point estimates for all Twitter users with more than 2,000 followers who belong to a political party, grouped by party. In each figure, I also show the ideological locations of each party on the left-right scale, estimated using expert surveys (Bakker et al., 2012). (Given the volatility of the Italian party system, expert surveys from previous elections are not directly comparable, so I only show how parties are scaled within each electoral coalition.)

As in case of the US, these results show that my estimation method is able to classify accounts according to the party to which they belong. With few exceptions, all Twitter accounts from the same party are clustered together. Furthermore, the order of the parties seems to be similar to that reported by different studies based on expert surveys for the "left-right" dimension.

The results show the lower degree of accuracy in the UK: both the Labour Party and the Liberal-Democrats are located to the left of the Conservative Party on average, but the latter is classified as right-wing, almost overlapping with the conservatives, perhaps indicating their status of coalition partners. In Spain, the Socialist Party (PSOE) is located to the left of the Conservative party (PP), with the recently-created Center party (UPyD) between them. However, the Communist party (IU) is misclassified: it is closer to the center than the PSOE. In Italy, the ideal points are not only clustered within party, but also within each electoral coalition, with all of them correctly classified on the left ("Bene Comune"), center ("Con Monti Per l'Italia"), and right ("Coalizione di Centro-destra"). Finally, in Netherlands all parties are accurately scaled, with the exception of the Socialist party (SP) and the right-wing Party for Freedom (PVV), which appear slightly more centrist in my Twitter estimates than in the expert surveys.

Figure 14: Ideological Location of Parties in the United Kingdom




It is important to note that the results for different countries are not directly comparable, as the estimation was performed independently, and the resulting dimension does not have a homogenous scale across countries. However, it would be possible to use Twitter users who follow accounts from more than two countries as "bridges", to use the term applied to legislators who serve in different chambers in the literature on ideal point estimation using roll-call votes. It is also necessary to explore why this method performs differently across countries. One possibility is that the emerging scale collapses different dimensions. For example, in Spain the outliers to the left of the PSOE and IU are all members of the catalan branches of each party.

Figure 15: Ideological Location of Parties in Spain

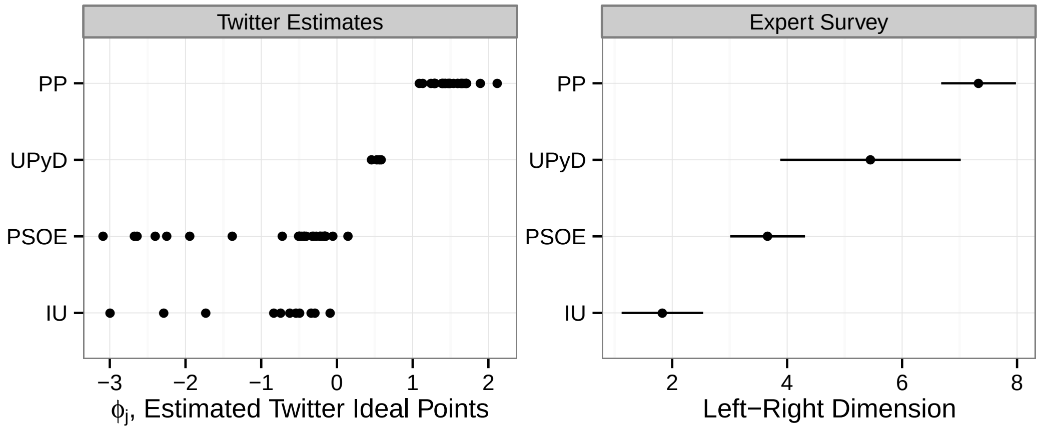



Figure 16: Ideological Location of Parties in Italy

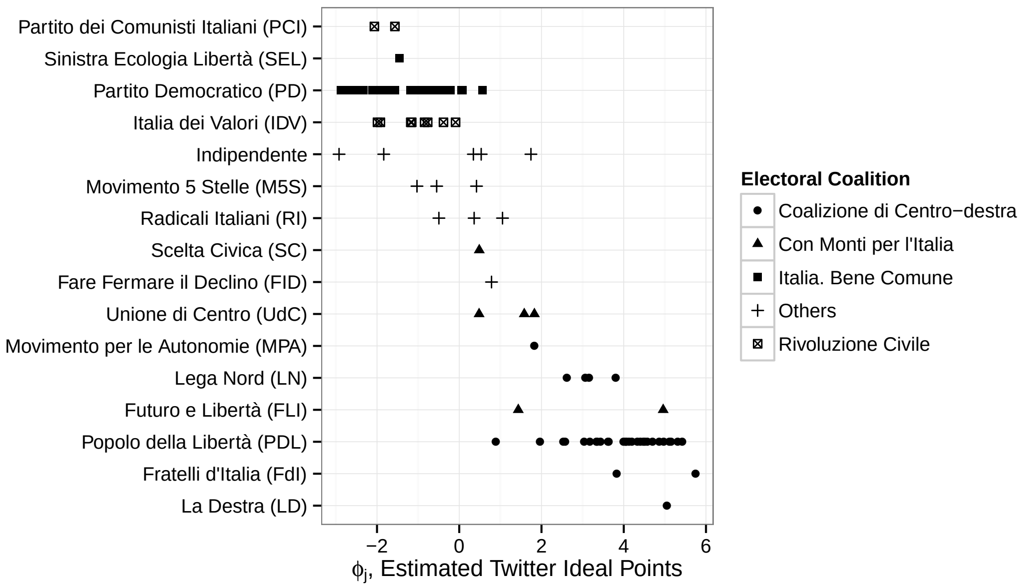
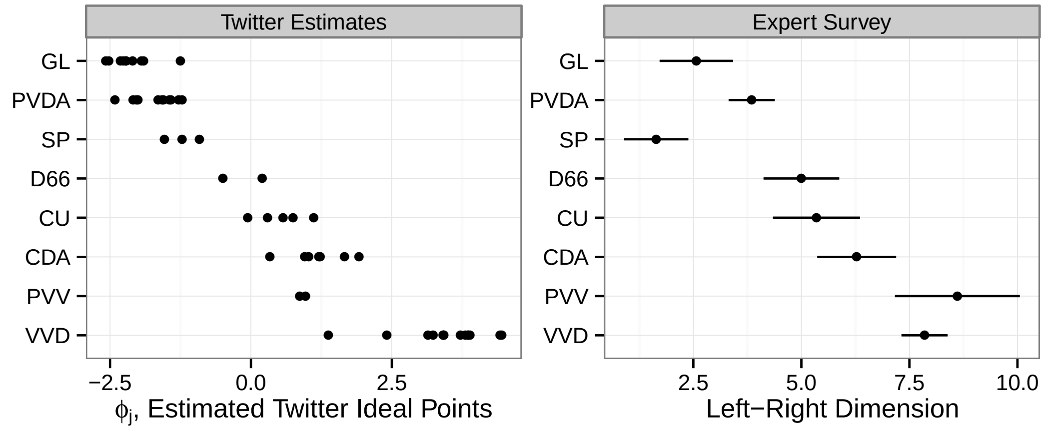


Figure 17: Ideological Location of Parties in the Netherlands




## 6 Social Media and Political Polarization: Echo Chamber or Pluralist Debate?

A recurring theme in the literature on internet and politics is how the increasing amount and heterogeneity of political information citizens have access to affects their political views (Farrell, 2012). Several authors argue that, as a result of this transformation, individuals are being increasingly exposed to only information that reinforces their existing views, thus avoiding challenging opinions (Sunstein, 2001; Garrett, 2009). This generates a so-called echo-chamber environment (Adamic and Glance, 2005) that fosters social extremism and political polarization. Given that a substantial proportion of citizens now rely mostly on the internet to gather political information[21], the policy implications of this issue are obvious: how individuals gather political information affects the quality of political representation, the policy-making process, and the stability of the democratic system (Mutz, 2002).

In the specific context of Twitter, this issue is also relevant because the extent to which users' behavior on this platform is polarized remains an open debate. On one hand, Conover et al. (2010, 2011, 2012) find high levels of clustering along party lines: "the network of political retweets exhibits a highly segregated partisan structure, with extremely limited connectivity between left- and right-leaning users" (Conover et al., 2011, p.89). Yardi and Boyd (2010) and Gruzd (2012) qualify this conclusion. While

[21] According to a survey conducted by the Pew Research Center in 2011, 31% of U.S. adults rely most on internet for political information.

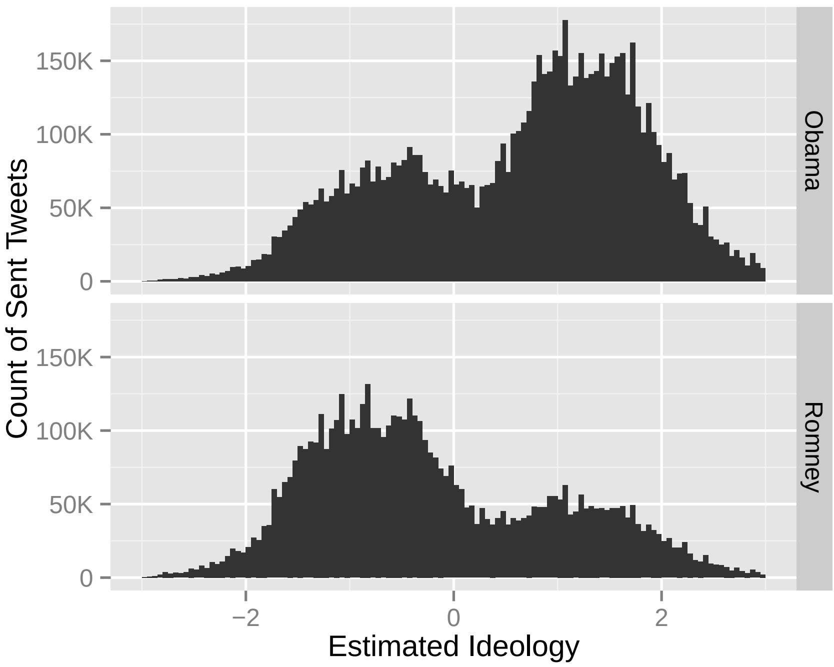
they also find that Twitter users tend to cluster around shared political views, their results show that open cross-ideological exchanges are very frequent, and individuals are exposed to broader viewpoints. Similarly, when examining other types of behavior on Twitter, Conover et al. (2011) also find that user-to-user interactions via "@-replies" between ideologically-opposed individuals take place at a higher rate compared to the network of retweets.

One possible reason for the variability in the results of these studies is that the source of information to estimate ideology is also then used to measure polarization. For example, Conover et al. (2010, 2011, 2012) apply network clustering algorithms to classify users by their tweeting behavior, and then see to what extent users that belong to the same group interact with each other. The problem with this approach is that these algorithms are trained precisely to maximize the distance between individuals across different communities, and are thus biased towards finding polarized networks.

As a substantive application of the estimation method I propose in this paper, I replicate the analysis of this set of studies. In contrast with their approach, I use two completely different sources of data to measure ideology and users' behavior of U.S. Twitter accounts. As it was explained in section 3, my ideal point estimates are based on the 'following' connections established between users. In parallel, I have captured all tweets mentioning any of the two presidential candidates ("Obama" or "Romney") from August 15th, 2012, to November 6th, 2012, and selected those (nearly 20%) that were sent by Twitter users in my sample of over 300,000 accounts. This dataset will allow me to measure to what extent political conversations on Twitter are polarized along ideological lines.

I show results of my analysis in Figures 18, 19 and 20. The first figure plots the number of tweets published in the interval of study by users along the latent ideological dimension (in bins of width 0.05). The top panel refers to tweets that mention "Obama", while the bottom panel refers to tweets mentioning "Romney". The pattern that emerges yields two results. First, I find that the conversation in Twitter is dominated by individuals with extreme views. Despite the fact that (by construction) ideology has a unit variance distribution, we find that the distribution of the number of tweets is highly bimodal, with the modes at approximately $-1$ and $+1$ – this is, one standard deviation away from the average Twitter user. Second, I find a very distinct pattern in the tweets mentioning President Obama: conservative Twitter users sent a substantively higher proportion of tweets than their liberal counterparts. The opposite result emerges in the sample of tweets mentioning Mitt Romney: liberal Twitter users appear to monopolize the conversation, although to a lesser extent. This finding suggests that most tweets sent during this period were negative, and is also consistent with the results of the analysis by Conover et al. (2012), who also discovered that right-leaning

Figure 18: Number of Tweets Mentioning Presidential Candidates, by Ideal Point Bin




Twitter users exhibit greater levels of political activity.

Figures 19 and 20 provide evidence for the existence of an "echo-chamber" environment on Twitter. Here, I use a heat plot to visualize the structure of the two most common types of interactions in Twitter: retweets and mentions[22]. The color of each cell (of size $0.2 \times 0.2$) represents the proportion of tweets in the sample that were retweets/mentions of users with ideal point $X$ to users with ideal point $Y$[23]. Therefore, if we were to find perfect polarization (this is, users interacting only with those of identical ideology), we would find a pattern that would resemble a line with slope one.

Both figures show very similar results. Cross-ideological interactions are rare, since

[22] A retweet consists on re-posting another user's content with an indication of its original author. It is used whenever the 'retweeter' wants to publicize the content of the original post, but it is not necessarily a sign of endorsement. In politics, candidates often encourage their followers to retweet their messages. A mention consists on including in a tweet the handle of another user (e.g. "@BarackObama"), so that the user that is mentioned can easily find the tweet. It is therefore an indication of a conversation between two Twitter users.

[23] For example, in the left panel of Figure 19 we can see that around 1% of all retweets mentioning Obama had an original author a Twitter user whose ideal point was in the interval between 1 and 1.2, and were retweeted by Twitter users in the same interval.

Figure 19: Political Polarization in Retweets Mentioning Presidential Candidates

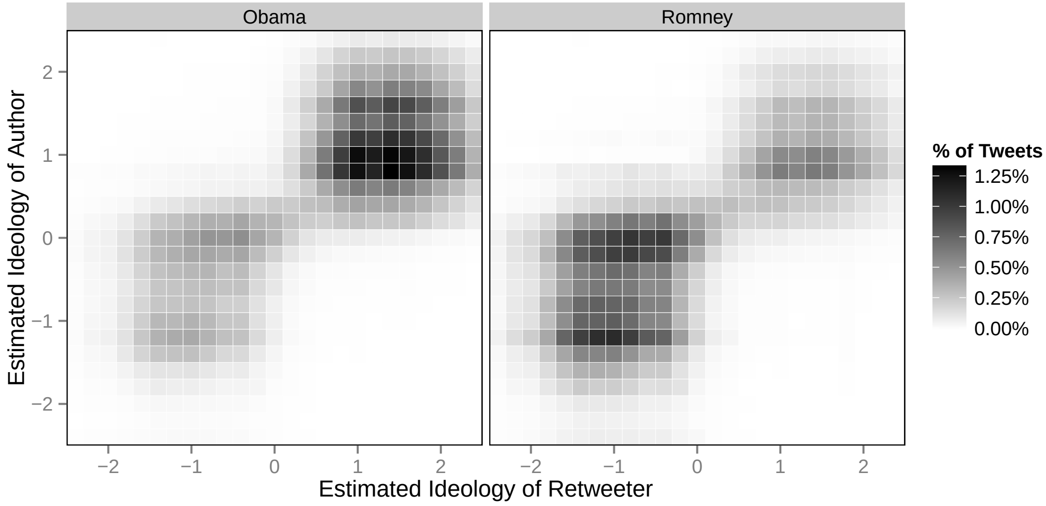
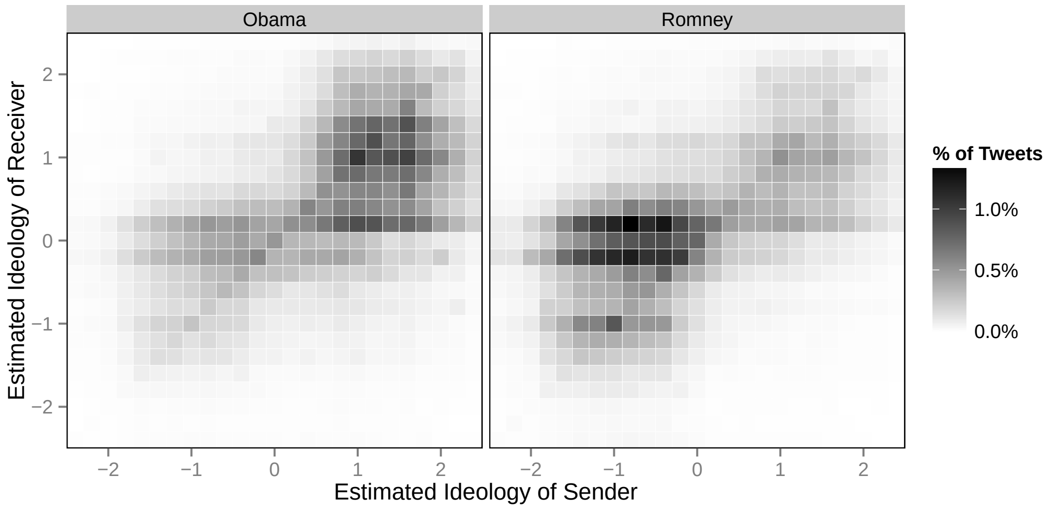



Figure 20: Ideological Polarization in Conversations Mentioning Presidential Candidates




most interactions tend to take place among Twitter users with similar ideological positions. This behavior is particularly predominant among right-leaning Twitter users, as indicated by the darker colors. While liberal users also present this pattern, they tend to engage more often in conversations all along the ideological spectrum. To sum up, the picture that emerges points towards a high degree of polarization, which is driven predominantly by conservative Twitter users. The strength of their ties has important implications during the electoral campaign. As Conover et al. (2012) argue, the topology of the network of right-leaning Twitter users facilitate the rapid and broad dissemination of political information. In a context in which interactions taking place through this platforms are increasingly covered in the traditional media, the cohesiveness of this group of users has the potential to manipulate the public agenda.

# 7 Conclusions

Millions of people are writing personal messages on Twitter everyday. Many of these "tweets" are either irrelevant personal experiences, replication of existing information or simply spam. However, given the number and heterogeneity of users, some valuable data can be extracted from this source. Recently, some scholars have started to examine whether specific patterns in the stream of tweets might be able to predict consumer behavior. But the literature on the measurement of public opinion using Twitter data is still underdeveloped.

One of the main reasons is the lack of certainty about any inference that we might draw from this data. Twitter users are younger, more interested in politics and have higher incomes than the average citizen. It is therefore necessary to know more about the distribution of key socioeconomic and political factors among Twitter users in order to be able to infer valid estimates from this data.

That was the motivation behind this paper. Addressing these concerns, I have proposed a new measure of ideology in Twitter that might be used to weight estimates of public opinion in future studies. In contrast with the existing content-based measures, I have argued that a more promising approach is to study the 'following' links between ordinary Twitter users and political actors with a strong presence on this platform. I have applied this measure in four different countries: United States, United Kingdom, Spain, Italy, and the Netherlands; and in a sample of voters from Ohio. My results show that this method successfully classifies most political actors and ordinary citizens according to their political orientation, with the locations along the ideological scale resembling positions estimated using roll-call voting, party manifestos, and expert surveys.

These results highlight the unexplored potential of Twitter data to generate ideology estimates that could prove useful in political science. To illustrate this possibility, in this paper I have presented an application that relies on such estimates. Using the 2012 US presidential election campaign as case of study, I have shown that public exchanges in Twitter take place predominantly among users with similar viewpoints, and also that right-leaning users form a cluster of highly-motivated individuals, who dominate public conversations on Twitter.

This application was just an example of many intriguing research questions that could be answered using this new source of information, ranging from studies of party competition, to analyses of public opinion, media slant, collective action, and electoral behavior.

## References


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

Adams, J., S. Merrill, and B. Grofman (2005): *A unified theory of party competition: A cross-national analysis integrating spatial and behavioral factors*, Cambridge Univ Pr.

Al Zamal, F., W. Liu, and D. Ruths (2012): "Homophily and latent attribute inference: Inferring latent attributes of twitter users from neighbors," in *Proceedings of the International Conference on Weblogs and Social Media*.

Ansolabehere, S. and E. Hersh (2012): "Validation: What Big Data Reveal About Survey Misreporting and the Real Electorate," *Political Analysis*.

Asur, S. and B. Huberman (2010): "Predicting the future with social media," in *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*, IEEE, vol. 1, 492–499.

Bafumi, J., A. Gelman, D. Park, and N. Kaplan (2005): "Practical issues in implementing and understanding Bayesian ideal point estimation," *Political Analysis*, 13, 171–187.

Bafumi, J. and M. C. Herron (2010): "Leapfrog representation and extremism: A study of American voters and their members in Congress," *American Political Science Review*, 104, 519–542.

Bakker, R., C. de Vries, E. Edwards, L. Hooghe, S. Jolly, G. Marks, J. Polk, J. Rovny, M. Steenbergen, and M. A. Vachudova (2012): "Measuring party positions in Europe: The Chapel Hill expert survey trend file, 1999–2010," *Party Politics*.

Barberá, P. (2013): "streamR: Access to Twitter Streaming API via R," R package available on CRAN.

Betebenner, D. W. (2012): "randomNames: Function for creating gender and ethnicity correct random names," R package available on CRAN.

Blais, A., D. Blake, and S. Dion (1993): "Do parties make a difference? Parties and the size of government in liberal democracies," *American Journal of Political Science*, 40–62.

Bond, R., C. Fariss, J. Jones, A. Kramer, C. Marlow, J. Settle, and J. Fowler (2012): "A 61-million-person experiment in social influence and political mobilization," *Nature*, 489, 295–298.

Boutet, A., H. Kim, E. Yoneki, et al. (2012): "Whats in Your Tweets? I Know Who You Supported in the UK 2010 General Election," in *Sixth International AAAI Conference on Weblogs and Social Media*.

Bradley, A. (1997): "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern recognition*, 30, 1145–1159.

Brier, G. (1950): "Verification of forecasts expressed in terms of probability," *Monthly weather review*, 78, 1–3.

Bryant, J. and D. Miron (2004): "Theory and research in mass communication," *Journal of communication*, 54, 662–704.

Chan, J. and W. Suen (2008): "A spatial theory of news consumption and electoral competition," *Review of Economic Studies*, 75, 699–728.

Clinton, J., S. Jackman, and D. Rivers (2004): "The statistical analysis of roll call data," *American Political Science Review*, 98, 355–370.

Congosto, M., M. Fernández, and E. Moro Egido (2011): "TWITTER Y POLÍTICA: INFORMACIÓN, OPINIÓN Y PREDICCIÓN?" *Cuadernos de Comunicación Evoca.*

Conover, M., B. Gonçalves, J. Ratkiewicz, A. Flammini, and F. Menczer (2010): "Predicting the political alignment of twitter users," Unpublished manuscript.

Conover, M., J. Ratkiewicz, M. Francisco, B. Goncalves, A. Flammini, and F. Menczer (2011): "Political polarization on twitter," in *Proc. 5th Intl. Conference on Weblogs and Social Media.*

Conover, M. D., B. Gonçalves, A. Flammini, and F. Menczer (2012): "Partisan Asymmetries in Online Political Activity," *EPJ Data Science*, 1, 1–19.

Dodds, P., K. Harris, I. Kloumann, C. Bliss, and C. Danforth (2011): "Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter," *Arxiv preprint arXiv:1101.5120.*

Dow, J. K. (2001): "A comparative spatial analysis of majoritarian and proportional elections," *Electoral Studies*, 20, 109 – 125.

Downs, A. (1957): "An economic theory of political action in a democracy," *The Journal of Political Economy*, 65, 135–150.

Enelow, J. and M. Hinich (1984): *The spatial theory of voting: An introduction*, Cambridge Univ Pr.

Farrell, H. (2012): "The Internet's Consequences for Politics," *Annual Review of Political Science, forthcoming.*

Franklin, M. (2004): *Voter turnout and the dynamics of electoral competition in established democracies since 1945*, Cambridge University Press.

Garrett, R. (2009): "Politically motivated reinforcement seeking: Reframing the selective exposure debate," *Journal of Communication*, 59, 676–699.

Gayo-Avello, D. (2010): "All liaisons are dangerous when all your friends are known to us," *Arxiv preprint arXiv:1012.5913.*

——— (2012): "I Wanted to Predict Elections with Twitter and all I got was this Lousy Paper A Balanced Survey on Election Prediction using Twitter Data," *Arxiv preprint* *arXiv:1204.6441*.

GELMAN, A., J. CARLIN, H. STERN, AND D. RUBIN (2013): *Bayesian data analysis, 3rd Edition*, Chapman & Hall/CRC.

GELMAN, A., J. HILL, AND E. CORPORATION (2007): *Data analysis using regression and multilevel/hierarchical models*, Cambridge University Press.

GELMAN, A. AND D. RUBIN (1992): "Inference from iterative simulation using multiple sequences," *Statistical science*, 7, 457–472.

GOLDER, S. AND M. MACY (2011): "Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures," *Science*, 333, 1878–1881.

GONG, A. (2011): "An automated snowball census of the political web," .

GONZALEZ, R., R. CUEVAS, A. CUEVAS, AND C. GUERRERO (2011): "Where are my followers? Understanding the Locality Effect in Twitter," *Arxiv preprint arXiv:1105.3682*.

GRUZD, A. (2012): "Investigating Political Polarization on Twitter: A Canadian Perspective," *unpublished manuscript*.

HALE, S., D. GAFFNEY, AND M. GRAHAM (2012): "Where in the world are you? Geolocation and language identification in Twitter," Tech. rep., Working paper.

HECHT, B., L. HONG, B. SUH, AND E. CHI (2011): "Tweets from justin bieber's heart: the dynamics of the location field in user profiles," in *Proceedings of the 2011 annual conference on Human factors in computing systems*, ACM, 237–246.

INGLEHART, R. (1990): *Culture shift in advanced industrial society*, Princeton Univ Press.

JACKMAN, S. (2012): "Estimates of Members' Preferences, 112th U.S. House and Senate," Retrieved on August 3rd, 2012.

JESSEE, S. (2009): "Spatial voting in the 2004 presidential election," *American Political Science Review*, 103, 59–81.

KEDAR, O. (2005): "When moderate voters prefer extreme parties: Policy balancingin parliamentary elections," *American Political Science Review*, 99, 185–199.

KING, A., F. ORLANDO, AND D. SPARKS (2011): "Ideological Extremity and Primary Success: A Social Network Approach," *Paper presented at the 2011 MPSA Conference.*

——— (2012): "The Social Primary: Measuring Momentum in the 2012 Republican Primary Campaign," *Unpublished Manuscript.*

KING, G., J. ALT, N. BURNS, AND M. LAVER (1990): "A unified model of cabinet dissolution in parliamentary democracies," *American Journal of Political Science*, 846–871.

KWAK, H., C. LEE, H. PARK, AND S. MOON (2010): "What is Twitter, a social network or a news media?" in *Proceedings of the 19th international conference on World wide web*, ACM, 591–600.

LAMPOS, V., T. DE BIE, AND N. CRISTIANINI (2010): "Flu detector-tracking epidemics on twitter," *Machine Learning and Knowledge Discovery in Databases*, 599–602.

Lau, R. and D. Redlawsk (1997): "Voting correctly," *American Political Science Review*, 585–598.

Laver, M. and K. Shepsle (1996): *Making and breaking governments: Cabinets and legislatures in parliamentary democracies*, Cambridge Univ Pr.

Lax, J. R. and J. H. Phillips (2012): "The democratic deficit in the states," *American Journal of Political Science*, 56, 148–166.

Lazarsfeld, P., B. Berelson, and H. Gaudet (1944): *The peoples choice: How the voter makes up his mind in a presidential election*, New York: Duell, Sloan and Pearce.

Lewis-Beck, M., H. Norpoth, W. Jacoby, and H. Weisberg (2008): *The American voter revisited*, University of Michigan Press.

McPherson, M., L. Smith-Lovin, and J. Cook (2001): "Birds of a feather: Homophily in social networks," *Annual review of sociology*, 415–444.

Metaxas, P., E. Mustafaraj, and D. Gayo-Avello (2011): "How (Not) To Predict Elections," Unpublished manuscript.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller (1953): "Equation of state calculations by fast computing machines," *The journal of chemical physics*, 21, 1087.

Mislove, A., S. Lehmann, Y. Ahn, J. Onnela, and J. Rosenquist (2011): "Understanding the Demographics of Twitter Users," in *Fifth International AAAI Conference on Weblogs and Social Media*.

Mutz, D. (2002): "Cross-cutting social networks: Testing democratic theory in practice," *American Political Science Review*, 96, 111–126.

O'Connor, B., R. Balasubramanyan, B. Routledge, and N. Smith (2010): "From tweets to polls: Linking text sentiment to public opinion time series," in *Proceedings of the International AAAI Conference on Weblogs and Social Media*, 122–129.

Oken Hodas, N. and K. Lerman (2012): "How Visibility and Divided Attention Constrain Social Contagion," .

Parmelee, J. and S. Bichard (2011): *Politics and the Twitter Revolution: How Tweets Influence the Relationship Between Political Leaders and the Public*, Lexington Books.

Pennacchiotti, M. and A. Popescu (2011): "A machine learning approach to twitter user classification," in *Fifth International AAAI Conference on Weblogs and Social Media*.

Poole, K. and H. Rosenthal (1997): *Congress: A political-economic history of roll call voting*, Oxford University Press, USA.

——— (2007): *Ideology and Congress*, Transaction Pub, 2nd edition ed.

Reynolds, A. (2002): *The architecture of democracy: Constitutional design, conflict* *management, and democracy*, Oxford University Press.

Sakaki, T., M. Okazaki, and Y. Matsuo (2010): "Earthquake shakes Twitter users: real-time event detection by social sensors," in *Proceedings of the 19th international conference on World wide web*, ACM, 851–860.

Sang, E. and J. Bos (2012): "Predicting the 2011 Dutch Senate Election Results with Twitter," *EACL 2012*, 53.

Sartori, G. (1994): *Comparative constitutional engineering: an inquiry into structures, incentives, and outcomes*, NYU Press.

Schelling, T. (1978): *Micromotives and macrobehavior*, WW Norton & Company.

Shor, B., C. Berry, and N. McCarty (2010): "A Bridge to Somewhere: Mapping State and Congressional Ideology on a Cross-institutional Common Space," *Legislative Studies Quarterly*, 35, 417–448.

Stan Development Team (2012): "Stan Modeling Language: User's Guide and Reference Manual. Version 1.0." .

Stokes, D. (1963): "Spatial models of party competition," *The American Political Science Review*, 57, 368–377.

Strom, K. (1990): *Minority government and majority rule*, Cambridge Univ Press.

Sunstein, C. (2001): *Republic. com*, Princeton University Press.

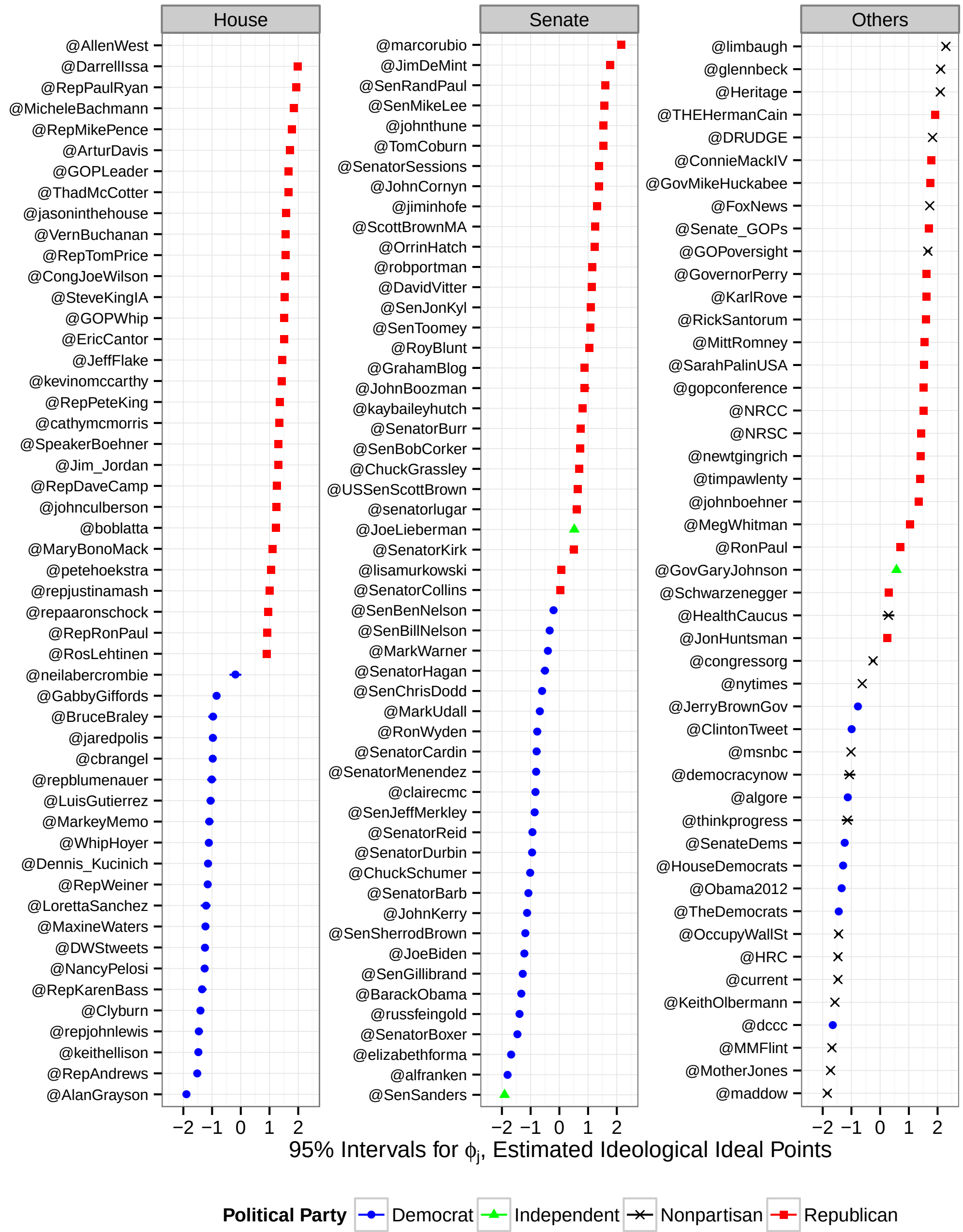
Tumasjan, A., T. Sprenger, P. Sandner, and I. Welpe (2010): "Predicting elections with twitter: What 140 characters reveal about political sentiment," in *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, 178–185.

Wu, S., J. Hofman, W. Mason, and D. Watts (2011): "Who says what to whom on twitter," in *Proceedings of the 20th international conference on World wide web*, ACM, 705–714.

Yardi, S. and D. Boyd (2010): "Dynamic debates: An analysis of group polarization over time on twitter," *Bulletin of Science, Technology & Society*, 30, 316–327.

# A Appendix. Additional Results

Figure 21: Estimated Ideal Points for Key Political Actors with 10,000 or more followers



# B Technical Appendix

## B.1 Estimation of the Bayesian Spatial Following Model

### B.1.1 The Model

The statistical model I assume to explain the decision of following a political account in Twitter is:

$$P(y_{ij} = 1|\alpha_j, \beta_i, \gamma, \theta_i, \phi_j) = \text{logit}^{-1}\left(\alpha_j + \beta_i - \gamma||\theta_i - \phi_j||^2\right) \tag{2}$$

where $y_{ij}$ equals 1 when user $i \in \{1, \ldots, n\}$ decides to follow political actors $j \in \{1, \ldots, m\}$, $\theta_i$ is the ideal point of user $i$ on the ideological dimension; $\phi_j$ is the ideal point of user $j$; $\alpha_j$ and $\beta_i$ are the "popularity" and "political interest" parameters of users $i$ and $j$, respectively; and $\gamma$ is a normalizing constant.

Given that none of these parameters is observed, the statistical problem here is inference for $\theta = (\theta_i, \ldots, \theta_n)'$, $\phi = (\phi_j, \ldots, \phi_m)'$, $\alpha = (\alpha_j, \ldots, \alpha_m)'$, $\beta = (\beta_i, \ldots, \beta_n)'$, and $\gamma$. Assuming local independence (individual decisions to follow are independent across users $n$ and $m$, conditional on the estimated parameters), the likelihood function to maximize is the following:

$$p(\mathbf{y}|\theta, \phi, \alpha, \beta, \gamma) = \prod_{i=1}^{n}\prod_{j=1}^{m} \text{logit}^{-1}(\pi_{ij})^{y_{ij}}(1 - \text{logit}^{-1}(\pi_{ij}))^{1-y_{ij}} \tag{3}$$
$$\text{where } \pi_{ij} = \alpha_j + \beta_i - \gamma||\theta_i - \phi_j||^2$$

The complexity of this equation makes direct estimation using maximum likelihood highly intractable. However, samples from the posterior density of each parameter in the model can be obtained using Markov-Chain Monte Carlo methods. More specifically, to make computation more efficient, I employ a hierarchical setup that considers each of the four set of parameters as drawn from common population distributions whose hyperparameters are also estimated:

$$\alpha_j \sim \text{N}(\mu_\alpha, \sigma_\alpha) \qquad \beta_j \sim \text{N}(\mu_\beta, \sigma_\beta)$$
$$\theta_i \sim \text{N}(\mu_\theta, \sigma_\theta) \qquad \phi_j \sim \text{N}(\mu_\phi, \sigma_\phi)$$

The full joint posterior distribution is thus:

$$
\begin{aligned}
p(\theta, \phi, \alpha, \beta, \gamma|\mathbf{y}) \propto\ & p(\theta, \phi, \alpha, \beta, \gamma, \mu, \sigma) \\
& \prod_{i=1}^{n}\prod_{j=1}^{m} \text{logit}^{-1}(\pi_{ij})^{y_{ij}}(1 - \text{logit}^{-1}(\pi_{ij}))^{1-y_{ij}} \\
& \prod_{j=1}^{m} \text{N}(\alpha_j|\mu_\alpha, \sigma_\alpha) \prod_{i=1}^{n} \text{N}(\beta_i|\mu_\beta, \sigma_\beta) \\
& \prod_{i=1}^{n} \text{N}(\theta_i|\mu_\theta, \sigma_\theta) \prod_{j=1}^{m} \text{N}(\phi_j|\mu_\phi, \sigma_\phi)
\end{aligned}
\tag{4}
$$

#### B.1.2 Identification

The increasing number of parameters that this method estimates[24] comes at a cost. As it stands in equation 1, the model is unidentified: any constant can be added to all the parameters $\theta_i$ and $\phi_j$ without changing the predictions of the model; and similarly $\theta_i$ or $\phi_j$ can be multiplied by any non-zero constant, with $\gamma$ divided by the same constant, or $\alpha_j$ or $\beta_i$ divided by its square root. These problems are sometimes called "additive aliasing" and "scaling invariance" (see e.g. Bafumi et al., 2005). In order to solve these two issues and identify the equation, I fix the mean of $\sigma_\alpha$ at one and apply a unit variance restriction on $\theta$. In the multilevel setting, this is equivalent to giving the $\theta_i$'s an informative $\text{N}(0, 1)$ prior distribution (Gelman et al., 2007, p.318).

An additional difficulty is reflection invariance: the resulting scale can be reversed (flipped left-to-right) without changing the prediction of the models. This is a problem for interpretation, but not for estimation, and can be easily solved by specifying starting values that are consistent with the expected direction of the scale (e.g. setting $\theta_i$ for a liberal Twitter user as $-1$), or by multiplying the estimated parameters by $-1$ after the chain has converged if the scale is not in the proper left-right direction.

#### B.1.3 MCMC algorithm

To improve the efficiency of the estimation procedure, I divide it in two stages. First, I use a No-U-Turn sampler, a variant of Hamiltonian Monte Carlo sampling algorithms (Gelman et al., 2013), to estimate the parameters indexed by $j$. This allows for very efficient sampling from the posterior distribution – 1,000 iterations from 2 chains with overdispersed starting values is enough to obtain effective number of simulation draws over 500. However, the current implementation of this algorithm in the Stan modeling

[24] In the case of the US, for example, I am estimating a total of over 600,000 parameters.

language (Stan Development Team, 2012), as other sampler libraries such as JAGS or BUGS, does not handle well large datasets, and therefore I use only a random sample of 25,000 $i$ users who follow at least 20 $j$ users. This limitation does not affect the estimation of the parameters indexed by $j$, whose posterior distribution remains essentially constant once the size of this random sample is higher than 10,000 $i$ users.

In the second stage, I use a Metropolis algorithm (Metropolis et al., 1953) with a uniform jumping distribution to estimate all parameters indexed by $i$. The only difference with respect to the original formulation of this sampling algorithm is that, in each iteration, a new value of the $j$ parameters is drawn from their posterior distributions (estimated in the first stage), in order to account for our uncertainty about their true value. This approach seems less efficient than the one used in the first stage – over 2,000 iterations and 2 chains are now necessary to obtain effective number of simulation draws around 200. However, note that each of the $i$ parameters can be estimated individually because we assume local independence[25], and therefore multi-core processors can be used to run multiple samplers simultaneously and dramatically increase computation speed[26].

The first stage is implemented using the Stan modeling language, while the second stage is implemented using `R`. I use flat priors on all parameters, with the exception of $\mu_\theta$, $\sigma_\theta$ and $\sigma_\alpha$, which are fixed to 0, 1, and 1 respectively for identification purposes. The samplers in both stages are run using two chains with as many iterations as necessary to obtain effective number of simulation draws (Gelman and Rubin, 1992) over 200, which is enough to estimate the parameter means with a precision of two decimal digits. Each chain is initiated with random draws from a multivariate normal distribution for $\phi$ and $\gamma$, the logarithm of the "indegree" of user $j$ or "outdegree" of user $i$ for $\alpha$ and $\beta$ (to speed up convergence), and values zero for $\theta$, with the exception of those who belong to a party, $-1$ for left-wing politicians and $+1$ for right-wing politicians. This choice is only necessary in order to interpret the results of the model (see section above). The results appear to be insensitive to the choice of priors and initial values. All relevant code is available upon request.

## B.2 Convergence Diagnostics and Model Fit

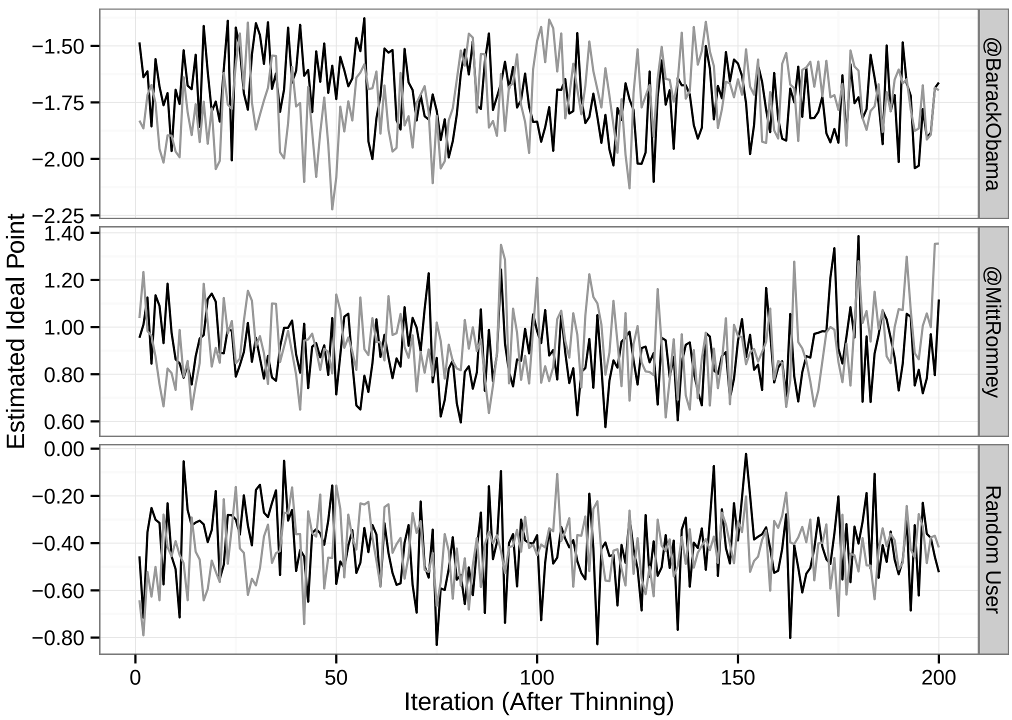
Despite the relatively low number of iterations, visual analysis of the trace plots, estimation of the $\hat{R}$ diagnostics, and effective number of simulation draws show high level of convergence in the Markov Chains. Figure 22 shows that each of the two chains used

[25] Note also that it is the independence assumption what allows us to divide the estimation method in these two stages.

[26] Samples from the $i$ parameters in the second stage can be compared with those obtained for the random sample in the first stage to ensure that there were no errors in the estimation. In all the examples in this paper, the correlation between these two sets of estimates is $\rho = .999$.

to estimate the ideology of Barack Obama, Mitt Romney and a random $i$ user have converged to stationary distributions. Similarly, all $\hat{R}$ values are below 1.05, which is considered the rule-of-thumb to monitor convergence of multiple chains, and the effective number of simulation draws is over 200 for all parameters. The results of running Geweke and Heidelberg diagnostics also indicate that the distribution of the chains is stationary.

Figure 22: Trace Plots. Iterative History of the MCMC Algoritm



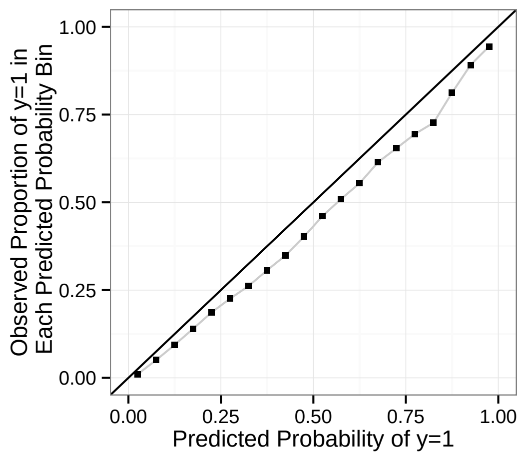

The results of a battery of predictive checks for binary dependent variables are shown in Table 2. All of them show that the fit of the model is adequate: despite the sparsity of the 'following' matrix (less than 8% of values are 1's), the model's predictions improve the baseline (predicting all $y_{ij}$ as zeros), which suggests that Twitter users' following decisions are indeed guided by ideological concerns. Besides the widely known Pearson's $\rho$ correlation and the proportion of correctly predicted values, Table 2 also shows the AUC and Brier Scores. The former measures the probability that a randomly selected $y_{ij} = 1$ has a higher predicted probability than a randomly selected $y_{ij} = 0$ and ranges from 0.5 to 1, with higher values indicating better predictions (Bradley, 1997). The latter is the mean squared difference between predicted probabilities and actual values of $y_{ij}$ (Brier, 1950).

Table 2: Model Fit Statistics.

| **Statistic** | **Value** |
|---|---|
| Pearson's $\rho$ Correlation | 0.592 |
| Proportion Correctly Predicted | 0.940 |
| PCP in Baseline (all $y_{ij} = 0$) | 0.924 |
| AUC Score | 0.915 |
| Brier Score | 0.046 |
| Brier Score in Baseline (all $y_{ij} = 0$) | 0.076 |

A visual analysis of the model fit is also shown in Figure 23, which displays a calibration plot where the predicted probabilities of $y_{ij} = 1$, ordered and divided into 20 equally sized bins (x-axis), are compared with the observed proportion of $y_{ij} = 1$ in each bin. This plot also confirms the good fit of the model, given that the relationship between observed and predicted values is very close to a 45-degree line (in dark color).

Figure 23: Model Fit. Comparing Observed and Predicted Proportions of $y_{ij} = 1$



# C Imputing Twitter Users' Gender

## C.1 Description

Gender is one of the most important predictors of electoral behavior and individual political attitudes in the United States (see e.g. Lewis-Beck et al., 2008). In this section I present a method to impute it for each Twitter user based on their self-reported full name. This method uses the frequencies of appearance of first names by gender in anonymized databases as the basis for a model that estimates the probability that a given user belongs to each gender group. This approach has been found to yield highly accurate estimates by different studies (see e.g. Mislove et al., 2011), and is a common procedure when matching and validating voting records (Ansolabehere and Hersh, 2012, p.6).

In short, the method has three different steps. First, each Twitter user's full name is pre-processed and the first name is extracted. Users who report their gender the description are pre-classified. Second, Twitter users' names are merged with a database of common first names using a fuzzy string matching method. Finally, I use Bayes' rule to compute the probability that each user belongs to a given gender, and use this vector of probabilities to impute gender.

In detail:

1. **Pre-processing.** Besides their screen names, Twitter users are also asked to provide their full name. Most of them accept to do so. Even if they choose pseudonyms, it is probably the case that the associated gender of their alternative names is the same as their own. The first step in the analysis is thus to split their reported names in first, middle (when available), and last name. In addition, before proceeding to the following step, I also classify those users who report their gender in their description. Individuals who define themselves as "mother", "woman", "girl", "wife", "grandmother"... are classified as women, whereas those who define themselves as "guy", "man", "dude", "dad", "father", "husband"... are classified as men. Around 18% of the users in the U.S. sample include any of these words in their description, and can therefore be classified with nearly perfect accuracy.

2. **Partial String Matching.** The second step is to match first names with their equivalent in a database that reports their use frequency among different gender and race groups. In this paper, for reasons of convenience, I use the anonymized database available in the RandomNames `R` package (Betebenner, 2012). I then compute the generalized Levenshtein distance between each name in my dataset and each name in this database. This statistic measures the distance between two

strings by computing the minimal number of insertions, deletions, and substitutions needed to transform one string into the other. This score is then divided by the number of characters in each name to obtain a score that ranges from 0 (perfect match) to the number of characters in the name (completely imperfect match). Matches whose score is above 0.25 are discarded. (This is threshold is arbitrary, but it seems to maximize precision in this application.) Around 80% of Twitter users have at least one of their names matched to a name in the `randomNames` database. Note that the partial string matching step is necessary because some names are misspelled or abbreviated.

3. **Computing Vector of Probabilities.** Once the names have been matched, it is known how frequent they are across gender groups, and the probability that a given user belongs to each group can be easily computed using Bayes rule. The probability that a given Twitter user belongs to a category $j$ is:

$$P(C_j|fullname_i) = \frac{1}{K_i}\sum_{k=1}^{K_i}\frac{P(name_{ik}|C_j)P(C_j)}{\sum_{j=1}^{J}P(name_{ik}|C_j)P(C_j)}$$

where $name_{ik}$ is each of the $K_i$ names of user $i$ that have been matched with the database, $P(name_{ik}|C_j)$ is the probability that an individual in class $j$ is named with $name_{ik}$, and $P(C_j)$ is the prior probability of belonging to class $j$. Note that each matched name is considered equally informative, since we assign equal weight to each match. Prior probabilities are obtained from the approximate distribution of gender in the US according to the 2010 census: 49% male, 51% female.

This procedure generates, for each Twitter user and characteristic, a vector of probabilities indicating how likely it is that he/she belong to each group. These vectors are used to sample from a binomial (or multinomial) distribution and thus impute a value for the race and gender variables for each Twitter user. The resulting distribution is reported in Table 1.

## C.2 Validation

In this section I present the results of a preliminary validation analysis of the method I employ to impute gender. A random sample of 500 users was drawn from the universe of Twitter users in the US I consider in this paper. Their gender was labelled manually by a third person, based on their profile pictures, description, and most recent tweets. This data is considered "ground truth" and is compared to the imputed gender for each user in the following two confusion matrices.

The results show that gender is correctly imputed for 75.6% of the users in the sample, as reported in Table 3. This level of accuracy is high given the relatively low level of information that the method requires. In comparison, Al Zamal et al. (2012) achieve a maximum accuracy of 80.2% using a complex machine learning model whose features are the complete text of each user's latest tweets, their network of friends and followers, their retweeting tendency, and other information. Note that this result is also significantly better than assigning gender at random – in which case the level of accuracy would be around 50% by construction.

Table 3: Confusion Matrix. Gender

| | | Imputed Gender | | |
|---|---|---|---|---|
| | | Male | Female | Unknown |
| | Male | **206** | 22 | 13 |
| Labelled Gender | Female | 14 | **117** | 18 |
| | Unknown | 37 | 18 | **55** |

# D Ideology and Self-Reported Votes

Table 4: 25 sample tweets and their imputed votes

| Tweet | Imputed Vote | Estimated Ideal Point |
|---|---|---|
| @ToddKincannon: South Carolina doesn't nominate losers. We didn't vote for Romney. Y'all should have listened. BACKTRACK much? | Obama | -0.19 |
| ll cast my vote tomorrow for Barack Obama. | Obama | -1.22 |
| RT @NolteNC: For four years I couldnt wait to vote AGAINST Barack Obama. But today I was proud to vote FOR Mitt Romney. | Romney | 1.32 |
| Just cause you love Kid Rock amp; hes all for Romney and is on his side, doesnt mean you absolutely have to vote for Romney now. VoteObama | Romney | -0.07 |
| RT @favorisonme: I cast my vote early for President Obama because I believe he truly cares for ALL people and he is the best man for the ... | Obama | -0.55 |
| Americans: "I'm voting for Romney because 24 million are unemployed. Rest of the world: If I was American I'd vote Obama cause he's cool. | Romney | -0.08 |
| Obamacare is a sham tho, thats one reason why my vote aint going to Obama either | Obama | -0.28 |
| RT @favorisonme: I cast my vote early for President Obama because I believe he truly cares for ALL people and he is the best man for the ... | Obama | -0.27 |
| back home for a few hours. Going to cast my vote for Mitt Romney in the morning romneyryan2012 | Romney | 1.21 |
| Lil old lady asks if I "voted properly". I said yes. Lady, "O good I'm so glad you didn't vote for Romney." #headdesk #smh #somepeoplestupid | Obama | 1.28 |
| RT @CFLtvEngineer: Why I Changed My Vote to Romney featuring Joshua K. http: t.co VDrJr3th | Romney | 0.80 |
| RT @favorisonme: I cast my vote early for President Obama because I believe he truly cares for ALL people and he is the best man for the ... | Obama | -0.11 |
| RT @UncleRUSH: Proud to cast my vote for Barack Obama!! http://t.co/fJpQXSkr | Obama | -0.25 |
| I have a friend that's a Republican based off his pocketbook but the hatred towards our President I must change my vote for Obama #VoteObama | Obama | -0.74 |
| @BarackObama Romney won Tx thanks to GOD my vote did not go to waste #teamObama. | Obama | -0.33 |
| RT @RelatableShiet: Mitt Romney is very inspirational. He inspired me to vote for Obama. #VoteObama | Obama | -0.02 |
| Clear eyes! Full hearts! Cant lose! Go vote for Romney! | Romney | 2.11 |
| As a centrist Repub, I yearn for good market-based policymakers to vote for. As a CleanTech proponent Im voting for Obama and Warren 2moro. | Obama | 0.02 |
| RT @chrisrockozfan: #IVoted just now. Could not be prouder to vote for Barack Obama. Not a day goes by that I don't feel lucky to have h ... | Obama | -0.60 |
| RT @favorisonme: I cast my vote early for President Obama because I believe he truly cares for ALL people and he is the best man for the ... | Obama | -1.36 |
| RT @jonlovett: Could not be prouder to vote for Barack Obama. Not a day goes by that I dont feel lucky to have him as our President. | Obama | -0.35 |
| RT @NolteNC: For four years I couldnt wait to vote AGAINST Barack Obama. But today I was proud to vote FOR Mitt Romney. | Romney | 1.52 |
| @AnnDRomney @MittRomney @seanhannity RomneyRyan2012 You got my vote and praying that we will see President Romney tonight. | Romney | 0.94 |
| @Tesssody to go vote for Romney..... Lol | Romney | 2.86 |
| And my vote didn't count, Romney wins LA #Election2012 | Romney | 0.25 |