Martins de Souza, Rafael; Guedes da Graça, Luís Felipe; dos Santos Silva, Ralph
Politics on the Web: Using Twitter to Estimate the Ideological Positions of Brazilian
Representatives

A R T I C L E

Politics on the Web: Using Twitter to Estimate the Ideological Positions of Brazilian Representatives*

Rafael Martins de Souza
Fundação Getulio Vargas, Brazil

Luís Felipe Guedes da Graça
Universidade Federal de Santa Catarina, Brazil

Ralph dos Santos Silva
Universidade Federal do Rio de Janeiro, Brazil

The use of social media has become increasingly widespread among citizens and politicians in Brazil. This means of communication served as a key arena for debate and propaganda during the 2014 legislative and presidential elections, when a very polarized political scenario emerged. New approaches have been developed that use information from the social network structure constructed by political actors on social media platforms, such as Twitter, in order to calculate ideal points. Can data from the decision to 'follow' a profile on Twitter be used to estimate politicians' ideological positions? Can approaches like this show the variance of political positions even within a very fragmented legislative body, such as the Brazilian Chamber of Deputies? This article presents and analyzes the successful application of a Bayesian spatial model developed by Barberá (2015), using data from Brazil. This method allowed to capture differences between parties and political actors similar to those found by means of roll call votes. It also makes possible to calculate ideal points for actors who participate in the public debate, but are not professional politicians.

Keywords: Bayesian inference; election campaigns; political participation; social media; Brazil.

The academic interest dedicated to the Internet and to its impact on political and social dynamics has been rising steadily. From an initial focus on the study of the Internet as a holistic phenomenon, researchers around the world are increasingly following Farrell's (2012) advice to consider the Internet as a bundle of casual mechanisms that can be disentangled, each one having different impacts on individuals and political discourse. Going even further, some researchers are looking at the Internet not because of its impact on society, but in search of data or new methods for social science investigation.

Social media platforms became crucial during the Brazilian presidential campaign of 2014. Not only did the main candidates make extensive use of these platforms to communicate and spread campaign videos, but members and supporters of various parties used the Internet to voice their positions, to debate and argue - sometimes quite ferociously - about the presidential election. As these media platforms played such an important role in the 2014 Brazilian presidential campaign, we advanced the idea of using online data to replicate methods already well established in the social sciences. Our focus in this article is the use of Twitter 'following' data to estimate the ideological positions of federal deputies in Brazil. We tested the method developed by Barberá (2015) in order to calculate the position of federal deputies on a latent government-opposition dimension. This study discusses whether data from the decision to follow a profile on Twitter can be used to estimate politicians' ideological positions in the Brazilian context. Can approaches like this show the variance of political positions even in a very fragmented legislative body such as the Brazilian Chamber of Deputies? Should the latent space produced by this method be interpreted as a representation of ideological positions or of government-opposition antagonism?

Brazil may be regarded as an interesting case to test this method due to its congressmen's high use of social media in a multiparty legislative body that has one of the largest party fragmentation in the world. In addition, applying this method to a case that does not belong to the usual United States/European Union axis allows for greater robustness of the findings.

Following this introduction, we present an overview of the debate about the Internet and political science methods (Section 02), a brief account of Brazilian politics around the time we gathered the data (Section 03), the method that was adopted (Section 04), the data used (Section 05), and the results (Section 06). We then conclude

the article with a brief discussion.

Our findings show a good fit between estimates using Twitter data and more traditional methods used in the literature. At first, the latent space produced using social network connections seems difficult to decipher: does it represent ideological positions or government-opposition dynamics? Because of the PT's (Partido dos Trabalhadores, Workers' Party) and the PSDB's (Partido da Social Democracia Brasileira, Brazilian Social Democratic Party) extreme positions in the political space we are led to believe there is a mixture between the two most common interpretations for latent space in politics: left-right and government-opposition. The fact that the method produces an unidimensional space adds to this confusion.

Once we consider the origin of the data inputted in the calculations, namely the social media platforms, it is possible to dismiss part of what led us to consider the existence of government-opposition interference. As we are dealing with connections in a virtual social network, the distance between the PT and the PSDB is greater in our estimation approach than in other ideological position estimation methods due to these parties' historically longstanding competition for the Presidency. For users of social media platforms, the historical political rivalry between the two parties affects their decisions: the cost of following PT's or PSDB's politicians on Twitter is higher than to follow a more ideologically-distanced politician from another party.

The Internet and political science methods

The use of the Internet to reach individuals or to harvest new types of data constitutes a novel influence on social science methods, bound to leave its mark on methodology textbooks. One of the fields that has experienced the possibilities opened up by Internet data is that of surveys and election prediction. Not only may reaching respondents through the web change survey methods (GROVES, 2011), but some have been discussing the possibilities and limitations of using Twitter to predict elections (GAYO-AVELLO, 2012; JUNGHERR et al., 2012; LAMPOS, 2012; TUMASJAN et al., 2011). Moreover, Beauchamp (2016) uses Twitter textual data to predict state-level polls in the US.

In the last few years, some scholars combined virtual social networks data and the estimation of ideological positions. Originally, Poole and Rosenthal (1997) consolidated the use of ideal point estimates in political science. Through the use of roll-

call votes from the first to the 100th US Congress, they were able to estimate the spatial positions of all members of Congress. The idea - present in every version of the method developed by Poole and Rosenthal (2001): Nominate, D-Nominate and DW-Nominate - is to represent ideology in a spatial model, assuming that ideology is a bundle of interrelated issues that tend to go together, providing individuals, the media, and politicians with labels that work as shortcuts to predict positions in a series of issues.

With the advance of personal computer processing capacities and the increase in the use of Bayesian statistics, other estimation procedures have been proposed. Jackman (2001) and Clinton et al. (2004) present models that calculate ideal points using the Bayesian approach, thus reducing the problems caused by a high number of parameters and facilitating the use of auxiliary information, such as expert judgments or the position of key actors to inform the model.

Despite being very important and influential in political science around the world, the use of roll-call data to estimate ideal points restricts the application of these types of methods solely to legislators. It also makes use of a source of data that is based on highly strategic moments and may suffer from selection problems (HUG, 2010). In order to expand the types of agents whose positions can be estimated, Bonica (2013) develops a method to recover candidates' ideal points using campaign contribution data. With this approach, the author is able to estimate ideology from successful and unsuccessful candidates.

Nonetheless, approaches that are based on roll-call votes or campaign donations are still only applicable to legislators or candidates. The escalating use of social media and access to the data created by users has opened new opportunities in estimating ideological positions. Ecker (2017) describes the literature that makes use of this new kind of data as being divided into three different approaches: 01. Those using the content of messages posted by social media platform users; 02. those using the communication patterns of these websites; and 03. those using social network structure to estimate the ideological positions of users.

Conover et al. (2011) test different ways to predict political alignment in social media with data from the US 2010 midterm elections. They compare methods that use support vector machines (SVM) to analyze hashtags and complete textual data of Twitter users to methods based on communication patterns such as retweets and mentions. The latter method performs better at predicting political alignment -

something the authors suggest is due to the homophylic characteristic of social networks.

King et al. (2011) focus on whether US parties have increased the nomination of extremist candidates in the primary process. They have produced ideological estimates using follow/friend data from Twitter and have found that their estimates correlate highly with W-nominate scores.

Ceron and Cremonesi (2013) use data from different origins (Facebook, Twitter, and blogs) to estimate intraparty heterogeneity in the 2012 Italian Democratic Party primary process. Through quantitative text analysis they are able to explain endorsements and critics inside the party nomination with ideological distance of politicians.

Analyzing the United Kingdom's 2010 general elections, Boutet et al. (2013) propose different methods to classify Twitter users ideologically. They put forward an estimative method that uses the number of messages with reference to a specific party and a semantic content analysis. Compared to other methods that use the volume of mentions, sentiment analysis, retweet structure, followers and support vector machines, they find that methods that use following/follower data have a better performance. Their baseline for comparison is Twitter users that identify themselves as connected to a party in their profile description. Despite these findings, they defend their own method because of its computation speed, as it could be used for a real time estimation process.

Bond and Messing (2015) use Facebook data to estimate ideology. They estimate the political inclination of more than 1,223 political pages on this social network and then use their estimates to produce an individual estimate for profiles that follow these pages. Each one of the 6.2 million users receives an estimate that consists of their following political pages estimates' mean. They go further and do a survey of some of these users. Facebook data could be a great source for academic analysis, but access to it is very limited. This is the main reason behind the wide use of Twitter data in research. Despite some rules that make access and download of Twitter data a time-consuming operation, this social network website provides an API for data access.

Barberá (2015) proposes a method to calculate ideological estimates based on the decision to follow a profile on Twitter. He argues that the social ties developed by Twitter users, political actors, and media outlets can be the basis for an estimation of ideal points for ordinary users and politicians alike. The argument for this method relies

on the fact that, just as roll-call votes or campaign donations represent a choice to allocate scarce resources, the decision on whether or not to follow another Twitter account is a costly signal of the user's ideological position, since it is an allocation of a scarce resource: their attention. Barberá's findings (2015) suggest that the method is robust, as he presents strong correlations with estimates based on roll-call votes for individual legislator level in the US House of Representatives and the Senate. Estimates also seem to fit well with accounts for the UK, Germany, Italy, Spain and the Netherlands. The model developed is similar to the one presented by Clinton et al. (2004). The use of data about decisions to follow a profile on Twitter, instead of roll-call votes, allows the estimation of ideal points for the Twitter profiles of individuals, media outlets, and politicians. As the author argues, this should be considered an additional measurement method to be used in political science. It is a well-based approach that conveys data produced online with methods developed earlier in political science, expanding the horizons of research.

Ecker (2017) is the first to replicate Barberá's work (2015) in a different case. Looking at Polish politicians in the lower House, in a baseline study like this one, he finds that party level estimates are similar to other estimation methods based on expert and elite surveys. However, individual legislator level estimates correlate at a medium 0.57 with estimates based on roll-call votes.

In this study, we follow the example of Barberá (2015) and Ecker (2017) and test whether the method used by them produces estimates that can be validated by more traditional approaches. This method has the great advantage of being applicable to a wide array of Twitter users, not only those that participate in the political process as professional politicians or campaign donors. We applied it to a small number of profiles comprising the accounts followed by more than 30 federal deputies. Nevertheless, the technical difficulties involved in this application are still quite high. First, in spite of its open access, Twitter's API has rate limits[1] that hinder fast access to large amounts of data. The popularity of some users implies a list of follower/following that comprises users in the hundreds of thousands, which - in combination with the API rate limits – means that several hours or days are needed for data collection. Second, the estimation process also takes a long period to be completed. This method presents some computational difficulties: the use of Markov chain Monte Carlo methods in the

_____

[1] See Twitter API rate limits - https://dev.twitter.com/rest/public/rate-limiting.

estimation process places high computational demands when the number of cases rises. It took Barberá (2015) 18 hours using a high performance computer at New York University to compute his estimates. Access to this kind of computer is still rare in the social sciences, and even rarer in the Brazilian political science community. Imai et al. (2016), in a replication with fewer cases than the original study, took 6.5 days to calculate estimates on an ordinary computer, and it took us a similar amount of time. These authors have created a quicker way to calculate this kind of point estimates, which produces results similar to Barberá's work (2015). However, their methods resort to a bootstrap technique for computing and approximate the credible intervals. Thus, depending on the number of bootstrap replications, they need to compute their (pseudo) point estimates several times, that is: their computing time may also be long. Confidence interval gives a measure of uncertainty on the latent space and it tells us when two or more individuals have or do not have differing ideological positions. We have kept with Barberá's (2015) methodology since, from a Bayesian point of view, it provides the right way to estimate the whole posterior distribution of the ideological position and any summary from it, such as posterior mean (our point estimate) and credible intervals. Moreover, our intention here is to check the validity of estimates calculated with social network data in comparison to usual estimates based on different methods.

A very brief overview of Brazilian politics

The 2014 presidential campaign in Brazil was one of the closest electoral races in the country's political history. Dilma Rousseff, the reelected candidate from the Workers' Party (PT, Partido dos Trabalhadores), won the election in the runoff with 51.6% of valid votes against 48.4% for the challenger, Aécio Neves, from the Brazilian Social Democratic Party (PSDB).

The result was another chapter in the dispute between the two parties for the Brazilian presidency. In the last six national elections, the PT and the PSDB have been the major protagonists in the electoral competition for the presidency. Since 1994, both parties have occupied either first or second position in the presidential race, led coalition governments, and have left their mark on public policies. Despite the recurrent attempts by third parties to break away from this dichotomy in presidential contests, the PT and the PSDB once again polarized voters' preferences in the 2014 election.

Regardless of what might be seen as a crystallization of preferences around two parties in the competition to lead government coalitions, this polarization in presidential campaigns has not been met by a concentration of preferences around the same parties in legislative elections. The concurrent 2014 election for the Chamber of Deputies in Brazil produced the most fragmented legislature in the country's history, with 28 parties represented and an effective number of parties of 13.2 (GALLAGHER, 2015).

The effect, both on political stability and on coalition management, of the combination of a proportional open list electoral system, multiparty legislatures, and direct election for president has been one of the major issues in the Brazilian political science debate (AMES, 2001; FIGUEIREDO and LIMONGI, 1995; SHUGART and CAREY, 1992; STEPAN, 2000). These resultant political structures, i.e. clear political poles but multiple legislative parties, pose an interesting challenge to ideology point estimations. As Zucco Jr (2009) argues, roll-call votes, the usual data source for the estimation of ideology, are the result of extensive political exchanges in order to gather legislative support. Estimation of ideology that is exogenous to legislative behavior is an important way to recover what could be understood as more genuine ideological positions.

The 2014 national election in Brazil was extremely competitive and reinforced the importance of web based communication. As a result, citizens as well as politicians running for different positions used social media intensively for campaigning and debating. In a survey of five metropolitan areas and of the Federal District, carried out by the Department of Public Policy Analysis of Getúlio Vargas Foundation shortly after the first round elections in Brazil (RUEDIGER, 2016), 45.3% of respondents said they had used the Internet to get information about politics, and 58.4% considered that the campaign on the Internet or on social media had had some influence on their voting decision.

Point estimation using twitter data

The method proposed by Barberá (2015) relies on the assumption that Twitter users prefer to follow politicians, political actors, or media outlets that have positions similar to theirs. Allocating time and attention to these profiles, by deciding to follow them on Twitter, should be understood as a costly signal about the user's preferences.

In order to test this ideal point estimation method, we decided to apply it to

Brazilian federal deputies. Although the biggest advancement brought by Barberá's (2015) article is its applicability to individuals other than political elites such as elected officials, it is still necessary to test the method's robustness in a population whose ideology we can estimate using different methods. The estimation of Barberá's (2015) method will be confronted with individual legislator level estimates using roll-call votes from the Núcleo de Estudos do Congresso for the period 2011-2014 (SANTOS and CANELLO, 2015), as well as party level estimates from Zucco Jr and Lauderdale (2011) and Power and Zucco Jr (2012)[2]. After making the comparison with these consolidated estimates, and since our application has allowed us to estimate ideal points for a series of political actors other than federal deputies, we present estimates for some journalists and commentators of political news, to illustrate how the use of Twitter data allows us to estimate comparable ideal points for users' profiles in general.

Assumptions and statistical model

Barberá's (2015) model is similar to latent space models, item-response theory models, and models that use roll-call votes or campaign contributions to estimate latent political dimensions. Starting with a matrix that consolidates data about user's choice on whether or not to follow a range of different Twitter accounts, we can estimate the position of a user in a single latent space.

Consider that a Twitter user $i \in \{1, \dots, n\}$ has a choice of whether or not to follow a Twitter profile $j \in \{1, \dots, m\}$, and let $y_{ij} = 1$ if the Twitter user i decides to follow the Twitter profile j, and $y_{ij} = 0$ otherwise. We assume that this decision is a function of the squared Euclidean distance in the latent ideological dimension between Twitter users i and j, i.e., $\gamma \|\theta_i - \phi_j\|$ where $\theta_i \in R$, $\phi_j \in R$, and $\gamma > 0$ is the normalizing constant. Moreover, the parameter vectors $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_m)$ and $\beta = (\beta_1, \beta_2, \dots, \beta_n)$ represent the level of political interest for each user and the popularity of politicians in Brazil, respectively. Now, let $\eta_{ij} = \alpha_j + \beta_i + \gamma \|\theta_i - \phi_j\|^2$. Then, conditional on the set of parameters $\Psi = (\theta, \phi, \alpha, \beta, \gamma)$, the observations y_{ij} are independent, by assumption,

---

[2] Tarouco (2011), as well as Tarouco and Madeira (2013), approach the ideological position of parties through their manifestos. Despite their sound methodology and interesting results, we decided not to use their data for comparison due to the fact that they have been produced at precise historical and sparse moments, and do not comprise most of the parties represented in the Chamber of Deputies during 2015.

such that the likelihood function is given by

$$p(y|\theta, \phi, \alpha, \beta, \gamma) = \prod_{i=1}^{n} \prod_{j=1}^{m} \pi_{ij}{}^{y_{ij}} (1 - \pi_{ij})^{1-y_{ij}}$$

with logit $(\pi_{ij}) = \eta_{ij}$, where y is the vector of observed values. Completing our Bayesian model specification, we assume independence a priori and set $^{a}\gamma \sim G(a_{\gamma}, b_{\gamma})$, $\alpha_j \sim N(\mu_{\alpha}, \sigma_{\alpha}^2)$ and $\phi_j \sim N(\mu_{\phi}, \sigma_{\phi}^2)$, for $j = 1, \ldots, m$, and $\beta_i \sim N(\mu_{\beta}, \sigma_{\beta}^2)$ and $\theta_i \sim N(\mu_{\theta}, \sigma_{\theta}^2)$ for i= $1, \ldots, n$. Due to identifiability, the hyperparameters are set as in Barberá (2015). The posterior distribution then follows by combining the likelihood function with the prior distribution.

A few computational details are given in Appendix A, available online on BPSR website.

Data

The first part of the research involved the definition of our cases: Twitter accounts of Brazilian federal deputies. A list of federal deputies was downloaded from the Chamber of Deputies on June 01, 2015. A search was made for an active[3] Twitter account for every one of the 513 deputies. Our database was then restricted to the Twitter accounts of 354 federal deputies.

In order to estimate the positions of these federal deputies we needed data about their decisions on whether or not to follow other profiles. The choice of this target group is a key aspect of the method. It is the user's decision - or, in our case, the federal deputy's - whether or not to follow the target Twitter accounts, and this provides us information about user preferences. With this in mind, we gathered every profile followed by them. This method would allow us to go on and gather the followers of these profiles and use them to estimate their ideal points. But, as we have mentioned earlier, this could escalate into a process that would require a massive amount of time for estimation. Due to this computational restriction, we made a list of 391 profiles being followed by at least 30 federal deputies. These profiles are designed hereafter as Actors. After obtaining this list of profiles we gathered their Twitter followers' lists and cross-checked account names focusing on creating a sparse relationship matrix. If a cell has a value of zero it means that there is no relation between a deputy and a Twitter profile. If

_____

[3] We decided to consider a Twitter account active if any posts had been made in 2014.

a cell has a value of one, it means that the federal deputy is following this profile on Twitter. This way we were able to estimate ideal points for the 354 federal deputies' profiles and for the 391 Actors' profiles[4].

As recommended by Barberá (2015), we set starting values for some of these Actors' Twitter profiles: for clear government supporters $\phi_j = -1$ and for government opponents $\phi_j = 1$. We assigned 32 profiles as more pro-government, 11 as opposition, while the others were not classified. The list of profiles and their assigned values are in Appendix B, available online on BPSR website. This first 'guess' is important in kick-starting our calculations, but does not restrict the results. As shown by our choice of initial guesses, we expect that the estimation should classify federal deputies on a latent space that shows either proximity to or distance from the government's ideology. This initial classification notwithstanding, the latent space can 'a priori' represent different characteristics; it tells us which Twitter profiles have a similar 'following' structure, not the meaning of the poles of the space.

Results

In order to better evaluate the results, the data is presented from an aggregate level down to specific cases. Our expectations are that the method will allow us to grasp latent positions in an unidimensional space. At first, we must see if the method did indeed estimate positions for Twitter profiles in a diversified range of scores. Figure 01 shows histograms for 01. all the Twitter profiles, federal deputies and Actors, 02. only for the federal deputies and 03. only for the Actors.

As Figure 01 shows, the method developed by Barberá (2015) allowed us to estimate latent space scores for federal deputies and Actors that range from -2.75 to +2.35. Federal deputies' Twitter profiles were more likely to have positive estimations, and Actors' Twitter profiles to have negative estimations.

Figure 02 shows the scores of federal deputies per party. Here we can observe some interesting patterns. First, all the federal deputies of the Workers' Party (PT) have negative scores, thus showing that our initial guess that ideological proximity with the government would be represented by negative scores was correct. The Workers' Party's most faithful coalition partner, the Communist Party of Brazil (PC do B, Partido

_____

[4] Because some politicians have central roles in their party or in Congress, they may appear in both lists.

Comunista do Brasil) also has its representatives scoring on the negative side of the latent dimension. The federal deputies belonging to parties that form the core opposition to government also score as expected. Most representatives from the Brazilian Social Democratic Party (PSDB), Democrats (DEM, Democratas) and Popular Socialist Party (PPS, Partido Popular Socialista) have positive scores.

Figure 01. Histograms for all Twitter profiles, federal deputies and Actors
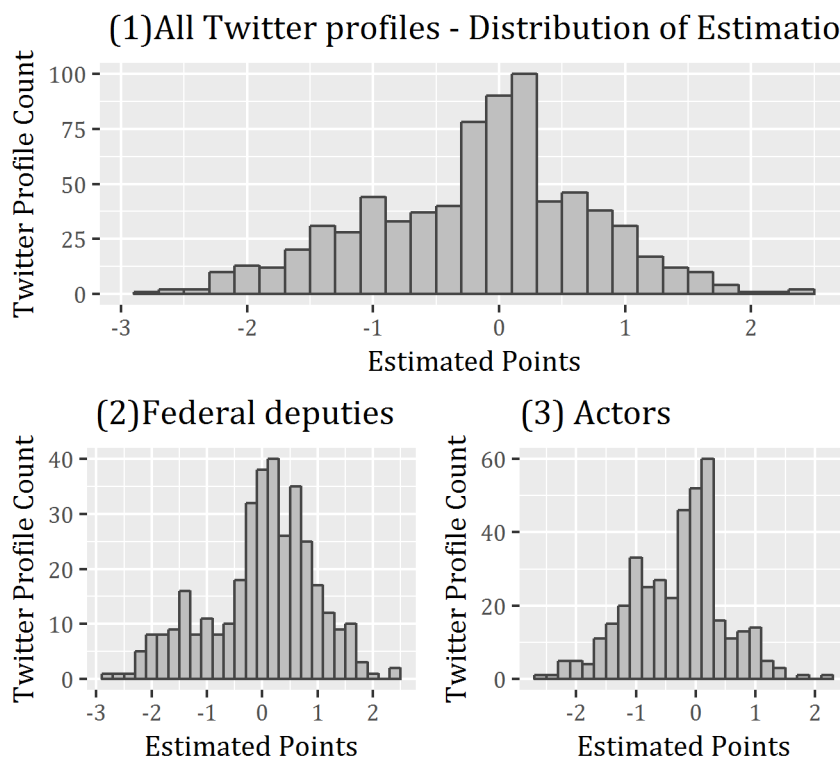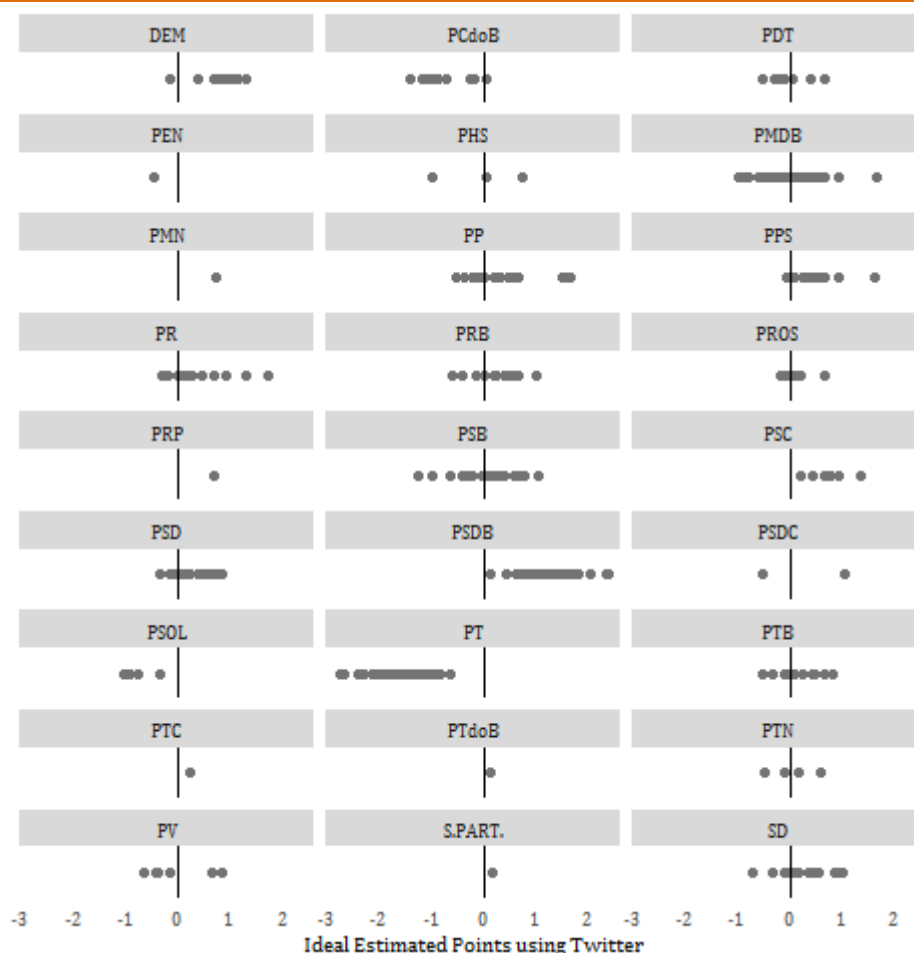

Source: Database -Twitter ideal points. See www.bpsr.org.br/files/archives/Dataset_Souza_Graça_ Silva.

We can also notice in Figure 02 that parties considered to be part of the government coalition due more to office-seeking strategies than to ideological proximity, such as the Brazilian Democratic Movement Party (PMDB, Partido do Movimento Democrático Brasileiro), the Progressive Party (PP, Partido Progressista), and the Republican Party (PR, Partido da República) have federal deputies scoring in the positive and negative parts of the latent space. The scores of federal deputies belonging to the PMDB are one of the most interesting results of the applicability of this method. They have shown a consistent division regarding government support, shown by the fact that some of them openly supported the opposition candidate's campaign for the presidency in 2014. This division apparently has signified their insertion in different

patterns of Twitter relations.

Figure 02. Scores for federal deputies per party

Another interesting aspect of the results is the scores estimated for the representatives of the Socialism and Liberty Party (PSOL, Partido Socialismo e Liberdade). These politicians consistently position themselves as an opposition party on the left. But their scores on the negative side of the spectrum, close to the PT federal deputies, seems to represent their immersion in a social network environment connected to leftwing activism more than their opposition status in the Chamber of Deputies. These results indicate that our latent space may also be reflecting a left-right division. The fact that this method calculates positions only on one latent space could allow for dimensions such as government-opposition and left-right to be scrambled in the representation of this single dimension.

At first, our approach seems to produce estimates that relate to common

expectations. We should look at other studies to validate these results. Zucco Jr and Laudedale (2011), Power and Zucco Jr (2012) and Santos and Canello (2015) provide us data from different approaches that help validate our estimates. The comparison between estimates is presented in Figure 03.

Figure 03. Comparison between estimates of political parties
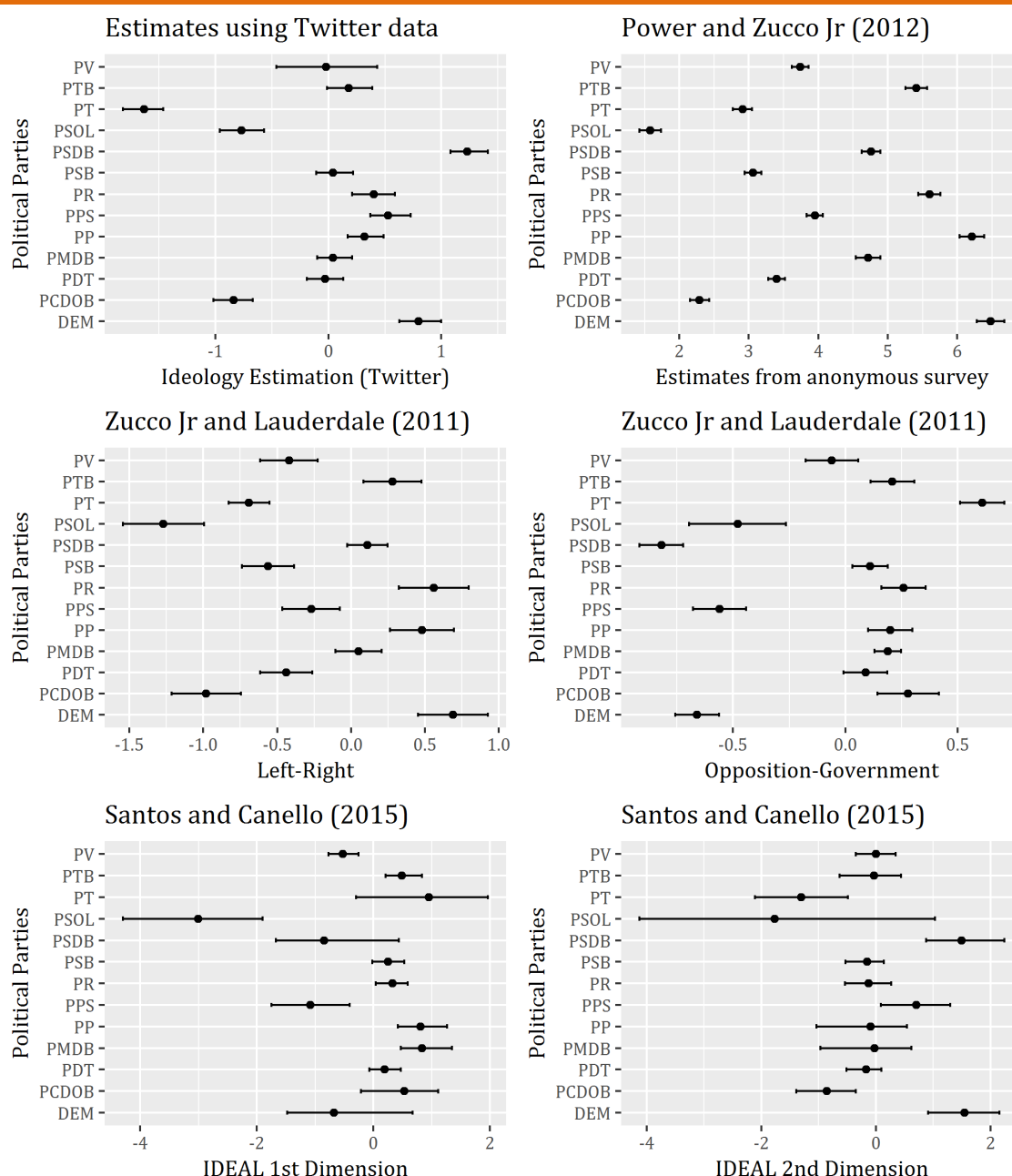
Source: Database -Twitter ideal points. See www.bpsr.org.br/files/archives/Dataset_Souza_Graça_Silva.

Zucco Jr and Lauderdale (2011) combine roll-call votes and an anonymous survey of Brazilian legislators to identify two dimensions. The first is interpreted as a

government-opposition space and the second as a left-right axis. The estimates we used are from the 2007-2010 period (53rd legislature). Santos and Canello (2015) provide estimates of two dimensional latent spaces using roll-call votes. Power and Zucco Jr (2012) estimate parties' ideological positions using an anonymous survey of Brazilian legislators. We used the estimates from the 2009 wave of surveys. The estimates of political parties for the method we test here are calculated as the mean of individual legislator's positions.

As in our estimation, the PT and the PSDB appear in different poles in every estimation in Figure 03, but their distance varies. This indicates that the latent space constructed by our estimation represents a divide between government and opposition, since such a distance between the PT and the PSDB appears only at the divide government-opposition of Zucco Jr and Lauderdale (2011) estimates, and in Santos and Canello (2015) IDEAL 2nd dimension. Generally, these parties appear closer to each other in latent spaces that may be interpreted as an ideological left-right axis and that are not based on strategic environment produced data, such as in Power and Zucco Jr (2012) and Zucco Jr and Lauderdale (2011) left-right space.

This inference, that the latent space estimated in this study represents a government-opposition divide, does not account for the fact that most of the parties that comprised the government coalition, such as the PMDB, the PP and the PR, are positioned closer to clearly opposition parties such as the DEM and the PPS, and are well distanced from the PT. This inference also does not account for the proximity between the PSOL and the PT. The positions of all these parties resemble more a left-right ideological divide.

Despite these remarks, the correlations between our estimates aggregated by party level and all the other estimates are not strong. Our estimate correlates at 0.249 with Power and Zucco Jr's (2012) ideological estimates based on an anonymous survey of congressmen, at 0.248 with Zucco Jr and Lauderdale's (2011) estimate of left-right position, at -0.454 with Zucco Jr and Lauderdale's (2011) estimate of government-opposition, at -0.297 with Santos and Canello's (2015) IDEAL 1st dimension and at 0.395 with Santos and Canello's (2015) IDEAL 2nd dimension.

Nevertheless, the main test for the robustness of this method is to compare its estimates with individual legislator's IDEAL estimates made available at the NECON webpage (SANTOS and CANELLO, 2015) for the years 2011-2014 roll-call votes. The

works of Zucco Jr and Lauderdale (2011) and Power and Zucco Jr (2012) rely, to some extent, on an anonymous legislators' survey and does not identify the respondents, deeming individual level comparison impossible. Since we are comparing data collected during the 2011-2014 legislature with data collected in 2015 from the Twitter profiles of federal deputies, we do not have comparable estimates for all our observations, although we were able to compare the estimators for 200 of them. The comparison between the estimates calculated with Twitter data and the first dimension of the IDEAL estimate for federal deputies shows a sound Pearson correlation of 0.799. As to IDEAL's 2nd dimension, the correlation is -0.299. A dispersion graphic is presented in Figure 04.

Figure 04. Comparison of estimations for individual legislators: IDEAL and Twitter
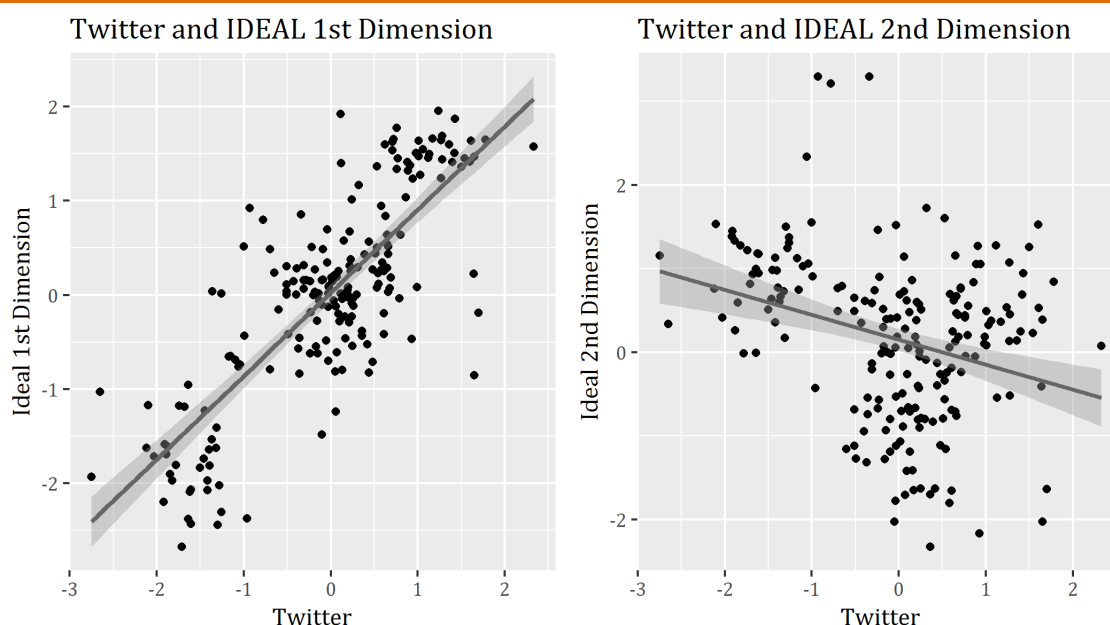
Source: Database -Twitter ideal points. See www.bpsr.org.br/files/archives/Dataset_Souza_Graça_Silva.

The estimates made using Barberá's method (2015) show a very promising application since it produced an estimate similar to the IDEAL first dimension without resorting to any data produced inside strategic environments, such as the ones in which congressional votes are immersed. The fact that there is a strong correlation between data from Twitter and roll-call votes is also found by Barberá (2015): his estimates correlate at an even higher level of 0.941 for politicians in the House of Representatives and of 0.954 for Senators. His comparison with data for other countries such as the UK, Germany, Italy, Spain and Netherlands are based on experts' surveys and party-level estimates. Ecker (2017) tests Barberá's method (2015) with data from the Polish lower

House. He finds a good fit between expectations and party aggregate level estimates, but a not so sound correlation of 0.57 for individual level estimator comparison with roll-call votes. This way our findings position themselves between these two tests, showing that the method appears to be robust.

Interpreting the latent space produced by our estimates is one of the main difficulties of this study. As it appears, the estimates comprise what might be understood as mixed signs. On the one hand, it resembles a left-right ideological distribution, as most of the coalitional partners of the PT classified as center-to-right parties by other authors – like the PR, the PP, the PMDB and the PTB – appear distanced from the Workers' Party. The proximity between the PT and the PSOL reinforces this interpretation. On the other hand, the extreme distance between the PT and the PSDB is similar to what could be interpreted as a government-opposition axis.

In order to better asses this confusion, we should take into account that the data used as input to our estimates is produced in a virtual social media platform. So it represents the immersion of these politicians and Actors in different networks of communication and fluxes of information. This network seems to be constructed mainly around ideological preferences; however, the electoral polarization between the PT and the PSDB in presidential elections seems to have influenced the formation of these networks also. Users that choose to follow politicians from one of these parties usually do not follow politicians from the other. This way, users with political preferences closer to the PT or to the PSDB are immersed in networks that do not connect and are exposed to diverse fluxes of information. This reinforces the idea that virtual social networks platforms are characterized by their homophylic nature.

Social media platform users – even professional politicians – are not immersed in such a strategic environment like Congress during votes. In roll-call votes, being in this environment (and expecting rewards and punishments) can influence federal deputies from different ideological positions to converge on the same vote. On a large scale, this dynamics can generate similar estimates for politicians with very different ideological preferences. It is through this that a government-opposition axis can appear as an interpretation of a latent space constructed using roll-call votes. The point is that the data we use in this article to calculate estimates are not under such a mechanism. Costs associated with following a Twitter account are related to being exposed to information produced at some ideological point that may differ from the follower's

perspective, as well as being associated with this specific account in one's personal profile. These costs rise as we choose to follow profiles that significantly differ from our political views. As the case of the PT and the PSDB shows, these costs can have a more intense relation to parties that are strong competitors at some level. As our estimate results suggest, the connection structure revealed by the social media platforms of politicians and users is characterized by political ideology, but this is exacerbated in cases of political rivalry.

After assessing some of the aggregate results of the estimation made using Twitter data, we looked at specific federal deputies and Actors to understand whether the variations estimated for them presented the expected characteristics and to show the best part of Barberá's (2015) method: its applicability to agents other than politicians who cast votes in Congress. In Figures 05, 06 and 07 we show the estimated points on this latent space calculated for key federal deputies, important political actors, and several of the best known journalists that cover or comment on daily politics in Brazil. As stated above, we collected data not only from federal deputies, but also from a population of 391 Twitter profiles being followed by at least 30 federal deputies. This list of influential Actors contains not only politicians who hold other offices, but also journalists, intellectuals, religious leaders, official party Twitter profiles, and even a couple of athletes. We estimated positions for all the Twitter profiles included in the process.

Figure 05 shows the estimated position of federal deputies occupying important positions in the Chamber of Deputies. We can readily see that there is a clear distance between the position estimated for the government leader (@guimaraes13PT) and the leader of the minority (@BrunoAraujo4511). It is also possible to observe the distance between the legislative leaders of the party of the President (@sibamachado13) and Jandira Feghali (@jandira_feghali), the leader of PC do B, and the legislative leader of the party of the opposition's presidential candidate, Carlos Sampaio (@carlossampaio_), and the leader of DEM, also an opposition party, Mendonça Filho (@mendoncafilho). Another interesting point highlighted by this method is how the structure of the social network in which important PMDB deputies are immersed, such as Eduardo Cunha (@DepEduardoCunha - Speaker of the Chamber of Deputies) and Leonardo Picciani (@DepLeopicciani - leader of PMDB), is closer to those of opposition politicians. The proximity of Eduardo Cunha with this network and his distance from the one in which the leader of the government and the PT are immersed is consistent with this politician's

actions as speaker and his argument during the year of 2015 that his party should leave the government coalition.

Figure 05. Estimated points for selected federal deputies
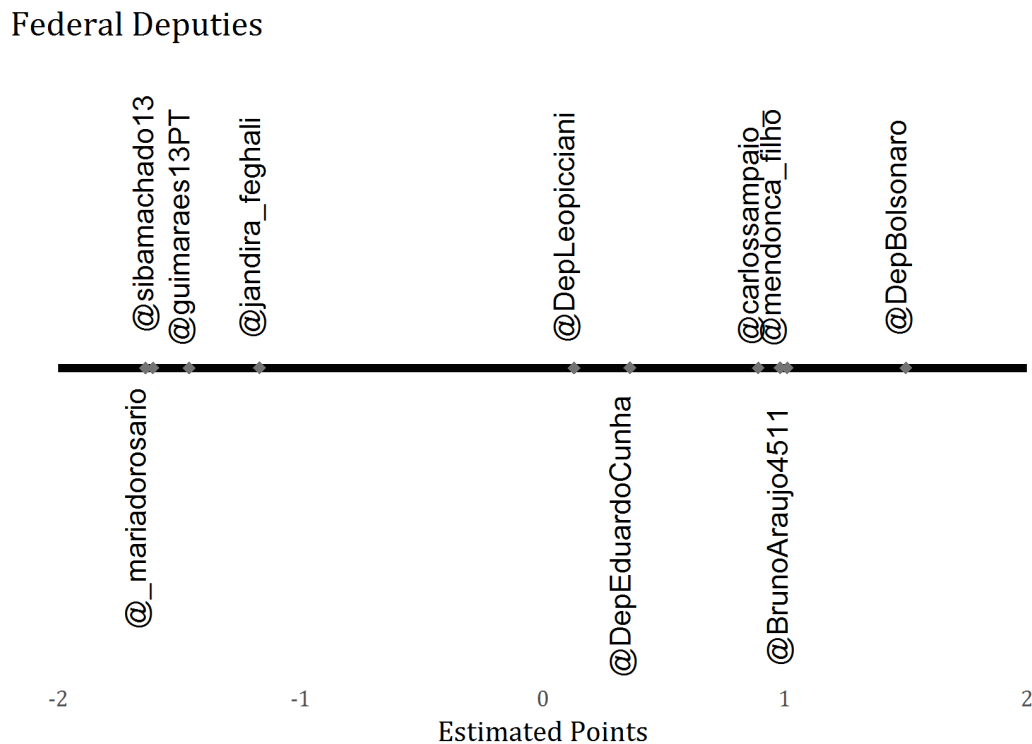
Source: Database -Twitter ideal points. See www.bpsr.org.br/files/archives/Dataset_Souza_Graça_Silva.

Figure 05 also shows the position of two federal deputies whose parties are part of the government coalition but have profound divergences in their political positions. The Workers' Party deputy Maria do Rosário and the PP's deputy Jair Bolsonaro have been the protagonists of some controversies in the debate about human rights. Our estimates positioned them at a very large distance from each other, showing that they participate in disparate digital networks.

Twitter data from twelve politicians' profiles shown in Figure 06 also follow the expectations that observers of the Brazilian political scene probably have. The official Twitter profiles of the President (@dilmabr) and of former President Lula da Silva (@presidente_lulal) score on the left side of the figure, while the profiles of Aécio Neves (@AecioNeves), the 2014 PSDB's presidential candidate, Geraldo Alckmin (@ geraldoalckmin_), governor of São Paulo and PSDB's 2006 presidential candidate, and of José Serra(@joseserra_), PSDB presidential candidate in 2002 and 2010, appear on the other side of the latent space. Senators Aloysio Nunes (PSDB - @Aloysio_Nunes) and

Ronaldo Caiado (DEM - @SenadorCaiado), important political figures in the opposition, also appear on the right side of the spectrum. The official Twitter profile of the PMDB Senator Eunício Oliveira (@Eunicio), and of the Vice-President, Michel Temer (@MichelTemer), also from the PMDB, have estimated scores near zero, signaling that they position themselves in intermediate positions. Sérgio Cabral, PMDB's former governor of Rio de Janeiro, who was one of the politicians supporting the PT-PMDB coalition, is positioned to the left of his party colleagues. Geddel Vieira Lima (@geddel_), another PMDB politician who opposed the PT-PMDB coalition, had an estimate that positioned him to the right of Michel Temer and Eunício Oliveira. We can also observe in the graph that Marina Silva (@silva_marina), the presidential candidate positioned in 3rd place in 2014, appears between her competitors Dilma Rousseff and Aécio Neves. Finally, Gilberto Kassab (@gilbertokassab_), former mayor of São Paulo and the founder of PSD, a party described as centrist and non-ideological, has an estimate around the center of the axis.

Figure 06. Estimated points for selected political Actors
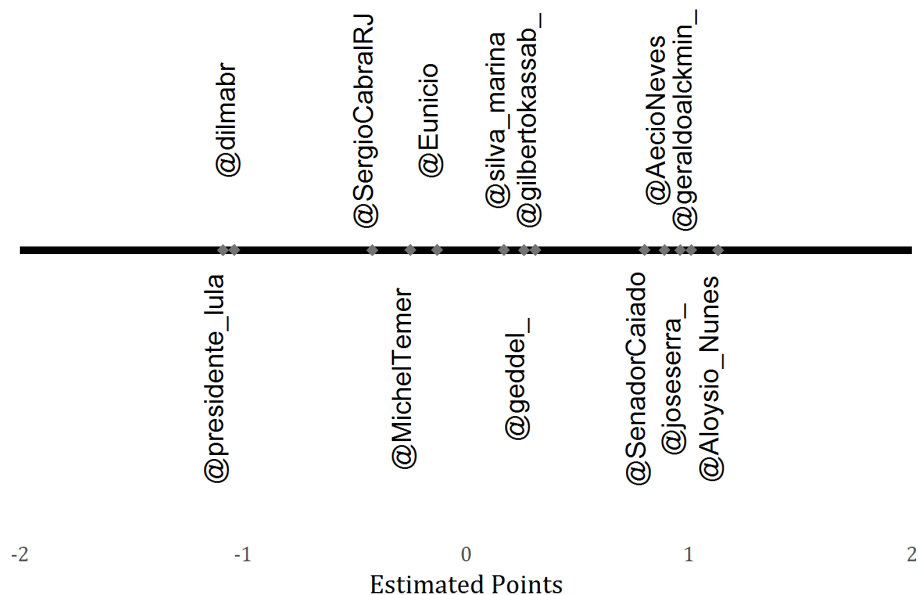
Source: Database - Twitter ideal points. See www.bpsr.org.br/files/archives/Dataset_Souza_Graça_Silva.

It is important to observe that some of these profiles, like those of President Dilma and of the opposition presidential candidate, Aécio Neves, were part of the list of Actors that had starting values assigned to −1 and +1. This, however, is only one way to

treat the statistical identifiability of the model, creating for us expectations about what the positive and negative results should mean. The values of each of the profiles used as priority information are estimated disregarding this initial information.

One of the most interesting aspects of the method developed by Barberá (2015) and the use of social media data in the estimation of positions on a latent space is that we can apply it to a sample of Twitter profiles that extrapolate personalities occupying political offices. The fact that it does not rely on roll-call votes or campaign contribution data allows us to estimate positions for a wide range of actors who participate in public political discourse. For instance, we can apply this methodology to estimate positions - on this latent scale - of political journalists that cover and comment on the news. Specialization in the media has brought about several journalists that cover national politics for large media outlets like television channels, newspapers, magazines and websites. A few of them are well known for their favorable or unfavorable positions towards the Workers' Party government, while others seek to maintain a more balanced profile. A group of them were followed by at least 30 federal deputies and became part of our group of Actors. We have selected 11 journalists from different media outlets and plotted the estimates made from their Twitter profiles in Figure 07.

In the 11 Twitter profiles of journalists presented in Figure 07 we can see that three are positioned on the left side of the spectrum, all of whom are journalists running blogs that either present a more pro-government editorial approach (@luisnassif and @pmoreiraleite) or a more left-oriented view (@blogdosakamoto). On the right side of the figure we observe columnists who write in Veja magazine, which has built an anti-PT reputation. Both Diogo Mainardi (@diogomainardi) and Reinaldo Azevedo (@reinaldoazevedo) are journalists that voice strong disagreement with PT preferences. Rachel Sherazade (@RachelSherazade) is a TV and radio anchor famous for her criticisms towards the Workers' party-led government. Close to them is the profile of Arnaldo Jabor (@Arnaldo_Jabor), who comments on politics on the night news programs of Globo television, and who has also expressed critical viewpoints about the federal government during the PT years. Closer to the middle of the figure we see the other four journalists' Twitter accounts, all of them with a more informational than analytical approach. It is interesting to note that the criticism directed at some of these journalists, in the sense that they have pro-opposition or

pro-government tendencies, seems to be confirmed by their Twitter 'follow' structure and their scores.

Figure 07. Estimated points for selected journalists
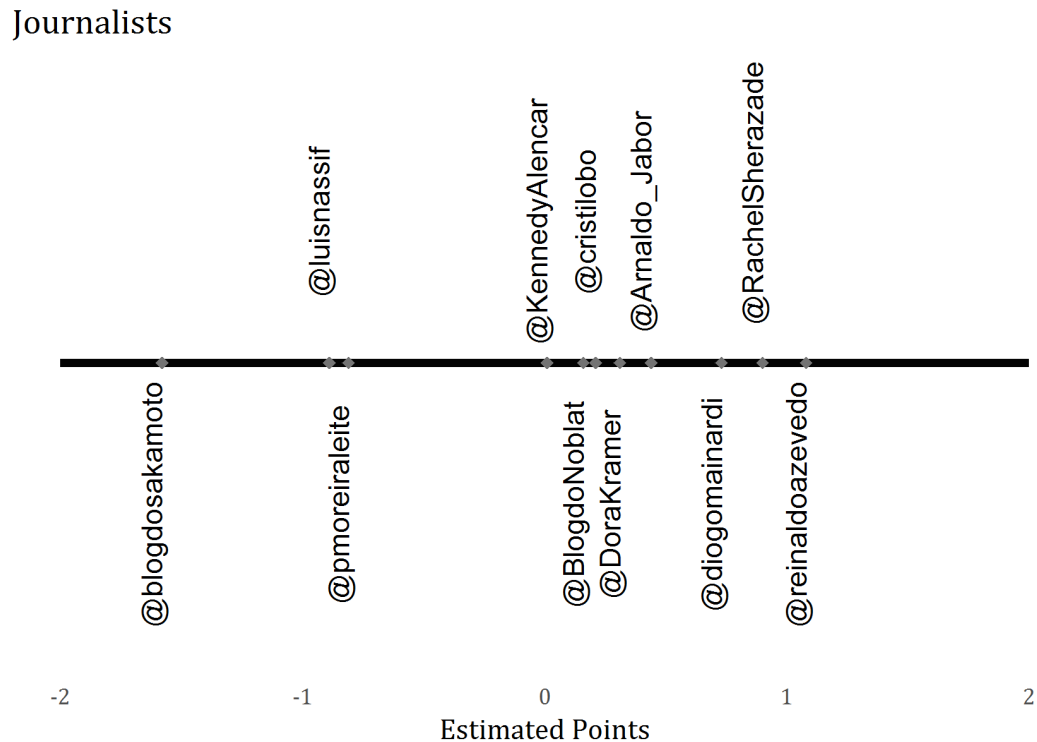
Source: Database -Twitter ideal points. See www.bpsr.org.br/files/archives/Dataset_Souza_Graça_Silva.

Concluding remarks on the proposed method to estimate ideological positions

This study was concerned with applying and discussing the method proposed by Barberá (2015) for a database of Twitter profiles comprising Brazilian federal deputies and Actors with a large number of followers amongst themselves. The main advantage of this method is that it would allow researchers to estimate ideal points in a one-dimensional latent space with data about relations between 'followed' Twitter profiles. In this way, it sought to extend common estimations of the ideology of political elites, making it applicable to any Twitter profile. We were interested in testing if this method could be applied to a multiparty democracy like Brazil, and if it would produce intelligible results.

Applied to a sample of accounts whose political positions we already have information about, such as those of federal deputies, the method has produced robust results, comparable at the individual legislator level to estimates produced from roll-call votes. Party level estimates did not perform as well, but they presented an

understandable positioning of parties, comparable to estimates produced via roll-call or legislator surveys. After presenting the robustness of the method to estimate ideal points for the federal deputies in Brazil, we looked at other actors - such as politicians who do not hold seats in the Chamber of Deputies and even journalists that cover the political scene in Brazil. The estimates corroborate expected variations in estimating distances between Actors who position themselves along different political poles, and showing proximity between those that sound similar in their political discourse.

One of the main difficulties with this method, as with any method that estimates ideal positions in any latent space, is how to interpret the distance between estimates. At first glance, it appears that there is an overlap between the latent space and the proximity with the parties that have constructed themselves as the main contenders for the presidential race. Profiles connected with the PT have consistent negative estimated positions, while profiles connected with the PSDB have positive scores. This seems to indicate that we might be looking at a government-opposition dimension. The fact that profiles connected to PSOL, a party that positions itself as a left opposition to the PT government, all have negative scores raises the question as to whether this latent space might also represent a left-right political dimension. This apparent confusion is somewhat diminished if we recall that the data used consists of connections in social media platforms. With this in mind we can argue that the impression that we might be looking, at some level, at a government-opposition axis is due to the political rivalry between the PT and the PSDB, which hinders the connections between politicians and users identifying themselves with any of these parties. Despite the distance shown between these parties, the method is prolific in showing that federal deputies and relevant actors are immersed in very different Internet social network patterns, which comprise a plethora of political discourses.

References

AMES, Barry (2001), *The deadlock of democracy in Brazil* (interests, identities, and institutions in comparative politics). Michigan: University of Michigan Press. 352 pp..

BARBERÁ, Pablo (2015), Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data. *Political Analysis*. Vol. 23, Nº 01, pp. 76-91.

BEAUCHAMP, Nicholas (2016), Predicting and interpolating state-level polls using twitter textual data. *American Journal of Political Science*. Vol. 53, Nº 01, pp. 107-121.

BOND, Robert and MESSING, Solomon (2015), Quantifying social media's political space: estimating ideology from publicly revealed preferences on Facebook. *American Political Science Review*. Vol. 109, Nº 01, pp. 62–78.

BONICA, Adam (2013), Ideology and interests in the political marketplace. *American Journal of Political Science.* Vol. 57, Nº 02, pp. 294-311.

BOUTET, Antoine; KIM, Hyoungshick, and YONEKI, Eiko (2015), What's in twitter: I know what parties are popular and who you are supporting now! *Social Network Analysis and Mining*. Vol. 03. Nº 04, pp. 1379–1391.

CERON, Andrea and CREMONESI, Alessandra Caterina (2013), Politicians go social. Estimating intra-party heterogeneity (and its effects) through the analysis of social media. *Paper* presented at NYU La Pietra Dialogues on Social Media and Political Participation. Florence.

CLINTON, Joshua; JACKMAN, Simon, and RIVERS, Douglas (2004), The statistical analysis of roll call data. *American Political Science Review*. Vol. 98, Nº 02, pp. 355–370.

CONOVER Michael D.; GONÇALVES Bruno; RATKIEWICZ, Jacob; FLAMMINI, Alessandro, and MENCZER Filippo (2011), Predicting the political alignment of twitter users. *Conference* presented at IEEE Third International Conference on Social Computing (SocialCom). Boston.

ECKER, Alejandro (2017), Estimating policy positions using social network data: cross-validating position estimates of political parties and individual legislators in the polish parliament. *Social Science Computer Review*. Vol. 35, Nº 01, pp. 53-67.

FARRELL, Henry (2012), The consequences of the internet for politics. *Annual Review of Political Science.* Vol. 15, Nº 01, pp. 35–52.

FIGUEIREDO, Argelina Cheibub and LIMONGI, Fernando (1995), *Executivo e legislativo na nova ordem constitucional*. Rio de Janeiro: Editora FGV. 232 pp..

GALLAGHER, Michael (2015), Election indices dataset. Available at *<https://www.tcd.ie/Political_Science/staff/michael_gallagher/ElSystems/Docts/ElectionIndices.pdf>*. Accessed on December 16, 2015.

GAYO-AVELLO, Daniel (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. *Available at <https://arxiv.org/abs/1204.6441v1>*. Accessed on February 27, 2015.

GROVES, Robert M. (2011), Three eras of survey research. *Public Opinion Quartely*. Vol.

75, Nº 05, pp. 861–871.

HUG, Simon (2010), Selection effects in roll call votes. *British Journal of Political Science*. Vol. 40, Nº 01, pp. 225–235.

IMAI, Kosuke; LO, James, and OLMSTED, Jonathan (2016), Fast estimation of ideal points with massive data. *American Political Science Association*. Vol. 110, Nº 04, pp. 631-656.

JACKMAN, Simon (2001), Multidimensional analysis of roll call data via Bayesian simulation: identification, estimation, inference, and model checking. *Political Analysis*. Vol. 09, Nº 03, pp. 227–241.

JUNGHERR, Andreas; JÜRGENS, Pascal, and SCHOEN, Harald (2012), Why the pirate party won the German election of 2009 or the trouble with predictions: a response to Tumasjan, A., Sprenger, T. O., Sander, P. G., Welpe, I. M., "Predicting elections with twitter: what 140 characters reveal about political sentiment". *Social Science Computer Review*. Vol. 30, Nº 02, pp. 229–234.

KING, Aaron Scott; ORLANDO, Framk J., and SPARKS, David B. (2011), Ideological extremity and primary success: a social network approach. *Paper* presented at Annual Meeting of the Midwest Political Science Association. Chicago.

LAMPOS, Vasileios (2012), On voting intentions inference from twitter content: a case study on UK 2010 general election. Technical report. Available at *<http://arxiv.org/pdf/1204.0423v11>*. Acessed on February 27, 2015.

POOLE, Keith T. and ROSENTHAL, Howard(2001), D-Nominate after 10 years: a comparative update to congress: a political-economic history of roll-call voting. *Legislative Studies Quarterly*. Vol. 26, Nº 01, pp. 05–29.

POOLE, Keith T. and ROSENTHAL, Howard (1997), *Congress*: a political-economic history of roll call voting. Oxford: Oxford University Press. 320 pp..

POWER, Thimothy J. and ZUCCO JR., Cesar (2012), Elite preferences in a consolidating democracy: the brazilian legislative surveys, 1990-2009. *Latin American Politics and Society*. Vol. 54, Nº 04, pp. 01–27.

RUEDIGER, Marco Aurélio (ed)(2016), Eleições 2014- comportamentos, valores e expectativas do eleitorado brasileiro. Vol. 05. Rio de Janeiro: FGV-DAPP. *No prelo*.

SANTOS and CANELLO, (2015), Pontos ideais estimados - (2011-2014). NECON -Núcleo de estudos do congresso. Available at *<URL: http://necon.iesp.uerj.br/ >*. Accessed on December 02, 2015.

SHUGART, Matthew Soberg and CAREY, John M.(1992), *Presidents and assemblies*: constitutional design and electoral dynamics. Cambridge: Cambridge University Press. 332 pp..

STEPAN, Alfred (2000), Brazil's decentralized federalism: bringing government closer to the citizens? *Daedalus*. Vol. 129, Nº 02, pp. 145–169.

TAROUCO, Gabriela da Silva(2011), Brazilian parties according to their manifestos: political identity and programmatic emphases. *Brazilian Political Science Review*. Vol. 05, Nº 01, pp. 54-76.

TAROUCO, Gabriela da Silva and MADEIRA, Rafael Machado (2013), Partidos, programas e o debate sobre esquerda e direita no Brasil. *Revista de Sociologia e Política*. Vol. 21, Nº 45, pp. 149-165.

TUMASJAN, Andranik; SPRENGER, Timm O.; SANDNER, Philipp G., and WELPE, Isabell M. (2011), Election forecasts with twitter: how 140 characters reflect the political landscape. *Social Science Computer Review*. Vol. 29, Nº 04, pp. 402–418.

ZUCCO JR, Cesar(2009), Ideology or what? Legislative behavior in multiparty presidential settings. *The Journal of Politics*. Vol. 71, Nº 03, pp. 1076–1092.

ZUCCO JR., Cesar and LAUDERDALE, Benjamin E. (2011), Distinguishing between influences on Brazilian legislative behavior. *Legislative Studies Quarterly*. Vol. 36, Nº 03, pp. 363–396.