Recovering the French Party Space from Twitter Data

François Briatte, Ewen Gallic

Recovering the French Party Space from Twitter Data

François Briatte[*] and Ewen Gallic[†]

May 19, 2015

Abstract

This study explores the possibility to retrieve information on partisan polarization from data generated by online social media users. The specific application that we pursue consists in placing a sample of over 1,000 French politicians on a unidimensional left-right scale by using their followers on Twitter as a proxy for their relative ideological positions. The methodology that we use to that end closely replicates that of Barberá (2015), who developed a Bayesian Spatial Following model to retrieve such ideal point estimates in the United States and in five European countries. Our results concur with existing measures of the French party space, and yield additional insights into the behaviour of ideologically extreme social media users.

Keywords Twitter, France, polarization, ideal point estimation

[*]PhD candidate, IEP Grenoble / ESPOL Lille; francois.briatte@univ-catholille.fr.

[†]PhD candidate, University of Rennes 1; ewen.gallic@univ-rennes1.fr.

Introduction

The ideological placement of political elites and ordinary citizens is central to many branches of political science, which have developed a large array of methods to estimate the relative distance that separates political actors such as legislators (Clinton, Jackman and Rivers, 2004) or parties (Laver, Benoit and Garry, 2003). The measurement of elite and mass ideology, however, remains problematic, insofar as only few methods exist to estimate such positions on a joint ideological scale.

Several recent studies have suggested that this difficulty can be overcome by resorting to social media. Because these media "enable ordinary citizens to endorse and communicate with political figures and elites" (Bond and Messing, 2015, p. 64), the patterns of interaction that they generate might contain rich information about the ideological preferences of their users, under the assumption that individuals tend to interact with like-minded people.

In a highly original paper, Barberá (2015) offers to empirically test that assumption using Twitter data. Hypothesizing that Twitter users are selective in whom they choose to "follow" online, he successfully shows how the follower base of politicians in the United States and in five European countries can be leveraged to uncover a single latent dimension that effectively mirrors existing measures of left-right ideological placement of these politicians.

In this paper, we closely replicate the research steps described by Barberá (2015) in order to achieve similar results on a large sample of French politicians. We then use the estimated ideological positions of their Twitter followers to verify whether those situated at the ideological extremes are more active at spreading information than others, as Barberá and Rivero (2014) found for a similar sample of American and Spanish Twitter users.

Our results confirm the validity of Barberá (2015) method to measure ideologi-

cal positions from Twitter data, as his Bayesian Spatial Following model recovers ideal point estimates that effectively reconstitute the contemporary French party space. Similarly, we find compelling evidence that ideologically extreme Twitter users differ from non-extreme ones in terms of the kind and quantity of information that they diffuse to their own audiences.

1 Background

The Twitter micro-blogging service is among the most used social networking platforms worldwide, with 288 million claimed monthly active users.[1] While the exact size of Twitter's audience in any country is difficult to determine precisely, an estimated 6.6 million of French residents visited Twitter in January 2015, at a time where many people turned to the Internet to follow the news on the terrorist attacks that struck the country (Médiamétrie, 2015).

In addition to its popularity among the general public, Twitter has also attracted many political parties, elected officials and candidates, who use it to communicate to each other and to larger audiences, such as party supporters, through short messages and links to additional media. In some contexts, politicians have also turned to Twitter in order to bypass the institutional constraints that limit their access to journalists (see e.g. Chibois, 2015).

The joint presence of political elites and ordinary citizens on social media such as Twitter creates interesting possibilities for the measurement of ideological positions. In similar fashion to campaign contributions (Bonica, 2014; Bonica and Woodruff, 2014), social media accounts exist for all sorts of politicians, regardless of their mandate or incumbent status. Estimates of ideological positions derived from these data thus have the potential to extend across and beyond political institutions.

[1] See https://about.twitter.com/company (last accessed 7 May 2015).

A further feature of social media data is that they "span the divide between politicians and individuals" (Bond and Messing, 2015, p. 64), which opens the possibility to estimate ideological preferences for both types of actors on a single scale. In this view, social media accounts operate as "bridges" between elite and mass positions, and might even serve as bridges between actors from different countries (Barberá, 2015, p. 83 ; Barberá, Popa and Schmitt, 2015).

The bridging capabilities of Twitter accounts take several forms, among which the strongest is the decision to "follow" a given account, in order to become exposed to all its future messages. If such decisions are influenced by ideological preferences, then it becomes possible to assume that "Twitter users [will] prefer to follow politicians whose positions on the latent ideological dimension are similar to theirs" (Barberá, 2015, pp. 77–78).

This key assumption is similar to the proximity hypothesis of spatial voting (Enelow and Hinich, 1984), which Bonica (2014, p. 369) extends to campaign contributions, and which Bond and Messing (2015) extend to Facebook "likes". What we offer in the rest of this study can be taken as an additional test of Barberá (2015) hypothesis that ideological proximity also governs over decisions to follow other users on Twitter.

Our empirical test case is France, a country governed by a multiparty system for which the existing literature does not currently provide ideal point estimates of politicians (see, however, Rosenthal and Voeten, 2004 for such measures under the Fourth Republic, and Godbout and Foucault, 2013 for estimates of party voting unity in the Fifth Republic up to 2012).

2 Data

2.1 Sample of politicians on Twitter

This study relies on a sample of 1,008 French politicians who have opened a Twitter account under their own name. The primary data source was the *Élus 2.0* website by SARL Ideose[2], from which we retrieved 788 of these accounts. As we collected the followers of these accounts, we identified 220 additional politicians holding Twitter accounts, either by inspecting the verified accounts[3] that included a mandate keyword in their description, or by adding a few accounts manually.

Slightly over one out of four of these politicians are women, which comes very close to the percentage of women in the current lower chamber of the French parliament (27%). Table 1 shows the number of male and female politicians by party affiliation and the relative weight of each party in the full sample.

Party	Accounts	Males	Females	Sample weight (%)
PS	322	222	100	31.9
UMP	272	216	56	27.0
UDI	87	66	21	8.6
FN	81	54	27	8.0
IND	54	40	14	5.4
EELV	51	34	17	5.1
PRG	49	39	10	4.9
MODEM	43	32	11	4.3
FDG	23	18	5	2.3
DVD	14	12	2	1.4
DVG	12	10	2	1.2

Table 1: Politicians sample, by party affiliation.

[2] See http://elus20.fr/elus-web-facebook-twitter/ (last accessed 7 May 2015).

[3] As Barberá and Rivero (2014) explain, "Twitter grants verification to public figures, including journalists and media outlets, in order to certify that their profile corresponds to their real identity."

Our sample of politicians is made for almost 60% of members affiliated with either the *Parti Socialiste* (PS) or with the *Union pour un Mouvement Populaire* (UMP). The next parties by order of their relative weight in the sample are the largest of the two centrist parties (UDI), the far-right *Front National* (FN), and the Greens (EELV). All parties are represented by over 20 politicians, except for the "mixed-left" (DVG) and "mixed-right" (DVD) groups, which host several smaller parties.

An important feature of this sample is that it includes politicians holding all sorts of electoral mandates: for instance, it contains over half of all members of the lower house of the French parliament, and a fourth of all members of the upper house. It also contains 60 Members of the European Parliament, which represents 80% of all French seats in that chamber. Finally, roughly half of the sample is made of politicians who are neither national or European parliamentarians, and who either hold mandates at the local level, such as councilors or mayors, or hold non-electoral mandates, such as ministers or party spokespersons.

Our sample contains 54 politicians with no partisan affiliation (IND), who are likely to be ideologically heterogeneous and who we thus discarded from our analysis. Similarly, in order to ensure that all politicians in our sample were still in involved in politics, we also discarded those for whom we could not clearly identify at least one ongoing mandate.

2.2 Sample of Twitter followers

Over 3.16 million Twitter users followed at least one member of our sample of politicians as of late March 2015, at a time where local elections were being held. A large amount of them, however, are inactive accounts, or even fake accounts created for the purpose of amplifying the visible audience of their targets.

For that reason, we imitated Barberá (2015, p. 81) by subsetting each politician's followers to those who 1) had sent at least 100 tweets, 2) had sent at least one tweet in the last six months, 3) had at least 25 followers, 4) were located in France,

and 5) followed at least four different politicians, which was the average number of politicians followed among all users. This approach left us with a final sample size of 84,279 Twitter users among the larger follower base of French politicians.

To assess the representativeness of that sample, we further analysed the contents of the "name" and "location" fields of their user profiles, and identified some biases that are consistent with those reported in other studies of Twitter users, such as Mislove et al. (2011) or Barberá and Rivero (2014): overall, we found the population of that sample most likely to be male, located in urban areas, and possibly younger than the general population (see Appendix A for details).

3 Methods

3.1 Exploratory approach

In this study, the data under examination are made of several thousands of Twitter users following the Twitter accounts of individual French politicians, which we might denote as the adjacency matrix $M(i, j)$, where $M_{ij} = 1$ if the Twitter user i follows politician j, and 0 otherwise. In order to offer some visual representation of that matrix, Figure 1 shows a force-directed network of its politicians, coloured by party affiliation and connected by a tie when either one of them shares over half of his or her followers with the other one.

This representation of the data conveys some useful insights, such as the high level of partisan homophily, insofar as two members of the same party are more likely to share a large proportion of their Twitter followers than two randomly selected politicians.[4] Figure 1 also suggests that the primary dimension contained in the follower patterns of French politicians reflects the left-right divide

[4] That probability can be estimated through an exponential random graph model that takes the likelihood of a tie to exist in the network as its dependent variable. See Appendix B for an estimation of that likelihood.

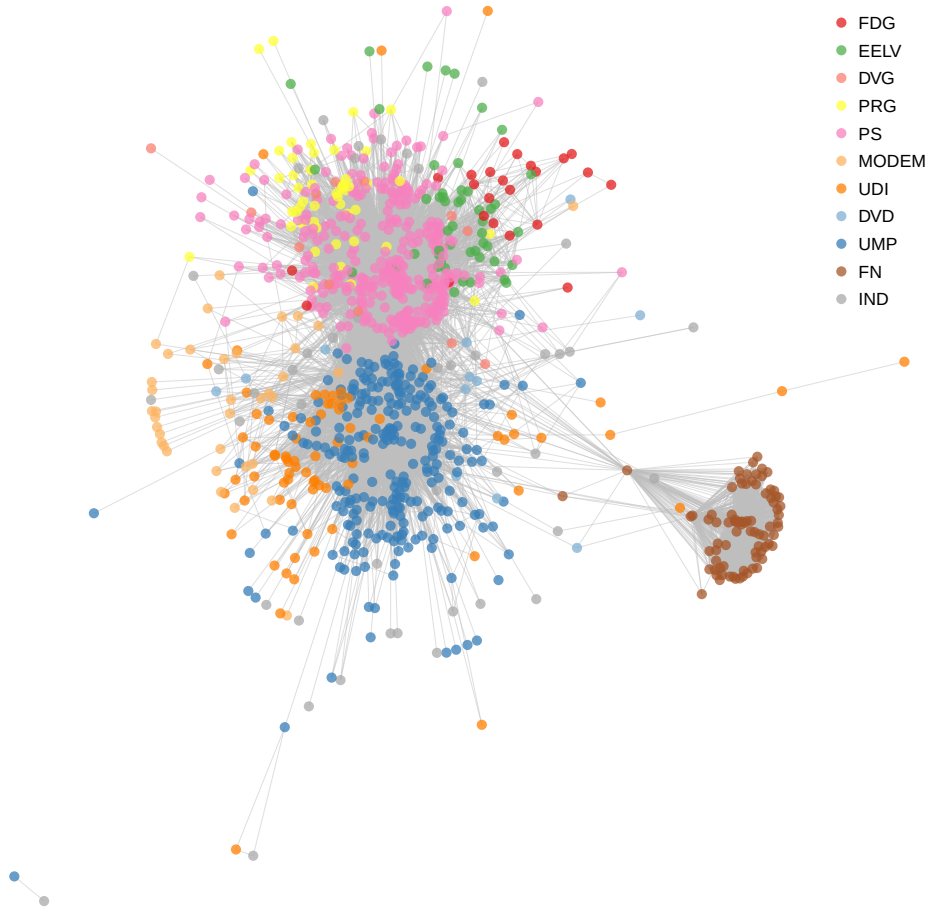

Figure 1: Network of shared Twitter followers among French politicians. The direction and weight of the ties are not shown to avoid overplotting.

of French politics, with left-wing politicians forming a distinct part of the network, right-wing politicians forming another one, and far-right politicians being more isolated from all others.

3.2 Spatial model

In order to uncover that primary latent dimension, we used the Bayesian Spatial Following model described in Barberá (2015, pp. 78–80), which includes random effects by holding the popularity of each politician and the political interest of their followers constant at the user-level. Letting $y_{ij} = 1$ be the probability that a Twitter user i independently chooses to follow a politician j, that probability can be modelled as

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

where α_j controls for the overall propensity of politician j to attract Twitter followers (equivalent to his or her popularity), β_i controls for the propensity of user i to follow politicians (a proxy for his or her political interest), γ is a normalizing constant, and $||\theta_i - \phi_j||^2$ is the squared Euclidean distance in the unidimensional latent space between follower i and politician j, $\theta_i \in \mathbb{R}$ being the ideal point of Twitter follower i and $\phi_j \in \mathbb{R}$ being the ideal point of politician j (Barberá, 2015, p. 79).

To implement that model, we made direct use of the replication material made available by Barbera (2014), using identical steps and settings on all aspects of the estimation strategy, including a similar set of starting values to identify the politicians' ideal points, and similar model diagnostics (see Appendix C for details).

4 Results

Working paper note: this section reports only on *intermediary* results that we obtained in April, after running the first stage of the model on a matrix of 418 politicians × 17,934 followers. The full results on the complete sample are still computing.

4.1 Results for politicians

Based on the ideal points that we obtained for each politician, Figure 2 shows the mean ideal point of each political party represented in our sample, compared to its Left/Right score in the ParlGov database (Döring and Manow, 2014), which is a time-invariant score computed as the weighted mean values of party positions taken from several expert surveys on political parties.[5]

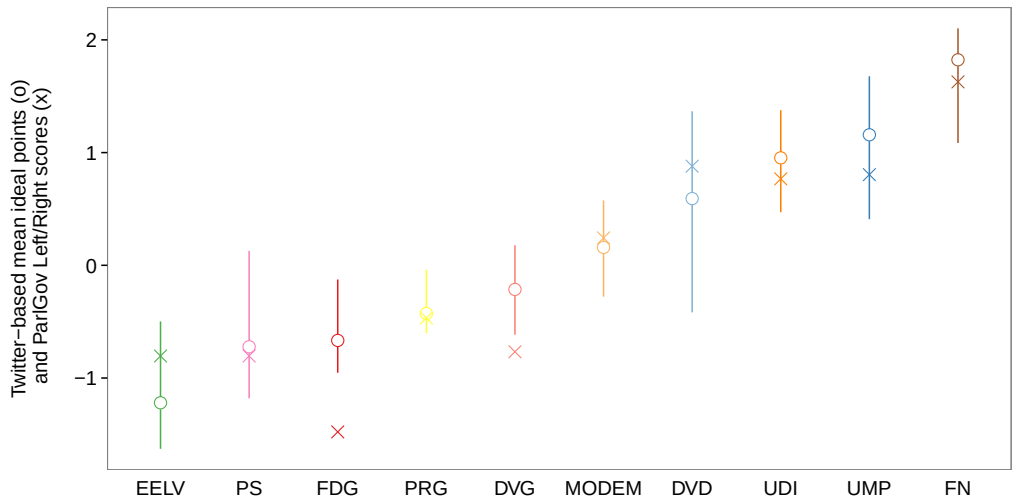

Figure 2: Comparison of party-level positions. Twitter-based ideal points are shown with their range and are normalized, like ParlGov scores, to a mean of 0 and standard deviation of 1.

Like the party scores from the expert surveys compiled by the ParlGov database,

[5] See http://www.parlgov.org/documentation/party-positions/ (last accessed 7 May 2015).

the Twitter-based ideal points of our politicians sample reflect the left-right divide of French politics. Furthermore, at the exception of the positions for the Communist (FDG) and mixed-party groups (DVG and DVD), both measurements are highly similar in their ordering of French political parties from left to right.[6]

In order to further illustrate the validity of the politicians' ideal points, Figure 3 shows the distribution of the estimated ideal points for the politicians of the two largest left-wing and right-wing parties, as well as for a restricted set of key political actors that were estimated as part of the followers sample. In this figure, the ideal points of official party accounts have face validity, as well as those of a left-wing civil society organization (@SOS_Racisme), of the currently left-wing presidency (@Elysee), and of the youth organization of the extreme-right party (@FNJ_officiel).

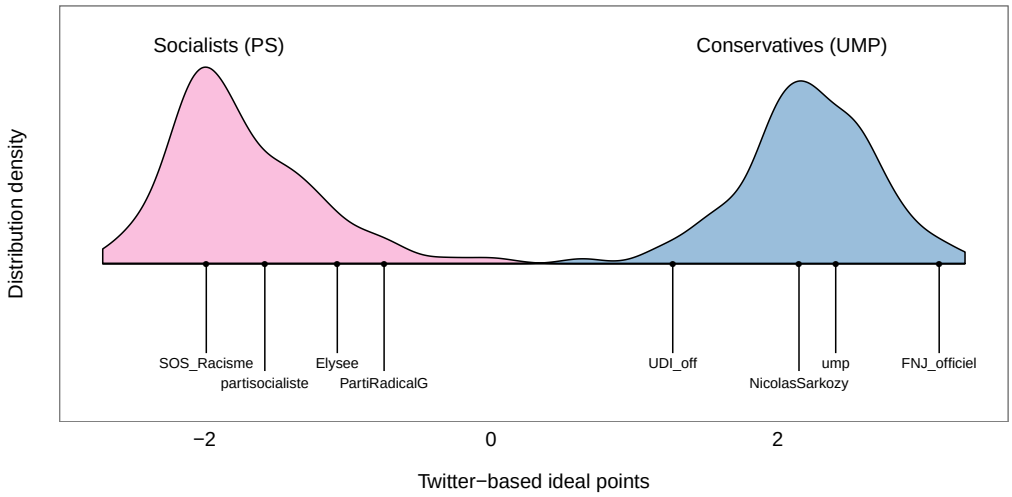

Figure 3: Ideal points for Socialist and Conservative politicians (top), compared to those of key political actors included in the followers sample (bottom).

[6] Possible explanations for these discrepancies are the low number of Communist politicians who hold Twitter accounts, and the heterogeneous nature of the mixed-left and mixed-right party groups.

4.2 Results for followers

In order to assess the validity of ideal points for the entire sample of followers, Figure 4 replicates Figure 4 in Barberá (2015: 85). In this figure, the left panel "compares the distribution of ideal points for the two types of Twitter users in the sample—political actors and ordinary citizens" (Barberá, 2015: 83), whereas the right panel compares the ideal points of individuals who included a left-wing or right-wing partisan hashtag in their profiles (such as "#PS" or "#MODEM") to all others.

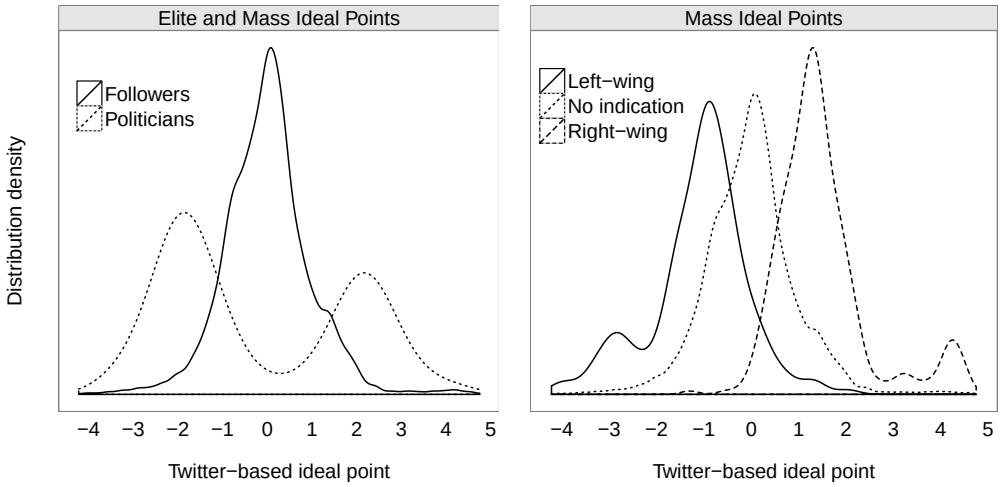

Figure 4: Ideal points for Socialist and Conservative politicians (top), compared to those of key political actors included in the followers sample (bottom).

The results of the left panel are highly consistent with those of Barberá (2015). The unimodal distribution of ideal points for ordinary citizens and the bimodal distribution of ideal points for politicians confirm that "political actors are more polarized than mass voters" (Barberá, 2015, p. 84), as do the longer tails on each end of the distribution of mass ideal points.

The results of the right panel are also consistent with those of Barberá (2015), insofar as the Twitter followers who self-identified with a political party in their profile descriptions lean either towards the left or the right, while those with

no partisan hashtag are centered to the middle of our ideological scale. Both the left-wing and right-wing user groups also clearly display the respective positions of far-left and far-right users, with greater distance on the far-right than on the far-left.

4.3 Political mentions

Finally, we take inspiration from Barberá and Rivero (2014) and analyze the activity of our sample of Twitter followers as a function of their estimated ideal points. To do so, we use 2.8 million tweets sent by over 95% of these users in the first 16 weeks of year 2015, a period that covers the aftermath of the terrorist attacks of January 2015 and both rounds of the local elections held in March 2015.

Table 2 reports the results of ordinary least squares regressions that successively take the number of messages ("tweets") sent by the users, the number of followers that they have attracted, the number of messages that they have forwarded to their followers ("retweets"), and the number of times that they have mentioned any of the politicians' accounts included in our sample in their tweets.

In each regression, we control for the gender and seniority of the user, with the latter measured as the number of days since the creation of his or her Twitter account. Finally, we introduce an ideological variable to identify users situated at the extremes of the ideological scale, and set the baseline category to that variable to users situated at the centre.[7]

The results show that, on average, Twitter followers situated at the extremes of the ideological spectrum uncovered by the model publish more tweets, especially for those at the far-right of that spectrum. The same users also attract more followers, retweet more information to them, and mention far more politicians

[7] Since the ideological scale produced by the model has a mean of 0 and a standard deviation of 1, we categorized as "far-left" or "far-right" all users whose ideal point was situated outside of the $(-1, +1)$ interval.

	Tweets	Followers	Retweets	Mentions
Intercept	1.84^{***}	1.46^{***}	-0.13^{***}	-0.40^{***}
	(0.02)	(0.02)	(0.02)	(0.02)
Far-left	0.13^{***}	0.09^{***}	0.04^{**}	0.23^{***}
	(0.02)	(0.02)	(0.01)	(0.02)
Far-right	0.05^{*}	0.06^{***}	0.05^{***}	0.41^{***}
	(0.02)	(0.02)	(0.01)	(0.02)
Account age	-0.00	0.00^{***}	-0.00^{***}	-0.00^{***}
	(0.00)	(0.00)	(0.00)	(0.00)
Male user	0.02	-0.05^{***}	-0.09^{***}	0.05^{***}
	(0.02)	(0.01)	(0.01)	(0.01)
Tweets (log-10)		0.37^{***}	0.91^{***}	0.61^{***}
		(0.01)	(0.01)	(0.01)
Num. obs.	8112	8112	8112	8112
RMSE	0.63	0.51	0.41	0.47

$^{***}p < 0.001, {}^{**}p < 0.01, {}^{*}p < 0.05$

Table 2: Regression estimates of Twitter activity by ideological status, with standard errors in brackets. All dependent variables are logged at base 10.

in their tweets.

These findings are fully consistent with those of Barberá and Rivero (2014), who also find, for instance, that women are slightly more active at retweeting information to their followers. Our results also underline the relationship between ideology and political activity among Twitter users, with ideologically extreme users being more prone to interact with politicians by mentioning them in their messages.

In order to illustrate this last observation, Figure 5 shows the average proportion of each political party within the mentions of nine groups of Twitter users, with the group of each follower i defined as the standardized value $\theta'_i = \left\lfloor \frac{\theta_i}{\sigma_\theta} \right\rfloor$, where θ_i is the Twitter-based ideal point of the follower and σ_θ the standard deviation of θ.

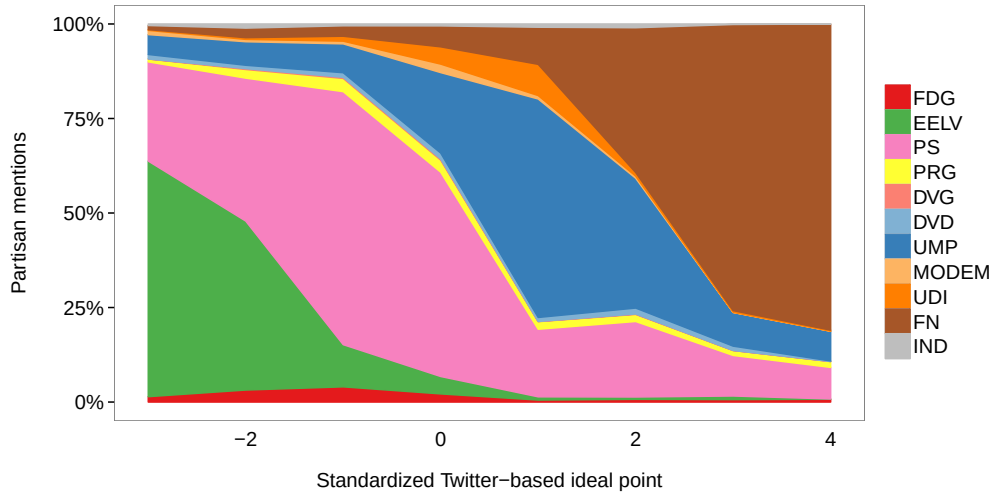

Figure 5: Proportion of partisan mentions according to the standardized ideal point of Twitter followers, based on the party affiliation of each politician mentioned.

The observed substitutions in partisan mentions show how Twitter followers estimated to the left of our ideological scale refer primarily to the Green (EELV) and Socialist (PS) Parties, while those estimated to the right refer primarily to the Conservatives (UMP), or for those situated at the most extreme end of the

scale at values of $\theta'_i > 3$, to the extreme-right party (FN). From that perspective, the estimated ideal points of our sample of Twitter users are matched by the content of their political messages.

Conclusion

This study provides a replication of Barberá (2015) methodology to estimate Bayesian ideal points from Twitter data. The results on our sample of French politicians and several thousands of their followers confirm the validity of that approach, which remains methodologically challenging due to the large amounts of data that it requires and to the computational costs of the estimation procedure.

Furthermore, the additional results that we provide with regards to the behaviour of ideologically extreme Twitter users match those of Barberá and Rivero (2014), insofar as these users are more actively involved in spreading information online than their non-extreme counterparts, as well as in interacting with politicians themselves.

These last findings lend support to the idea that social media are likely to add to political polarization by operating as an "echo chamber" of party politics (see Barberá, Popa and Schmitt, 2015). In the specific context of France, they also suggest that citizens holding extreme-right political views are particularly active on such media, in line with the common (mis)perception that they are being ostracized by "mainstream" outlets.

* * *

The replication code for this study is available at https://github.com/briatte/elus, and the raw data are available from the authors upon request.

References

Barbera, Pablo. 2014. "Replication data for: Birds of the Same Feather Tweet Together. Bayesian Ideal Point Estimation Using Twitter Data.".

Barberá, Pablo. 2015. "Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data." *Political Analysis* 23(1):76–91.

Barberá, Pablo and Gonzalo Rivero. 2014. "Understanding the political representativeness of Twitter users." *Social Science Computer Review* .

Barberá, Pablo, Sebastian Adrian Popa and Hermann Schmitt. 2015. "Prospects of Ideological Realignment(s) in the 2014 EP elections? Analyzing the Common Multidimensional Political Space for Voters, Parties, and Legislators in Europe." Working Paper. Available at http://pablobarbera.com/static/eu-dimensionality.pdf.

Bond, Robert and Solomon Messing. 2015. "Quantifying Social Media's Political Space: Estimating Ideology from Publicly Revealed Preferences on Facebook." *American Political Science Review* 109(01):62–78.

Bonica, Adam. 2014. "Mapping the ideological marketplace." *American Journal of Political Science* 58(2):367–386.

Bonica, Adam and Michael J Woodruff. 2014. "A Common-Space Measure of State Supreme Court Ideology." *Journal of Law, Economics, and Organization* p. ewu016.

Chibois, Jonathan. 2015. "Twitter et les relations de séduction entre députés et journalistes." *Réseaux* 188(6):201–228.

Clinton, Joshua, Simon Jackman and Douglas Rivers. 2004. "The statistical analysis of roll call data." *American Political Science Review* 98(02):355–370.

Döring, H and P Manow. 2014. "Parliaments and governments database (ParlGov): Information on parties, elections and cabinets in modern democracies." Working Paper. Stable version 14-12. Available at http://www.parlgov.org/.

Enelow, James M and Melvin J Hinich. 1984. *The spatial theory of voting: An introduction*. CUP Archive.

Godbout, Jean-François and Martial Foucault. 2013. "French legislative voting in the Fifth Republic." *French Politics* 11(4):307–331.

Laver, Michael, Kenneth Benoit and John Garry. 2003. "Extracting policy positions from political texts using words as data." *American Political Science Review* 97(02):311–331.

Médiamétrie. 2015. "L'audience de l'Internet en France en janvier 2015." Press Release. Available at http://www.mediametrie.fr/internet/communiques/.

Mislove, Alan, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela and J Niels Rosenquist. 2011. "Understanding the Demographics of Twitter Users." *ICWSM* 11:5th.

Rosenthal, Howard and Erik Voeten. 2004. "Analyzing roll calls with perfect spatial voting: France 1946–1958." *American Journal of Political Science* 48(3):620–632.