Political Astroturfing and Twitter Following in the 2011 Spanish Campaign

Michael J. Jensen
Department of Political Science
Autonomous University of Barcelona
MichaelJames.Jensen@uab.cat

Brian Kesser
IGMAS Technologies, Inc.
Woodland Hills, CA USA
briankesser@gmail.com

Political Astroturfing and Twitter Following in the 2011 Spanish Campaign

The 2011 electoral campaign was the first national election in Spain that saw widespread use of social media as a campaign tool. Though internet use in Spain lags behind its northern neighbours, it ranks ninth worldwide in terms of the number of Twitter accounts registered per person and third in terms of Twitter accounts that are active (Jensen and Anduiza 2012; Omicrono 2012). The frequency of candidate retweets, mentions, and follower growth have often been used as measures of a candidate's online support, some evidence going so far as to suggest that candidates with more Twitter followers are more likely to win an election (Kravets 2010). Over the course of the campaign for the November 20, 2011 elections in Spain, the leaders of both the main parties saw sizeable gains in their Twitter and Facebook followers. Mariano Rajoy, leader of the conservative Partido Popular (PP) boasted the largest following with nearly 120,000 Twitter followers at the time of the election -- roughly 40,000 more than Alfredo Rubalcaba, the leader of the centre-left Partido Socialista Obrero Español (PSOE) despite the fact that Rajoy,was a relative newbie, having first opened a Twitter account on July 27, 2011 and he kept his Tweets private until September 15th of that year. Although it is normal that political figures are considered relevant and therefore can expect to see their following grow over time, there are some indications that in Rajoy's case this does not stem from a majority or even plurality of support among those active on Twitter: data shows that Rubalcaba had a higher frequency of retweets and mentions – traditional metrics of online support – over the course of the electoral campaign than Rajoy (El País 2011). For this reason, Rajoy's relatively rapid growth in Twitter followers is somewhat puzzling. This paper examines his growth of followers using a combination of survey data and data mined the profiles of Rajoy's Twitter followers.

Within the short history of the internet, Twitter and other online social network platforms are relatively recent innovations. Though these platforms enable researchers to access large quantities of data, there is little theory building based on empirical work in the field. In part this may be a consequence of the fact that computer scientists rather than political scientists are best trained to extract and process the data, much of which relies on working knowledge of one or more programming languages. However, these researchers often lack the ability to integrate their observations within the operations of political life. Because the behavioural dimensions of following political candidates via social media implicate both practices of political behaviour and political communication as well as the technical properties of these media and various associated online economies, we examine the Twitter following using a combination of insights from both fields.

The dynamics of network construction and the implication of human as well as machinic "actors"[1] in the construction of follower networks have significant consequences for the wider study of communications and patterns of behaviour within social media, as well as political practice. Not only may there exist good theoretical reasons for distinguishing discrete biographical human actors from machines and assemblages of machines and humans, astroturf – i.e. the centrally controlled use diverse communication sources to create false impressions regarding the support for individuals, objects, and events (Ratkiewicz et al. 2011) – is not without consequences. To the extent social media communications become decoupled from representative human actors, the consequences may be significant distortions in beliefs and discourses (Ratkiewicz et al. 2010). For this reason, in this paper we concentrate on distinguishing Twitter accounts in terms of whether their behaviour is indicative of a correspondence between a single account and a single human actor or feigned individuality. The results suggest a combination of political factors as well as evidence of

1 Quotation marks are used here only to denote the contentious nature of the term actor as applied to nonhuman entities. We need not resolve this debate here.

astroturf may help account for differences in the growth of Twitter followers between the two candidates.

Theories of the Twitter Class

It is a matter of considerable debate the extent to which candidate mentions over social media sites can reliably predict electoral outcomes, in part a consequence of the lack of standardized procedures for sampling and measuring support (Gayo-Avello 2011; Jungherr, Jurgens, and Schoen 2011; Tumasjan et al. 2011). Candidate mentions implicate a complex field of semantic and semiotic relationships that may belie straightforward conclusions based on candidate mentions. Despite or because of Twitter's 140 character limit, tweets have carried a diverse and creative range of meanings that exceed the content contained within the character limit (Chadwick 2011). By contrast, the act of following a candidate carries more interpretive closure as it is linked to a relatively discrete set of propositions. We identify three potential meanings signified by following a candidate. First, following may indicate a level of support for candidates, at least within the relevant universe of Twitter users. Second, apart from electoral support for a candidate, a person may choose to follow a candidate on Twitter for the informational or entertainment value, or some other gratification or use. Finally, Twitter followers may not represent discrete biographical humans but instead be automated bots or purchased accounts. While there may be wider cultural or structural explanations for social media following, these analyses are focused on the individual level as our primary interest is in differentiating individual characteristics and behaviours that may shed light on distinguishing the overall presence of human and machine-controlled Twitter accounts. We will develop the theories behind each of these positions before turning to the data analysis.

Social Media Following as an Indicator of Electoral Support and Information Value

The question of whether social media followings correspond to the levels of electoral support has received the most attention in the literature. This is in part because if this connection can be drawn, it may supplement polling data as a measure of the flow of momentum in the course of an electoral campaign(Conover et al. 2011a; Jungherr, Jurgens, and Schoen 2011; Tumasjan et al. 2011). The prevailing social practices of network construction, software architecture of the Twitter website and assorted clients for using Twitter, and the interdependencies that obtain between the two may impact whether or not a particular user follows a candidate. On the one hand, Twitter like most online communication flows enables a great degree of control over one's information environment as following a campaign requires that one opt in in to that particular stream of communication.[2] On the other hand, the Twitter website and many of its clients recommend followers based on the profiles of those already in a user's current network (Golder et al. 2009). For those interested in artificially increasing their online profiles on Twitter may choose to purchase followers of a high profile figures in order to trick the Twitter algorithm into recommending these accounts to others. As a result, these accounts will initially have a large number of “friends” but few followers. Although users do not necessarily follow these recommendations, to the extent that they do, the software code that creates the interface with Twitter may help reinforce a user's overall patterns of following like-minded or diverse political actors.

Whether or not following a political campaign on Twitter is a veritable indication of candidate support is an open question at both the level of individuals and in the aggregate. The selective exposure debate raises significant questions as to whether there is a tendency towards homophily in networks. This debate is particularly intense regarding exposure to political information (Bennett and Iyengar 2008; Garrett 2009; Holbert, Garrett, and Gleason 2010; Sunstein

2 This is not to say that one has full control. Given that many social networks are created outside of political preference articulation, social media platforms such as Twitter may be a source of byproduct learning (Chadwick 2012).

2001). Though other evidence suggests that selective exposure does not always obtain as Twitter interactions transcend partisan divides (Conover et al. 2011a). To the extend these findings hold generally, it would undermine the linkage between following a candidate and support for that individual as it would indicate a tendency for persons to consume information independent of its opinion reinforcement value. However, an analysis of Spanish Twitter users over the course of the 2011 electoral campaign generally finds support for the selective exposure hypothesis as follower networks of those engaged about the election exhibit common ideological positions and similarities in policy positions (Barberá 2012). Switching the focus to followers of political campaigns, some limited evidence suggests that at the individual level persons tend to follow candidates that they support (Golbeck and Hansen 2011). Yet this does not necessarily translate into the proposition that candidates with more followers tend to have more electoral success as recent systematic research indicates there is no clear pattern between social media followers and electoral results (Vaccari and Nielsen 2012).

Turning to the question of operation of human motivations in the patterns of candidate following, we have two potential explanations that may help account for our observations during the Spanish 2011 campaign. The first explanation is that a person supports the campaign and is engaging in selective exposure. The electoral implications of that would mean that such a person was more likely to vote for the party of the candidate followed given the system of closed ballots and proportional representation in Spain. The alternative explanation is that there is an informational or other value associated with following a campaign that is exogenous to electoral intentions. Given the lack of a clear direction in the literature, we pose this as a research question: is the practice of Twitter following reasonably characterized as an expression of candidate support or is there another consummatory function that drives Twitter following?

Sentient analysis

We distinguish three conditions of human relationships to Twitter use. First, there is the account controlled by a unique, individual human being or a group of human beings in the case of an account corresponding to an organizational entity such as a media organization or an interest group. This is the conventional mode of Twitter use. Second, there are accounts that are semi-automated, involving a delegation of permissions to either a manchic, individual or organizational entity to post status updates in addition to the formal account operator. For instance, in the wake of the 2011 general election, both the PP and PSOE asked supporters to hand over their Twitter passwords to the party so that the parties could automatically relay mass messages (Barberá 2012). Alternatively, persons can link their Twitter accounts to Google alerts or automatic updates conditioned website updates or blog posts via Twitterfeed (Jantsch 2008). Finally, accounts may be created in bulk for purchase. Although creating bulk accounts for sale or using them to send automated and duplicate messages is a violation of the Twitter terms of service, accounts are commonly available for sale on Ebay and other websites and automated tweeting is not uncommon.[3]

There are two categories of of bots: follower bots that are relatively inactive and purchased to increase follower totals and tweet bots that are used to spread certain types of communications. Follower bots typically have few followers or statuses posted since that is not their function. The creation of follower bots is facilitated to the extent that one not even need be the verified user of the account or hand over a password as only the name of the user followed is required. In contrast, tweet bots often have larger number of statuses that are automatically updated and they are more likely to have large numbers of friends as a way to increase visibility relative to the number of accounts that follow them back.

The diffusion of astroturf in a medium like Twitter may have significant negative political consequences by distorting political communications within this medium. The use of bots or automated accounts detaches Twitter communications from persons. This allows certain persons to

3 See the Twitter terms of service: https://support.twitter.com/forums/26257/entries/18311#.

hijack the flow of communications or create a false illusion of support for individuals and ideas. This has implications for the flow of political support and the maintenance of legitimacy the authorities in a regime (Easton 1965, 278). Astroturfing has a propagandistic function that can be used by incumbent authorities to overwhelm the flow of critical communications about their policies, personalities, or prevailing ideologies within a regime. Alternatively, it can be used by challengers to undermine those objects through the transmission of critical communications. Previous research shows that false information is more likely to diffuse and that such information can become widespread seeded by a small number of bots (Ratkiewicz et al. 2011). Although the transmission of memes goes beyond the scope of this paper, artificially inflating the number of followers a candidate has can also create the illusion of legitimacy among certain segments of the population.

Each category of Twitter account has unique characteristics. First, as humans are embedded in varied networks and relationships with family, friends., co-workers and colleagues, they typically have more than a handful of Twitter followers. Likewise, they on average can be expected to have roughly comparable numbers of followers and friends – that is, accounts that they follow. A purchased bot need only to follow the account it is purchased to follow. These bots need not be active and may have few status updates as greater levels of activity often cost more to maintain. Tweet bots on the other hand often have a great number of friends and are fairly active as they regularly provide automated updates. Humans, in contrast to bots, have a greater need to customize their accounts: the presence of a unique avatar, a URL with a website or a blog link, information on the user's location, or a description of the user. Leaving a profile with its default conifugration would be more indicative of a bot for which many of these attributes are unnecessary. Additionally, the date on which an account was created can provide insights into whether accounts have been purchased or created for the purpose of inflating the number of accounts. Follower bots are typically created around the time they are purchased to follow an account. Finally, the uniqueness of the statuses of each account may shed some light as bots often tweet in bulk in an effort to artificially diffuse messages or links. Though these are all circumstantial conditions, the occurrence of one or more of these factors raises suspicion that an account does not correspond to a biographical human being. That is not to assign responsibility for astrotuf as both campaigns or other interested parties may be culpable as well as have independent reasons for engaging in these activities.

Data and Methods

We use two sources of data to study the Twitter followers of the candidates for the two leading parties in Spain. The first is a survey of 1979 Spanish internet users conducted in November 2011, just weeks before the Spanish general election. This survey is part of panel of online survey takers conducted in conjunction with the Spanish Centre for Sociological Investigation (CIS). Respondents were recruited from multiple Spanish internet portals that utilize audited customer databases. Quotas were used in the sample construction to further reduce the attendant sample biases of an online survey. However, given the age stratification in Spanish internet use, the survey sample is composed of persons 16-45. This dataset enables us to analyse the behavioural dimensions of candidate following on Twitter. Our dependent variable is a dichotomous variable denoting whether or not the individual follows a candidate or party participating in the general election. The question reads: “Thinking of your normal activities in online social networks (Facebook, Tuenti, Twitter, etc.), are you a friend or follower of a group or political party or one of the candidates in the general election in the November 20th general election?” Candidate support is operationalized in terms of an individual's propensity to vote for either Rubalcaba's party, the PSOE, or Rajoy's party, the PP. If the propensity to vote for the PP is more stronly linked to following a candidate, this would provide a selective exposure explanation for the differential in Twitter followers. The alternative, the informational explanation is operationalized as electoral interest: if a person is following a

candidate not necessarily as an expression of support but as a way to stay informed about the campaign, then this person is likely to be more interested in the election.

The second category of data are extracted from the Twitter application programming interface (API). First, the number of followers were recorded on a daily basis from September 30, 2011 through the election on November 20, 2011. The data spans nearly the entire unofficial campaign period after the parliament closed as well as the official campaign which began November 7, 2011. The data were collected every evening at exactly the same time to ensure comparability in measurement intervals.[4] Second, the profiles of Rajoy's 235,940 Twitter followers and Rubalcaba's 103,570 were downloaded. Both the daily counts of followers and the follower IDs were extracted via a script written in the R language using the twitteR module (Gentry 2012).

Third, using a program written in the Python language, the follower profiles for all of each candidates' followers were downloaded. Given the volume of profiles and Twitter's API linitations, the process took nearly twenty-four hours using computers located on two continents in order to reduce the server load. Twenty-one parameters were archived from each of the profiles including the user name, ID, date the account was created, location, profile language, the last status from this account, the platform from which the last status was communicated, and descriptive statistics regarding the number of statuses, followers, friends and related metrics of Twitter activity. These operations occurred on February 13, 2012, nearly three months after the elections to enable the analysis of these accounts during a period of electoral peace time.

Given the puzzle posed at the outset, we focus particularly on Rajoy's followers. To differentiate accounts controlled by unique sentient beings, we identify nine different characteristics. None of them are definitive in separating bots or purchased batches of followers from biographical human beings, but they provide circumstantial evidence corresponding to the completeness and level of activity of the account. Each of these factors aid in distinguishing machinic versus sentient Twitter followers. These factors include:

1. The number of followers each account has
2. The number of statuses posted on the account
3. The number of friends, i.e. the number of accounts that follow a use.
4. The number of favourites an account has
5. Whether the account contains location information
6. Whether or not an account has the default profile or has customized it in some way
7. Whether or not the account has a description of the user
8. Whether the user has posted a unique URL corresponding to a wider web presence
9. The ratio of a followers friends to followers

Additionally, we examine trends in when the accounts were created, the last account statuses, and the relationships between the date the account was created and patterns surround the creation of friend and follower networks.

Results

We begin by examining the overall growth of Twitter followers for the leaders of both of the major parties over the course of the campaign. The growth of Rubalcaba and Rajoy's Twitter following are presented in Figure 1. Rajoy started with roughly 15,000 more followers than Rubalcaba, 53,263 to 39,419. Both candidates have significant growth in their followings over the course of the campaign. Surges appear in both candidates' followings at the same times in response to the November 7th debates; however, Rajoy's follower growth slope is more inclined than Rubalcaba's. Rajoy's final tally on election day was 119,530 for Rajoy compared with 79,173 for Rubalcaba.

4 Due to the switch from European summer time, October 30, 2011 is a 25 hour interval.

[Figure 1 here]

Next we investigate the actors that predict whether or not someone followed one of the parties or candidates on Twitter over the course of the Spanish general election campaign. The dependent variable dichotomous indicating whether or not the respondent followed a party or a candidate via an online social network site. Although 81.6% of internet users in the sample regularly use social media platforms, only 15.3% reported following a candidate or campaign in this medium. We use a binary logistic regression to identify the factors influencing whether or not a person friends or follows a political actor related to the campaign. There are two models. The first model includes the propensity to vote for each of the major parties (PSOE and PP), gender, age, and an index of nonpolitical internet use. Internet use is aggregated as an additive index including purchasing products online, commenting on blogs or forms, maintaining a blog or website, downloading software or media, or making a video call online. Each item was couched in terms of regular use with either a yes or no response, resulting in a scale from 0 to 5. The second model introduces interest in the election. The results are presented in Table 1.

The first model indicates that those with greater propensity to vote for the PSOE, Rubalcaba's party, were more likely to follow a candidate on a social media site though the same does not hold for supporters of the PP, Rajoy's party. The propensity to vote for the PSOE drops out when electoral interest is introduced; the propensity to vote for the PP remains insignificant, however, the coefficient becomes negative. In both models the overall use of the internet and gender (male) are the most significant factors predicting whether or not a person followed a political actor during the campaign. Finally, both models find age is inversely related to following a campaign via social media.

[Table 1 here]

We now turn to the Twitter profile data. During observation window, Rajoy's following grew by 75,180 followers to 119,530 in total by the day after the election. Table 2 contains descriptive information concerning Rajoy's followers. The means and medians of each of the following statistics are provided: the number followers each of his followers has, the number of friends they have (i.e. the number of accounts each of Rajoy's followers follow), the total number of messages a user has posted, and number of the of posts by each follower flagged as "favourites" by other users.[5] These data indicate a wide degree of variance along each of these dimensions. On average, these accounts have a large number of both followers and friends and are quite active in posting to Twitter. However, of the 235,940 accounts, there are 40,123 (17.0%) with five or fewer followers, 21,856 (9.3%) have only one follower, there are no accounts without followers; 27,116 (11.5%) have never posted a status while 42,645 (18.1%) have posted 5 or fewer statuses; and 5,569 with five or fewer friends (2.4%).

We next consider the dates on which the accounts were opened. The dates range from September 11, 2006 up through February 13, 2012. The distribution of dates on which Rajoy's followers opened accounts are presented in Figure 2. These data present aggregated daily totals for the number of accounts opened. The figure shows a significant spike in new accounts throughout most of the month of November 2011 when the election was held. During this month, the accounts of 18,865 of his followers were opened, more than three times as many as in October 2011 when there were 6,143. These two months constitute by far the months when the largest concentration of Rajoy's Twitter followers opened accounts.

5 There is no limit on the number of "favourite" posts a user can flag therefore this metric does not denote a favourite post in exclusive terms as much as it denotes a "like."

[Figure 2 here]

Because this spike only corresponds to the date on which the accounts were opened, it may well be the case that some of these accounts that were opened in November began following at at later date. We are able to gain some greater insight into the timing between when the accounts were opened and when they began following: the original list of follower ID's returned from the Twitter API are in chronological order according to the date which they began following. Combined with the daily follower totals, this makes possible a rough correspondence between the order in which a user began to follow a candidate and the calendar date. This correspondence is not perfect due to unfollowing and account closures.[6] Close to 6,300 of the accounts that began following during the observation window had stopped by the time data was collected. Taking these correspondences as an approximation of the date on which each account began following Rajoy, we plot the date the account was created, for those that began following during the campaign, in Figure 3.

[Figure 3 here]

Among the accounts that began following during the campaign, the concentration of new accounts in the month of November is even more profound. Given the imprecision in the estimation, the last account included in this list was created on November 26, 2011. However, that shows there was a short lag in time between when the accounts were created and when they began following. To systematically evaluate this relationship, we regress the date the account was created on against the estimate of the date it began following. These results are displayed in Figure 4. Each point corresponds to an account located in terms of its date created on the X axis and the date it began following on the Y axis. The line in middle is the linear regression slope. If the same average lag between an account's creation and following Rajoy maintained, we should see a consistent cluster of accounts on both sides of the linear slope line however it continues to shorten over time.

[Figure 4 here]

Among the account level characteristics we identify the distribution of six characteristics that raise suspicion about an account. Here we consider only the 66,235 accounts corresponding to the growth of followers over the electoral campaign period. These measures pertain to signs of activity on Twitter as well as the degree to which the account profile is complete. The activity measures include the number of statuses, friends, and followers associated with the account. Account customization measures include whether the account has location information, a user description, and URL. The account activity metrics flagged accounts with fewer than five friends, followers, or statuses while the account customization measures are binary indicating the presence or absence of each feature. The measures were summed together producing an additive scale. The distribution of these elements are indicated in Figure 5. Over 97% of these accounts contain at least one suspicious element, only 1,805 exhibited none and the median value is two. At the high end, there were 12 cases exhibiting all six elements. These results are presented in Figure 5.

[Figure 5 here]

Overall there is little customization in these accounts: 36,876 (55.7%) have a default profile; 57,184 (86.3%) have no URL, of those that do, 4% have duplicate URLs; and only 28,438 (42.9%) have a unique description (the majority of those without a unique description have entered nothing in this part of the profile). In addition to these descriptive attributes, another indicator of sentient or

6 Forced closures usually occur when accounts are inactive or in violation of the terms of service.

machinic accounts is the ratio of an account's friends to followers. Those who are quite well known in an area of specialization will have very large followings relative to their friends, most people grow their friends and followers in roughly equal margins, automated bots typically have large numbers of friends but few followers. Turning to our data, Table 4 indicates the ratio of friends to followers. The median is 1.1 but the mean is 195.6 (s.d. 488.0) indicating a high degree of variance.

[Table 4 here]

More telling is the change in the ratio of friends to followers over time. Though older accounts will likely have more of both friends and followers, this ratio should not change as much if an account represents a biographical human being. By contrast bots are characterized by a large imbalance of friends to followers. The graph of this ratio over time based on the date each account was opened is presented in Figure 6. The smoothing line is a linear regression line. The Y intercept is quite low at around -500 indicating that the newest accounts contain a large number of outliers. This suspicion is verified in Figure 7 which plots the residuals over time. The downward slope confirms the presence of significant outliers while the bulk of the observations cluster around the regression line.

[Figure 6 here]

[Figure 7 here]

The final evidence we examine concerns the last statuses of these accounts. Again we confine our analysis to the accounts that were added roughly over the course of the campaign. There are 13,764 which lack a visible last status update. This includes 5,612 accounts that are protected, and thus the status is not publicly available, and an additional 8,152 accounts that have never posted a status. The frequency of each last status was calculated and categorized. There were a total of 51,639 unique last statuses indicating a relatively small subset are identical, though this figure underestimates the total number of duplicated last statuses as any status with a shortened URL would appear to have unique content.

We analysed the last statuses to classify the repeated statuses. The top five categorizations are listed in Table 5. The most common last message was “hola” or some elongated or shortened form of the Spanish word for “hello” (102). The next most common references were mentions of automated retweets either praising Rajoy or his policies or criticising the unions that had mobilised against him at that time (86). All of these tweets ranged from 4-9 identical iterations. The third most common set of duplicative tweets mentioned Rajoy by name or by his Twitter handle (47). Fourth, was a character symbol such as # and nothing else, ellipses, or a variety of emoticons expressing happiness or sadness (42) ; finally, the hash tag, “Twitteroff” signalling the end of a day of tweeting was the fifth most common last status update. The modal source for these tweets are the web (20,225) and the majority of the remaining tweets are communicated using mobile device Twitter clients. Looking at automated sources of tweets, even if we use an expansive definition of automated tweets, amount to at most 381 tweets counting those driven by Google Alerts, Twitterfeed, and other websites that automatically post tweets to an otherwise unique human-controlled account.

Discussion

These data suggest a mixed picture regarding the incidence of astroturf among the Spanish president's Twitter following. The survey evidence finds support for the informational explanation to account for Twitter following. and there is evidence that political interest combines with support

for the centre-left party may have motivated the growth in Rubalcaba's Twitter following. However, there is little evidence in the survey data that can account for Rajoy's relative rise. The analysis of Twitter profiles finds some evidence that a sizeable portion of Rajoy's followers are in some way suspicious as either bots or inactive accounts. Five categories of analysis suggest the presence of political astroturf. Nevertheless these figures do not reach the scandalous proportions of astroturfed followers seen recently in the followings of other prominent political figures.[7]

The first evidence considered looked at the behavioural patterns of campaign following on social media. This data only further deepened the puzzle as there was no link between support for the PP and following a campaign online, thereby eliminating the hypothesis that his follower growth could be explained by political support. This would also go against the selective exposure hypothesis if we bracket the presence of astrotuf.

A second possible explanation we considered was the differential informational value of followign Rajoy versus Rubalcaba. It is true that Rajoy's PP were expected to win by large margins and that may have depressed overall support for Rubalcaba's party: in the 2011 general election the PSOE received 4.3 million fewer votes than in 2008, however, the PP only managed to grow its vote share by roughly 500,000.[8] Given this context, there may be informational reasons explaining why individuals followed Rajoy in greater numbers rather than Rubalcaba given that the PSOE were so far behind in the polls. However, this is only a partial explanation: even critics of the selective exposure hypothesis note there is motivated information consumption which would move persons to follow other accounts in addition to Rajoy and none of the smaller parties or their leaders had nearly the level of follower growth as Rajoy.

For this reason we took a closer look at the accounts following Rajoy. Twitter accounts that are created to boost follower totals or favourable communications have distinct signatures. These include the the dates on which the accounts are created in relation to the growth in followers, manner in which the Twitter profiles are elaborated, the practices of friending and following, and content of their tweet communications. We focused specifically on the accounts added to Rajoy's total during the course of the electoral campaign as a way to identify the implication of astroturf in the growth of followers over this time frame. First, our data shows that among Rajoy's followers, there is a spike in the dates on which they were created corresponding to the same period of time as the campaign. Indeed, 25,008 accounts out of 66,235 followers added during the campaign – 37.8% – were created in that same time window. Over the course of the campaign, the lag between when an account was created and when it was added dropped considerably. This finding is consistent with the use of follower bots.

Second, we analysed Rajoy's Twitter followers in terms of six characteristics corresponding to the customization of the profile and activity on Twitter. These are factors that aid in distinguishing follower bots from human followers. The more complete a profile is in terms of its description, location information, and evidence of additional web presence, the more likely it is to be a distinct human being. Although it is possible to create bots with well-elaborated profiles, these typically are more expensive (Ding 2012). Almost all the accounts had at least one suspicious feature and the majority had at least two. In addition the level of profile customization, a sizeable portion had few followers or statuses indicating low level of Twitter activity.

Follower bots, however, are not the only kind of bot we found evidence for. Tweet bots may be used to spread certain memes online. These accounts have fully automated status updates rather than the semi-automated updates of users who link an account to a website or blog via Twitter feed or Google alerts, the latter for which we found little evidence. Tweet bots tend to be distinguished by the ratio of friends to followers. We identified the presence of Tweet bots in two ways: in terms

7 Cases include studies suggesting that 92% of Newt Gingrich's Twitter following has been purchased or recent studies on sudden jump in Mitt Romney's Twitter following. Both men disavow any knowledge.

8 The electorate was actually slightly smaller in 2011 than in 2008 but not by much so these figures provide roughly comparable base points to make a comparison on the overall support for the parties in the electorate.

of the ratio of friends to followers and evidence of duplicate tweets. Concerning the ratio of friends to followers, though the median Rajoy follower had a roughly 1 to 1 ratio between friends and followers, the standard deviation on this metric is quite large. Furthermore, we found that over time, the outliers grew larger indicating that more of these suspicious accounts were created as the campaign progressed.

The second indication of the presence of tweet bots is evidence of identical tweets. The second largest category of duplicate tweets were tweets praising Rajoy, an attempt to confer legitimacy stemming from his personal qualities, or tweets in support of his policies. Previous research suggests there is a low threshold for meme cascades within the wider Twittersphere indicating that these tweets may very well be widely diffused (Ratkiewicz et al. 2010). The remaining categories of identical tweets in the top five are all consistent with follower bots in that they convey little informational content. These include mentioning Rajoy by name or his Twitter handle, variations on common greetings in Spanish without a recipient, and tweets containing a single grammatical symbol or emoticon, undirected at a recipient. All of these tweets had between three to nine repetitions.

These results suggest two general conclusions for studies concerning the political consequences of social media communications. The first general conclusion is that astroturf may very well be inflating Twitter follower counts as well as the transmission of communications. For this reason, there may be limits to the ability to draw general conclusions about the distribution of preferences or beliefs about human actors using Twitter. Studies that aim to discern human networking and communication patterns from the distortions of machinic accounts may need to develop procedures to systematically exclude such communications.

Second, even fake accounts may nevertheless be real in their effects. One may hire a staff to spread favourable messages via social media or one might choose to automate the process. Either way, the consequences can induce distortions in the overall distribution of discourse within social media platforms. These abilities are not limited to authorities within governments as those outside of governments may wish to create false impressions and associations. Hence it is not necessarily an advantage for a political actor to have large numbers of bots in their social media following as this can be used by others to associate them in undesirable ways. Additionally, just as tweet bots may be harnessed to create large amounts of legitimating tweets, they can be harnessed to flood Twitter communications with delegitimating critical claims. The same conditions obtain for investment firms or other companies that that use various algorithms to mine data from social media and other online sources as a way of predicting market behaviour. Therefore the development of strategies to identify astroturf is an important task not only for those who study social media behaviour but also practitioners who make use of these data, journalists who report on these topics, and consumers.

Barberá, Pablo. 2012. "A New Measure of Party Identification in Twitter. Evidence from Spain." pablobarbera.com (August 14, 2012).

Bennett, W. Lance, and Shanto Iyengar. 2008. "A New Era of Minimal Effects? The Changing Foundations of Political Communication." *Journal of Communication* 58(4): 707–731.

Chadwick, Andrew. 2011. "Britain's First Live Televised Party Leaders' Debate: From the News Cycle to the Political Information Cycle." *Parliamentary Affairs* 64(1): 24 –44.

Chadwick, Andrew. 2012. "Recent Shifts in the Relationship between the Internet and Democratic Engagement in Britain and the United States." In *Digital Media and Political Engagement Worldwide: A Comparative Study*, eds. Eva Anduiza, Michael J. Jensen, and Laia Jorba. Cambridge University Press, p. 39-55.

Conover, M. D, 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*,.

Ding, Jason. 2012. "The Twitter Underground Economy: A Blooming Business." *The Barracuda Labs Internet Security Blog*. http://www.barracudalabs.com/wordpress/index.php/2012/08/03/the-twitter-underground-economy-a-blooming-business/ (August 15, 2012).

Easton, David. 1965. *A Systems Analysis of Political Life*. New York: John Wiley & Sons.

El País. 2011. "Rubalcaba fue más influyente que Rajoy en Twitter durante la campaña." *EL PAÍS*. http://politica.elpais.com/politica/2011/11/22/actualidad/1321988832_920902.html (August 14, 2012).

Garrett, R. Kelly. 2009. "Politically Motivated Reinforcement Seeking: Reframing the Selective Exposure Debate." *Journal of Communication* 59(4): 676–699.

Gayo-Avello, Daniel. 2011. "Don't turn social media into another 'Literary Digest' poll." *Communications of the ACM* 54(10): 121.

Gentry, Jeff. 2012. "Twitter client for R."

Golbeck, Jennifer, and Derek Hansen. 2011. "Computing political preference among twitter followers." In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, New York, NY, USA: ACM, p. 1105–1108.

Golder, S. A., S. Yardi, A. Marwick, and D. Boyd. 2009. "A structural approach to contact recommendations in online social networks." In *Workshop on Search in Social Media at ACM SIGIR*, http://research.microsoft.com/pubs/102204/SIGIR2009-Workshop.pdf (August 14, 2012).

Holbert, R. Lance, R. Kelly Garrett, and Laurel S. Gleason. 2010. "A New Era of Minimal Effects? A Response to Bennett and Iyengar." *Journal of Communication* 60(1): 15–34.

Jantsch, John. 2008. "Google Alerts to Twitter." *Small Business Marketing Blog from Duct Tape Marketing*. http://www.ducttapemarketing.com/blog/2008/11/07/google-alerts-to-twitter/ (August 14, 2012).

Jensen, Michael J. and Eva Anduiza. 2012. “Online Political Participation in the United States and Spain.” In *Digital Media and Political Engagement Worldwide: A Comparative Study*, eds. Eva Anduiza, Michael J. Jensen, and Laia Jorba. Cambridge University Press, p. 80-101.

Jungherr, A., P. Jurgens, and H. Schoen. 2011. “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* 30(2): 229–234.

Kravets, David. 2010. “Analysis: Bigger Twitter, Facebook Flock Boosts Election Odds.” *Wired.com*. http://www.wired.com/threatlevel/2010/11/election-odds-twitter/ (August 12, 2012).

Omicrono. 2012. “España es el tercer país del mundo que más usa Twitter.” *Omicrono*. http://www.omicrono.com/2012/01/espana-es-el-tercer-pais-que-mas-usa-twitter-del-mundo/ (August 12, 2012).

Ratkiewicz, Jacob, Michael Conover, Mark Meiss, Bruno Gonçalves, Snehal Patil, Alessandro Flammini, and Filippo Menczer. 2010. “Detecting and Tracking the Spread of Astroturf Memes in Microblog Streams.” http://adsabs.harvard.edu/abs/2010arXiv1011.3768R (August 14, 2012).

Ratkiewicz, J., M. D Conover, M. Meiss, B. Goncalves, A. Flammini, and F. Menczer. 2011. “Detecting and tracking political abuse in social media.” In *Fifth International AAAI Conference on Weblogs and Social Media*, , p. 297.

Sunstein, Cass R. 2001. *Republic.com*. Princeton: Princeton University Press.

Tumasjan, Andranik, Timm O. Sprenger, Philipp G. Sandner, and Isabell M. Welpe. 2011. “Election Forecasts With Twitter How 140 Characters Reflect the Political Landscape.” *Social Science Computer Review* 29(4): 402–418.

Vaccari, Cristian and Rasmus Kleis Nielsen. 2012. “What Drives Politicians' Online Popularity?” Presented at the Midwestern Political Science Association Conference, Chicago Illinois.

Figure 1: Growth of Twitter Followers Among the Two Major Party Leaders

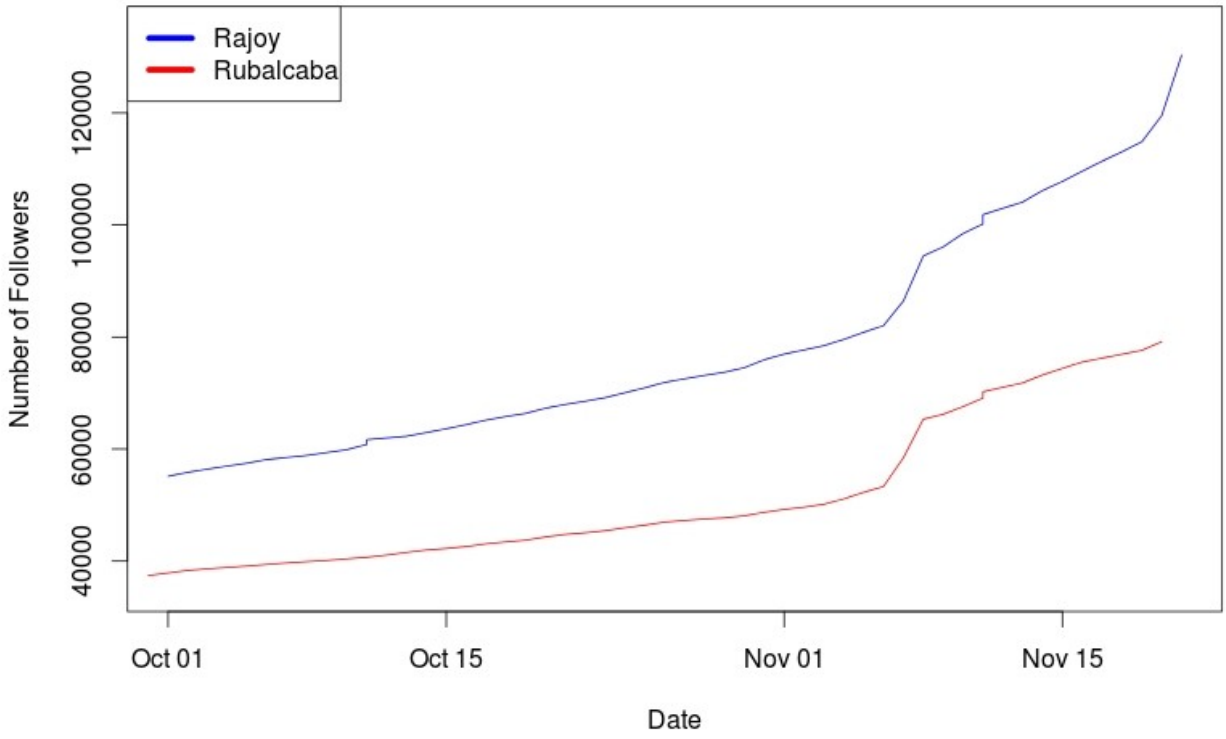

Table 1: Predicting Candidate Following on Social Media

	Estimate (std. error)	p-value	Estimate (std. error)	p-value
Intercept	-1.527 (0.0433)	0.000	-3.401 (0.485)	0.00
Vote PP	0.019 (0.017)	0.266	-0.028 (0.018)	0.127
Vote PSOE	0.050 (0.018)	0.007	0.019 (0.020)	0.322
Sex (female)	-0.537 (0.129)	0.000	-0.427 (0.136)	0.002
Age	-0.035 (0.009)	0.000	-0.016 (0.010)	0.090
Internet use	0.319 (0.054)	0.000	0.256 (0.056)	0.000
Election interest			0.860 (0.080)	0.000
AIC	1631.8		1507.9	
N	1979		1979	

Table 2: Followers' Descriptive Statistics (Rajoy Full Sample)

	Mean	Std. Deviation
Followers	1346	971.7442
Friends	1304	671.4364
Statuses	3572	2876.689
Favourites	172.3	303.5149

Figure 2: Dates on Which Twitter Accounts Following Rajoy Opened

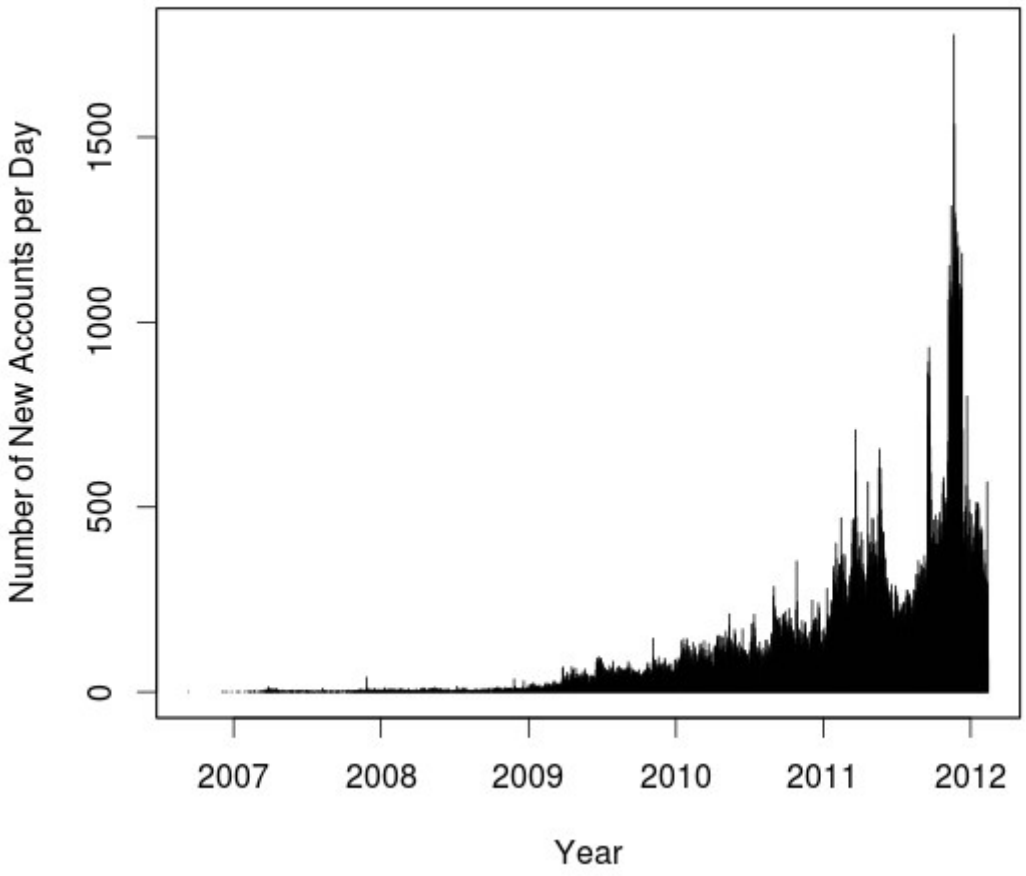

Figure 3: Distribution of Dates Rajoy Followers Account Opened Daily

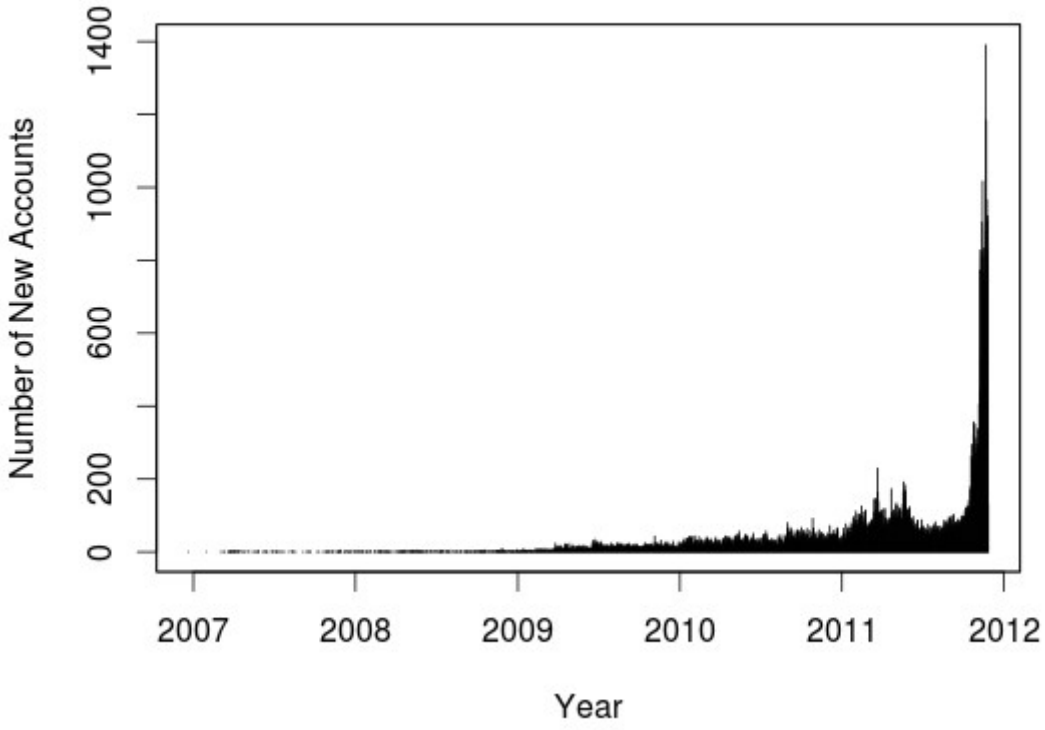

Figure 4: Date Account Created Compared to Follow Dates

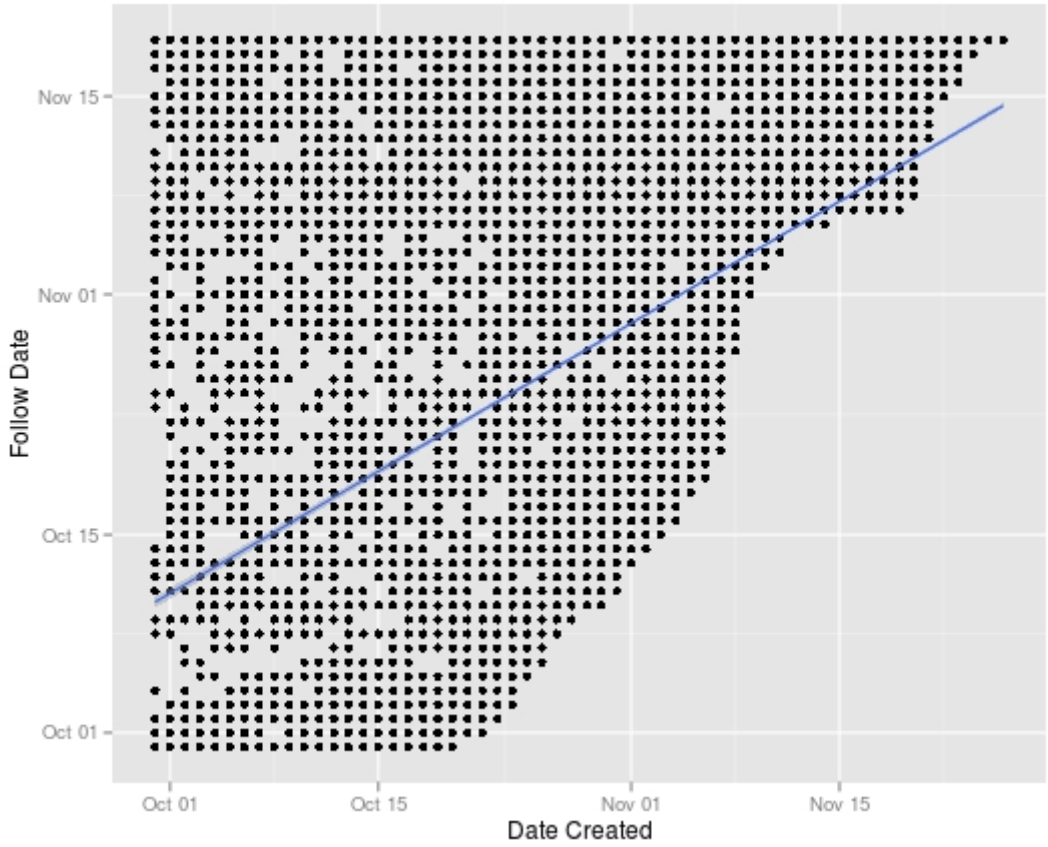

Figure 5: Number of Suspect Elements Per Account

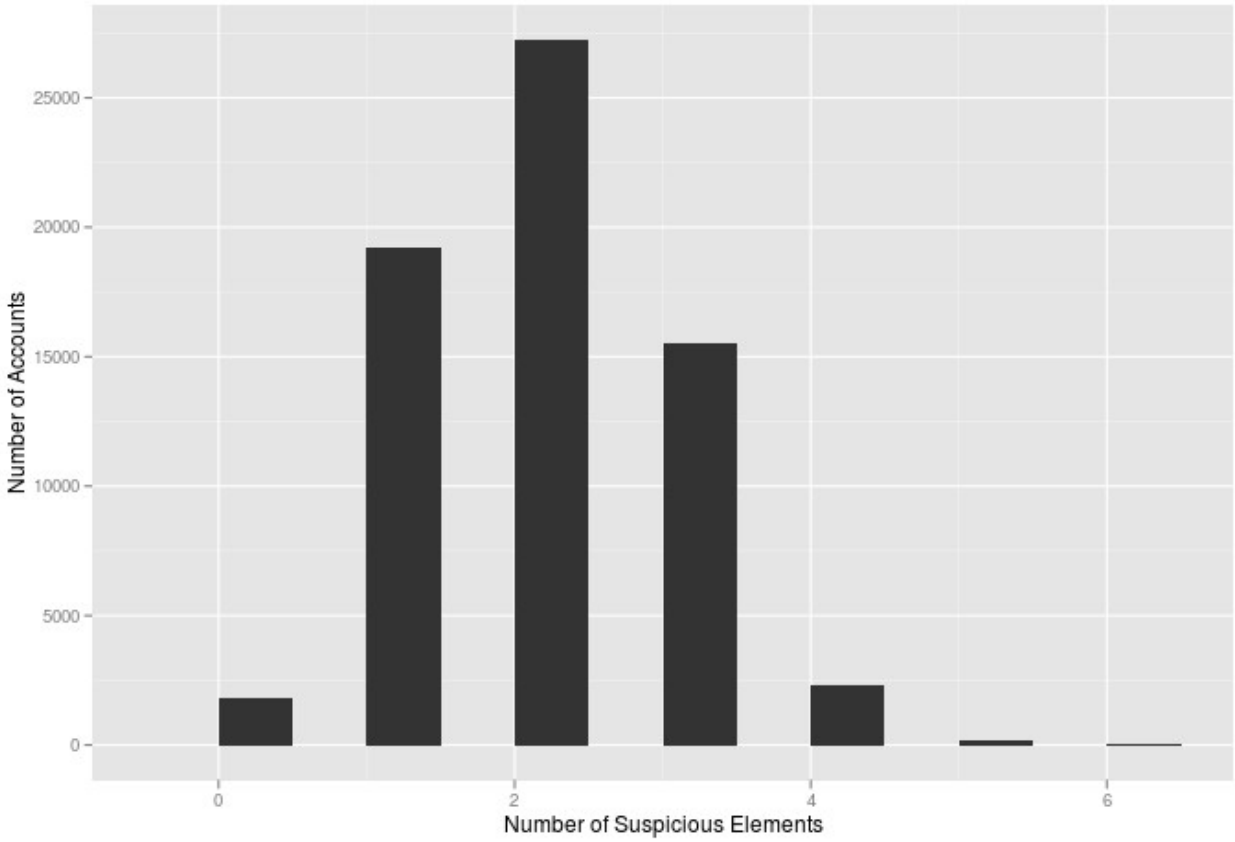

Table 3: Followers' Account Profile Characteristics (Added During Campaign)

Dimension	Number of Cases
Followers (<6)	13205
Friends (<6)	1797
Statuses (<6)	13129
No Location	50521
No Unique Description	37797
No URL	57184
Default Profile	29359
n	66235

Table 4: Ratio of Friends to Followers

	Mean (std. Dev.)	Median
Ratio Friends: Followers	195.6 (488.0)	1.01

Figure 6: Change in the Ratio of Twitter Friends to Followers Based on the Date Created

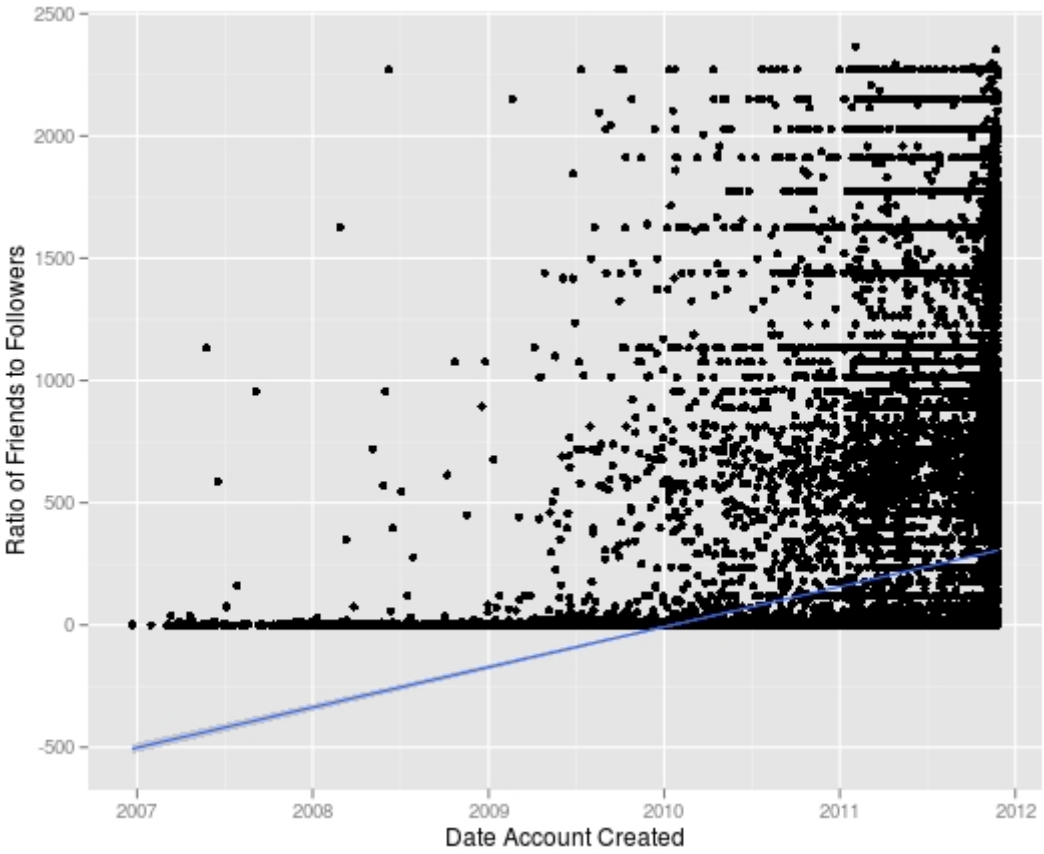

Figure 7: Friends to Followers Regressed on the Date Created Residuals Plots

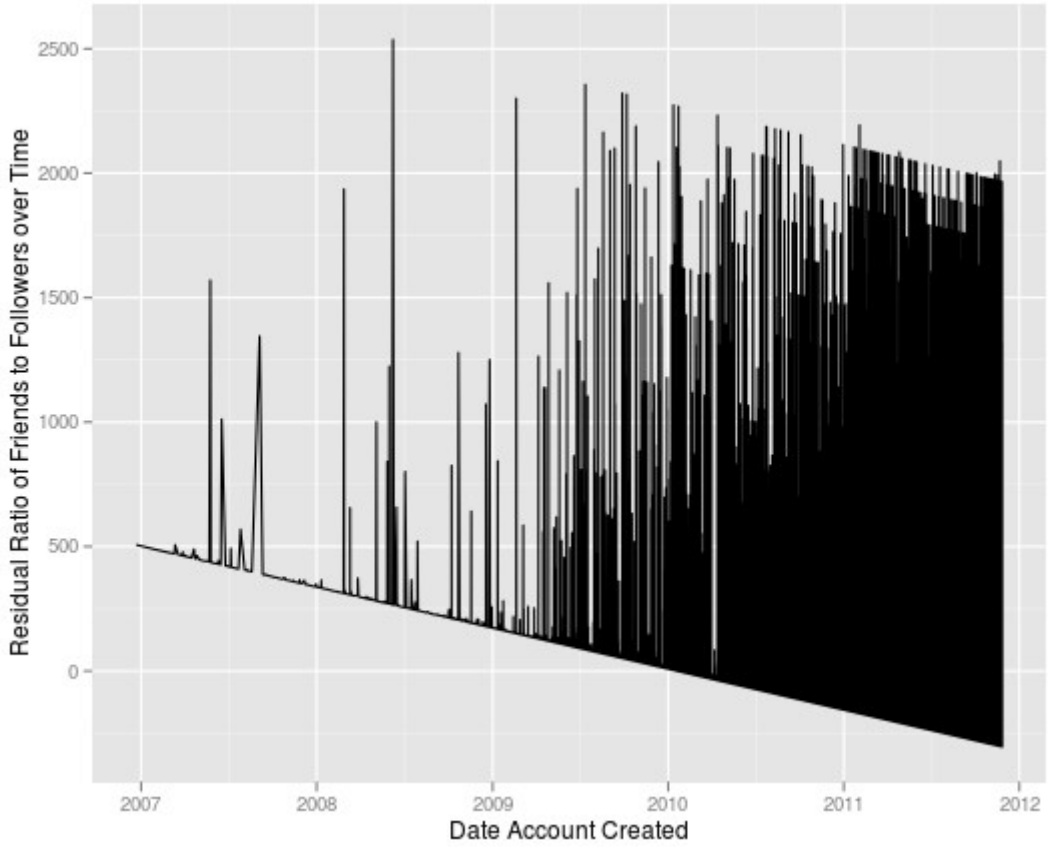

Table 5: Last Status Categories

Category	Number of Last messages
Hello	94
Political Support	86
Mariano Rajoy	47
Character symbol/Emoticon	42
Twitteroff	17