I. INTRODUCTION

The electoral campaign is a period preceding elections where political parties do an organized effort so that their candidates garner supporters. Maximizing the influence of their messages over voters is the main objective. In this way, politicians use different techniques to transmit their messages in the most effective way to their potential voters, such as mass meetings, rallies, hustings or media management. Understanding and exploiting in a more efficient way the available resources for information flow than your opponent can make the difference.

Over the last century mass media has been monopolized by “old media”, such as televisions or newspapers. However nowadays we are attending to a transition where a new interactive online social media world is settling its bases. Online social networks, such as Twitter with over 200 million users, have become ideal platforms for information flows. This has been noted in [1] where they reported that these tools may serve as a framework for discussion. Other studies have been directed towards identifying influential users [2] or discovering its commercial usage [3]. Moreover the percentage of population using online social networks has increased in recent years, reaching in Spain a 42% of the population, quantity that is almost duplicated (82%) for young adults between 18-29 years old [4].

Following the idea “one must be where people are”, politicians are now present in the most popular online social networks. However some politicians do not have a defined strategy for the usage of these tools and the rest are still far of exploiting all the available potential. The importance and popularity of social media in politics became clear with Obama’s campaign for the 2008 U.S. Presidential elections and his famous tweet: "This is history...", posted just after winning the elections. This fact attracted not only popular, but also scientific attention, making political conversations in Twitter a popular subject for research. Lately, the data gathered from Twitter has been used as a “social sensor” to predict election outcomes [5]. Other studies have focused in analyzing the interactions between different political communities [6], and finally a proof-of-concept-model has been developed [7] to predict candidate’s victory.

In this article we introduce a new parameter that measures the ratio of the support in Twitter between two candidates, which we call the Relative Support, and how it can be used to indicate and quantify which candidate and in which proportion is getting more benefits from events occurring offline. We further study the dynamical patterns emergent from the Twitter mention and retweet networks within the framework of complex networks theory [8–10]. We also interpret politicians behavior by filtering these networks and analyzing the interactions going on between the different political parties. Finally we introduce a model based on the heterogeneous preferential attachment formalism [11] capable of growing political conversations and illustrate it by reproducing the mentions and retweets taking place in Twitter among politicians.

II. SYSTEM

The present work is based on data collected from the online social network Twitter. This web application allows people to post and exchange text messages limited by 140 characters. There are several interaction mechanisms in Twitter to transfer information. The first of these is the ability of people to follow and be followed by other users. This is a passive mechanism that allows users to receive the messages written by their followers in real time. The Twitter’s followers network is a directed graph where non-reciprocal relations are admitted and it states
the social substratum through which information may flow. Previous studies have reported a high heterogeneity in the followers distribution [12]. Another important mechanism to interact is the retweet or message retransmission. This mechanism allows individual messages to propagate throughout the network and serves as a way for people to endorse their point of view over specific subjects [13]. In addition to this, another relevant way for direct interaction is the mentions mechanism. By mentioning someone’s username in the message text, people is able to send directed messages to the mentioned user’s inbox. This mechanism is often used to establish conversations between users, through the exchange of messages, or just to refer somebody in the message’s text [1].

Our dataset is constructed from public access messages posted in Twitter, related to the 2011 Spanish Presidential elections. We downloaded all the messages that included the keyword 20N, using the Twitter API interface, in a three week period including the election day. We chose this keyword because it is an ideologically neutral identifier, used by all the political parties during the campaign and voting day. In summary we analyzed over 370,000 messages, written by over 100,000 users. We found that 40% of the messages were retweets, and over 25% contained at least one mention. This fact makes the event quite relevant, since it has been reported that retweets represent about 4% of the overall messages [14].

III. RESULTS AND DISCUSSION

A. Time Series

We can begin to understand how the Spanish political landscape is reflected in Twitter by comparing the number of times that each political party has been mentioned during the 20N discussion and the number of votes it obtained in the elections. Previous studies show that there is a correlation between the number of times a political party is mentioned during an electoral campaign and voting day. In summary we analyzed over 370,000 messages, written by over 100,000 users. We found that 40% of the messages were retweets, and over 25% contained at least one mention. This fact makes the event quite relevant, since it has been reported that retweets represent about 4% of the overall messages [14].

Since in Spain there are two main political parties that outstand on top of the others, we focused our study on them. We analyze the time series of the accumulated tweets mentioning at least one of these parties, PP and PSOE, or their candidates, Rajoy and Rubalcaba. Looking at Figure 1A we can state two things: Firstly tweets contain more mentions to the political parties rather than to their candidates; secondly the more conservative party, PP and its candidate Rajoy, were much more mentioned than PSOE and Rubalcaba.

One of the most important results we have obtained studying the 20N Twitter conversation, is that the time series of the accumulated tweets mentioning political parties or candidates present piecewise linear growth, as it is showed in Figure 1B. On top of that the points where the slope increases coincides with important events occurring outside Twitter that fuel the activity of the conversation, which occur at the same time for both parties. This fact makes us think about the ratio between the rate at which the cumulative mentions to two political parties grow as a better indicator of the outside political support to each party, than just the raw number of mentions. Following this idea we define an instant indicator of the support in Twitter between two political parties, which we call the Relative Support parameter \( RS_A^B \), given by the following expression:

\[
RS_A^B = \frac{m_A}{m_B} 
\]  

(1)

where \( m_A \) and \( m_B \) are the slopes for the accumulated mentions to the A and B political parties.

From our point of view there are two days of special relevance in our study: the debate between the two main candidates that took place on November 7th, and the voting day on November 20th. We did a further analysis of these two days.

During the debate, people’s attention was completely focused on the two candidates, Rajoy and Rubalcaba. This provoked that, contrary to what happened during the whole campaign, the candidates were more mentioned than their corresponding political parties, as it can be seen in Figure 1A. Therefore, for this period, we studied the time series of the accumulated mentions to the candidates rather than the parties. The majority of tweets about 20N posted on this day are concentrated on the two hours that lasted the debate, with a total of 2,733 messages mentioning Rubalcaba and 4,150 mentioning Rajoy. In Figure 1C we present a detail of the time series of the accumulated tweets for the candidates during the debate. We can observe that, in accordance with what we said before, both series present linear growth, being the slopes for both candidates constant during the whole debate, and changing its value at the end of it. The Relative Support during the debate was \( RS_{Rub}^{Raj} = 1.53 \), Rajoy over Rubalcaba. This value is pretty close to the relation between the votes (1,55) obtained by the two candidates thirteen days after (see Table II).
Table I. Results by political party for the votes obtained, the mentions on tweets and the messages sent from official accounts (Activity).

| Political Party                      | Acronym | % Votes | % Tweets | Activity |
|--------------------------------------|---------|---------|----------|----------|
| Partido Popular                      | PP      | 44.62   | 39.92    | 1228     |
| Partido Socialista Obrero Español    | PSOE    | 28.73   | 26.33    | 1819     |
| Izquierda Unida                      | IU      | 6.92    | 5.03     | 451      |
| Unión Progreso y Democracia          | UPyD    | 4.69    | 11.8     | 1852     |
| Convergencia i Unio                  | CIU     | 4.17    | 4.51     | 208      |
| AMAIUR                               | AMAIUR  | 1.37    | 2.76     | 11       |
| Partido Nacionalista Vasco           | PNV     | 1.33    | 2.20     | 11       |
| Ezquerra Republicana de Catalunya    | ERC     | 1.05    | 1.47     | 113      |

Table II. Comparison between the ratio of votes and the Relative Support parameter for the two main political parties.

| Votes Ratio | Debate | Voting Time | Waiting for Results | Results Release |
|-------------|--------|-------------|---------------------|-----------------|
| $PP_{PSOE}$ | 1.55   | $RS_{PSOE}^{PP}$ = 1.53 | $RS_{PSOE}^{PP}$ = 2.31 | $RS_{PSOE}^{PP}$ = 1.64 |
| $PSOE_{PP}$ | 1.00   | $RS_{PSOE}^{PSOE}$ = 1.00 | $RS_{PSOE}^{PSOE}$ = 1.54 | $RS_{PSOE}^{PSOE}$ = 1.54 |

The election day survey is one of the most relevant and reliable surveys to predict election outcomes. This makes us believe that a further study on this day must be done when analyzing election results. It can be seen in Figure 1A, the major increase of political mentions occurred during the voting day, what reinforces our idea about the importance of this day. In Figure 1B we show a detail of the time series of the accumulated messages from 8:00 to 21:20 for the Spanish political parties. In correspondence with our theory of piecewise linear growth, in Figure 1B we can distinguish three important regions (D, E, F) of the space-time for this day: “Voting time”, “Waiting for results”, “Results release”, that we further discuss, and present in panels D, E and F.

VOTING TIME (8:00-19:00). This panel covers the entire voting period, from the opening of the electoral colleges to the closure. Over 7,500 tweets containing either PP or PSOE were posted. From the four panels studied in detail (Figure 1A, D, E, F), this one presents by far the lowest activity per hour, what makes it the less representative sample. The Relative Support took a value of $RS_{PSOE}^{PP} = 2.31$ in favor of PP, indicating that PP users were much more enthusiastic than PSOE.

WAITING FOR RESULTS (19:00-20:00). This period lasts only one hour, starting with the closure of polls and ending when the first news were released. This first news informed about the participation statistics, and gave provisional results for a 5% scrutiny. Over 5,000 tweets mentioning either of the two main parties were posted during this hour. During this period the Relative Support parameter estimated quite accurately the upcoming results, taking a value of $RS_{PSOE}^{PP} = 1.64$ in favor of PP.

RESULTS RELEASE (20:00-21:30). This region covers the entire period in which the results were given, starting with a 5% of scrutiny and ending with an 85%, point at which the politicians made their first speeches. It was the period with more activity per hour in Twitter of the whole study, with more than 13,000 tweets posted mentioning PP or PSOE. The measure of the Relative Support while results were given was of $RS_{PSOE}^{PP} = 1.41$, pretty close to the relation between the votes of the two parties, as it can be seen in Table II.

Summarizing, we have centered our study on the two dominant parties of the Spanish political landscape and observed that in this system the relation in votes and tweets between them coincides quite precisely (Table I). We introduce a new measure to study political support in Twitter, the Relative Support between two parties $RS_B$, which we see as a useful tool to study future elections or to determine how Twitter users react to external events, and who gets more popular with them. In our study we identify the debate between the two candidates (7th of November) as the key point in Twitter. This was the point where users of the social network began to actively participate in the 20N conversation, and during the two hours of debate people reflected their preferences in Twitter, $RS_{PP} = 1.54$. The lack of external critical political events during the campaign and the firmness of people’s vote intention, maintained the ratio of tweets constant around this value along the whole campaign. Although future work should be done in applying the RS parameter to other elections, we believe that this parameter is capable of revealing election outcomes even when offline events occurring at the last minute change voter support. In this way it would have detected Zapatero’s victory against forecast in the 2004 Spanish Presidential election, that took place four days after the 11M terrorist attack.

B. User Interactions

So far we have seen that the user activity is correlated with the election results. In this section we will further analyze such activity and characterize its emergent structural and dynamical patterns based on the Twitter
interaction mechanisms. First of all, we have analyzed the cumulative probability distribution for the user activity, that we define as the number of messages posted by user. This distribution follows a power law in the form $P(x > x^*) = x^{-\beta}$, where $\beta = 1.275 \pm 0.002$, as shown in Figure 2A. Such exponent is within the expected values for scale-free human activity phenomena [15]. This implies a very high heterogeneity level in user behavior. In fact, we found that half of the messages were posted by only 7% of the participants, who were the most active users and posted from 8 to over 4,000 messages each, while the other half of the messages were posted by the remaining 93% of users, who posted less than 8 messages each. Similar results were obtained in the study of the 2009 German elections [8] and in the study of the 2005 Canadian elections [16], who concluded that the political discussions during the campaign in social media were controlled by a very small fraction of the participants. However, it is unclear whether this activity really represented an actual discussion or debate. To answer this question we will next analyze the user activity taking into account the way participants interacted with each other, either by the mention or retweet mechanisms. Therefore we have built two networks according to “who mentioned who” and “who retweeted (or retransmitted) who”. Both networks have directed and weighted edges [8, 17], whose weight is directly proportional to the number of times that a user has mentioned or retweeted another user. In total, the mention network has over 39,631 nodes and 86,029 links, while the retweet network has over 75,546 nodes and 153,549 links. In Table III we present the networks’ main properties, some of which we will discuss next.

In Figure 2B we present the in strength cumulative distribution for both networks. The in strength indicates the number of mentions received by user, and the number of retransmits gained by user, respectively. Both measures are related to the level of collective attention that users may gather along the conversation. The in strength distributions follow power laws in the form $P(x > x^*) = x^{-\beta}$, where $\beta_M = 1.14 \pm 0.01$ and $\beta_R = 1.051 \pm 0.008$. Once more, such distributions display a high heterogeneity level found in the users profiles. As a matter of fact, we found that just 1.04% of the users were target for half of the total mentions and 2.24% of the users wrote the messages that caused half of the total retransmissions. These results show that both mechanisms are highly elitist, since a remarkably small fraction of users, mainly compound by media and politicians, concentrate half of the collective attention, while the large majority individually attracted only a few. Such collective attention is built out of adding individual efforts, which are characterized in the out strength distributions, shown in Figure 2C. These distributions indicate the amount of mentions...
where the hubs that concentrate much of the incom-
form better to an exponentially truncated power law in the
and for the retweet network we found that the data fit
for both networks (19). As shown in Table III, we found
this measure by splitting it into combinations of in and
out degree pairs (18). As our networks have directed edges, we have calculated
sortativity by degree coefficient for both networks (18).

In order to unveil how such heterogeneous users interacted with each other, we have also calculated the asso-
sortativity by degree coefficient combined by in and out degrees.

| Property          | Mentions | Retweets |
|-------------------|----------|----------|
| Nodes             | 39.631   | 75.546   |
| Edges             | 86.029   | 153.549  |
| Bidirectional Edges | 2.17%    | 0.99%    |
| \( r_{\text{out, out}} \) | -0.039   | 0.087    |
| \( r_{\text{out, in}} \) | -0.141   | -0.107   |
| \( r_{\text{in, in}} \) | -0.021   | -0.043   |
| \( r_{\text{in, out}} \) | -0.005   | 0.017    |

or retransmissions made by user, respectively. For the mention network we found a power law distribution in
the form \( P(x > x^*) = x^{-\beta_M} \) where \( \beta_M = 1.438 \pm 0.001 \), and for the retweet network we found that the data fit
better to an exponentially truncated power law in the form \( P(x > x^*) = x^{-\beta_R} e^{-x/c} \) where \( \beta_R = 1.479 \pm 0.005 \) and \( c = 130 \pm 30 \). As we found on the overall user activity distribution, the out strength distributions show that a
small fraction of users (7.71% for mentions and 12.41% for retweets) concentrated over half of the activity, while the majority of users who concentrated the other half (92.29% for mentions and 87.59% for retweets), mentioned less than 8 users and retweeted less than 4 mes-

In order to unveil how such heterogeneous users interacted with each other, we have also calculated the asso-
sortativity by degree coefficient for both networks (18).

or retransmissions made by user, respectively. For the mention network we found a power law distribution in
the form \( P(x > x^*) = x^{-\beta_M} \) where \( \beta_M = 1.438 \pm 0.001 \), and for the retweet network we found that the data fit
better to an exponentially truncated power law in the form \( P(x > x^*) = x^{-\beta_R} e^{-x/c} \) where \( \beta_R = 1.479 \pm 0.005 \) and \( c = 130 \pm 30 \). As we found on the overall user activity distribution, the out strength distributions show that a
small fraction of users (7.71% for mentions and 12.41% for retweets) concentrated over half of the activity, while the majority of users who concentrated the other half (92.29% for mentions and 87.59% for retweets), mentioned less than 8 users and retweeted less than 4 mes-

Figure 2. User cumulative distribution for user activity (A), mention and retweet networks in strength (B) and mention and retweet networks out strength (C). The solid lines represent the best fitted curve for each distribution. The dashed lines indicate the percentage of users that posted 50% of the messages (A), received 50% of the mentions or retweets (B) and made 50% of the mentions or retweets (C).

Table III. Topological properties of the mention and retweet networks. \( r \) represents the assortativity by degree coefficient combined by in and out degrees.

Table III. Topological properties of the mention and retweet networks. \( r \) represents the assortativity by degree coefficient combined by in and out degrees.

Previous works on network assortativity (18), state that social networks tend to be assortative, as popular
people want to be friend with popular people, and regular
people are usually friends among the regular people. However our measures indicate the opposite. This was already reported by Hu and Wang (21), who detected that most online social networks are disassortative and in the same order as the networks of our study. The reason for this result is that online relations are different from real life ones. For example in Twitter, regular people are now able to relate and communicate with popular accounts, either by following, mentioning or retweeting their mes-

We have also carried out a community structure analy-
sis for both networks based on a random walk algorithm (22) and found that this conversation also presents a modular structure. In fact the largest modules in the mention network are formed around politicians, while mass me-

edia accounts centered the largest modules in the retweet network. In an effort to further discover the role that politicians and mass media accounts have played during the election campaign, we have analyzed how users men-
Experimental modes were mainly used to campaign rather than debate, as we will see in the following section. This lead us to suggest that users do not turn to the political landscape in Twitter, we have analyzed the political filtered mention and retweet networks. Our target is to find patterns that help us understand how politicians interacted during the campaign. For this matter we have filtered the previously mentioned networks, only remaining the official politician’s accounts from those parties with over 10 participants in the conversation. We found that politicians do not relate to each other randomly and that they use the different interaction mechanisms with specific purposes. The cumulative strength function distribution of both networks are shown in Figure 4.

In order to understand the way politicians used both mechanisms, we have analyzed the mention and retweet networks by measuring assortative mixing patterns according to discrete characteristics. Therefore we have classified all nodes according to the political party they belong to, and calculated the assortativity coefficient $r$ over the matrix $e_{ij}$, that defines the fraction of edges going from one party to another, as described by Newman in [18]. We found $r$ to be close to 1 at both networks ($r_M = 0.905$ and $r_R = 0.990$), as presented in Table IV which means that politicians mentioned and retweeted mostly their own partisans. However, mentions tend to happen across parties a little more than retweeting, which is the most segregative interaction found. This result indicates a considerable lack of debate between the politicians and reveals some of the strategies followed by them during the campaign. A previous work on Korean elections [22], reported that mentions between politicians reflect the political alliances between candidates. However we find retweets to be a more overwhelming interaction to map the political endorsements, as it presents the highest assortative mixing coefficient. This issue has already been pointed out during the 2010 U.S. Congress elections, where retweets were found to be more ideologically polarizing than mentions among regular users [6].

To further explain the structural features found in the interactions between politicians, we propose a model based on the heterogeneous preferential attachment formalism [11]. The idea behind it, is that the probability of a node $i$ interacting with a node $j$ not only depends on their respective degree, but also of the affinity between them. In our model nodes (politicians) are classified according to discrete characteristics (political parties). Thus the probability of appearance of a new interaction from any politician, $i$, belonging to party $A$, to a politician $j$, who belongs to a party $B$, is given by the following expression:

\[ P_{ij} = \frac{S_j}{\sum_{j\in B} S_j} f_{AB} \]  

where $S_j$ is $j$’s strength, and $f_{AB}$, is the affinity value

\[ \begin{array}{|c|c|c|} \hline 
\text{Network} & \text{Experimental } r & \text{Modeled } r \\
\hline 
\text{Mention} & 0.905 & 0.86 \pm 0.03 \\
\text{Retweet} & 0.991 & 0.989 \pm 0.005 \\
\hline 
\end{array} \]
power law distributions of exponents: $\gamma_{M} = 1.3$ for the mention network and $\gamma_{R} = 1.6$ for the retweets one. In this way we modeled the resulting distributions by simulating the heterogeneity found in the users behavior and using the same microscale connection rule for both interaction mechanisms. In Figure 4 we present the resulting cumulative strength function distribution for both networks, after having averaged over 1,000 realizations. It can be noticed that the model reproduces perfectly the strength function distribution for both networks, and maintains the assortative mixing levels as presented in Table IV.

IV. CONCLUSIONS

The perfect political campaign strategy has been eternally chased by politicians. To that effect we have tracked voter sentiment and uncovered the underlying structure of the campaign in Twitter, measuring the impact that different events have produced on politicians popularity and analyzing the roles played by the various users. For this matter, in this paper we propose a parameter that measures the relative support in Twitter between two candidates, and apply it to our case of study: the 2011 Spanish Presidential elections. Furthermore we have analyzed the graph structural and dynamical patterns emergent from interactions taking place among users, finding out that the collective attention is driven by a very small fraction of users, who dominate the interaction mechanisms. We have also analyzed politicians behavior finding a profound segregation and lack of debate among them. Finally we propose a network growth model based on heterogeneous preferential attachment, to explain the emergence of such segregated modules in the politician’s networks. Despite we can’t assure that the campaign on Twitter determined the election outcomes, our results suggest that there is a strong correlation between the activity taking place in Twitter and election results. This fact suggests that further research should be done on identifying the most efficient techniques to influence voter sentiment in Twitter.

ACKNOWLEDGMENTS

Support from MICINN- Spain under contracts No. MTM2009-14621, and i-MATH CSD2006-32, is gratefully acknowledged.

[1] C. Honeycutt and S. C. Herring HICSS IEEE Computer Society pp. 1-10 (2009)
[2] D. M. Romero, W. Galuba, S. Asur, and B. A. Huberman Influence and passivity in social media. WWW’11 (2010)
[3] B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury Journal of the American Society for Information Science and Technology 60, 2169 (2009)
[4] Pew Research Center. Global Digital Communication: Texting, Social Networking Popular Worldwide (2011)
[5] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe Predicting elections with Twitter: What 140 characters reveal about political sentiment. ICWSM The AAAI Press (2010)
[6] M. Conover et al. Political polarization on Twitter. ICWSM The AAAI Press (2011)
[7] A. Livne, M. P. Simmons, E. Adar, and L. A. Adamic The party is over here: Structure and content in the 2010 election. ICWSM The AAAI Press (2011)
[8] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez and D.-U. Hwang Phys. Rep. 424, 175 (2006)
[9] R. Albert and A.-L. Barabasi Rev. Mod. Phys. 74, 47-97 (2002)
[10] M. E. J. Newman SIAM Review 45 2, 167-256 (2003)
[11] A. Santiago and R. M. Benito Europhysics Letters 82, 58004 (2008)
[12] H. Kwak and C. Lee and H. Park and S. Moon WWW ’10 ACM, 591-600 (2010)
[13] D. Boyd, S. Golder and G. Lotan HICSS IEEE Computer Society, pp. 1-10 (2010)
[14] pearanalytics. Twitter study (2009)
[15] M. E. J. Newman Contemporary Physics 46, 323 (2005)
[16] R. Koop and H. J. Jansen Social Science Computer Review 155 (2009)
[17] A. Barrat, M. Barthlemy, R. Pastor-Satorras, and A. Vespignani Proc Natl Acad Sci USA 101, 3747 (2004)
[18] M. E. J. Newman Physical Review E 67, 026126 (2003)
[19] J. G. Foster, D. V. Foster, P. Grassberger, and M. Paczuski Proc Natl Acad Sci USA 107, 10815 (2010)
[20] K. A. Wojciech Galuba, D. Chakraborty, Z. Despotovic Outtweeting the Twitterers - predicting information cascades in microblogs. Microblogs 3rd Workshop on Online Social Networks, WOSN (2010)
[21] H.-B. Hu and X.-F. Wang Europhysics Letters 86, 18003 (2009)
[22] M. Rosvall and C. T. Bergstrom Proc Natl Acad Sci USA 105, 1118 (2008)
[23] H. Y. Yoon and H. W. Park Social media information flow and public representation: A case of s. Korean politicians on Twitter 9th International Triple Helix Conference (2011)
[24] A.-L. Barabasi, R. Albert, and H. Jeong Physica A: Statistical Mechanics and its Applications 272, 173 (1999)

http://www.pearanalytics.com/blog/wp-content/uploads/2010/05/Twitter_study_2009.pdf