Political Opinion Dynamics in Social Networks: the Portuguese 2010-11 Case Study

António Fonseca and Jorge Louçã

ISCTE - Lisbon University Institute
antoniofilipefonseca@gmail.com
Jorge.L@iscte.pt

Abstract

The research on opinion dynamics in social networks and opinion influence models often suffer from a lack of grounding in social theories as well as deficient empirical data validation. The current availability of large datasets, and the ease we can now collect social data from the Internet, makes validation of theoretical social models a less difficult task. Starting by a state-of-the-art of the research and practice concerning political opinion dynamics in social networks, we identify the main strengths and weaknesses of this domain. We then propose a novel method for uncovering political opinion dynamics using on-line data gathering. The method includes three distinct phases: (1) data collection, (2) multi-agent modelling (3) validation. Specifically, we tested the significance of both Social Impact Theory, originally proposed by Latané (1981), and Brownian Agent modelling, proposed by Schweitzer (2002), for characterizing political opinion formation during electoral periods. These two models were tested using more than 100.000 tweets collected during the periods from the 30th of October to the 21st of January 2011 and from the 27th of March to the 6th of June 2011, concerning the Portuguese presidential and legislative elections occurred in 2011. Following the data collection, two distinct on-line communities were inspected: the general Twitter user community, and the traditional news media Twitter feeds. The opinion dynamics was simulated with grid adjustment of model parameters. This operation was performed on separate empirical series, respecting the talk about the six electoral candidates and parties. The complete process allowed concluding about the explanatory power of Social Impact Theory and Brownian Agents, and, on the other side, allowed characterizing opinion dynamics in this specific case study. This article details each phase of the method, illustrated using the dataset available at http://work.theobservatorium.eu/presid2011.

Keywords: Opinion Dynamics, Web 2.0, Social Networks, Multi-Agent Simulation, Sociophysics, Econophysics

1 Introduction

Internet social networking has recently become a major subject of research. Structural properties of social networks, namely the balance between users, network centrality, paths of information diffusion, network dynamics, user influence or even crawling methodology have become a focus of research which has provided an evolving set of conceptual tools in order to support the analysis of online networking.[1] Recent studies based on emotion analysis techniques introduced effective prediction and trend detection methods, not only of stock markets activity but also of political elections, and even movie.[2]

All these studies have profited from the ease by which large amounts of social data may automatically be collected from online social network API's. Castellano et al. [Castellano et al., 2009] succinctly state what is at stake, from the sociological research point of view, in this era of high-speed electronic information flow between people, where social platforms could become a laboratory for social sciences. In particular, the web could to have a strong impact on the studies of opinion formation, political and cultural trends, globalization patterns, consumer behaviour, and marketing strategies. Modern online social networks like Twitter; Facebook or YouTube provide relatively standard interfaces by which simple data collecting software may gather large amounts of data in real-time. This API's take part of a large technical innovation effort that each day online social network companies implement in order to enrich their web pervasiveness. Analytically, social media has a far reaching potential. Qualitative and quantitative sensing of online social content is revolutionizing online consuming. On January 2012 Netflix, the leader of online video rental in the US, announced a partnership with Facebook in order to show its clients which movies their friends were watching. This ability of social media feeds to automatically and in real time monitor population expression and sensibility will allow scientists to devise new forms of social data analysis and research[3].

When in 2008 presidential elections large scale of online networking helped the US president Barack Obama set records in terms of donations and grass root mobilization, the ability for online social networking to influence election results, the effective potential of online social media was made clear. Twitter had become a legitimate communication channel in the political arena. The relation between the online world and politics comes, however, much earlier.

1.1 Online Social Media and Politics

Ahead of the rest of the world, in 1996, US political candidates had already websites. In 1998 Jesse Ventura, a US Reform Party candidate, used email to win the Minnesota gubernatorial election. John McCain fund-raised money online in 2000 presidential campaign, and from 2004 on, blogs debate and discuss politics online, twenty four hours a day, on a planetary scale [Tumasjan et al., 2010]. This activity has the purpose of exchanging and publicizing political views, from personal or

[1] Since 2006 there are several dedicated annual conferences that approach this subjects, of which ICWSM - International AAAI Conference on Weblogs and Social Media (`http://www.icwsm.org`) is the principal venue.

[2] A very recent example of this is the new trend in American online newspapers to post news trends of politics based on Twitter feeds. Mention Machine (`http://www.washingtonpost.com/mention-machine`) by The Washington Post is a very popular example.

[3] The most promising project on this subject is the one million euro European join effort - FuturICT (`http://www.futurict.eu/`) to build an earth population simulator.

institutional perspectives. With accelerating velocity, our age's new technologies are herding crowds of people into new groups, through distant communication and raising new levels of collective power [Shirky, 2008]. On the week of January 17, 2001, during the impeachment trial of Philippine President Joseph Estrada, seven million text messages brought thousands of Filipinos to the streets angry with their corrupt president. Since then, several public street movements have been triggered worldwide by electronic online communication: Spain, 2004; Belarus, 2006; Moldova, 2009; Iran, 2009; Thailand 2010; The Arab Spring, 2011.

Social media have become coordinating tools for nearly all of the world's political movements, just as most of the world's authoritarian governments (and, alarmingly, an increasing number of democratic ones) are trying to limit access to it [Shirky, 2011]. One of the most significant of these controlling efforts is the Golden Shield Project, colloquially referred to as the Great Firewall of China, which involves thousands of workers and computer facilities on a huge effort to censor all the potential dangerous websites at the People's Republic of China.[4] Several attempts have been made to control political participation on the Internet, but all of them overestimate the value of broadcast media while underestimating the value of media that allow citizens to communicate privately among themselves. The value of access to information, particularly information hosted in the West, overestimates the importance of computers while underestimating the importance of simpler tools, such as cell phones [Shirky, 2011]. As the issues facing government have become more complex, social technologies have emerged that enable citizens to self-organize easily. These technologies may eventually enable democracies to scale and become more adaptable and direct [Ito, 2004]. Governing of the commons [Ostrom, 1990], specifically the knowledge commons [Hess, 2006] exchanged in the Internet, and at the same time allowing free speech, assuring property rights, security and privacy isn't at all a trivial task. A slowly developing public sphere, where public opinion relies on both media and conversation, is the core of the environmental view of Internet freedom [Shirky, 2011]. The adoption of this new space as a natural political environment constitutes a great challenge to modern democracies.

1.2 Online Polls

Polls in our society and particularly in the world of politics and the media have become both ubiquitous and enormously influential. The growth and widespread adoption of digital media have increased not only the amount and diversity of information available to citizens but also the opportunities for political participation and opinion expression, and also for opinion polling. We can think of opinion expression as a rational form of political participation in which individuals express opinions when the benefits of doing so outweigh any associated costs. In the Internet, political communication it is most often occurring within the relatively friendly confines of existing networks, far more than would be possible through face-to-face communication [Goidel, 2011]. Research has found that higher-status Internet users are more likely to use the Internet for "capital-enhancing activities," thus increasing the knowledge gap. This pattern of participation, referred to as the digital divide, remains an important component of new media use. Also, recent research has

[4] Significantly also, is the role of commentators hired by the government of the People's Republic of China (both local and central) or the Communist Party to post comments favourable towards party policies in an attempt to shape and sway public opinion on various Internet message boards. These commentators are said to be paid 0.5 Yuan for every post that either steers a discussion away from anti-party or sensitive content on domestic websites, and pejoratively called 'the 50 cent party'.

emphasized differences in use across levels of political engagement. The result is that online political talk is polarized, with highly educated and ideologically users talking more loudly and more often [Kirzinger, 2011]. For purposes of public opinion measurement this fact constitutes a severe low coverage bias problem on political polls. It is similar to the one found in phone polls with cell phones being normally associated with younger and low income respondents. It can however have different expression on distinct online media. As in classical polling, it can be attenuated with sample weighting and multi-mode surveying. The same happens when respondents give incomplete answers or don't respond at all to the inquiries. However, this problem may altogether be circumvented if instead of prompting respondents with inquiries, the information is directly grabbed from online participation on blogs or in social networks[5]. People are hard to being reached and persuaded to cooperate, but social data gathering can provide the kind of "it's nothing personal" inquiry in which invaluable information may be collected. New horizons are open for opinion studies and opinions polling in the Internet society. As an example, online polling through segmentation can reach special populations that would be prohibitively expensive or even impossible to reach over the phone. In not distant future, web-based gathering will allow capturing trends in public opinion, which may guide political debates as well as public policies.

Proponents of deliberative opinion polling contend that meaningful public opinion emerges only after a deliberative process in which the public carefully considers competing perspectives and weighs policy alternatives[6]. The data we analyze in this paper reveals, from a quantitative perspective, the result of a deliberative process. Surprisingly, the simple magnitude of mentioning candidates and parties on Twitter feeds is closely correlated with the deliberative result of the elections.

1.3 Data Gattering

The work we present constitutes an effort to validate social network research work with real data [7], as well as contributing to the ongoing research on population sensing through online media collecting.

Our work may be read as twofold:

- Firstly we examine the social network usage of Twitter during the January 2011 presidential elections and June 2011 legislative elections in Portugal. This from the electoral prediction perspective given by classical media polls collected during the campaigns and by the final electoral result.

---

[5] In 2009, US ABC television broadcast network polling director Gary Langer reported that nearly half of all market research spending was in online data [Keeter, 2011]

[6] Some scholars defend the opposite - opening up the number of venues for public participation appears to attract the intensely partisan more interested in promoting a particular view than in hearing from the other side. They say that hearing competing perspectives increases tolerance and understanding but decreases political involvement. A more polarized electorate may be more engaged but not necessarily more informed, deliberative, or thoughtful. The upshot is that polarization may increase participation particularly among the base [Mutz, 2006]. The authors have experienced the same phenomenon on browsing the Twitter Portuguese community.

[7] This type of work has been already reclaimed by some authors "One of the major problems with 'social physics' or sociophysics literature, especially the exploration and understanding of social processes by means of computer simulation is the lack of connection to real life examples" [Sobkowicz, 2009][Moss, 2005]

- Secondly, as we explain on the next section, we try to understand the structure of the opinion dynamics of the Twitter community supported on the analysis of multi-agent model simulations - a Brownian agent model and a social impact model. These two computer simulations allowed us to better understand the fact that online discussion magnitude tend to emulate election results.

We've collected Twitter feeds, over a period of three months preceding the elections. This gathering was done both from general users and also from newspaper, radios and television Twitter feeds.

At the first stage we will show that the quantity of news produced by the media about each candidate or party, closely estimates the final elections results. This finding was based on counting of newspaper articles and was already reported in 2007 by Véronis [Véronis, 2007]. Many recent studies also have showed similar findings [Connor et al., 2010][Tumasjan et al., 2010], this time based in online media usage. Usually, to this purpose, two main methods of social network examination are employed: counting of friends or followers of candidates and sentiment analysis. We use a different method of agent expression quantification by counting the number of tweets naming each presidential candidate or party. Some recent work [Gayo-avello, 2011] warns about the effectiveness of this method of electoral pooling. Although the results oddly confirm the reality, we agree that they lack some theoretical justification. This fact however brings up the important need for more in field data confirmation, so that a more robust opinion dynamics theory may be built. To this purpose we try to study inter-agent expression influence with multi-agent model simulation. In this process we examine with more detail the collected data about the elections from the Twitter user's perspective. We show that there is some correlation, although not direct, between the media news feeds and the general population's conversation over candidates.

1.4 Schweitzer's Brownian agents

Many large-scale phenomena observed in social systems constitute a large-scale "macroscopic" complex effect of the "microscopic" simple behaviour of a large number of interacting agents. This fact has led social scientists to the introduction of elementary models of social behaviour. Many of these models closely relate to models that have been introduced in modern traditional statistical physics. It is natural to approach them using the same concepts and tools that have been successfully applied in physics. From *opinion dynamics* to *cultural dynamics*; from *language dynamics* or *crowd behaviour* into *formation of hierarchies*, many of these models have already been analyzed and ameliorated in order to somehow evaluate and to provide insights on the phenomena they represent [Castellano et al., 2009].

In the particular field of *opinion dynamics* several models were devised which have different levels of representation:

- The Voter model [Clifford, 1973] - each agent takes the opinion of the majority of its neighbours.

- Majority rule model [Galam, 2002] - a random group of *r* agents in the community take the opinion of its majority at turns.

- Sznajd model [Sznajd, 2000] - A pair of neighbouring agents determines the opinions of their nearest neighbours.

- Deffuant model [Deffuant, 2000]- Agents discuss at pairs its continuous valued opinions and tend to converge to a compromise value

- Hegselmann-Krause model [Hegselmann, 2002]- An agent discuss with its neighbours its continuous valued opinion and tend to converge to a compromise value.
- Brownian models [Schweitzer, 2003] - Agents interact as Brownian particles according to a statistical equation.
- Social impact theory [Nowak et al., 1990] - each agent is influenced by its neighbours having each some degree of influence and supportiveness.

In this paper we will test the real data collected on the elections, against a Brownian model of agents as was initially proposed by Schweitzer [Schweitzer, 2010]. We chose this model because it was the first effort, as it is our knowledge on scientific literature, of one implementation of a multi-agent representation of opinion dynamics in online communities. We validate the model by feeding it with news media information, in order to observe agent information consumption. By varying three cognitive parameters of the agents, we show how they can mimic real population expression behaviour. We then draw conclusions from these experiments about possible parameterization.

1.5 Social impact theory

Having tested the influence of media, through a Brownian coupling, on the global expression of all the agents, we then proceed to analyze the influence of inter-agent connections using a multi-agent model based on Social Impact Theory from Latane [Nowak et al., 1990]. To this purpose we examine how diverse network topologies can determine different time patterns of agent's expression that we confront and analyse against the real voting expression observed in the Twitter network. Based on this comparison we find that the connection geometry between debaters, as generally modelled through standard social network models, is determinant factor to the overall community voting trends. Particularly we confirm the proximity between the Barabasi-Albert model of preferential attachment and the real Twitter network, not only from the social network standard metrics comparison, but also from the multi-agent opinion contamination modelling perspective. Also we show how poorly linked individuals are more susceptible to media influence and how richly connected individuals influence another. Finally we will show how a simulated private individualistic election, that takes place at each agents mind and which is expressed through individual opinion, can be propagated at a collective level.

1.6 Agenda

Our work will be reported in three steps:

- Firstly we will describe the social data we've collected, its significance and limitations, and also some conclusions it immediately may lead by comparison with classic political pools.
- Next we will describe a multi-agent model based on Schweitzer's Brownian Agents model and we will examine how different parametrical configurations of *arousal*, *political valence* and *informational noise* could influence a similarity between media news and debate over candidates.
- Finally we use a Social Impact theory multi-agent model, with its assumptions and simplifications, in order to mimic political debate inside a online community. We will draw conclusions about the influence of network topology on the overall agent expression.

2 Twitter Data

The data that we've collected in order to support and validate our model was obtained in real-time using two Python scripts running over the Twitter API interface. We choose to collect tweets from two distinct groups of users: news media and general users. These two groups present quite distinct online behaviours. The news media group almost exclusively posts the headlines of the news with a web link to their site. The general users usually tweet their strong opinions, cite other users or eventually they re-tweet the news posts. Although the messages keep flowing in a more or less constant pace they're mostly posted during the afternoon or by the evening. Instead of looking into the content of the messages, as was early performed by several authors [Tumasjan et al., 2010] [Connor et al., 2010] [Pak and Paroubek, 2010] we choose to keep a simple accounting of the number of tweets that refer to a single candidate name. This approach shows itself to be quite accurate in distinguishing the theme of the talk without further complication. The option for some type of sentiment classification would imply a significant amount of neutral sentiment tweets that could not be accounted for as useful data. This way we manage to obtain an absolute term of comparison between news talk and population talk.

2.1 News and Tweets

Comparing the evolution of the news media and general population tweets during the campaigns that preceded the elections day, we found surprisingly that, the percentage magnitude of tweeting about each candidate or party followed relatively closer the trends signalled by classical telephone pools.
Figure 1 depicts the comparison between news media tweets and these pools. There were six candidates on the race to win the elections which final results are reported in Table 1. For sake of clearness we opted not to draw on the chart all of the accounting of tweets but instead a spline interpolation that quite closely follows the trend.

Candidate	Final Result
Cavaco Silva	53,14%
Manuel **Alegre**	19,67%
Fernando **Nobre**	14,04%
Francisco **Lopes**	7,05%
Defensor **Moura**	4,52%
Manuel **Coelho**	1,58%

Party	Final Result
PSD	41,19%
PS	30,42%
CDS	12,72%
PCP	8,61%
BE	5,69%

Table 1: Final results of the elections.

In Figure 2 the same pool of data is depicted, this time the magnitude of the tweeting of users is depicted. Figure 3 and Figure 4 represent the same analysis for the legislative elections. In order to evaluate the degree of similarity between the four time series, a weighted average of Pearson correlation coefficients was calculated between pools and tweets (news and population).

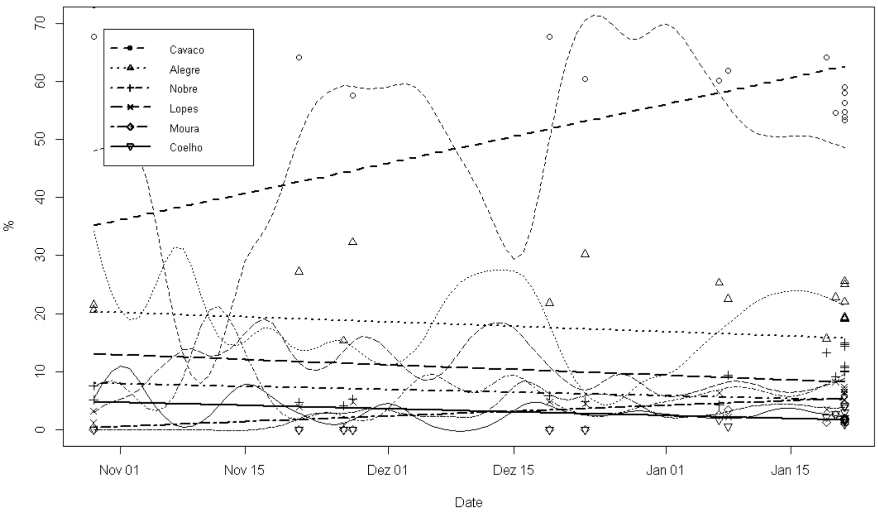


Figure 1: Spline interpolation of percentage magnitude of the total tweets for each candidate on presidential elections from news media feeds during the campaign. First order trend line and campaign polls as dots. Total of 44 news media users.

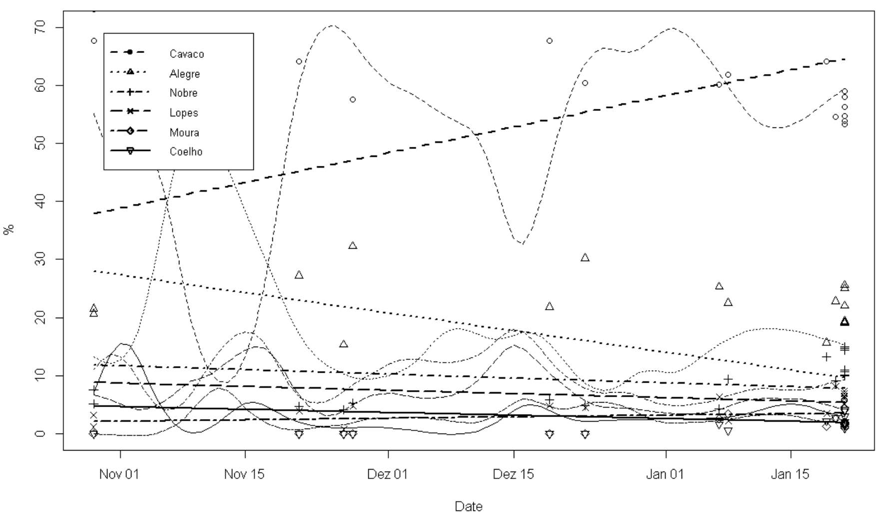


Figure 2: Spline interpolation of percentage magnitude of the total tweets for each candidate on presidential elections from regular population users during the campaign. First order trend line and campaign polls as dots. Total of 1903 users.

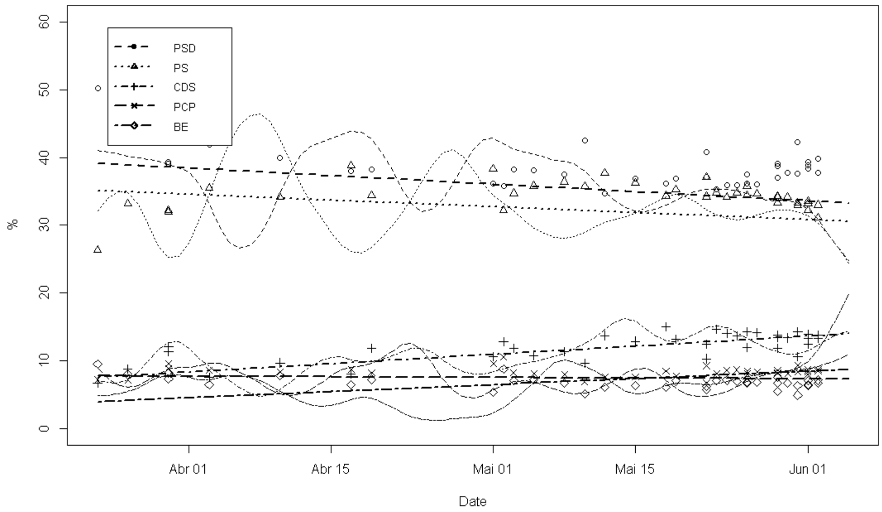


Figure 3: Spline interpolation of percentage magnitude of the total tweets for each party on the legislative elections from news media feeds during the campaign. First order trend line and campaign polls as dots. Total of 44 news media users.

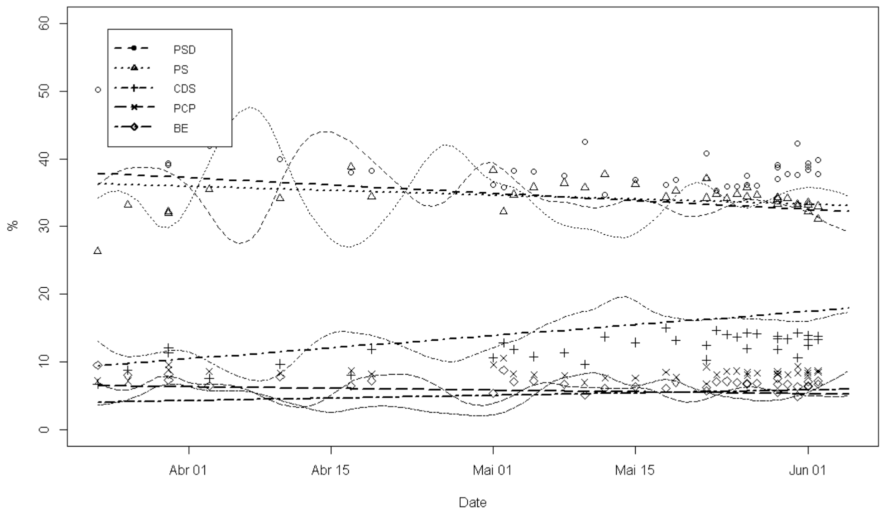


Figure 4 : Spline interpolation of percentage magnitude of the total tweets for each party on the legislative elections from regular population users during the campaign. First order trend line and campaign polls as dots. Total of 1903 population users.

We've considered a weighted average given by:

$$\langle \rho \rangle = \frac{1}{N} \sum_i k_i \frac{E[(V_i - \mu_{V_i})(P_i - \mu_{P_i})]}{\sigma_{V_i} \sigma_{P_i}} \qquad (1)$$

Here k^i are constant weighting coefficients doubling each passing week and having a maximum at the elections day; is the vector representing the percentage magnitude of tweets for the day i and P_i is the last polls vector before that day. The results were the following:

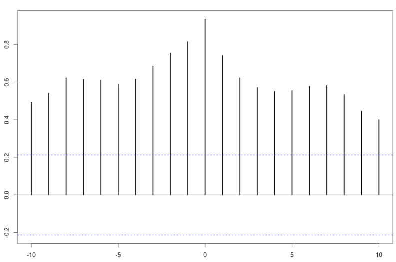
Series	$\langle \rho \rangle$
presidential news tweets	0.866
presidential user tweets	0.877
legislative news tweets	0.846
legislative user tweets	0.925

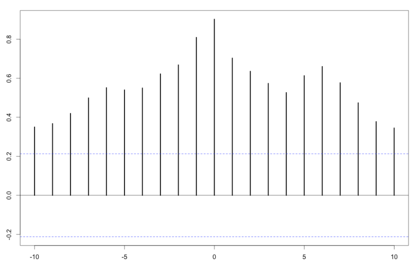
Table 2: Weighted average of Pearson correlation coefficients.

We can verify that the tweets percentage magnitude constitute a good approximation to the real-time voting intention. It is reasonable to assume that the expression intention on Twitter follows the real voting intention of the population. Assuming also that this intention is reflected in the polls. Particularly significant is the results obtained at the legislative elections respecting the population tweeting. The final vote intention is also close to the relative percentage, with some remarkable exceptions. These occur mostly in the presidential elections with the second and third candidate, keeping however their relative positions in the final result. We may conclude that people tend to talk about the most prominent candidates or parties in consonance with their voting intention. We may verify that although present several oscillations on the chatting flow, presumably due to the talk done by the media, the final result, both by media proportions as also from population proportions, tend to emulate the final vote. In order to examine the temporal relation between population tweeting versus media tweeting we've also computed the cross correlation between the two streams. The results are the following:

(a) Cavaco

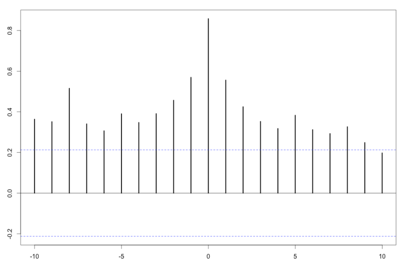
(b) Alegre

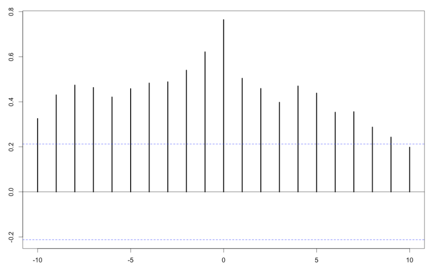




(c) Nobre

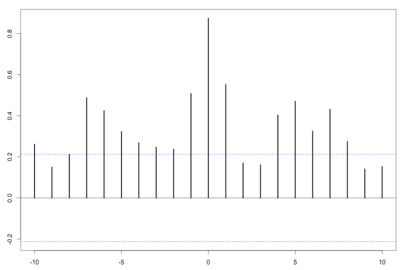
(d) Lopes





(e) Moura

(f) Coelho



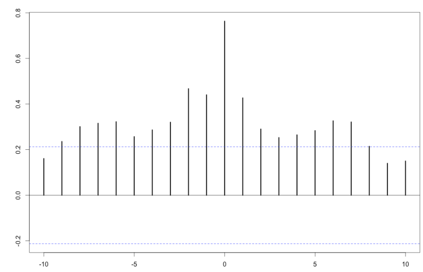
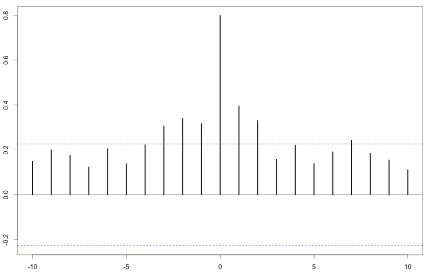
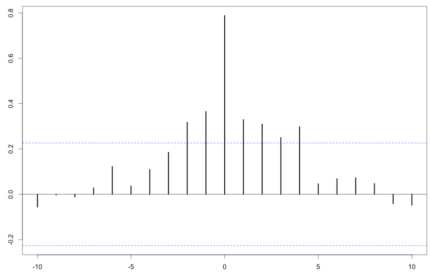


Figure 5: Covariance between time series of news and population tweets in presidential elections, lag of between -10 and 10 days.

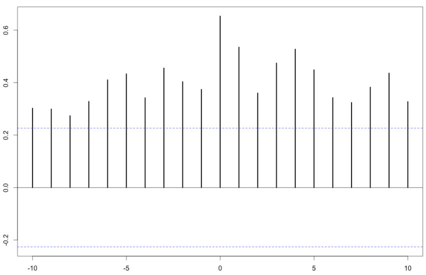
(a) PSD



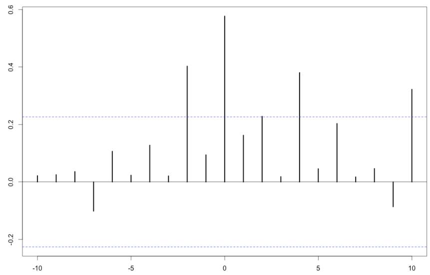
(b) PS



(c) CDS



(d) PCP



(e) BE

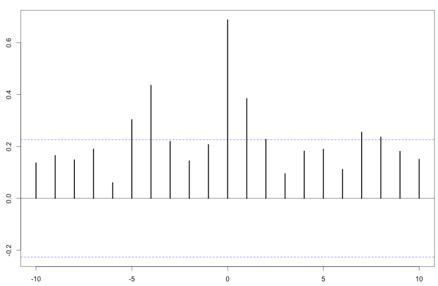


Figure 6: Covariance between time series of news and population tweets in legislative elections, lag of between -10 and 10 days.

We can notice that for every entity that there is a strong correlation each day between the news that are propagated by the media and the relative magnitude of tweeting. For the less talked candidates or parties this correlation is even stronger, which may signify that the discussion over their campaign is mostly motivated by the news. In fact, watching the campaigns more closely in Twitter we can in fact confirm this result, as the discussion about lesser candidates is almost circumscribed to their pronunciation on the media or to campaign events which are publicized on the news.

3 Multi-agent models

3.1 Brownian Agents model

In order to provide some insights into the strong temporal correlation between news and tweeting by the general population we've build a multi-agent model, based on Schweitzeir's Brownian Agents Model [Schweitzer, 2003] of opinion dynamics. More precisely the model was inspired by the approach originally proposed in [Schweitzer, 2010] which is an application of Brownian agents to online communities. In this approach, each *ith* agent has two main state variables: *arousal* , which characterizes the predisposition of the agents to be aware and to act, and *valence* , which represents the choosing of each agent for a particular candidate or party, $k \in$ {'*Cavaco*' , '*Alegre*' , '*Nobre*' , '*Lopes*' , '*Moura*' , '*Coelho*' } and $k \in$ { '*PSD*' , '*PS*', '*CDS*' , '*PCP*' , '*BE*' }. The voting intention of each agent is given by the k that corresponds to the maximum valence v^k.

Following the Schweitzer and Garcia's model, we chose both of these variables to have a time evolution given by the equations:

$$\dot{v_i}^k = -\gamma_v v_i^k(t) + F_{v_i}^k + A_{vi}^k \xi_v^k(t) \qquad (2)$$

$$\dot{a_i} = -\gamma_a a_i(t) + F_{a_i} + A_{ai} \xi_a(t) \qquad (3)$$

The first term on the right side of both equations is associated with the time response of the variables to general step stimulus with associated time constants γ_a and γ_v . The second terms F_{ai} and $F_{vi}{}^k$ reflect the deterministic characteristic response of the agents. The third term weighted A_{ai} and $A_{vi}{}^k$ respectively, ξ_a and $\xi_v{}^k$ represent the random individuality of each agent. We followed Schweitzer's model with some minor changes. In order to have a steady state zero value response of the agents on the absence of information, and also to reach real solutions for the differential equation response, we chose the term $F_v{}^k$ considering each candidate valence to be given by:

$$F_v^k = h^k(t)(b_1 v_i^k(t) + b_3 v_i^{k^3}(t)) \qquad (4)$$

The field $h^k(t)$ is associated with the quantity of news introduced into the community concerning each candidate or party that the agents have access to. This quantity is also inversely modulated by a general field $h(t)$ which represents the voting debate within the community generated by the expression of the agents as they talk back about candidates or parties. It is modelled by:

$$h^k(t) = \frac{'news\ about\ k'}{h(t)} \qquad (5)$$

And $h(t)$ is modeled by:

$$\dot{h}(t) = -\gamma_h h(t) + N_a \qquad (6)$$

N_a represents the number of agents having enough magnitude of arousal to express their political opinion. This condition is attained if $a_i > \tau$ (with $\tau_{min} \leq \tau \leq \tau_{max}$) and assume that the agent i will express itself. This way each agent has a notion of the media news about the candidates or parties but also the moderated interference of information field $h(t)$ of debating that lowers the impact of the news. The field $h(t)$ obeys a dynamic evolution, incorporating a time constant γ_h , and a dependence on the number of agents. When $a_i \leq \tau$ we assume the agent i does not convey any meaning to the debate so it is reset to zero.

Having chosen a uniform probabilistic distribution for τ, the deterministic term for the arousal is given by:

$$F_a = \hat{h}(t)(d_0 + d_1 a_i(t)) \qquad (7)$$

Where ĥ characterizes the average of $h^k(t)$ that influences each of the v^k:

$$\hat{h}(t) = \langle h^k(t) \rangle \qquad (8)$$

Figure 7 depicts the flow of information between the agents and the outside data collected from the Twitter stream in the multi-agent model. The agents receive the small signal of news $h^k(t)$ mixed with informational noise $h(t)$ that the community produces (Equation 5). The level of $h(t)$ depends on a damping constant γ_h and also on the number of agents with high *arousal* N_a (Equation 6). The signal h^k is injected in the deterministic component of *valence* F_v^k (Equation 4) which determines the voting of the agents in the community (Equation 3). By its turn, *arousal* dynamics (Equation 2) depend also on this signal through their deterministic component F_a (Equation 7 and 8). The output valence of the community, represented by a voting, is compared with the real tweets by the population collected from the Twitter stream using an average over the ensemble of collected tweets of a cosine similarity measure C_r (Equation 9).

$$C_r = \langle \overrightarrow{\arg \hat{\max_k} \{v_i^k(t)\}} \cdot \overrightarrow{T^k(t)} \rangle \qquad (9)$$

The vector $\overrightarrow{T^k}(t)$ represents the number of tweets collected from Twitter with the six or five dimensions of each of the elections, six presidential candidates or five parties.

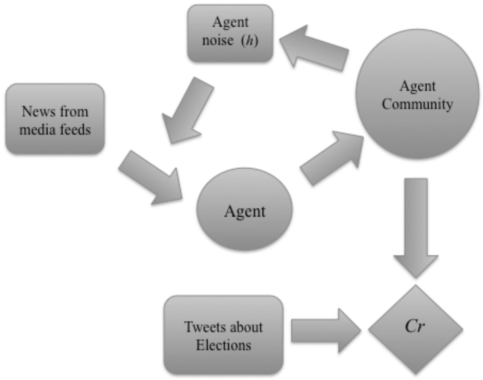


Figure 7: Flow of information on multi-agent model.

Arousal

In Figure 8 we examine the dependence between the similarly measure C_r and the time constants γ_h and γ_a. The constant γ_h determines, through Equation 6, the impulse time response of the information field $h(t)$ to bursts of *arousal* of the agents. Also the constant determines, through Equation 2, the impulse time response of this *arousal* in each agent i. We have chosen values for the parameters of the simulation so the system would be in a excited regime where at least one agent should have arousal greater then τ in the community at any time (b_1=+1.0,b_3=-1.0,d_0=0.05,d_1=0.5, τ_{min}=0.1, τ_{max}=0.9) [Schweitzer, 2010]. We made A_{ai} and A_{vi}^k negligible so the system is leveraged only lightly into a dynamic regime. Using this configuration of parameters, the community of agents, as can be examined on monitoring their *arousal*, expresses bursts of emotion equivalently to bursts of debate between Twitter users. The time profile of these bursts is mainly determined by γ_a with lower values increasing the frequency of the bursts. In the two charts of Figure 8 depicting the relation between γ_a and γ_h, we may see there is significant dependence under this constant. When the frequency of bursts is low, the similarity tends to be better. Also when the smoothing in the community informational field is higher, determined by higher γ_h also the similarity is better. These results might be expected. If we attend to Equation 5, low bursts and smooth informational noise are favourable to the clearer impact of news in $h^k(t)$. In presidential elections we notice that the dependence in γ_h is less pronounced. It is in fact almost absent. This means that the oscillations in $h(t)$ don't impact much in the importance of news, which may be due to the particular profile of the stream.

Valence

While *arousal* measures the degree in which the emotion encourages or unencourages activity, *valence* expresses the positivity or negativity of that emotion. Based in valence the agents chose the candidate or party they will vote. In Figure 9, the dependence between the time constant of valence γ_v and γ_a is depicted. As we've already pointed out, the increase in the smoothing for larger γ_h is favourable to a good fit between the simulation, news and real tweets. This is because we presuppose that the tweets in the community are highly correlated with the news, as we've seen in the previous section. As with the other two variables, γ_v represents a time constant of the impulse response for the valences v_i^k of each agent given by Equation 3. Low γ_v means more oscillating valence of the agents. High γ_v is associated with a more constant opinion. In both our models, for the presidential and the legislative elections, we confirm as depicted in Figure 9 that lower values are associated with a better correspondence between simulation and real tweets. Looking into the details of the stream of tweets, a weekly periodic oscillation is observable. We chose a ratio of 1/7 between days and ticks of the simulation in order to detect any harmonic correlation, without any significant result. However we found, as it was expectable, that less smoothing of *valence* is associated with better similarity. In fact a more oscillatory expression of *valence* better translates the influence of news in the voting of the community. It seems that *valence's* volatility allows a more diverse expression of the agents and consequently a more fine grained emulation of news impact. This interpretation is a field approach; each agent is exactly like any other. It is only the average influence of news in the agent community dynamics that happens to modulate the voting. Is this later mechanism, and the high similarity obtained (0.8~0.95) with the real stream of tweets, that happens to be quite remarkable.

The anonymity of the system, somehow confirms the intuition that it is not the individual choosing of each agent/'Twitter user' that determines the similarly between news and tweets, but instead the aggregate result. People tend to tweet in synchronicity with the news stream, and after selecting the impact on the news about each candidate or party, as much news arrive simultaneously, with the same magnitude as the final elections results.

3.2 Social Impact Model

Social Impact Theory created by Bibb Latané [Nowak et al., 1990] in 1981 is defined over three fundamental rules:

- Social impact is the result of social forces including the strength of the source of the impact.
- The amount of the impact tends to increase as the number of sources increases.
- The more targets of impact exist, the less impact each individual target feels.

After analysing the temporal influence of news with the Schweitzer Brownian model, we analyse next the spatial influence of news in terms of network topology using the known Twitter community of users. In the next section we will use a social impact model to examine in more detail the influence of ego network topology in the diffusion of news. We will also examine the influence of agent's memory in the emulation of real election results.

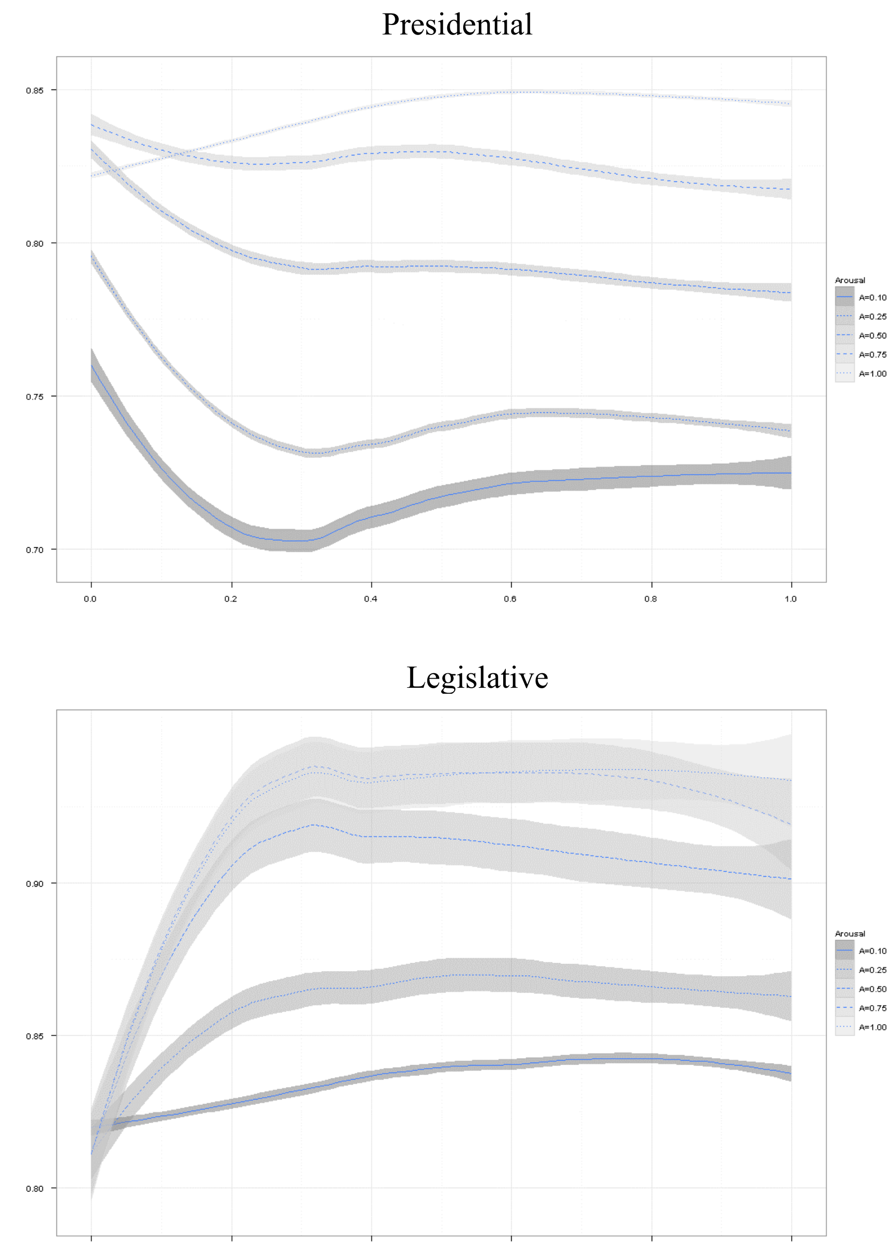


Figure 8: Chart γ_h of versus γ_a. Spline interpolation of average function of γ_a of cosine similarity on Brownian Model. Average of 20 runs, N=1000 agents, b_1=+1.0, b_3= -1.0, d_0=0.05 , d_1=0.5 , τ_{min}=0.1, τ_{max}=0.9, A_{ai} and A_{vi}^k negligible.

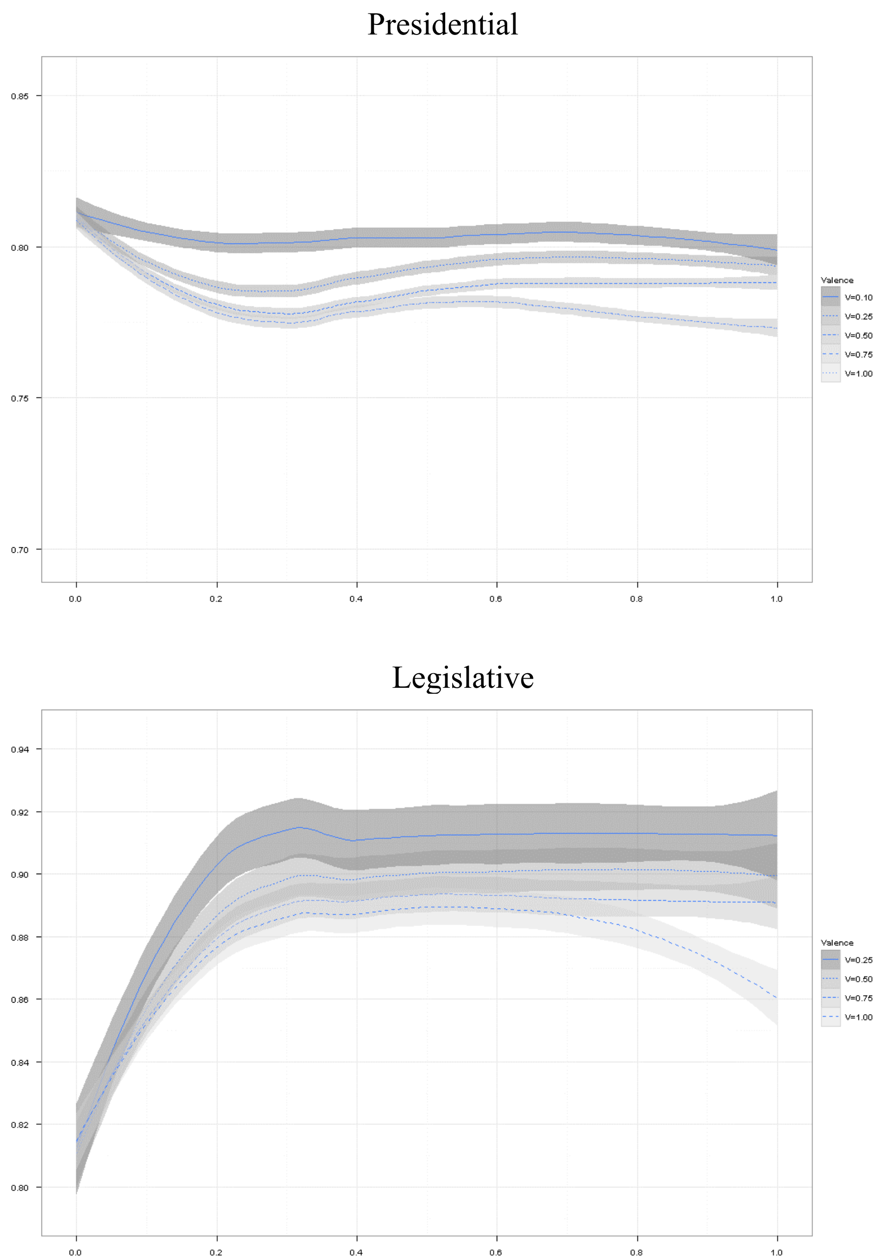


Figure 9: Chart γ_h of versus γ_v. Spline interpolation of average function of γ_v of cosine similarity on Brownian Model. Average of 20 runs, N=1000 agents, b_1=+1.0, b_3= -1.0, d_0=0.05 , d_1=0.5 , τ_{min}=0.1, τ_{max}=0.9, A_{ai} and A_{vi}^k negligible.

3.2.1 The network

Using the same collection of user data from Twitter we've extracted the graph of their original online social network. We've also synthesize several other networks having in common the same number of nodes. In Table 3 most common metrics over these networks are reported for comparison with the original.

	Twitter	B-A	E-R	K	W-S	Lattice
Number of Nodes	1903	1903	1903	1903	1903	1904
Average Degree	48,790	3,997	184,703	27,014	8,000	3,858
Diameter	5	8	4	6	8	133
Average Path Length	2,292	4,371	2,442	3,786	4,776	45
Density	0,026	0,002	0,020	0,008	0,004	0,002
Modularity	0,173	0,517	0,142	0,568	0,714	0.526
Number of Communities	5	26	13	8	14	441
Average Clustering Coefficient	0,270	0,019	0,020	0,260	0,339	0
Total triangles	335496	78	9059	18195	5937	0

Table 3: Table Comparison of common metrics between the Twitter network and five other synthetic network models.

It has been claimed that social networks generally present node degree distributions exhibiting power law behaviour. In Table 3 we can see that there is no other generative synthetic network that completely resembles our collected social network. Some studies report that Barabasi-Albert networks, implemented with mixed generative models of *preferential attachment* and *uniform growth*, may present a free scale power law distribution [Albert et al., 2002]. Apparently *preferential attachment* generative process is present in many natural and social networks [Newman, 2010]. From networks of routers in the internet, Hollywood actors, scientific paper citations, size of cities and many other phenomena, *preferential attachment* mechanism seems to be present and adequately explain the process of growth. However, looking at the node degree distribution of our Twitter network depicted in Figure 10 we see there is no large extend correspondence between the power law behaviour of a scale-free network and the real topology of our Twitter community. Sala et al. [Sala et al., 2011] recently showed that the degree distribution of social networks is better characterized by a mixture of power law and lognormal degree distribution - a Pareto-Lognormal distribution:

$$f(x) = \beta x^{\beta-1} e^{(-\beta\mu + \frac{\beta^2\tau^2}{2})} \Phi^c(\frac{\log x - \mu + \beta\tau^2}{\tau}) \qquad (10)$$

Using the values: β=1.2; μ=4.0 and τ=1 we've confirm a significant better fit to the complementary cumulative degree function of the Pareto-Lognormal, which may presuppose that this generative model (Equation 10) is far more adequate to explain the process of growth of Twitter networks. The authors devised a two-phase iterative algorithm that integrates fundamental properties from the *law of proportional effects* and *preferential attachment*. The algorithm alternates between adding new nodes to the network using a *preferential attachment* model, and growing the connectivity among nodes using the law of *proportionate effects*. Somehow it seems also intuitive as a human behaviour: popular people tend to attract connections, but also when connecting tend to connect more than less connected people. When browsing any community of people in Twitter we may in fact testify this process. With some remarkable institutional user exceptions, general users with many followers, tend also to follows many more people than less connected users. This model seems to fit quite well with our experimental data.

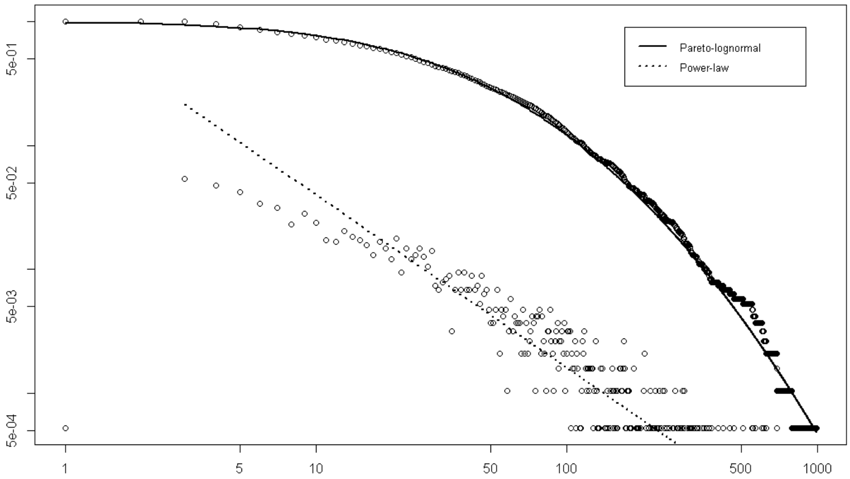


Figure 10: Degree distribution of Twitter network and complement cumulative Degree Distribution. Comparison with correspondent power law fit (α=1.396) and with CCDF of Pareto-Lognormal fit.

3.2.2 Multi-Agent Model

One of the possible mathematical models of social impact theory, which can easily be implemented in a multi-agent platform, is given by the set of equations [Holyst et al., 2001]:

$$I_i = \sum_{j=1}^{N} \frac{p_j}{d_{ij}}(1 - \sigma_i \sigma_j) - \sum_{j=1}^{N} \frac{s_j}{d_{ij}}(1 + \sigma_i \sigma_j) \qquad (11)$$

$$\sigma_i(t+1) = -sign(\sigma_i(t) I_i(t) + h_i(t)) \qquad (12)$$

Here I_i is the impact suffered by agent i from the community. Each agent i has a certain degree of *persuasiveness* p_i and a degree of *supportiveness* s_i reflecting the strength of interactions with individuals holding opposite or the same opinion respectively. The opinion variable is σ_i, which is bipolar, having the values +1 and −1. The variable d_{ij} corresponds to the distance between the agents, and the variable $h_i(t)$ represents some kind of community opinion noise that interferes in the social impact phenomenon.

In order to implement the voting, equation 12 is changed to:

$$\sigma_i^{k'}(t+1) = \sigma_i^k(t)(I_i^k(t) + h_i^k(t)) \qquad (13)$$

with additional step of normalization:

$$\sigma_i^k(t+1) = \begin{cases} +1 \; k = \arg\max_k(\sigma_i^{k'}(t+1)) \\ -1 \quad \text{otherwise} \end{cases} \tag{14}$$

This normalization substitutes the $-sign()$ operation in 12. The variable h_i^k in 13 represents the flow of news information that enters the community. The modelling for the voting process consisted in a adaptation of the original opinion model, having each agent a vector of opinions with $k \in$ { '*Cavaco*' , '*Alegre*' ,'Nobre' , 'Lopes', 'Moura', 'Coelho'} or $k \in$ { '*PSD*' , '*PS*' , '*CDS*' , '*PCP*' , '*BE*' }.

3.2.3 Experiment

The multi-agent model was implemented over a connection network identical to the Twitter network that was sampled. As well as with the Brownian agent simulation, the flux of information noise in equation 13, that influences the overall community, was derived from the stream of news collected. On our implementation we've tried different approaches for the values of the two parameters p_i and s_i. Having no special criteria to distinguish one agent from the other, we chose to value these parameters as random numbers over a Normal distribution with mean 1 and standard deviation 0.5. The distance parameter d_{ij} was considered of one hop only. We've tried larger reach for the impact I_i, with severe computational time degradation. The main purpose of our experiment was to examine topological influence; so more diffuse and global reach would compromise structural resolution. We opted to simulate influence over just 1 hop, leaving to other research work the study of long-range influence.

The experiment was divided into three distinct tests:

- In the first test, we evaluate the impact of varying the media coverage in the community. For this purpose an additional agent was created that partly disseminates news into the network.
- In the second test we evaluate, in both elections, the impact of varying the time lag between debate and news. For this purpose a damping factor δ on I_i was introduced in the model.
- In the final test we evaluated the impact of changing the network topology. For this purpose a percentage of the links original network are randomized.

Media coverage

In order to test the influence of media coverage in the network opinion dynamics, an additional agent is randomly wired to the community. The number of agents with which this agent is connected varies between 10% and100% of the total. The stream of news is directly injected into the agent, being its opinion defined by Equation 14 relative to the news flow, as a social impact. The agent then propagates its opinion to the community as a real tweet. Figures 11 and 12 depict the difference between the votes in the community and real final election results. The votes for each candidate/party are counted in the community, and then a maximum likelihood estimation of the mean, for 30 complete runs (campaign runs) of the simulation is performed. The charts represent the difference between this estimate and the elections in the population. There is a notorious increase in similarity, particularly in the legislative elections, when media coverage is increased. Notably there is a pronounced error reduction around 60% of coverage, which is considerably stronger in the legislative elections. This reducing of error of agent simulation against real election results by increasing media coverage in the agent community, in both elections, seems to parallel and complement identical correlation already found in the Brownian model between and cosine similarity.

In order to better test this hypothesis, a similar essay was devised, this time within a simple regular lattice. In this essay, which results are depicted in Figures 13 and Figure 14, we notice that the dependence on media coverage is extremely more linear. There seems that somehow the simplicity of the network, with his monotonic degree distribution, eliminates the sophisticated behaviour that the real network presents. In fact, as the random re-wiring of the additional news agent in each run of the simulation is geometrically identical, as all agents present the same ego network topology, we should expect that the increase in the percentage of wiring of this agent would have a linear proportionate impact. This is the case. A linear increase in news reach translates in a linear reduction in voting error, thus confirming that the network topology has a determinant effect on the voting of the agents.

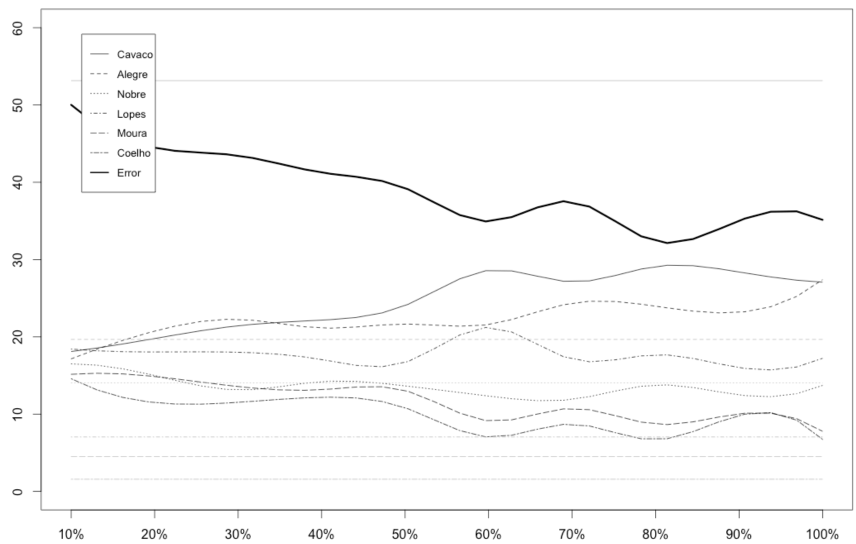


Figure 11: Chart of the difference error (darker line) between the MLE of the mean and final election results, of 30 runs of the simulation, function of percentage of media coverage. Original network. Presidential elections. Estimating errors : 0.464, 0.417, 0.328, 0.430, 0.301, 0.267. N=1903 agents, δ=0.

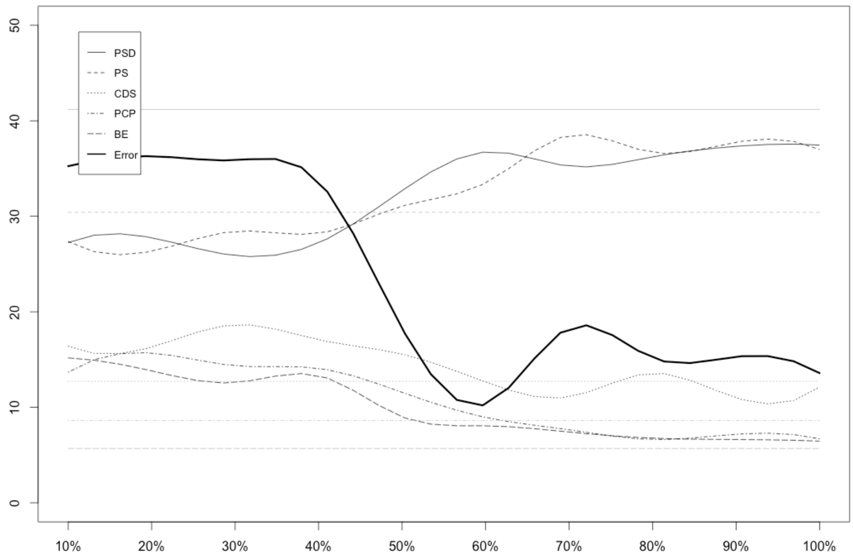


Figure 12: Chart of the difference error (darker line) between the MLE of the mean and final election results, of 30 runs of the simulation, function of percentage of media coverage. Original network. Legislative elections. Estimating errors : 0.501, 0.472, 0.378, 0.305, 0.246, N=1903 agents, δ=0.

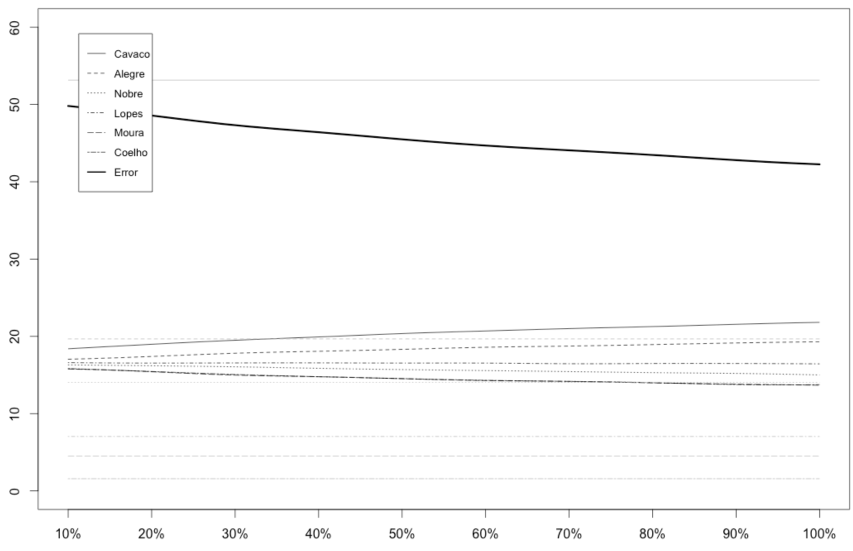


Figure 13: Chart of the difference error (darker line) between the MLE of the mean and final election results, of 30 runs of the simulation, function of percentage of media coverage. Lattice network. Presidential elections. Estimating errors : 0.173, 0.057, 0.050, 0.053, 0.048, 0.044. N=1903 agents, δ=0.

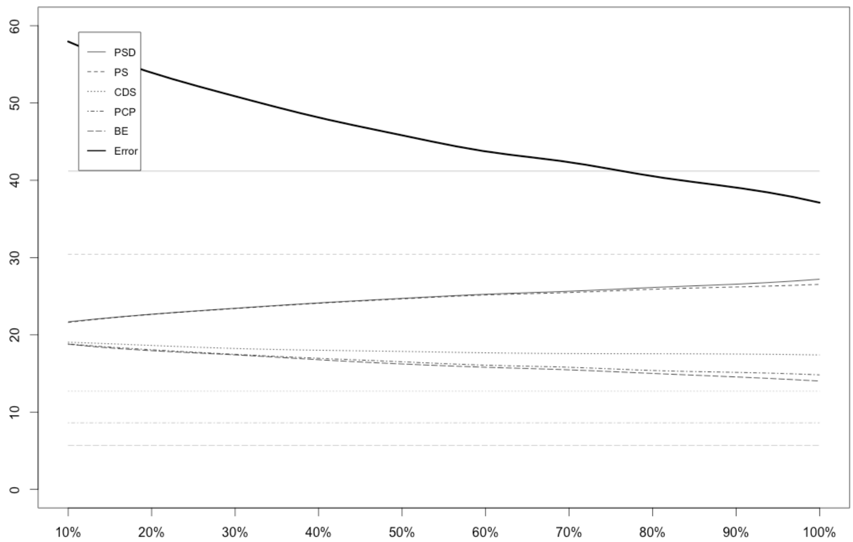


Figure 14: Chart of the difference error (darker line) between the MLE of the mean and final election results, of 30 runs of the simulation, function of percentage of media coverage. Lattice network. Legislative elections. Estimating errors : 0.121, 0.090, 0.083, 0.065, 0.070. N=1903 agents, δ=0.

Network topology

The second test we've performed with the social impact model was to partly change the network topology of the original network in order to detect how significant its impact is in the propagation of news. In Figures 15 and 16 the error against final election results is compared, function of the proportional randomization of the network links. A dependence on the original network topology is noticeable. In the legislative election case, a randomization of only 2% of the network seems to better adjust the influence of news in the network with the final results. This fact should not have any particular significance other than as noticeable large modifications of original network, towards an Erdós-Renyi random graph, tend in fact to influence the patterns of news propagation towards less realistic results. In the presidential election this difference is not so strong. We may conclude that the free-scale topology of the network, may impact the actual news propagation, having as given opinion dynamics model the social impact theory.

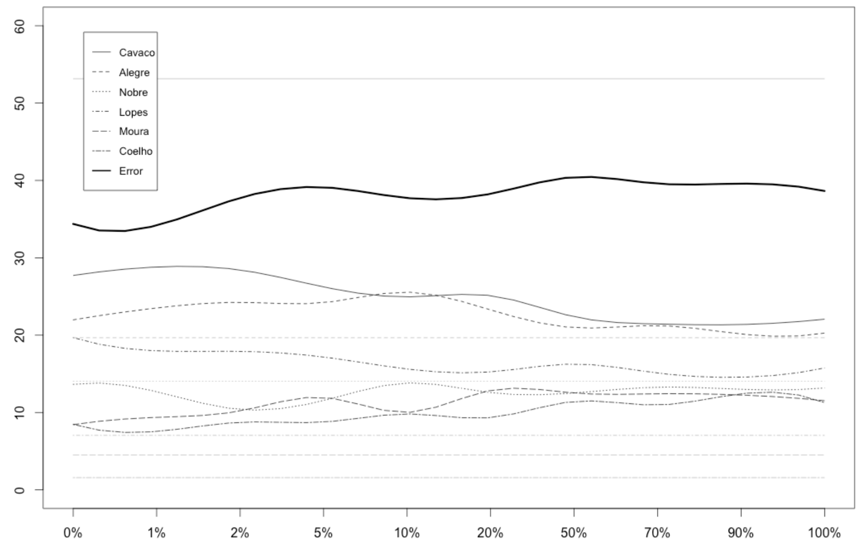


Figure 15: Chart of the difference error (darker line) between the MLE of the mean and final election results, of 30 runs of the simulation, function of percentage randomization on original network. Presidential elections. Estimating errors : 0.527, 0.487, 0.365, 0.405, 0.372, 0.325. N=1903 agents, media coverage of 60%, δ=0.

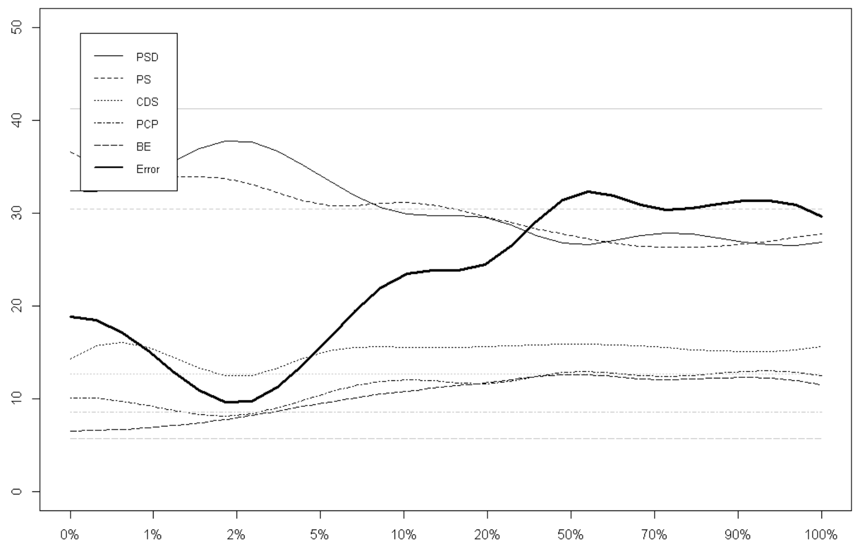


Figure 16: Chart of the difference error (darker line) between the MLE of the mean and final election results, of 30 runs of the simulation, function of percentage randomization on original network. Legislative elections. Estimating errors : 0.563, 0.529, 0.395, 0.343, 0.265. N=1903 agents, media coverage of 60%, δ=0.

Time lag

Contrary to the Brownian model, the classical social impact model doesn't convey any memory effects. In order to test the model in which concerns the memory of the agents, an additional variable δ was introduced that provides a delayed social impact in the actual agent's opinion. Equation 13 is now re-written as:

$$\sigma_i^{k'}(t+1) = \sigma_i^k(t)[I_i^k(t) + \delta I_i^k(t-1) + h_i^k(t)] \quad (15)$$

Figures 17 and 18 show that the impact of this new configuration of the model on the error of prediction. In the absence of lag, the model has significantly less error. This effect is far more pronounced in the case of the legislative elections. Taking into account the other charts already examined, this fact seems to point out that the collected tweets corresponding to the legislative elections are significantly more correlated with election results, than in the presidential case. Examining the process of recollection, it happened that in the case of the presidential, the tweets corresponding to the candidate 'Cavaco' are related not only with the presidential campaign, but also with current affairs of 'Cavaco Silva' as president of Portugal. This may explain the less correlation within an election context with the other candidates.

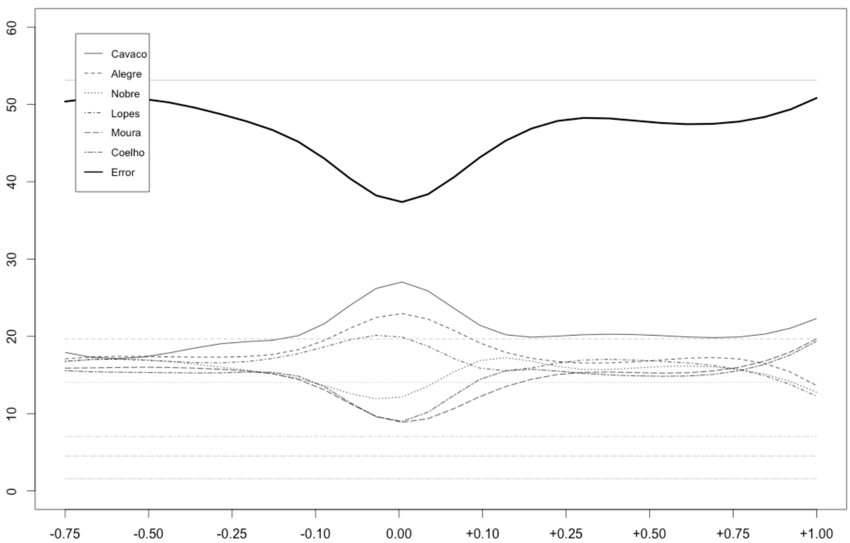


Figure 17: Chart of the difference error (darker line) between the MLE of the mean and final election results, of 30 runs of the simulation, function of δ. Lattice network. Presidential elections. Estimating errors : 0.442, 0.416, 0.302, 0.388, 0.248, 0.245. N=1903 agents, media coverage of 60%.

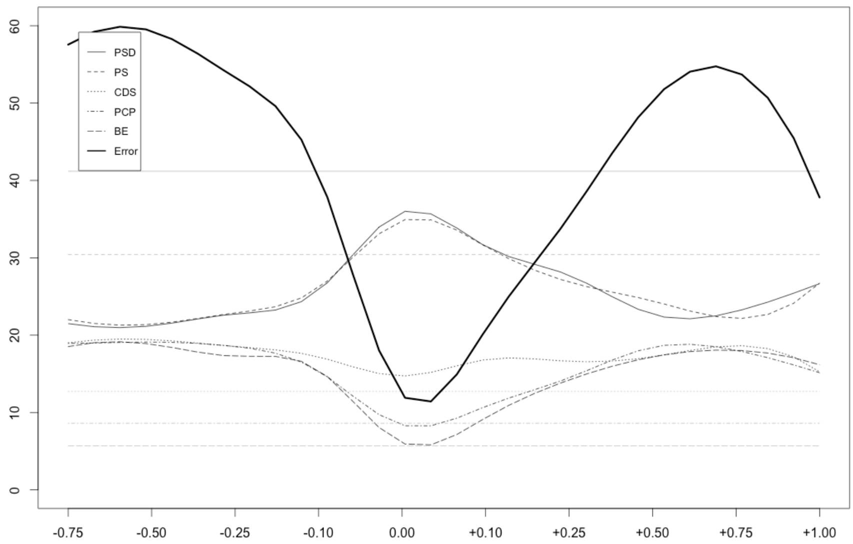


Figure 18: Chart of the difference error (darker line) between the MLE of the mean and final election results, of 30 runs of the simulation, function of δ. Lattice network. Legislative elections. Estimating errors : 0.520, 0.518, 0.432, 0.352, 0.323. N=1903 agents, media coverage of 60%.

4 Conclusions

The twitter communication on political elections we studied in this paper has several properties worth pointing out:

1. The flow of tweets talking about each candidate or party during the three months that preceded the elections has a relative magnitude that follows closely the quantity of daily news produced about that candidate or party. This fact, allow us to conclude that, when talking about politics, people tend to talk about news events, and particularly about the news of the day. Examining the particular Twitter content just confirms this fact. More significantly, we also found that the proportion of tweets and news for each candidate or party, will be good estimative of the elections final result. Although with significantly less precision than classical pools and whenever close ties aren’t present, the relative position of candidates or parties can nevertheless be predicted.

2. Simulating the opinion dynamics of the voting expression with a multi-agent Brownian model, we found that low levels of arousal of the agents are associated with more sporadic bursts of tweeting in the community, and consequently also with a better fit with real tweeting of the agents. As previously mentioned, we could testify this phenomenon, as tweeting about any subject in the Twitter community tends to be intermittent. People must not be highly responsiveness to the Twitter flow in order to mimic this pattern. In fact in very uncommon to have many people at once listening to

each other. Twitter isn't particularly a chatting platform although several dialogs may exist simultaneously. Also we noted that the level of informational noise about each candidate/party, that affects the agents as a whole, should normally be low. This is to say that discussion about each particular candidate or party should not be blurred by community discussion. This is also apparent in the Twitter chatting. We also found that de *valence* in the opinion of each agent should have low resilience, which is to say that each agent should discuss every candidate/party in brief periods. Within the resolution of one day this finding is also consistent with the normal characteristics of Twitter, which contrary to other social media like blogs, Facebook or YouTube tend to extend the duration of themes for several days or even months.

3. The degree distribution of the Twitter community network, for which we've collected online data, has a close fit with a Pareto-Lognormal probability density function. This distribution has a generative model reasonably fit to the one observed in the Twitter community. People tend to connect with famous people obeying a preferential attachment schema but at the same time having a proportionate effects weighting. People tend to follow people with many in links but the most followed also like more to connect with other people.

4. The social impact theory simulations we've implemented have shown that the media coverage is essential to the result similarity between tweets and elections results. The social impact model we implemented had a one hop network reach. With this pre-condition, larger reaching of the news is associated with more similarity between tweets and voting. The topology of the network also impacts the relation with real voting. Examining the opinion dynamics in a lattice network we may find that network complexity is determinant with real tweet emulation. Finally, the result obtained by time correlating news and tweets is confirmed through the introduction of a time lag constant. Confirming the above results we found, when opinion is determined by social impact influence, that best fit with experimental results is obtained when opinion within the community is expressed synchronously with news.

More work on online communication and topology influence on communication still remains to be done. Namely, a more detailed analysis of opinion influence patterns that can easily be implemented using more computational resources in the social impact model.

Also more work remains to be done in the fascinating subject of online democracy. Political marketing, online opinion pooling and Sociology in general, will all benefit from the studies of online social networking as a novel framework for structured social data gathering. The recent advances on areas of 'Big Data' information systems and 'Crowdsourcing' constitute the early steps of this fast growing sociologic revolution.

Multi-agent modelling can provide valuable intuitions in the comprehension of sociological phenomena. Since the early works of Thomas Schelling, with his segregation model, many socio-mechanisms have been hypothesised from the basic setting of simple agent rules. With the present work, we hope to have contributed to the less studied subject of validation of multi-agent modelling with real data, which constituted the main goal of our paper.

References

[Albert et al., 2002] Albert, R., and Barabási, A. (2002). Statistical mechanics of complex networks. In *Reviews of Modern Physics, Vol. 74, 1, pages 1–58*.

[Castellano et al., 2009] Castellano, C., Fortunato, S., and Loreto, V. (2009). Statistical physics of social dynamics. In *Reviews of Modern Physics, Vol. 81, 2, Page: 591–646*.

[Connor et al., 2010] Connor, B. O., Balasubramanyan, R., Routledge, B. R., and Smith, N. A. (2010). From Tweets to Polls : Linking Text Sentiment to Public Opinion Time Series. In *4th International AAAI Conference on Weblogs and Social Media, number May*.

[Clifford, 1973] Clifford, P., and A. Sudbury, (1973). A model for spatial conflict In *Biometrika 60(3), 581*.

[Deffuant, 2000] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch,(2000). Mixing beliefs among interacting agents In *Advances in Complex Systems, 3 (1-4), 2000*.

[Galam, 2002] Galam, S., (2002). Minority Opinion Spreading in Random Geometry In *European Physical Journal B 25,403-406*.

[Gayo-avello, 2011] Gayo-avello, D. (2011). Limits of Electoral Predictions using Twitter. In *ICWSM-11 Barcelona, Spain*.

[Goidel, 2011] Goidel, K. (2011). Public Opinion Polling in a Digital Age: Meaning and Measurement. In *Political Polling in the Digital Age*Louisiana State University Press, 11–27.

[Greenberg, 2011] Greenberg, A. (2010). Alternatives to Polling. In *Political Polling in the Digital Age*Louisiana State University Press, 11–27.

[Hess, 2006] Hess, C., Ostrom, E.(2006). Understanding Knowledge as a Commons From Theory to Practice. *The Mit Press*, Cambridge Mass.

[Hegselmann, 2002] Hegselmann, R., Krause, U. (2002). Opinion dynamics and bounded confidence: models, analysis and simulation. In *Journal of Artificial Societies and Social Simulation, Vol 5, 3*.

[Holyst et al., 2001] Holyst, J. A., Kacpersky, K., and Schweitzer, F. (2001). Social impact models of opinion dynamics. *Annual Reviews of Computational Physics*, 48(22):253–273.

[Ito, 2004] Ito J. (2004). Emergent Democracy. *Extreme Democracy*, http://www.extremedemocracy.com/.

[Keeter, 2011] Keeter, S. (2011). Public Opinion Polling and Its Problems. In *Political Polling in the Digital Age*Louisiana State University Press, 11–27.

[Kirzinger, 2011] Kirzinger, A., Goidel, K., Xenos, M. (2011). Too Much Talk, Not Enough Action? . In *Political Polling in the Digital Age*Louisiana State University Press, 11–27.

[Moss, 2005] Moss, S. and Edmonds, B. (2006). Towards Good Social Science. *Journal of Artificial Societies and Social Simulation*, http://jasss.soc.surrey.ac.uk/JASSS.html.

[Mutz, 2006] Mutz, D. (2006). Hearing the Other Side: Deliberative versus Participatory Democracy. *Cambridge University Press*, Cambridge, USA.

[Newman, 2010] Newman, M. (2010). Networks: An Introduction *Oxford University Press*, Oxford, USA.

[Nowak et al., 1990] Nowak, A., Szamrej, J., and Latané, B. (1990). From private attitude to public opinion: A dynamic theory of social impact. *Psychological Review*, 97(3):362–376.

[Ostrom, 1990] Ostrom, (1990). Governing the Commons: The Evolution of Institutions for Collective Action *Cambridge University Press*.

[Pak and Paroubek, 2010] Pak, A. and Paroubek, P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In *7th conference on International Language Resources and Evaluation (LREC'10)*, pages 1320–1326.

[Sala et al., 2011] Sala, A., Gaito, S. , Rossi, G. P., Zheng, H. and Zhao, B., Y.(2011). Revisiting Degree Distribution Models for Social Graph Analysis. *arXiv:1108.0027v1 [cs.SI]*, July 2011.

[Shirky, 2008] Shirky, C.(2008). Here Comes Everybody: The Power of Organizing Without Organizations. *Penguin USA*, February 28, 2008.

[Shirky, 2011] Shirky, C.(2011). The Political Power of Social Media. In *Foreign Affairs*, January/February 2011.

[Schweitzer, 2003] Schweitzer, F.(2003). Brownian Agents and Active Particles. *Springer Series in Synergetics*, 2003.

[Schweitzer, 2010] Schweitzer, F.(2010). An agent-based model of collective emotions in online communities. In *European Physical Journal B, vol. 77, no. 4 (2010) pp. 533-545*.

[Sznajd, 2000] K. Sznajd-Weron, J. Sznajd (2000). Opinion evolution in closed community. In *International Journal of Modern Physics*, C 11, No.6 2000, 1157.

[Sobkowicz, 2009] Sobkowicz, P. (2009). Modelling Opinion Formation with Physics Tools : Call for Closer Link with Reality. *Journal of Artificial Societies and Social Simulation*, 12(1).

[Tumasjan et al., 2010] Tumasjan, A., Sprenger, T. O., Sandner, P. G., and Welpe, I. M. (2010). Predicting Elections with Twitter : What 140 Characters Reveal about Political Sentiment. In *Word Journal Of The International Linguistic Association*, pages 178–185.

[Véronis, 2007] Véronis, J. (2007). Citations dans la presse et résultats du premier tour de la présidentielle 2007 Technical report.