

How fast societies reach consensus over elections?: a methodology to study the evolution of Twitter conversations before election day

Daniel Cárdenas-Sánchez

Universidad del Rosario

Andrés Miguel Sampayo

Universidad del Rosario

Maykol Rodríguez-Prieto

Universidad del Rosario

Alejandro Feged-Rivadeneira (✉ alejandro.feged@urosario.edu.co)

Universidad del Rosario

Research Article

Keywords: democratic elections, Twitter, consensus

Posted Date: July 6th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-660016/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

How fast societies reach consensus over elections?: a methodology to study the evolution of Twitter conversations before election day

Daniel Cárdenas-Sánchez^{1,*}, Andrés Miguel Sampayo², Maykol Rodríguez-Prieto³, and Alejandro Feged-Rivadeneira^{2,*}

¹Escuela de Ingeniería Ciencia y Tecnología de la Universidad del Rosario

²Facultad de Estudios Internacionales, Políticos y Urbanos de la Universidad del Rosario

³Facultad de Economía Aplicada de la Universidad de Santiago de Compostela

*daniel.cardenass@urosario.edu.co

*alejandro.feged@urosario.edu.co

ABSTRACT

Technological developments in media and communications such as press, radio, and television have disrupted electoral processes and reshaped political landscapes. Similarly, the development of surveys enabled the social sciences to study the voting processes and measure consensus in a population prior to democratic elections. Literature on social networks and elections has focused on predicting electoral outcomes rather than understanding how the discussions between users evolve over time. As a result most studies focus on a single election and few comparative studies exist. In this article, a methodology to analyze Twitter conversations about election candidates is proposed. Using DeGroot's consensus model -an assumption that all users are attempting to persuade others to talk about a candidate- the methodology allows to identify the structure and strength of connections of the mention networks on each month prior to an election day. It also helps to make comparisons between elections and identify patterns on different contexts. In the end, an analysis on the elections where the incumbent was running and the political regime is presented.

Please note: Abbreviations should be introduced at the first mention in the main text – no abbreviations lists. Suggested structure of main text (not enforced) is provided below.

Introduction

The advent of surveys professionalized the measurement of consensus prior to elections in the early 1936 [16], but the methodology remained controversial in the political science sphere for over a decade since its introduction[16]. Years later, surveys have been established as the quintessential tool for studying elections [15], even though controversy about their relevance in current political climate and social contexts has been risen [7].

Similarly, a significant body of literature exists about the role of mass media technologies on electoral processes involving press, radio and television [3] [14]. These technological disruptions of electoral processes were deemed of such importance that governments around the world imposed strict regulations about the type of contents that could be shown, and, most importantly, the type of actors that were allowed to broadcast [3] [12]. The use of mass media to manipulate and form public opinion leading to electoral outcomes has been studied widely and in diverse contexts: the first electoral failure and subsequent rise to power of the *Nationalsozialistische Deutsche Arbeiterpartei* in Germany has been partly attributed the shift from strict regulations imposed on radio broadcasting in the country, and later to the appropriation of the technology by the party [1]. Similar examples of the United States of America [3], Japan [9], Greece [8] and Brazil [11] suggest that radio regulation by the establishment played an important role in the formation of public opinion for political matters that often benefited government views [3].

With the advent of internet as a public service available for everyone in the decade of the 1990s, and the possibility of peer-to-peer communication at a massive scale, scholars warned about the possibility of the *Balkanization* of the sciences and other spheres of human communication [18]. Just two decades later, in the 2010s, the rise to power of candidates with extreme views in countries such as the US, Brazil, the UK, among others, social media and disinformation campaigns were seen as the culprit of the radicalization of voters [13][2]. Traditional mass media outlets such as press, radio and television were outspent by digital campaigns in Facebook, Twitter and YouTube, which allowed for micro-targeting of voters to trigger emotions associated with specific group profiles [19]. At the same time, surveys and voting intention analysis were perceived

by the public opinion to be unable to adequately predict the outcome of elections with the same degree of certainty that they exhibited just a decade earlier, however, detailed analysis suggests that error margins have not changed significantly over the years. As a result, the attention was drawn to social networks as a mean of predicting electoral outcomes to overcome the shortfall of surveys, with mixed results [17].

The emphasis on using social media as a prediction mechanism made researchers overview the role of these platforms as an electoral disruption. Just as the press, radio and television changed the way citizens formed opinions and preferences about an election process, social media has meant a new framework for electoral studies. Nowadays, the sources of information are not monopolized by any group, instead, opinions about facts are constantly published and shared among citizens. In this research, we look at the discussions on Twitter in the context of electoral processes and propose a methodology to analyze how the general conversation evolves towards one of the candidates.

The main objective of this investigation was to develop a framework to understand the evolution of political discourse from a perspective that combines aspects of mass media studies and consensus analysis prior to electoral processes under diverse contexts. As such, our first aim was to analyze the convergence/divergence of political debate as time to elections decreases from a modelling perspective that enabled us to understand and estimate the degree of consensus in a society. The second was to compare the degree to which the evolution of consensus in the political debate under different contexts is associated with one specific aspect of electoral results: the difference in votes between the winner and the runner up.

Results

To analyze how societies reach consensus over the candidates, a network of mentions was formed with the characteristic that at least one of the official candidate's account had to be mentioned. Then, the resulting network was assumed to be a naive individual network, as DeGroot[4] defines it, where each node is trying to persuade the others on a one-dimensional decision, in this case an election. This allows the methodology to use the second eigenvalue as a measure of the speed at which the network is converging. In other words, that measure shows how fast the people on Twitter agree on which candidate should be discussed. This last characteristic is easily perceived on this social network, and the mentions construct is a proxy for a direct attempt to change or support the attitudes and opinions towards one of the candidates.

In order to understand the behavior of the second eigenvalue on this particular type of graphs, a series of simulations of networks with different topologies and the same number of nodes were performed. The objective of this exercise was to determine how the eigenvalue changes according to different scenarios of candidate perception.

The results show that the second eigenvalue depends on two variables, one is the number of connections that the network has, also known as density (percentage of edges present in proportion with the maximum number), and another variable is the node structure, known as hubs. A node is a hub when it groups together a large number of nodes in a network, where each hub for the case study is a candidate. This nomenclature was adopted because it focuses on the candidates and, therefore, is different from the conventional communities.

Table 1 shows descriptive statistics for a density graph where the number of interactions that take place in the entire network is observed depending on the number of hubs. We will refer to the number of nodes as the structure of the network.

Table 1 shows a comparison between the density and the second eigenvalue change for different structures of the network and different link strengths. For the density dynamics, more hubs means more connections in the network, therefore, the density will increase.

The trend for the second eigenvalue is not linear: when there are two and three hubs, the value is similar, greater (slower convergence to consensus) than having only one hub. This results show that the structure of the network influence the speed of the convergence. The second eigenvalues here correspond to extreme scenarios of the simulation, determining boundaries of the second eigenvalues.

The networks were simulated creating the adjacency matrix from a series of random binary numbers using a Bernoulli distribution. This means that for every row of the matrix a random array of zeros and ones was created and the probability of the ones varied to create different scenarios. As Table 1 shows, there were two main variables used to create the scenarios: the structure and the probability of the links. The first one refers to the number of hubs, sources of information, that the graph had. The later is the probability that any nodes are connected following two rules. The probability of connection of any nodes must be smaller between different communities than inside the same one, which means that it is more likely that supporters of the same candidate are connected. And the other is that the probability of connection with the candidate and her supporters should be bigger than with the ones of other candidates.

One of the objectives of the simulated networks was to be as close as possible with the ones obtained from real life. However, the used scenarios reflected extreme situations in comparison with the real ones. To achieve this, a set of parameters were determined to get second eigenvalues inside the range of the ones from real elections and with a similar edge density. The

values determined were:

- probability between supporters of different candidates = strength value
- probability between supporters of the same candidate = strength value * 5
- probability between candidate and her supporters = strength value * 15

As a result, different scenarios were created changing the probabilities of the links by the same factor. This allowed to keep the two characteristics defined on every scenario. The factors used are describe in the Table 1, which in sum is a reduction on the size of the probability in a scale of ten. To avoid noise from a single simulation a bootstrap was conducted and the second eigenvalue was the average of a hundred repetitions.

Table 1 shows that an increase in the probability between the connections of the nodes the greater the density on the network, which means that the second eigenvalue is bigger in magnitude in general. However, it is evident that when comparing the presence of hubs in the different simulations setting a connection probability, a single hub in the network establishes a lower magnitude with respect to the presence of two or more hubs.

On the other hand, the simulated network that represents the existence of two hubs that are connected to each other has a smaller magnitude in the second eigenvalue compared to the networks that have two or three hubs (Table 1, Figure 1). As the hubs increase, the magnitude of the second eigenvalue increases. To compare with the real elections for every election a graph was modeled for each one of the months prior to the election, starting from five months before the election day.

Table 2 shows the second eigenvalue for each of the elections studied from 151 days before the election to days close to the Election. Mention networks for a selection of the observed elections are shown in Figure 2. The magnitude of the second eigenvalue translates speed into the convergence of electoral decisions, a greater magnitude represents slower convergence.

For the elections in Argentina in 2015, the second eigenvalue shows that as election day approached the convergence was slower (15.8), unlike 151 days before the election where the second eigenvalue was at 5.0. For the 2019 elections the situation was different, the range in which the second eigenvalue fluctuated was between 1.0 and 3.0.

For the Indian elections in 2014, convergence was slow reaching a second eigenvalue of 24.2 30 days after the election and for the previous days the eigenvalue was not below 9.3. For the year 2019, the second eigenvalue had a significant growth from 6.0 to 120 days before the election but after that it had a rapid convergence, remaining between 1.0 and 2.0.

As mention before, the second eigenvalue is a measure of how fast is a network converging. However, it is unknown which one of the candidates was more mentioned on each month. To take this into account a clustering using Topological Data Analysis was done. This analysis shows, on each month, how the Twitter users mentioned each candidates. In Figure 4 a sample of Argentina 2015 and France 2012 is presented on a color scale. These figures show how Twitter users tend to talk more of a single candidate or two or as the elections approach. Ideally, if consensus is increasing as voting day approaches, all the groups would tend towards a single color. Figure 4 shows a depiction of observed and simulated networks using TDA.

To see if characteristics of the political system influence the convergence, two aggregations were made. The first one was by type of regime. Figure 3, shows that the second eigenvalue in presidential regimes was generally lower. Similarly, the tendency to mention the winning candidate is more common in parliamentarians than in presidential ones. Towards the end of the election, the second candidate in the presidential elections is mentioned more often. In the presidential elections, as time passed, the mentions of the third candidate began to fall.

If the incumbent president or prime minister was a candidate for re-election, the second eigenvalue was less than when the incumbent executive did not participate. The tendency of the winner was similar whether there was a possibility for reelection or not. However, the trends of the other candidates differed considerably. For reelections, there was more talk about the second candidate, but the third it was almost nonexistent (in terms of the Twitter network). While the elections in which reelection was not sought, the trend of the runner-up was similar to that of the third candidate, as shown in Figure 3.

Discussion and conclusion

The most important finding of our study was that social network data in the context of electoral campaigns can be used to measure a novel variable based on theoretical models of consensus to describe the months prior to suffrage [4]. Our findings are consistent with literature as follows: First, we found that elections where the incumbent is up for reelection have a different variable profile than those without the incumbent. This finding is consistent with reports of literature of mass media in elections, where the incumbent is found to have a competitive advantage because she has more presence in radio and television[3]. Second, we find that total volume of mentions for a particular candidate is indicative that they pose a chance to win the elections, but that it not a robust method to forecast elections, as seen in Figure 5[6]. Furthermore, while our simulations show that a denser network will have higher lower levels of consensus, observed electoral networks can be convergent or divergent (decreasing or increasing consensus, respectively), evidence that participation in social networks does not necessarily polarize. Third, because our method is a systematic depiction of structural processes of consensus formation, we find that phenomena that has been difficult to define with surveys, such as the *band wagon effect*, can be defined precisely with this method in terms of two variables: mentions for each candidate, and change in the second eigenvalue as the elections approach. We found that elections that resulted in a narrow margin between candidates also presented the lowest values for consensus (measured by the 2nd eigenvalue of the DeGroot model).

In Argentina 2019 and India 2014, for example, the volume of mentions per candidate differed from the outcome of the election, as shown in Figure 3. In Argentina 2019, there was more talk about the president, who finished second. On the other hand, India 2014 shows a typical case where the winner is further discussed. On the former, the methodology allows to see that the evaluation of the presidency dominated the topics of discussion, leaving little room for the other candidates. In contrast, in India 2014 the most discussed candidate was in fact elected. Our methodology, therefore, shows both who is dominating the discussion whether consensus exists, rather than predicting the outcome.

Therefore, the evolution of the conversation on Twitter shows how the society reaches a consensus around each candidate with chances of winning. This research shows that the characteristics of the political regime (presidential elections faster than parliamentary) and the fact that the incumbent is running (incumbent elections faster than non-incumbent's) changes the speed at which the consensus is achieved. In other words, information on the social network will flow faster or slower between potential voters. That speed is measured by the second eigenvalue of the mention matrix and depends on the number of edges that the network has (more connection leads to faster convergence) and the structure of the connections (more connections between communities means faster convergence). By construction, the structure of the network reflects an "unwise society" in the terms described by [10], there are three or two nodes -the candidates- deeply connected to the rest of the network, which creates a structure of communities for each candidate. As a result, users that tweet about more than one candidate help the network achieve consensus faster.

Most studies about elections and social networks focus on single elections, this research wanted to see the patterns present in different contexts. For this reason, a variety of years and countries were considered and, specially, with variation in the fact that the incumbent was running for reelection and the political regime. The results show that the participation of the incumbent is associated with more mentions of the winner and the runner up, specially in the months closest to the election. In comparison, in non-incumbent elections the first candidate is mentioned around 20% more, and the second 30% less (Figure 3). On the other hand, the differences in the regime shows that parliamentary elections have on average a bigger second eigenvalue than presidential (Figure 3), which means that in the later convergence is achieved slower.

According to Golub and Jackson[10], the second eigenvalue reflects how fast a society is converging over a one-dimensional decision, in this case an election. This means that the interactions between people in Twitter, assuming that this social network is representative of the entire society, should evolve to reflect the same preferences the citizens have in real life. Therefore, three different measures should be aligned, the distance between the two main candidates of the election -as a measure of the level of agreement on one of the candidates-, the second eigenvalue -as a measure of the speed of convergence-, and the percentage of Twitter mentions -as a measure of the agreement on the social network-. As seen in Figure 5, in some elections the three indicators are aligned, a higher second eigenvalue election has close real life results and mentions, and vice-versa. However, others did not follow this logic because, as seen in the literature, the Twitter mentions do not reflect accurately the preferences of society, and due to the fact that densities in the networks influence the behavior of the second eigenvalue, as seen before.

As a result, it should be established that these results have limitations. The first one is that the sample of elections used is very limited to determine a definitive difference between regime types and incumbent participation. This analysis was performed anyway because of the difficulties of gathering the Twitter data (there are millions of tweets per election) and the need to have a first glimpse on the differences on social network conversations between contexts and years.

In conclusion, our study proposed a framework for analysing election conversations on different contexts on Twitter with empirical examples. Even when convergence is discussed, this methodology does not aim to predict winners; rather, it attempts to get a better understanding of the opinion formation process in the electorate and the effects that campaigns achieve on social networks. Comparison between elections allows analyses of polarization and campaign effectiveness to be done in a more standard manner. Since elections are a one-dimensional event, the DeGroot model works with the assumption that every tweet that mentions a candidate is trying to persuade the other members of the network to discuss one of the choices. Finally, these results are open for improvement, adding sentiment analysis could help establish the way broadcasting effects work, also adding campaign milestones could help get a better understanding on the changes of the speed of convergence. More elections data is needed to establish stronger claims on differences between incumbent and non-incumbent elections and the political regime.

Data and Methods

Data. Elections are a one-dimensional decision process, ideal to check with naive learning social network because it can be assumed that all the people in the network are interacting to persuade others to vote for one of the candidates, or avoid voting for another. Therefore, the data used from Twitter had to be related only to the candidates and the election, which is why this research focused only tweets where at least one of the three main candidates appeared.

This decision has several implications, the first one is that it does not capture the entire interactions about the election that happened on Twitter. A current trend in the platform is to avoid mentioning an account to prevent it from getting more attention, which could happen if the person wanted to express a negative perception of a candidate. Also, because it does not get absenteeism movements and opinions that prevents citizens from getting involved or from supporting any of the candidates. The decision of collecting only tweets with mentions was made, however, to reduce the amount of noise that otherwise would be present on the sample.

The elections chosen helped get the most diversity in two key aspects: different contexts -including different kind of electoral institutions- and different years. Having a diversity of countries allow the analysis of similarities on the interactions in the social platform and on the comparison with the results of the election. On the other hand, having two points in time revealed the trends in usage of the platform in each country and to see if there were differences due to the new election in the same country or to the changes in usage of Twitter. For the purposes of this research, in each election the tweets collected were the ones that mention at least one of the three main candidates and that were posted 150 days before or after the first election. As a result, the authors could see the changes in the networks of the elections on an equal time frame. Data were collected using an academic account on the Twitter API from 5 months prior to each election, searching for the handles (usernames) of the most important candidates. All data (dehydrated) and code are available in the GitHub repository (<https://github.com/dcardenas11831/Twitter-elections-consensus>).

Methods. Theoretical properties described in graph theory are useful to analyze how the efficiency of learning and diffusion depend, in sensitive ways, on the way the social network is organized [10]. Using DeGroot's model [4] defines a finite set $N = \{1, 2, \dots, n\}$ of agents or nodes interacting according to a social network. The interaction patterns are captured through an $n \times n$ non-negative matrix \mathbf{T} , where $T_{ij} > 0$ establishes the relationship interaction and indicates that i pays attention to j . Matrix \mathbf{T} may be asymmetric. \mathbf{T} is stochastic therefore its entries across each row are normalized to sum to one. Agents update beliefs by repeatedly talking weighted averages of their neighbors' beliefs. Each agent has a belief $P_i^{(t)} \in \mathbb{R}$ at time $t \in \{0, 1, 2, \dots\}$

It is possible to calculate how fast the system converges (if it does). DeMarzo et al. [5] developed results from spectral theory that relate rates of convergence to the size of eigenvalues. Specifically the second largest eigenvalue of the stochastic matrix \mathbf{T} denoted as $\lambda_2(\mathbf{T})$, arguing that if $|\lambda(\mathbf{T})|$ is small, then all the agents quickly reach an agreement. If $|\lambda(\mathbf{T})|$ is large, then convergence can take a long time.

The model considers a stochastic matrix that satisfies some convergence assumptions. For our study we built the stochastic matrix (Twitter mention network) which surely satisfies these assumptions. To ensure that the second eigenvalue is capturing the idea of speed in convergence, matrices with different topologies were simulated and the second eigenvalue was found for each of them.

Golub and Jackson [10] present the following results on the second eigenvalue:

- If $|\lambda_2(A)|$ is small, political beliefs and decisions converge rapidly.

- If $|\lambda_2(A)|$ is large, political beliefs and decisions converge slowly.

These results show that convergence is slow if society is divided into several factions that distrust each other. However, to compare different networks it is necessary to take into account two variables that affect the second eigenvalue, as seen in the simulated networks -explained in the results section-, which are: the structure of the network (the factions mentioned before) and the density of the links (the strength of the connections). A more connected network will have a smaller $|\lambda_2(A)|$ than a less connected one. In the same way, a network with more factions will have a bigger $|\lambda_2(A)|$, which means that will converge slowly.

Topological Data Analysis

We conducted a Topological Data Analysis using the Mapper algorithm, with a distance matrix of the nodes in the network as a measure of distance, and mentions of the winning candidate and the runner-up as filters. This resulted in a graph of interconnected clusters where each node is a group of Twitter accounts, grouped together by the number of mentions of the two main candidates.

Author contributions statement

D.C., A.S., M.R. and A.F. conceived the research, D.C. and M.R. conducted the experiments, D.C., A.S., M.R. analysed the results. All authors reviewed the manuscript.

Additional information

Competing interests

The authors declare no competing interests.

References

- [1] Maja Adena et al. “Radio and the Rise of the Nazis in Prewar Germany”. In: *The Quarterly Journal of Economics* 130.4 (2015), pp. 1885–1939.
- [2] Alexandre Bovet and Hernán A Makse. “Influence of fake news in Twitter during the 2016 US presidential election”. In: *Nature communications* 10.1 (2019), pp. 1–14.
- [3] Robert J Brown. *Manipulating the Ether: The power of broadcast radio in thirties America*. McFarland, 2004.
- [4] Morris H DeGroot. “Reaching a consensus”. In: *Journal of the American Statistical Association* 69.345 (1974), pp. 118–121.
- [5] Peter M DeMarzo, Dimitri Vayanos, and Jeffrey Zwiebel. “Persuasion bias, social influence, and unidimensional opinions”. In: *The Quarterly journal of economics* 118.3 (2003), pp. 909–968.
- [6] Daniel Gayo-Avello. “A meta-analysis of state-of-the-art electoral prediction from Twitter data”. In: *Social Science Computer Review* 31.6 (2013), pp. 649–679.
- [7] D Sunshine Hillygus. “The evolution of election polling in the United States”. In: *Public opinion quarterly* 75.5 (2011), pp. 962–981.
- [8] Petros Iosifidis and Stylianos Papathanassopoulos. “Media, politics and state broadcasting in Greece”. In: *European Journal of Communication* 34.4 (2019), pp. 345–359.
- [9] Masami Ito. *Broadcasting in Japan: Case-studies on Broadcasting Systems*. Vol. 8. Routledge, 2010.
- [10] M. O. Jackson and B. Golub. “Naive Learning in Social Networks and the Wisdom of Crowds”. In: *American Economic Journal: Microeconomics* 2:1 (2010), pp. 112–149. DOI: <http://www.aeaweb.org/articles.php?doi=10.1257/mic.2.1.112>.
- [11] Carolina Matos. “Media and democracy in Brazil”. In: *Westminster papers in communication and culture* 8.1 (2011).
- [12] Pippa Norris et al. *A virtuous circle: Political communications in postindustrial societies*. Cambridge University Press, 2000.
- [13] Francesco Pierri, Carlo Piccardi, and Stefano Ceri. “Topology comparison of Twitter diffusion networks effectively reveals misleading information”. In: *Scientific reports* 10.1 (2020), pp. 1–9.
- [14] John Street. *Mass media, politics and democracy*. Macmillan International Higher Education, 2010.

- [15] Michael W Traugott. “The accuracy of the national preelection polls in the 2004 presidential election”. In: *Public Opinion Quarterly* 69.5 (2005), pp. 642–654.
- [16] Michael Traugott et al. “Problemas relacionados con las encuestas preelectorales desde una perspectiva comparada”. In: *Estudios Públicos* 138 (2015).
- [17] Andranik Tumasjan et al. “Predicting elections with twitter: What 140 characters reveal about political sentiment”. In: *Proceedings of the International AAI Conference on Web and Social Media*. Vol. 4. 1. 2010.
- [18] Marshall Van Alstyne and Erik Brynjolfsson. “Could the Internet balkanize science?” In: *Science* 274.5292 (1996), pp. 1479–1480.
- [19] Wayne Xin Zhao et al. “Comparing twitter and traditional media using topic models”. In: *European conference on information retrieval*. Springer, 2011, pp. 338–349.

Simulated networks second eigenvalue				
Structure/Strength of links	10^{-4}	10^{-3}	10^{-2}	10^{-1}
One hub	1.00	1.93	3.83	10.79
Two hub	1.00	2.00	4.09	21.73
Three hub	1.41	2.07	4.30	21.78
Joined hubs1	1.73	2.46	4.95	14.19
Two v One hub2	1.00	2.23	4.93	25.22

Table 1. Second eigenvalue to different simulated networks. The value for the strength of links was determined so that the resulting graphs have a similar density as the real ones. As expected, the second eigenvalue increases with network density and with the number of hubs.

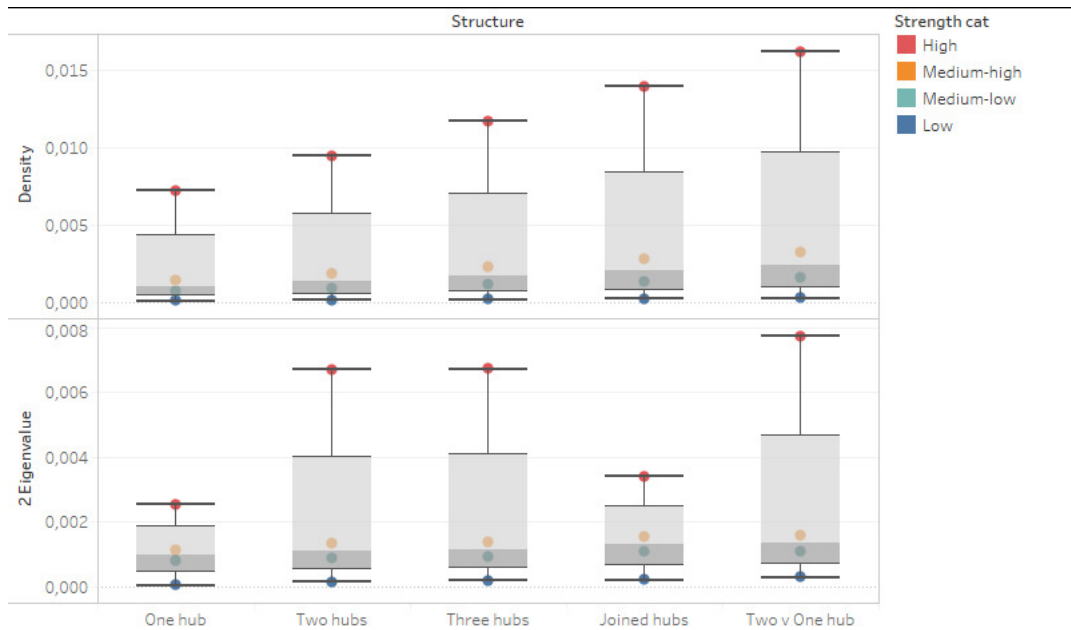


Figure 1. Behaviour of the second eigenvalue and density at different structures and simulation scenarios. Each scenario represents an increase on the probability of connections.

Second eigenvalue by election							
Election	Incumbent running	Regime	151-121	120-91	90-61	60-31	30-0
Argentina 2015	No	Pr	5.0	6.2	10.4	12.5	15.8
Argentina 2019	yes	Pr	1.0	3.0	2.0	3.0	1.0
India 2014	No	Pa	11.6	9.3	24.0	17.4	24.2
India 2019	yes	Pa	1.0	6.0	1.0	1.4	2.0
France 2012	yes	Pr	6.8	5.2	6.5	8.8	4.3
France 2017	No	Pr	3.0	8.6	2.8	3.0	1.7
Ecuador 2017	No	Pr	2.0	2.3	4.9	1.4	4.0
Kenya 2017	yes	Pr	3.0	2.0	3.0	2.0	3.0

Table 2. Second eigenvalue by days to voting day for all the observed. Argentina 2015, India 2014, and Ecuador 2017 show an increasing pattern of second eigenvalues, while the others show the opposite. This suggests that in the cases where the second eigenvalue was increasing, the consensus about the elections was decreasing.

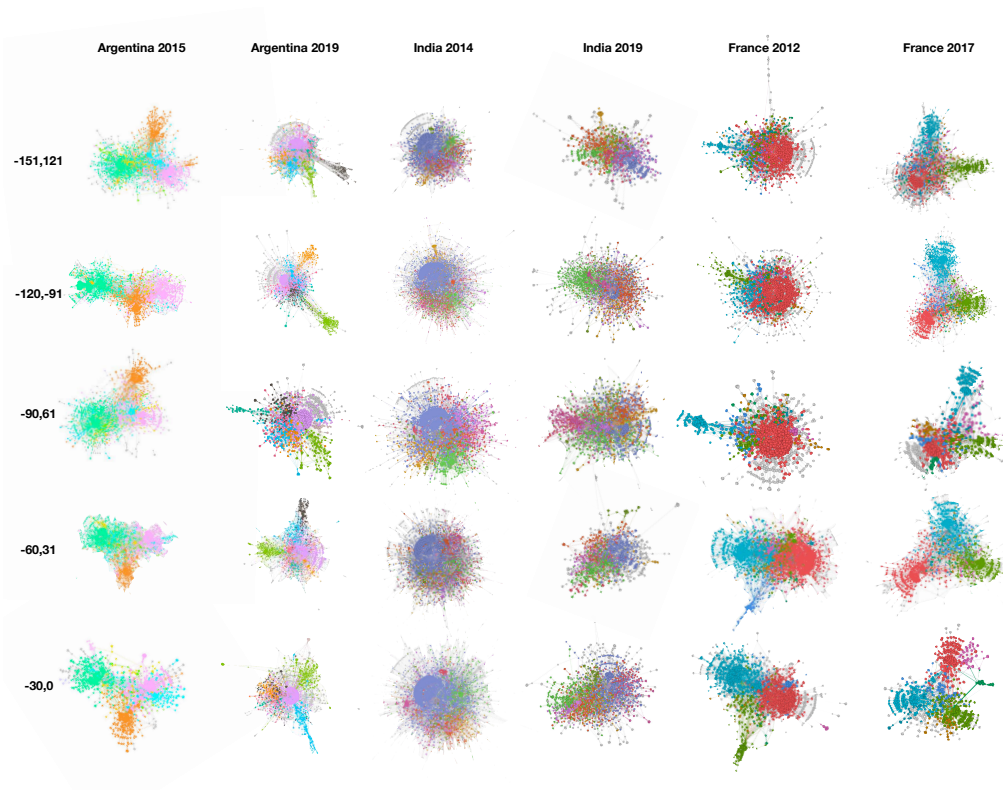


Figure 2. Twitter mention networks by election. A selection of observed elections and their progression over time (days to suffrage). Communities were colored using a walktrap algorithm for visualization purposes.

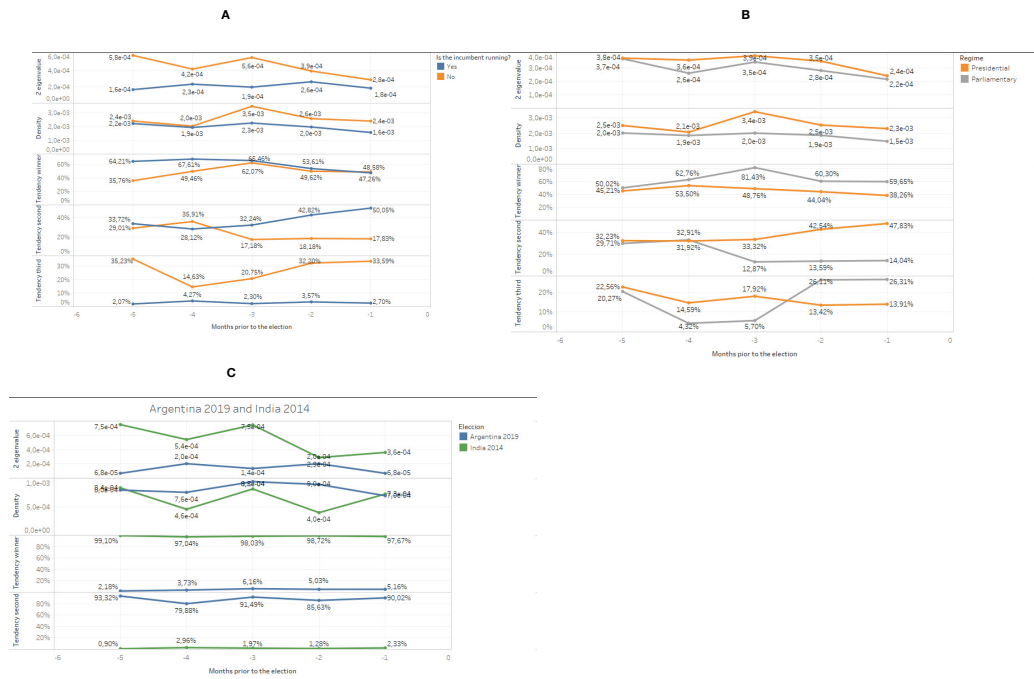


Figure 3. A: Elections separated if the incumbent was running for re-election or not. The trends describe the average values of the elections that belong to one of the categories. B: Elections by regime. The trends describe the average values of the elections that belong to one of the categories. C: Second eigenvalue to elections Argentina 2019 and India 2014. The image shows how the second eigenvalue evolves as well as its components. The proportion of mentions in the network was used as a proxy for the structure.

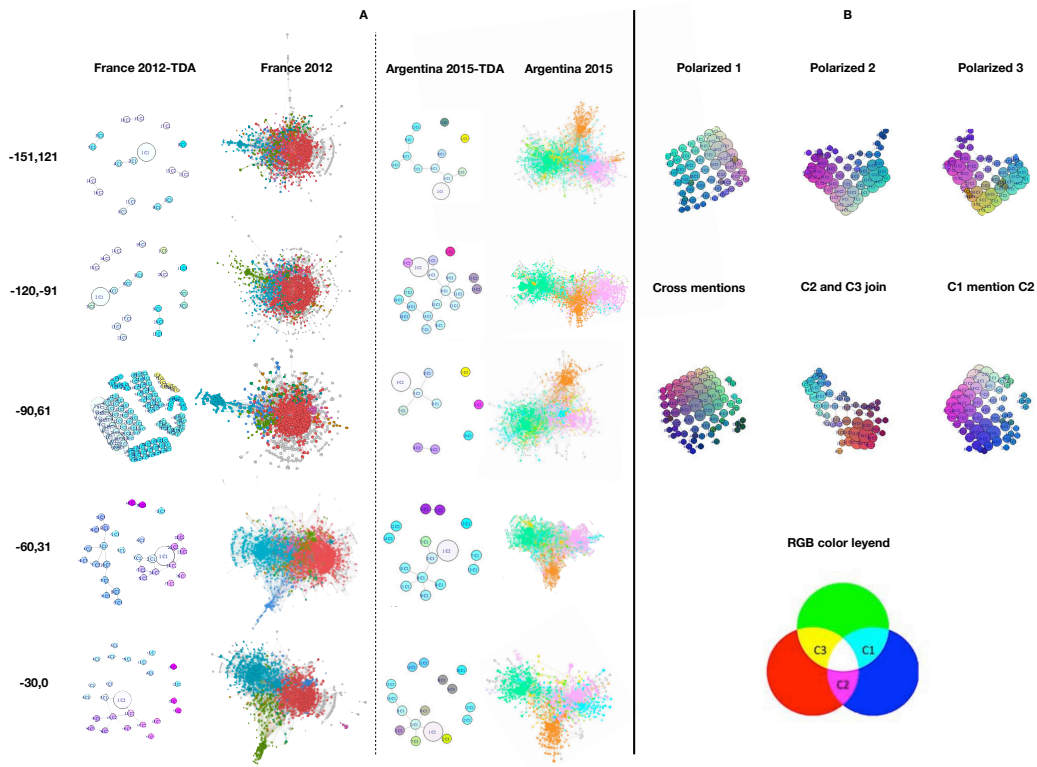


Figure 4. Topological Data Analysis of observed and simulated networks. A, Left: Topological Data Analysis of mention networks for France, 2012. Each of the five graphs represents a month prior to election. We observe candidate 1 (C1, blue) dominating the conversation, and the runner-up (C2, pink) becoming more important 60 days before the election. Mention networks, visualized for comparison purposes. A, Right: Topological Data Analysis clustering results and mention networks for Argentina 2015. B. Different simulated networks with archetype political scenarios with different degrees of polarization, cross mentions between clusters, and one scenario of C2 and C3 joining forces. The last one shows the legend to understand the images. More teal represents a group of people more inclined to talk about the winning candidate, more purple to the runner-up and more yellow more to the third one.

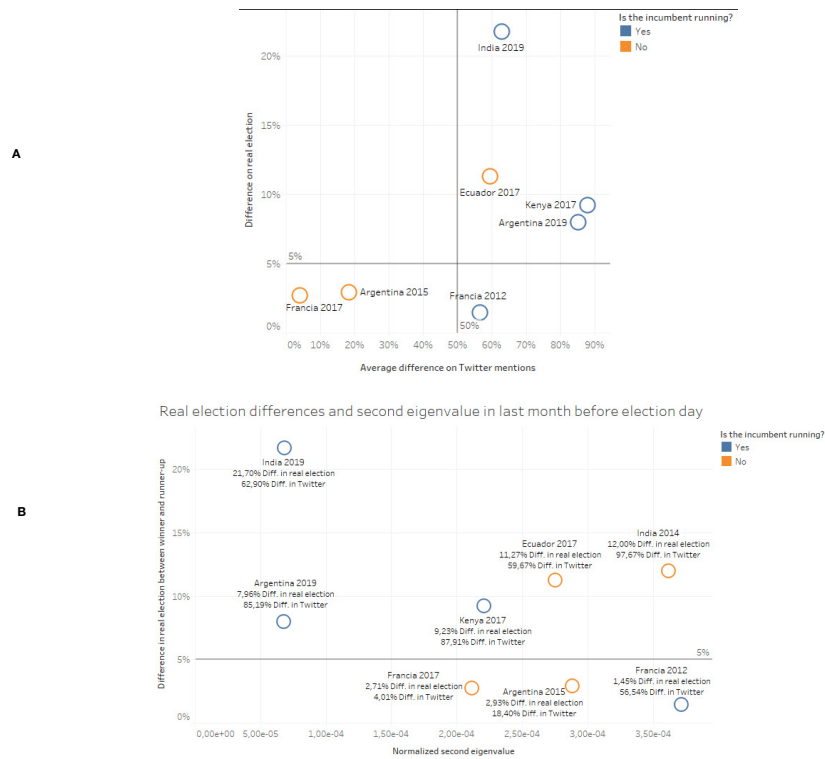


Figure 5. A: Comparison between the difference in mentions on Twitter in the last month before the election and the outcome. B: Relationship between the normalized second eigenvalue of the last months prior to each election day and the difference in the real election between the two main candidates. The color determine if an election had the incumbent running. The label on each election has the difference in the election and the difference on Twitter mentions (between the most mentioned candidate and the one that follows) in the last month.