Opinion Polarization during a Dichotomous Electoral Process

G. Olivares, J. P. Cárdenas, J. C. Losada, and J. Borondo

1Grupo de Sistemas Complejos, Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas, Universidad Politécnica de Madrid, Avda. Complutense s/n, 28040 Madrid, Spain
2Net-Works, Angamos 451, Reñaca, Viña del Mar, Chile

Correspondence should be addressed to J. Borondo; jborondo@gmail.com

Received 30 June 2018; Revised 17 October 2018; Accepted 17 January 2019; Published 10 February 2019

1. Introduction

A perfectly polarized society is divided into two groups of the same size and with totally confronted opinions. Around this definition, the authors of [1] proposed a methodology to study and measure the emergence of polarization from online social interactions. However, social polarization can emerge from diverse scenarios, for example, during electoral campaigns. Moreover, electoral process where there are only two political sides increases the chance of finding a high political polarization [2]. For example, the US citizens face a polarized scenario during the presidential elections, where they have to vote to make a dichotomic decision between two parties: republicans and democrats. Previous studies have shown that the US elections polarize Internet, where both Twitter [3] and blogs [4] manifest a high political polarization.

Another political scenario that requires a dichotomic decision is that of electoral processes that require a second round. While on the first round citizens can vote from a wide range of diverse political parties, on the final round they can just vote for the two final candidates. In fact, some voters may not be totally identified in neither side, but still must take a side. Previous papers have shown that this second round increases the political polarization of the country [5].

During the last years, social media platforms, such as Twitter, have become an effective media for political parties to propagate their messages and campaign for their candidates, reaching a high fraction of the population at a low cost [6–8]. Since Twitter data is publicly available, scientists have analyzed these electoral campaigns characterizing political strategies, analyzing the users’ influence [9–11], and modeling the human behavior in these processes [12–17].

In this paper, we will analyze the final week of the second round of the Chilean presidential elections, where voters had to choose between candidates Piñera and Guillier or abstention. We will focus on analyzing the political polarization that emerges and tracking its evolutions during the week preceding the elections and the final voting day. To this end, we first estimate the opinion of Twitter users from a minority of elite users, whose opinion was known. Next, we measure the resulting political polarization from the cited opinion distributions and track its evolution during a full week that includes the voting day. We found the conversation to be highly polarized, reaching its maximum during the previous day to the election and significantly decreasing the day after due to the presence of new users who only participated during that day.
the conversation. However, the decrease in the polarization observed on the election day was caused by new users not so engaged to the political debate that entered the conversation acting as bridges between the two sides.

2. System

2.1. Chilean Electoral System. The Chilean Presidential elections are held every four years. Any Chilean who fulfills the requirements that the Chilean Electoral Service imposes can stand as a candidate to be elected as the next president of the Republic of Chile.

A candidate can belong to a political party or a list of parties, being enough that this community inscribed his name as a candidate. If there is no consensus within a political community respect to the candidate, it is possible to hold “primaries” to define a unique name as a candidate. In the case of independent candidates, they have the obligation not to belong to any political party, at least two months before the registration of candidacies. They also need to have at least 0.5% of signatures of the voters of the last election of deputies, being 33,493 signatures in 2017.

On the election day, each qualified citizen can vote for only one candidate. When a candidate obtains an absolute majority (>50%), the candidate will be appointed “President of the Republic of Chile”. Otherwise, it will proceed to make a second presidential round, where the two candidates with more number of votes pass to a new election process.

In the elections of 2017, the left-wing parties did not hold primaries and several candidates were presented, denoting a clear division inside the political parties. This was not the case of the right-wing party.

The first round of the election of 2017 was held on November 19 with an abstention rate of 53.3% and the results did not give an absolute majority to any candidate (see Figure 1).

The second round was held one month later, on December 17, with a slightly higher participation rate (51%). The results gave as winner of the elections to Sebastián Piñera, the candidate of the right wing (see Figure 2)

2.2. Data Set. To build the dataset analyzed on this paper, we downloaded all the Twitter posts related to the 2017 Chilean Presidential Elections during the last week of the second round electoral campaign (this week includes the voting day). All the tweets were retrieved using the Twitter Streaming API. This API allows downloading tweets matching a set of keywords. In order to build the dataset, we have
chosen the following keywords: #chiledecide, #eleccionespresidenciales2017, #chileelige, #eleccionesenChile, #elecciones2018, #chilevota, #elecciones2017, #eleccionesChile2017, #piñera and #guillier.

The final dataset is composed of 203,612 messages posted by 68,048 different users between December 11th and December 17th, 2017. 83.5% of the total tweets actually contained a retweet.

3. Methodology

3.1. Retweet Networks. We can consider retweets as a proxy of influence between users. Hence, when user A retweets a tweet originally posted by user B, we can infer that B is influencing A [16, 18, 19]. In the retweet networks, each node corresponds to a user, and a link from node A to node B is established when user A retweets a message posted by user B. This link is directed with direction from A to B. Thus, the indegree \( k_{\text{in}} \) of a given node measures the total number of different users who have retweeted him, and the outdegree, the number of retweets that all the tweets he posted have received.

In Figure 3(a), we show the corresponding retweet network for the period of the electoral campaign, excluding the voting day (December 11-16). The network is composed of 26,175 nodes and 48,141 links. The corresponding network for the full electoral campaign including the voting day is plotted in Figure 3(b) for comparison. This network has a total of 68,048 nodes and 170,045 links. In the figure, the size of nodes is proportional to its indegree.

The degree distribution of both networks is very heterogeneous with a small number of users accumulating a high percentage of the retweets (see Figure 4). Most users (57%) only retweet a single tweet during the full analyzed period \( (k_{\text{out}} = 1) \) while a very small fraction of users (0.025%) retweeted more than 100. A similar trend is observed when analyzing the indegree distribution. The majority of users do not receive any retweets; we observed that 83.5% of users have
$k_{in}=0$ and $7\%$ of users have $k_{in}=1$. In contrast, these users live together with a minority of extremely influential users who received many retweets. In fact, $0.075\%$ of the users received more than 500 retweets and we even found a user whose tweets received over 5000 retweets. These highly retweeted users are the real influential ones ruling the conversation.

In order to properly analyze this conversation, we consider several periods of time. Thus, we build a retweet network for each day $D$, where the time window considered for the network of each day $D$ is the accumulated from day 0 up to day $D$.

### 3.2. Inferring the Opinion of the Users

To infer the opinion of the users participating on the Twitter conversation and measure the resulting polarization we use the methodology introduced in [1]. In that paper, we introduced a model to estimate the opinions of users who interact on a social network from a minority of hubs whose opinion is known. In the model we have two types of users, elite and listeners. The model assumes that we know with certainty the opinion of the elite. Thus, elite users have a fixed opinion that remains constant and act like seeds of influence. In contrast, the opinion of listeners is unknown and will be estimated from their social interactions. The model is initialized by assigning an opinion value $X$ to each elite and an initial value $X=0$ to each listener. The opinion of each listener is iteratively updated as the mean opinion value of her outgoing neighbors. Thus the opinion at time step, $t$, of a given listener, $i$, is given by the following expression:

$$X_i(t+1) = \frac{\sum_j A_{ij} X_j(t)}{k_{out}^i}$$

where $A_{ij}$ represents the elements of the network adjacency matrix, which is 1 if there is a link from $j$ to $i$, and $k_{out}^i$ corresponds to her outdegree. The process is repeated until all nodes converge to their respective $X_i$ value, lying in the range $-1 \leq X_i \leq 1$. Thus, the results of the model are given in a density distribution of nodes’ opinion values $P(X)$. In our case the network of the model is represented by the retweet network. We generate a network for each day ($D$) of the analyzed period; this network is the accumulated retweet network from day 0 up to day $D$. The corresponding adjacency matrix for each day $D$ is given by the following formula:

$$A_{ij}(D) = \begin{cases} 1 & \text{if } i \text{ retweet } j \in t \leq D \\ 0 & \text{otherwise} \end{cases}$$

#### 3.4. Elite Users

To select the elite users that will be used to infer the opinion of the remaining users we have first selected the top central nodes according to their indegree: $k_{in} > 200$. After applying this filter, we obtain 159 nodes (see Figure 5). From those we discarded 44 accounts because they belonged to institutional accounts and neutral mass media and accounts that were deleted by Twitter during the campaign.

We assigned an opinion of 1 to the remaining users that have been marked as elite (17 in favour of Piñera and 14 supporting Guillier). Next, by means of text mining techniques, we identified a group of users asking for the vote for one of the candidates (16 of them for Piñera and 22 for Guillier). Finally, in order to assign a value to the remaining 34 users we have trained a machine learning algorithm based on the texts used by the already classified users. To this end we trained a Support Vector Machine (SVM) with linear kernel [20, 21]. We have validated the training of the model by means of a $k$-fold cross validation process and obtained an average accuracy of $82\%$. Finally, we used the trained model to classify the remaining 34 users. In the end, a total of 64 users were classified in favor Guillier, to whom we assigned a value of -1; and 51 users in favour of Piñera, to whom we assigned a value of +1.

![Figure 5: Subgraph of the retweet network for the period 11-17 December, 2017, corresponding to 160 nodes with $k_{in} > 200$. Elite nodes associated with Piñera (red) and Guillier (blue) preferences. The rest of the nodes (grey) are not included in the elites. Size of nodes corresponds to $k_{in}$.](image-url)

This formula gives $\rho = 1$ when the distribution is perfectly polarized. In this case the opinion distribution function is two Dirac delta centered at -1 and +1, respectively. Conversely, $\rho = 0$ means that the opinions are not polarized at all.
Once the elite is selected, the opinion of the rest of the nodes is generated from (1). Figure 6 shows the fast convergence of this process for 100 random nodes.

4. Results and Discussion

From the distributions obtained with the model for each of the days in our dataset, we can track the evolution of the opinions of users who participated on the conversation over the 2018 Chilean electoral campaign. Thus, we will use these distributions to analyze the opinions of users and the evolution of the polarization over the electoral campaign.

We found that during the analyzed period two peaks on the distribution around the two confronted opinions -1 and +1 appear. However, we also identify significant differences in the distributions between the voting day (D 17) and the previous week. To further explore these differences, we have visualized both opinion distributions on the top panels of Figure 7. In this figure we compare the distribution for the pre-voting day (D 16) and the voting day (D 17). The distributions show that on the previous day to the elections there is a majority of users with an opinion value of $X<0$ (opinions totally in favor of candidate Piñera). However, we observe a shift on this behavior during voting day with a majority of users emerging around opinion $X=-1$. Moreover, on the voting day, we also find an increase of the moderate opinions around candidate Guillier, with a significant increase of users with $P(-1 < X < 0)$. Thus, our results show a clear shift in the opinions expressed by the Twitter users between the previous day to the elections and the voting day. While the opinion expressed on Twitter remained approximately constant during all the previous week to the election day, this was not the case for the voting day.

Despite this change in the opinions, we cannot still state that the users individually changed their opinion during the voting day, as the participation usually increases during the election day engaging new users [13]. Thus, to shed light on whether the distribution obtained on the voting day is due to a significant change on the opinion of users who actively participated during the previous week or to the participation of new users, we have analyzed separately the opinion of both groups. In the bottom panel of Figure 7, we show the resulting distribution of opinions for the voting day including only those users who already participated during the previous days. The opinion distribution of these users differs significantly from the one plotted on the top panels that includes all of the users. Thus, the population of users who participated during the previous week still maintains a majority of extreme values around 1, highlighting that the majority of opinions around -1 comes from the new users that joined the conversation the final voting day. However, a fraction of users who already participated during the campaign week slightly varied their opinion as there is an increase on moderate negative opinions for both populations.

To further understand the evolution of the opinions during the analyzed period we have visualized the time series for the negative and positive opinion in Figure 8. In this figure, triangles represent opinions in favour of candidate Piñera and circles opinions in favour of candidate Guillier.

This figure reflects that the opinions remained constant over the previous days to the voting day, with a majority of opinions in favor of Piñera. This trend changed during the voting day where more users participated on the conversation and the fraction of moderate opinions in favor of Guillier surpassed the ones in favor of Piñera. However, the significant increase of moderate opinions is not explained by new users, but by the change of mind during day 17 of fraction users. If we limit the analysis to the radical opinions (+1 or -1) we observe this same pattern, with a transition from a majority of +1 to a majority of -1. However, this shift in the radical opinions is due to new users commenting for first time on the hashtag during the last day. The vast majority of these new users that joined the conversation on the voting day showed clear support for Guillier (-1<X<0).

Next, we analyze how the polarization evolved during the electoral campaign week. To this end, we have computed the polarization index for each day and visualized the time series on the bottom panel of Figure 8. We found the conversation to be polarized over the full electoral period, with values around 60% that increased up to 70% just before the voting day. This fact manifests that during the pre-voting day the conversation reached its maximum confrontation. In contrast, on the voting day, we find an abrupt decrease where the polarization index reached a minimum value of 40%. However, if we do not consider new users (who only participated on day 17) and limit the polarization analysis to the users who participated during the full week we also observe this transition, but the decrease of the polarization in the voting day is not so abrupt and recovers its typical value. Thus, the extreme minimum value observed on the voting day is explained by the appearance of new users that participated on the conversation this one day and by the decrease of extreme opinions that gave place to more moderate opinions that concentrated on the negative side (see Figure 7). Hence, in the light of our results, we can conclude that new users joining the conversation on the voting day significantly varied the debate making it more rational with an increase on the flux of messages between the two sides.
To further understand the evolution of the polarization during the debate, we analyzed the profile of the users and tweets posted during the week. More particularly, we analyzed the use of mentions on tweets. Mentions are a Twitter interaction mechanism that is usually used to direct tweets to specific users, to refer them or to exchange information [16]. In the top panel of Figure 9, we show the time series of the fraction of tweets that contained at least a mention. The percentage of tweets containing mentions remained constant during the week with a value of approximately 80%. However, for the voting day, this value decreased reaching a minimum of 76%. This decrease in the number of mentions during the voting day shows that during that day the debate broadens its scope, not directing so much tweets to specific users. The tweets without mentions are not directed and they can be used to express feelings. A reduction in the number of mentions usually implies a transition from rational to more emotional, which can explain the decrease in the polarization index described before.

Next, we explore the reasons that explain the decrease in the percentage of tweets containing mentions to unveil whether it is due to new users that joined the conversation on the voting day or not. We found that indeed it is explained by new users that engaged to the debate this last day, as if we do not consider them, the fraction of mentions per tweet remains constant over the full period. Following, we analyze the diversity of the users participating in the conversation by computing the heterogeneity $\sigma$. This measure quantifies the
diversity of users according to the ratio of followers/friends \[22\]. We also observe a change on the heterogeneity of users on the voting day (see bottom panel of Figure 9) showing a significant increase. This means that on the last day more diverse users participated in the conversation, and these same users caused the polarization decline.

5. Conclusions

On this paper, we have used the methodology described on \[1\] to analyze the 2017 Chilean presidential elections and measure the resulting political polarization. From a minority of influential users, we have been able to estimate the opinion of the vast majority, showing that during the period preceding the voting day most of the users were supporting Piñera.

The Twitter users discuss, form their opinion, and propagate messages from their preferred candidate over the electoral period, showing their final decision on the day preceding the election. In this day, the polarization reaches its maximum value as users already decided their vote and limit their behavior to support their leader and amplify the messages supporting him. This propaganda behavior limits the discussion increasing radical opinions and the polarization. After this point, the polarization decreased during the voting day attracting new and more diverse users that opened the conversation and increased the flux of information across the two sides.

The high correlation among the opinions of users on day 16 and the voting behavior of the Chilean population that face a dichotomous approach to elections (vote candidate A or B) could indicate that until that day the opinion formation process of Twitter users is rational and mediated, reaching a final vote decision (vote A, B or do not vote). Moreover, this decision is consolidated during the pre-voting day where users who actively participated during the week tend to abandon the debate and just support their decision causing a maximum peak on the political polarization. In addition, during this day, the opinions inferred from Twitter correlate almost perfectly with the election outcomes.

Since on this paper we estimated the opinion of a large number of Twitter users we will discuss the possibilities and
limitations of using this method to predict election outcomes. If we associate the distribution of opinions for Twitter users during the electoral campaign to their voting intentions, we can compare them to the final election outcomes. We can associate opinions near $X=+1$ with an intention to vote Piñera and values around $X=-1$ with potential voters of Guillier. In such a scenario, on day 16, the percentage of opinions associated with Piñera is around 54.4% and the ones associated with Guillier around 45.6%. The real result was a victory of Piñera with 54.5% of votes over Guillier with 45.4% of votes. Since the results coincide quite accurately with the final voting outcomes, we could think that this methodology could be useful to infer election outcomes. Thus, despite the authors introducing in [1] this methodology to measure political polarization, in the light of our results the methodology could also be used to infer election outcomes. However, to do this task, we should do a deeper analysis and take into account more facts, such which profiles are over- or infra-represented on Twitter. Thus, further research still needs to be done in order to extrapolate the methodology to effectively predict election outcomes from social media conversations.

**Data Availability**

The data can be requested from the authors through personal communication.

**Conflicts of Interest**

The authors declare that they have no conflicts of interest.

**Acknowledgments**

The authors thank Rosa Mª Benito (Universidad Politécnica de Madrid) for her help in the realization of this paper. The authors acknowledge support from Projects of the Spanish Ministerio de Economía y Competitividad no. MTM2015-63914-P. Gastón Olivares acknowledges support from CONICYT Doctorado Becas Chile 72180383. This work was supported by the Office of Naval Research Global (ONRG) and the Air Force Office of Scientific Research (AFOSR) under Grant no. N62909-17-1-2010.

**References**

[1] A. J. Morales, J. Borondo, J. C. Losada, and R. M. Benito, “Measuring political polarization: Twitter shows the two sides of Venezuela,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 25, Article ID 033114, 2015.

[2] J. L. Sundquist, *Dynamics of the Party System: Alignment and Realignment of Political Parties in the United States*, Revised Ed, The Brookings Institution, Washington, DC, USA, 1983.

[3] M. D. Conover, J. Ratkiewicz, M. Francisco, B. Gonçalves, A. Flammini, and F. Menczer, “Political polarization on twitter,” in *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, pp. 89–96, 2011.

[4] L. A. Adamic and N. Glance, “The political blogosphere and the 2004 U.S. Election: Divided they blog,” in *Proceedings of the 3rd International Workshop on Link Discovery (LinkKDD ’05)*, pp. 36–43, ACM, 2005.

[5] G. Serra, “Should we expect primary elections to create polarization? a robust median voter theorem with rational parties,” in *Routledge Handbook of Primary Elections*, R. G. Boatright, Ed., pp. 72–94, 2018.

[6] B. A. Huberman, D. M. Romero, and F. Wu, “Social networks that matter Twitter under the microscope,” *First Monday*, vol. 14, no. 1, 2009.

[7] A. Herdağdelen, W. Zuo, A. Gard-Murray, and Y. Bar-Yam, “An exploration of social identity: The geography and politics of news-sharing communities in twitter,” *Complexity*, vol. 19, no. 2, pp. 10–20, 2013.

[8] J. H. Parmelee, “Political journalists and Twitter: Influences on norms and practices,” *Journal of Media Practice*, vol. 14, no. 4, pp. 291–305, 2013.

[9] F. I. Solop, “Rt @barack obama we just made history: Twitter and the 2008 presidential election,” in *Communicator-In-Chief: How Barack Obama Used New Media Technology to Win The White House*, J. A. Hendricks, J. Robert, and E. Denton, Eds., pp. 37–49, Lexicon Books, Plymouth, UK, 2010.

[10] A. J. Morales, J. C. Losada, and R. M. Benito, “Users structure and behavior on an online social network during a political protest,” *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 21, pp. 5244–5253, 2012.

[11] M. D. Conover, B. Gonçalves, A. Flammini, and F. Menczer, “Partisan asymmetries in online political activity,” *EPJ Data Science*, vol. 1, no. 6, pp. 1–17, 2012.

[12] B. Gonçalves, N. Perra, A. Vespignani, and M. Perc, “Modeling users’ activity on twitter networks: validation of Dunbar’s number,” *Plos One*, vol. 6, no. 8, Article ID e22656, 2011.

[13] J. Borondo, A. J. Morales, J. C. Losada, and R. M. Benito, “Characterizing and modeling an electoral campaign in the context of Twitter: 2011 Spanish Presidential election as a case study,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 22, Article ID 023138, 2012.

[14] A. Jungherr, “‘Tweets and votes, a special relationship; The 2009 federal election in Germany’,” in *Proceedings of the 2nd Workshop on Politics, Elections and Data, PLEAD 2013 - Co-located with CIKM 2013*, pp. 5–14, New York, NY, USA, 2013.

[15] G. Caldarelli, A. Chessa, F. Pammolli et al., “A multi-level geographical study of italian political elections from twitter data,” *Plos One*, vol. 9, no. 5, Article ID e95809, pp. 1–11, 2014.

[16] J. Borondo, A. J. Morales, R. M. Benito, and J. C. Losada, “Multiple leaders on a multilayer social media,” *Chaos, Solitons & Fractals*, vol. 72, pp. 90–98, 2015.

[17] U. Yaqub, S. A. Chun, V. Atluri, and J. Vaidya, “Analysis of political discourse on twitter in the context of the 2016 US presidential elections,” *Government Information Quarterly*, vol. 34, no. 4, pp. 613–626, 2017.

[18] D. Boyd, S. Golder, and G. Lotan, “Tweet, tweet, retweet: conversational aspects of retweeting on twitter,” in *Proceedings of the 43rd Annual Hawaii International Conference on System Sciences (HICSS ’10)*, pp. 51–60, IEEE Computer Society, January 2010.

[19] A. J. Morales, J. Borondo, J. C. Losada, and R. M. Benito, “Efficiency of human activity on information spreading on Twitter,” *Social Networks*, vol. 39, no. 1, pp. 1–11, 2014.

[20] D. K. Srivastava and L. Bhambhu, “Data classification using support vector machine,” *Journal of Theoretical and Applied Information Technology*, vol. 12, no. 1, pp. 1–7, 2010.
[21] T. Joachims, “Text categorization with Support Vector Machines: Learning with many relevant features,” in *Proceedings of the Machine Learning: ECML-98*, C. Nédellec and C. Rouveiro, Eds., vol. 1398 of *Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence)*, pp. 137–142, Springer, Berlin, Heidelberg, 1998.

[22] J. P. Cárdenas, G. Vidal, C. Urbina, G. Olivares, P. Rodrigo, and M. Fuentes, “Social crises: signatures of complexity in a fast-growing economy,” *Complexity*, vol. 2018, Article ID 9343451, 11 pages, 2018.
Submit your manuscripts at
www.hindawi.com