Love and Hate During Political Campaigns in Social Networks

Juan Carlos Losada[1], José Manuel Robles[2], Rosa María Benito[1],
and Rafael Caballero[2(✉)]

[1] Complex System Group, Universidad Politécnica de Madrid, Madrid, Spain
{juancarlos.losada,rosamaria.benito}@upm.es
[2] Universidad Complutense de Madrid, Madrid, Spain
jmrobles@ccee.ucm.es, rafacr@ucm.es

Abstract. Social networks have become central for public debate. However, this debate occurs, in many cases, in a polarized and socially fragmented way. The importance of a specific type of polarization marked by the expression of emotions (affective polarization) has recently been pointed out. This work seeks to analyze affective polarization through the study of opinions shared by users of Twitter during competitive events such as a political election (United States electoral campaign, 2016). To operationalize the affective polarization we propose, first, to consider the relationship between the user opinions about each contender. This approach allows us to describe what is the opinion of the supporters of one contender about the rest of them. Secondly, we combine sentiment analysis techniques, new diagrams describing graphically the tension between the positive and negative opinions, and numerical measurements obtained employing techniques borrowed from physics such as the gravity centers to identify and measure affective dispositions of users participating in this discussion. In summary, a way of measuring an emergent and central type of polarization in competitive contexts is proposed here.

Keywords: Social networks · Sentiment analysis · Love-hate diagram

1 Introduction

Initially proposed by Barnes [1] in the field social science, the concept of *social network* acquired a strong impulse with the raise of Internet, and in particular with the creation of social network web sites. According to Ellison *et al.* [2], a social network web site (Twitter in our case) is a web-based service that allow individuals to publish information under a public or semi-public profile, manage the list of users with whom they are connected, and check information shared by other users.

Social networks become particularly active during competitive events, where several contenders struggle to achieve some goal. Examples of such events include political processes [3–5], sport championships [6], or even song contests [7].

In these events, the users post their messages, often expressing their support or disagreement about each candidate. These opinions contain valuable information that can contribute to a better understanding of the process. In order to analyse the large amount of user opinions, techniques such as *sentiment analysis* [8] are employed. These techniques generate values representing the opinion of a post about a particular subject, in our case about each candidate. In our paper we assume that each post can be labeled as -1 (negative, which we call sometimes 'hate'), 0 (neutral) or $+1$ (positive also called in this paper 'love'), obtaining a different label for each candidate. Next, the message polarities from the same user are averaged to obtain the user opinion about each contender.

This process yields information about the support of each contender in the social network. However, this says little about the *polarization* of the debate, which constitutes an important question in the academic literature [9]. For this reason, this work proposes a methodological instrument devoted to analyze the level of polarization in these events.

In this paper we focus on the case of events with two contenders, in order to adequately describe the methodology, but it can be easily generalized to processes with more than two participants. The next section shows the diagrams and measurements that we propose to study this process. Then, in Sect. 3 we apply these tools to the particular case of the USA 2016 elections in Twitter. The paper ends showing some conclusions and future work.

2   Detecting Love and Hate

We are interested in inferring the opinion that Twitter users have towards two candidates in an electoral process, which we will call candidates C and T due to similarity to the case study that will be presented later.

At the top of the Fig. 1 is shown the process of assigning sentiment values to users. We assign a value X_i to the user u_i, (where X can be T or C) given by the average of the opinion values assigned to X in each message of u_i. The values $X_i = -1$ indicating totally negative sentiment or hate, $X_i = +1$ totally positive or love, and $X_i = 0$ a neutral sentiment for the candidate X.

We call U^X the set of users with sentiment for candidate X. In general, there may be users who never post about a candidate and also messages where the users post about both. This implies that, in general, the sets U^C and U^T have a different number of elements. We denote by $U^{CT} = U^C \cap U^T$ the set of users with sentiment for both candidates. These users can be represented in a two-dimensional space, as shown in the bottom of Fig. 1.

For each set of user, U^X, we calculate the probability distribution $p(x)$, considering the probability of finding a user with a given sentiment x.

$$p(x) = \frac{\#X_i = x}{\#X_i} \tag{1}$$

From this distribution of sentiments we obtain the polarization indices described in the Ref. [10].
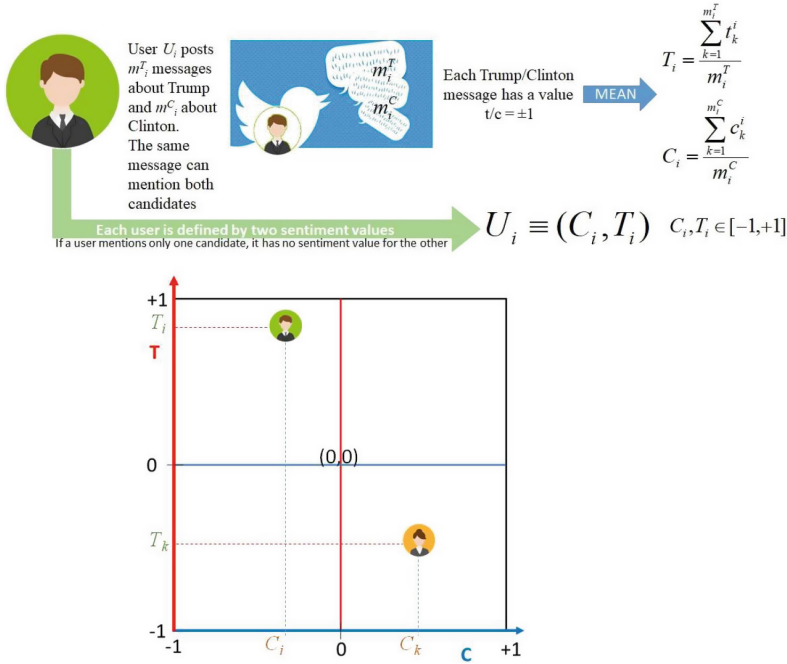
Fig. 1. Top: Construction scheme of user sentiment about C and T candidates. Bottom: Mapping of user sentiment values about both candidates in the (C, T) space.

The proportion of users in each interval of sentiments, $x < 0$ and $x > 0$, is given respectively by:

$$A^- = \int_{-1}^{0} p(x)\, dx \tag{2}$$

$$A^+ = \int_{0}^{1} p(x)\, dx.$$

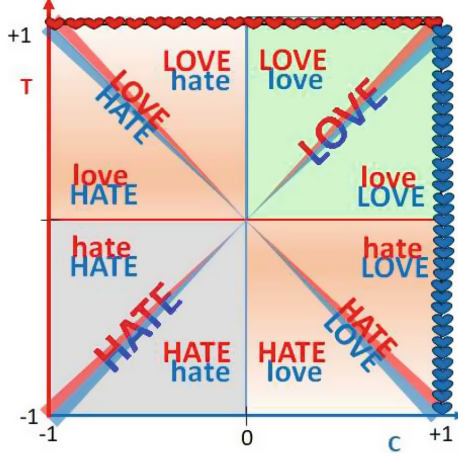The difference between these indices

$$\Delta A = A^+ - A^- \tag{3}$$

gives information on which sentiment is predominant about candidate X.

The average value of the sentiments of love (gc^+) or hate (gc^-) gives information on how extreme these sentiments are. These values can be obtained by calculating the centers of gravity of the function $p(x)$ in each interval:

$$gc^- = \frac{\int_{-1}^{0} p(x)x\, dx}{A^-} \tag{4}$$

$$gc^+ = \frac{\int_{0}^{1} p(x)x\, dx}{A^+}.$$

Fig. 2. Structure of a love-hate diagram for two candidates C and T. The axes that represent the maximum love for each candidate, is represented by blue hearts (C = +1) and by red hearts (T = +1).

A mean value close to ± 1 implies extreme positive/negative sentiments, while if the mean value is close to 0, the sentiments are more neutral.

From these indices, we can define a global polarization index using the following expression:

$$\mu = (1 - |\Delta A|)d, \tag{5}$$

where

$$d = \frac{|gc^{+} - gc^{-}|}{2} \tag{6}$$

it is half the distance between the centers of gravity. In this way μ is a magnitude between +1 and 0, where +1 implies extreme polarization and 0 minimum polarization.

In the case of users who have sentiment towards the two candidates ($u_i \in U^{CT}$), we can also analyze the distribution of sentiments in two-dimensional space (C, T), obtaining what we have called the love-hate diagram, whose scheme is shown in the Fig. 2. Each point in this diagram can be colored based on the number of users with that pair of sentiments. The users located on the line $X = +1$ are those who love candidate X, marked in blue in Fig. 2 for the case X = C and in red for X = T.

In the quadrant $C > 0$ and $T > 0$ users have feelings of love towards both candidates, while in the quadrant with $C < 0$ and $T < 0$ users have feelings of hatred towards both of them. On the line with slope +1, there are users who love or hate both candidates in the same way. In polarized situations it is expected that these two quadrants are practically empty, since users do not usually have the same type of feeling towards the two options.

In the quadrant with $C > 0$ and $T < 0$ are the users who love C and hate T. On the line with slope -1 are the users who love C the same as they hate T. By below this line users hate T more than they love C and above are users who love C more than they hate T. Analogously in the quadrant with $C < 0$ and $T > 0$ there are users who love T and hate C. On the line with slope -1 are those users who love T the same as they hate C. Below this line, users hate C more than they love T and above, they love T more than they hate C.

3   A Case Study: The 2016 USA Elections

3.1   The Dataset

As usual in social network research, our study adopts an observational (also known as correlational or non-experimental) rather than an experimental app-roach point of view. That is, we just collect and analyse data that already exists. We consider two criteria to characterize the subset of the data from the social network that can be relevant for our experiment:

Temporal Criterion. The fact that these events usually have a deadline, estab-lishes an upper temporal bound for the period of time we wish to consider. On the contrary, there is no rule of thumb about the lower temporal bound, that must be determined arbitrarily. In practice, filtering by this criterion is often allowed by the social media application program interfaces. In our case we downloaded tweets in starting at 0 h (UTC) of 2016-11-02 and ending at 9 am (UTC) of 2016-11-13.

Topic Criterion. We aim at collection messages related to our event, but the particular way of filtering the messages will depend on the particular network. In our case we used as topic criterion the Twitter names of both contenders, *@HillaryClinton* and *@realDonaldTrump*, represented by the letters C and T in the rest of the paper.

The initial dataset consisted of 13,358,353 messages or *tweets* following the argot in Twitter, corresponding to 2,967,701 users.

As usual in this kind or research, the raw data needs to be cleansed [11]. In our case, an important phase in our cleaning process involved removing all the users whose language was not English and their tweets, since we chose a training set of English tweets for our polarity classifier. For the sake of brevity we do not discuss here the details, but all the work ow scripts and a documents detailing each step, together with the initial dataset, can be found at https://rafaelcaballero.github. io/projects/electionsUSA16/. After this phase we obtained a working dataset of 10,638,997 tweets and 1,937,854 users.

Of these users, 906,422 posted messages including only *@realDonaldTrump*, 491,402 only mentioned *@HillaryClinton* and 540,030 mentioned both. Figure 3 depicts this information graphically in the form of a Venn diagram.
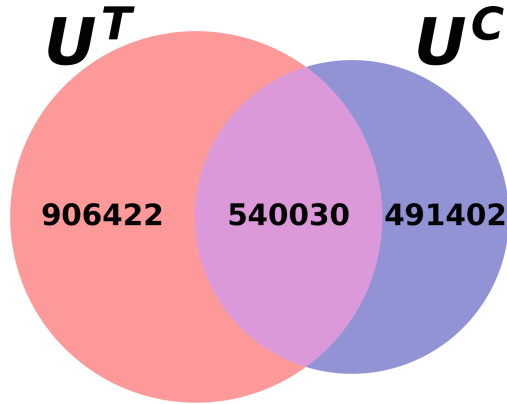
Fig. 3. Venn diagram representing the users that posted messages about each candidate, with U^T representing the set of users mentioning D. Trump, and U^C the set of users that mention H. Clinton.

3.2 Sentiment Analysis of Tweets

Once we have collected the data, the next step is to classify the messages, according to the opinion they contain toward each contender. Since the number of total messages is expected to be high, we need the help of an automatic or semi-automatic technique, such as sentiment analysis [12]. In our use case, we have decided to use a *naïve Bayes* classifier [13] implemented using the Spark ML library [14].

Although we have tried other more complex classifiers, this simple approach has provided the best results. The main reason is that the messages (tweets) considered in our case are highly unstructured, with little inner grammar complexity. Many messages were in fact lists of words (often campaign terms), used to expressed their feelings about a contender, and naïve Bayes classifiers work very well in these contexts. Naïve Bayes classifiers are supervised probabilistic classifiers, that is, they provide a probability of belonging to each class. In our case, we have used this feature to mark as neutral (label 0) those tweets with probability under a certain threshold (calculated automatically using cross-validation during the parameter tuning phase). Thus, the classifier distinguishes between negative polarity (represented by label -1), neutral polarity (0), or positive polarity ($+1$) with respect to the contender. As explained above, we in fact need two models, both based on the same classifier but with different training sets, one for each contender.

The model was trained with 3000 tweets manually labeled for two people. The Cohen's kappa yielded a result of 0.91 agreement in the case of candidate T and 0.88 in the case of candidate C. The trained model was tested on 200 additional tweets manually labeled, giving the following results on precision and recall:

	Precision			Recall		
	−1	0	1	−1	0	1
C	0.80	0.86	0.86	0.76	0.88	0.87
T	0.73	0.81	0.82	0.74	0.88	0.72

The overall F-Score for candidate C is 0.84 and for candidate T 0.78. The results are acceptable, both in precision and recall for both candidates.

3.3   User Opinion About Each Candidate Separately
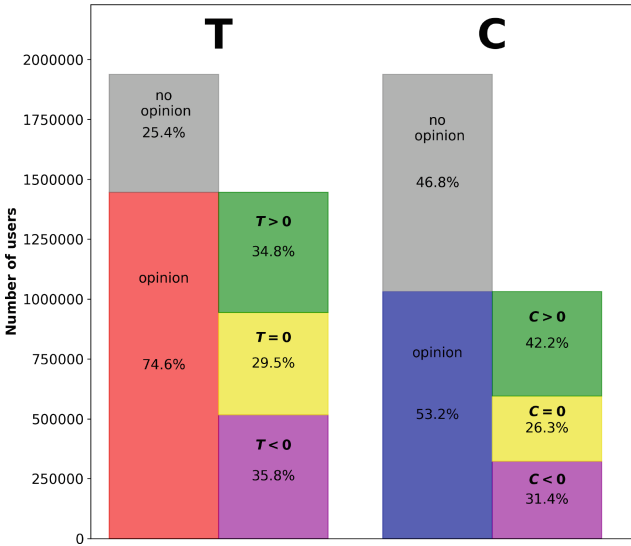
Fig. 4. Distribution of user's opinion for candidate T (left) and C (right). In each case the left-hand side bar represents the proportion of total users with and without opinion about the candidate, while the right-hand side bar divides those with opinion according to their particular sentiment: negative, neutral and positive

From the evaluation of the individual tweets we obtain the average opinion for each user, as explained before. A first glimpse of the results can be observed in Fig. 4. The left hand-side part of the figure shows the distribution of opinions with respect candidate T, while the right-hand side to candidate C. The diagram shows that a 25.4% of the users do not emit any opinion about candidate T, while a 46.8% do not emit any opinion about candidate C.

Among those users with some opinion about C, a majority (42.2%) has a positive opinion, while a 31.4% has a negative opinion and 26.3% are neutral. In the case of candidate T, the proportion of positive opinions is smaller (34.8%),

while the neutral and negative opinion about the candidate is bigger than in the case of candidate C. Thus, we can summarize this information pointing out that:

1. More users emit some opinion about candidate T than about candidate C.
2. In the case of T the neutral and negative opinion represents a bigger proportion than in the case of candidate C.

Thus, if we consider the opinions about each candidate separately we conclude that T obtains more attention than C, but with more negative opinions.
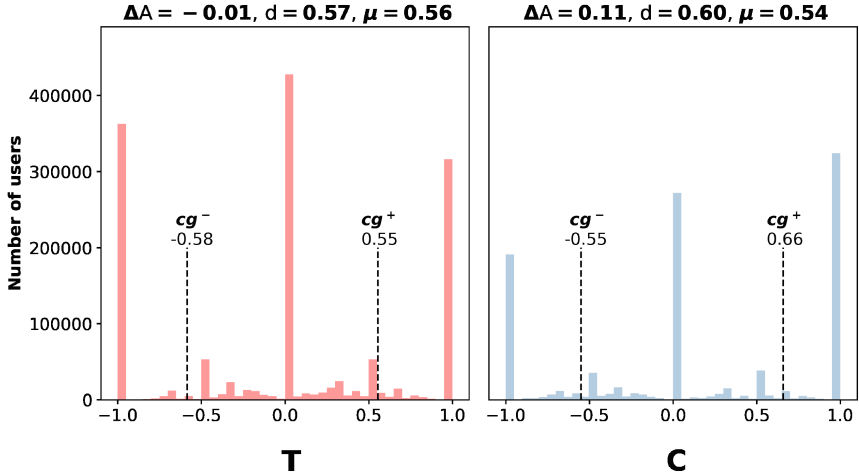
Fig. 5. Distribution of opinions about both candidates, including the gravity centers, the predominant sentiment ΔA, the global polarization index μ and the semi-distance between gravity centers d.

3.4   Polarization

Figure 5 shows the user distributions about each candidate. In the case of candidate T we confirm, as observed above, the negative opinion $(\Delta A < 0)$. Moreover, the numer of users with strong negative opinion $(T = -1)$ is greater than the number of users with extreme positive opinion $(T = +1)$. This is confirmed by the gravity center of the negative part which more shifted to the extreme. The overall polarization in the sentiment about candidate T is 0,56.

In the case of candidate C, the positive sentiments are predominant $(\Delta A = +0.11)$, and there are more users with extreme positive $(C = +1)$ than with extreme negative $(C = -1)$ sentiment. In this case we obtain a slightly less polarized context.

However, this is only part of the picture. Indeed, one of the main points of this paper is that this view that considers each candidate separately is not enough, and relating the sentiments about both candidates provides a deeper insight of the situation, as shown in the next sections.

3.5   Relating Opinions About Both Candidates

The Fig. 6 and 7 have two parts:

– The bottom bars contain the same information as Fig. 4, with the difference that in this case the figures represent the proportions of the four parts (negative, neutral, positive and no opinion), instead of separating the proportion of the users with no opinion from the total as in Fig. 4.
– The upper part further shows the opinion about the other candidate assumed that opinion about the first candidate is given in the lower part.
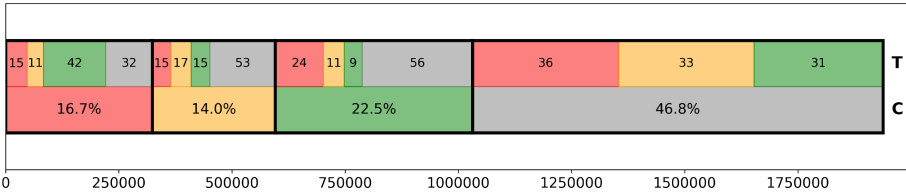
Fig. 6. Distribution of user's opinion for candidate C (below), and each group of opinions in C further divided according to the opinion about T (above). The red color means an overall negative opinion, the yellow color a neutral opinion, the green color a positive opinion and the grey color no opinion about the candidate.
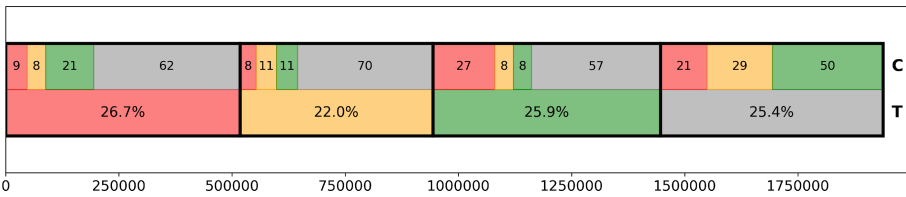
Fig. 7. Distribution of user's opinion for candidate T (below), and each group of opinions in T further divided according to the opinion about C (above). The red color means an overall negative opinion, the yellow color a neutral opinion, the green color a positive opinion and the grey color no opinion about the candidate.

That is, the figures try to show information about questions such as: if we consider the set of supporters of candidate C (green part of the bottom bar in Fig. 6), what opinion do they have about candidate T? Examining the figure we find that:

– As expected, the supporters of each candidate mostly ignore or are against the other candidate, as shown by the sections over the green section of the bottom bars in Figs. 6, 7 and in similar proportion for both candidates.

– In the case of the neutral opinion (yellow bottom bars) both figures show that most of the users do not express a opinion about the other candidate, but this is more accentuate in the case of T, where a 70% of those that present a neutral opinion about T have no opinion about the other candidate.
– Considering the users that have a neutral opinion about one candidate and some opinion about the other (that is, yellow below, not grey above) there are no big differences in the opinion about the second candidate (14% negative, 16% neutral and 15% positive in the case of users with neutral opinion about C, and 8% negative, 10% neutral and 10% positive in the case of those users neutral with respect to T.
– Regarding the users with a negative opinion about one of the candidates, we observe a noticeable difference: of those against T, 62% show no opinion about C, and only a 20% supports the other candidate. However, of those with a negative opinion about C, a 42% support T, the double proportion that in the other case.

Observe that the last point qualifies and even changes the sign of the result of the previous section: while more users express a negative opinion about T, not a great proportion of them consider this disagreement a reason to show support candidate C. On the contrary, those that are against C, show support in a greater proportion to candidate T.
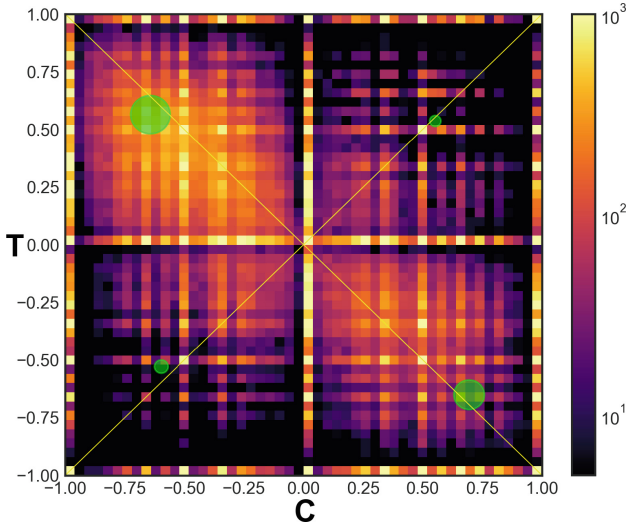
3.6   Love-Hate Diagram

If now we consider the sentiment about each candidate we can depict the love-hate diagram described in Fig. 2, but applied to our particular case. The result can be seen in Fig. 8. Remember that in this diagram support or 'love' for T increases to the upper part while support for candidate C increases to the right.

As expected, the quadrants representing users that support both candidates (right upper) contains the smaller number of users, followed very closely by the quadrant associated to those users disagreeing with both candidates (left bottom). Thus, most of the users are in the left upper quadrant (users supporting T and disagreeing with C) and in the right lower quadrant (positive for C, negative for T).

Of these two quadrants clearly the most crowded is the left upper, associated to 'love' for T and 'hate' for C. Observe how this results presents a new perspective to the result of Sects. 3.3 and 3.4 that examined the opinion about each candidate separately and found that users have a more negative opinion about T than about C. The reason for this difference is that the love-hate diagram considers only users that have opinion about both candidates.

Moreover, we observe that the gravity center of the left upper quadrant is below the diagonal, showing that in the case of users supporting T and disagreeing with C, the 'hate' for the opponent is stronger than the 'love' for the supported candidate, while in the case of the right lower quadrant the opposite occurs. This confirms the results of Sect. 3.5 in the sense than candidate T seems to obtain more support of those disagreeing with candidate C than the other way round.

Fig. 8. Love-hate diagram for candidates T and C. The gravity centers of each quadrant are shown as green circles with diameters proportional to the number of users in the quadrant

4   Conclusions

This paper has proposed tools to examine the relationship between the user opinions about two contenders in a competitive context. We think that this is important to get a deeper insight of the process, understanding not only the characteristics of users that support each candidate, but also what is the opinion about the other candidate of these supporters. This allows us to measure the polarization of the event in the social networks, and determine if the disagreement with one candidate implies support to the other candidate.

In particular, we have seen that in the case of the 2016 United States presidential elections, although in a first view users seem to be mostly against candidate Trump, examining the opinions about the two candidates simultaneously we find that this disagreement with Trump does not imply support for Clinton, while this happens to a major extent the other way round. It is worth noticing that our approach can be extended to the case of more than two contenders considering an approach of 'one against the others' and comparing the results.

We also think that to have polarization indexes clearly defined can be very valuable when comparing similar events. Thus, an interesting line of future work could be to observe the evolution of the polarization indexes in different contexts:

- During one election, observing if the polarization increases when the election is closer in time.
- Consider different elections and compare the evolution of the polarization indexes.

References

1. Barnes, J.A.: Class and committees in a Norwegian island parish. Hum. Relat. **7**, 39 (1954)
2. Ellison, N.B., et al.: Social network sites: definition, history, and scholarship. J. Comput.-Mediat. Commun. **13**, 210 (2007)
3. Bennett, L., Segerberg, A.: The logic of connective action: digital media and the personalization of contentious politics. Inf. Commun. Soc. **15**, 739 (2012)
4. Borondo, J., Morales, A.J., Losada, J.C., Benito, R.M.: Characterizing and modeling an electoral campaign in the context of Twitter: 2011 Spanish Presidential election as a case study. Chaos: An Interdisc. J. Nonlinear Sci. **22**(2), 023138 (2012)
5. Martín-Gutéerrez, S., Losada, J.C., Benito, R.M.: Recurrent patterns of user behavior in different electoral campaigns: a Twitter analysis of the Spanish general elections of 2015 and 2016. Complexity **2018**, 2413481 (2018)
6. Kassing, J.W., Sanderson, J.: Fan-athlete interaction and Twitter tweeting through the Giro: a case study. Int. J. Sport Commun. **3**, 113 (2010)
7. Highfield, T., Harrington, S., Bruns, A.: Twitter as a technology for audiencing and fandom: the #eurovision phenomenon. Inf. Commun. Soc. **16**, 315 (2013)
8. Yadollahi, A., Shahraki, A.G., Zaiane, O.R.: Current state of text sentiment analysis from opinion to emotion mining. ACM Comput. Surv. **50**, 1–33 (2017). Article No. 25
9. Fiorina, M.P., Abrams, S.J.: Political polarization in the American public. Ann. Rev. Polit. Sci. **11**, 563 (2008)
10. Morales, A.J., Borondo, J., Losada, J.C., Benito, R.M.: Measuring political polarization: Twitter shows the two sides of Venezuela. Chaos: Interdisc. J. Nonlinear Sci. **25**, 033114 (2015)
11. Osborne, J.: Best Practices in Data Cleaning: A Complete Guide to Everything You Need to Do Before and After Collecting Your Data. SAGE Publications, Thousand Oaks (2012)
12. Yue, L., Chen, W., Li, X., Zuo, W., Yin, M.: A survey of sentiment analysis in social media. Knowl. Inf. Syst. **60**, 617 (2019)
13. Rish, I., et al.: An empirical study of the Naive Bayes classifier. In: IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, vol. 3, pp. 41–46 (2001)
14. Meng, X., et al.: MLlib: machine learning in apache spark. J. Mach. Learn. Res. **17**, 1235 (2016)