Political Polarization on Twitter after the COVID-19 Outbreak in Spain

Juan Antonio Guevara – Complutense University of Madrid

This study aims to measure levels of public opinion polarization over time during the "state of alarm" (Estado de alarma) in Spain due to the outbreak of COVID-19. To do this, machine learning algorithms have been trained to detect which tweets are 'supporters' or 'detractors' of the Spanish government and its measure to tackle the pandemic. The "JDJ polarization measure" (Guevara et al., 2020) has been applied to measure polarization. Overall, this study has found high levels of polarization during the state of alarm. However, these levels are lower in the first half of the crisis and have increased over time. In conclusion, the attitudes underlying the contents of online social networks seem to have radicalized over time, as users tend to stick to their positions and take them to the extreme.

Introduction

In recent decades, with the appearance of Online Social Networks, society has fully adapted its forms of communication to this new environment. On the grounds of this new situation, the digital sphere provides adequate resources for the rapid dissemination of information, as well as the affordances for people to reach main social or political figures. Although the Internet and Online Social Networks provide innumerable facilities, they also present some risks. Different phenomena have appeared in our society with the arrival of this new scenario while others have increased. This is the example of polarization. This phenomenon might present greater risk in critical scenarios such as a global pandemic. Thus, this study aims to measure the levels of polarization during the state of alarm in Spain due to COVID-19, in which the Government of Spain imposed a confinement. On the grounds of the nature of the crisis, online political communication played an important role on the Internet where thousands of people discussed the crisis management of the government. This scenario represents a crucial situation in which society is at high risk of splitting and therefore polarization could occur.

1. Theoretical background

The facilities provided by Online Social Networks emphasize phenomena intrinsic to human nature such as homophily (Lazarsfeld and Merton, 1954). Due to this phenomenon, people are more likely to interact and communicate with those who are more similar to them. In this sense, society is made up of several groups with a high degree of homogeneity within groups. In those cases in which homogeneity – homophily based on values – plays an important role and those formed groups begin to separate from each other when polarization can occur.

Polarization is understood as the splitting of a certain society into two opposite and extreme groups with significant and similar sizes (Guevara et al., 2020). According to Sartori (2005), polarization does not necessarily mean negative consequences for society. This author states that a centrifugal polarization is, in fact, the one that supposes a risk, understanding it as the breakdown of communication between groups. Therefore, polarization should not be understood as a static phenomenon but as a dynamic one (Guevara et al., 2022).

When scenarios like the ones mentioned above occur, some phenomena appear along with polarization, such as incivility and flaming. Herbst (2010), considers as incivility the use of vulgar or ironic expressions, where some of the interlocutors are shown in an impolite way. In addition, it should be added that as Boxell et al. (2017) affirmed, in those situations in which the uncivil message comes from a source of authority, the messages produced around it are those that contain highest levels of inappropriate content. In this sense, the appearance of the phenomenon of polarization might bring some negative consequences for the proper use of Online Social Networks and, therefore, for society.

2. Methodology

- *Aim of the study.*

In this study I focus on detecting and measuring polarization in the digital debate on Twitter during the state of alarm in Spain. Since Polarization is not static but dynamic (Guevara et al., 2022), I intended to detect the development of this phenomenon by measuring polarization at two different moments in order to detect the increase of polarization levels, being the first half of the state of alarm and the second half.

- *Case of Study and data sources*

From March 15, 2020 to June 21, 2020, the government of Spain imposed the State of Alarm due to COVID-19. This imposition had a great impact on Spanish society due to the lockdown. During this period, Online Social Networks acquired a greater role in society.

I downloaded data from the Twitter API using the R package *"rtweet"* (Kearney, 2019). The data was downloaded in five rounds during the state of alarm because of the limitations of the app. The download phase was implemented through a bag of words composed of the names of the main Spanish political parties, their main representatives, and the words: *"gobierno", "España", and "coronavirus"* and *"estadodealarma"*. After this phase, 4 895 747 tweets were obtained.

- *Data sources and filtering*

Although tweets are downloaded according to a specific bag of words, it is usual that some non-desirable tweets are included in the dataset. To filter the data some machine learning algorithms were trained to exclude these non-desirable tweets. To do so, 1 500 tweets per round were labeled as "*desirable"* or *"non-desirable"* tweets depending on their content. The Support Vector Machine (SVM) algorithm was the one that showed better performance (see Table 1).

Table 1. SVM results for the filtering task by SVM

Round	Precision	Sensibility	Kappa	F-Score	AUC
1	0.8017	0.9322	0.3670	0.8620	0.6583
2	0.8167	0.5476	0.5077	0.6556	0.7344
3	0.8267	0.7027	0.6187	0.7596	0.8010
4	0.7867	0.7090	0.564	0.7457	0.7791
5	0.7659	0.8758	0.5216	0.8171	0.7567

After the filtering phase, 1 208 631 tweets remained.

- *Data encoding*

To find out what is the implicit position in the content of each tweet, 1 500 tweets per round were labelled as "supporter", "detractor" or "neutral" towards the government of Spain. This criterion not only included the position of a certain tweet towards the government of Spain but also its radicalization, Thus, a "supporter" is a tweet that presents extreme opinions that support the government, a "detractor" is a tweet that presents extreme views that are against the government and "neutral" those tweets that present (1) neutral information or (2) are published neutrally. Finally, I trained a Machine Learning model using the Natural Language Processing methodology. The results can be seen in Table 2, in which good levels of performance can be seen.

Table 2. SVM results for the labelling task by SVM

Round	Precision	Sensibility	Kappa	F-Score	AUC
1	0.8492	0.9854	0.4816	0.9122	0.6950
2	0.8960	0.9619	0.7761	0.9277	0.8780
3	0.8392	0.8488	0.6675	0.8439	0.8366
4	0.9133	0.9048	0.8225	0.9090	0.9121

5	0.8318	0.8600	0.6638	0.8456	0.8335

The entire data set was split into two halves to have two different times to compare with each other. The first half includes all tweets posted between the 15th of march and the 1st of May, while the second half includes the tweets published between the 1st of May and the 21st of June.

- *Measurement of Polarization: JDJ measure.*

To measure polarization I apply the measure proposed by Guevara et al. (2020). This measure is based on the fuzzy approach that maintains that reality is not black or white but there are some nuances. With this approach, the authors understand that a certain individual should be a supporter of a certain political party but also feel identified by some proposals from other political parties. This measure is based on the radicalization of a certain element, using the membership degree of a given element to belong to both poles of the attitudinal variable at the same time. Thus, it compares (1) how element *i* belongs to the extreme position A (e.g.: being a detractor) and how element *j* belongs to the extreme position B (e.g.: being a supporter) and (2) how element *i* belongs to the extreme position B (e.g.: being a supporter) and how element *j* belongs to the extreme position A (e.g.: being a detractor). Authors add these two-way comparisons to the computational risk of polarization between two individuals. Then, all the possible comparisons are computed on the population to compute a final polarization value.

$$JDJ(X) = \sum_{i,j \in N, i \leq j} \varphi\left(\phi\left(\mu_{X_A}(i), \mu_{X_B}(j)\right), \phi\left(\mu_{X_B}(i),\ \mu_{X_A}(j)\right)\right)$$

Where ϕ is an overlapping aggregation operator and φ is a grouping function. In this study, we use the product for ϕ and the maximum for φ.

To ensure that JDJ is an index – values between 0 and 1 – to improve its interpretability, we make the next transformation:

$$JDJ\_INDEX = \frac{JDJ}{N} * 2$$

Where *N* represents all the possible comparisons in the population. The value 1 is given when 50% of the elements present have a degree of belonging to one pole equal to 1 and 0 to the other pole and the other 50% of the elements is the opposite. In contrast, 0 occurs when not only do all the elements show the same membership towards both poles, but also when these membership degrees are extreme.

We use the probabilities of a certain user of belonging to both categories (poles), provided by the machine learning algorithm, as membership degrees to compute the JDJ measure. Note that since we are labelling tweets, we focus on measuring the polarization of the debate and not on users.

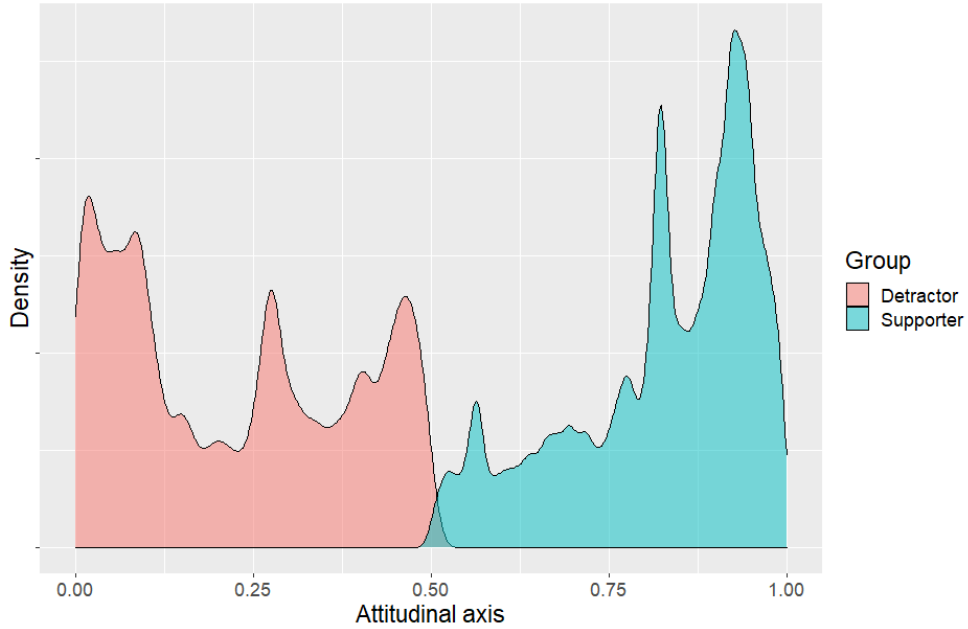
3. Results

Once the machine learning algorithm provides the probabilities that a given element – *tweets* in this case – belongs to the two opposite categories (*supporter* or *detractor* towards the government of Spain), we use them as membership degrees for each element to compute JDJ. Probabilities near to 0 represent the "*detractor"* position while probabilities close to 1 represent the "*supporter"* position. Since I have split the dataset into two halves, it can be seen the density functions for these probabilities in the following figures for each part of the data set (see Figure 1 and Figure 2). Since these figures represent the density of these probabilities, we can interpret the attitudinal position of all the tweets. High densities near 0.5 indicate a large number of tweets with fuzzy positions while high densities near the poles indicate a large number of extreme tweets.

On the other hand, since the calculation of polarization values includes the comparison between all elements, 1 208 631 tweets suppose $1\,208\,631^2$ comparisons. To avoid computational costs, I computed the JDJ values for 1 500 iterations of a random sample N = 200 for each half of the data set.

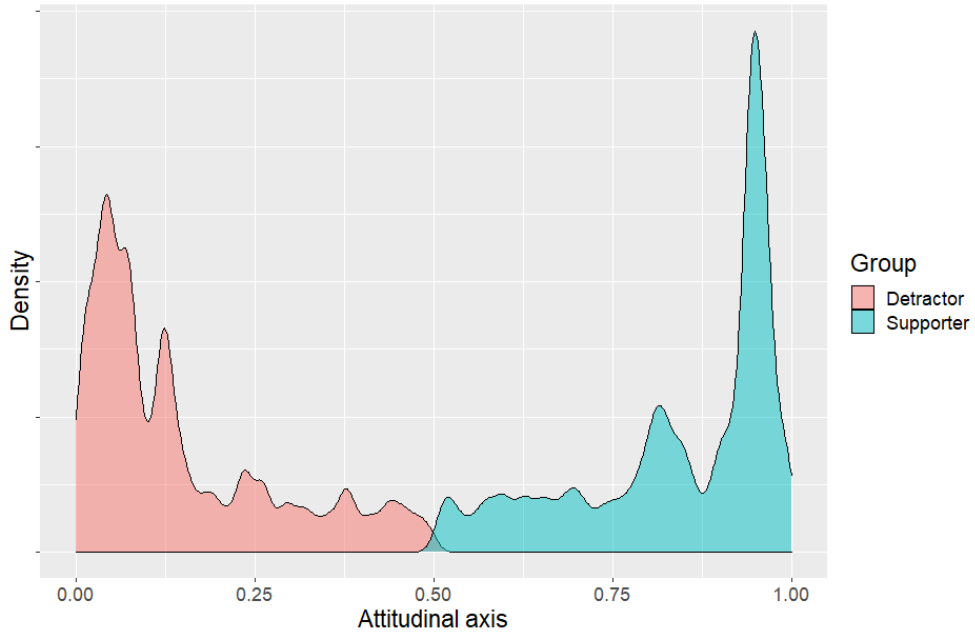
Thus, in the first half (Figure 1) it can be seen that those tweets classified as *supporters* of the government of Spain tend to hold extreme positions whilst the *detractor* group tend to be more moderate. The mean polarization values for the 1 500 iterations were computed being $JDJ_1 = 0.733; sd = 0.033$.

Figure 1: Density functions for the probabilities provided by SVM according to being a *supporter* or *detractor* towards the government of Spain for each tweet (First half: 15th of march and the 1st of May).



In contrast, in the second half (Figure 2) it can be seen how these probabilities have moved to the extremes, disappearing intermediate positions and radicalizing. This fact is a strong signal that opinions became radicalized over time. In fact, when the average values for the 1 500 iterations are computed, high levels of polarization can be found, being $JDJ_2 = 0.912; sd = 0.006$.

Figure 2: Density functions for the probabilities provided by SVM according to being a *supporter* or *detractor* towards the government of Spain for each tweet (Second half: 1st of May and the 21st of June).



Conclusion

As can be seen in the previous section, this study has found high levels of polarization in both situations (first and second half of the state of alarm). Moreover, it can be seen that the levels of polarization have increased over time. These findings support the idea that polarization is dynamic and there is a high probability that it will increase as a crisis situation becomes more dramatic. According to Figures 1 and 2, it can be seen that opinions become more radicalized over time during crisis scenarios such as the COVID-19 pandemic. In fact, tweets that have moderate content (higher density around 0.5 in Figure 1) tend to spread to the extreme positions, flattening the density of moderate positions and becoming wider and more radical. This could be a consequence of homophily and the action of filter bubbles and echo chambers during a severe crisis such as a pandemic. It can be concluded from this study that the role played by Online Social Networks and political communication on the Internet has facilitated the radicalization of the content posted by users.

References

Boxell, L., Gentzkow, M., & Shapiro, J. M. (2017). *Is the internet causing political polarization? Evidence from demographics* (No. w23258). National Bureau of Economic Research.

Guevara, J.A., Gomez, D., Robles, J.M., Montero, J. (2020). "Measuring Polarization: A Fuzzy Set Theoretical Approach". *Communications in Computer and Information Science*, vol 1238. Springer, Cham. https://doi.org/10.1007/978-3-030-50143-3_40

Guevara, J.A., Gómez, D., Castro, J., Gutiérrez, I., Robles, J.M. (2022). "A New Approach to Polarization Modeling Using Markov Chains. Communications in Computer and Information Science", vol 1602. Springer, Cham. https://doi.org/10.1007/978-3-031-08974-9_12

Herbst, S. (2010). *Rude Democracy: Civility and incivility in American politics*. Temple University Press.

Kearney, M. W. (2019). rtweet: Collecting and analyzing Twitter data. *Journal of open source software*, *4*(42), 1829.

Lazarsfeld, P.F. and Merton, R.K. (1954). "Friendship as a social process: A substantive and methodological analysis". in *Freedom and control in modern society*. New York: Van Nostrand, 18‑66.

Montalvo, J. G., & Reynal-Querol, M. (2003). Religious polarization and economic development. *Economics Letters*, *80*(2), 201-210.

Sartori, G. (2005). *Parties and party systems: A framework for analysis*. Colchester: ECPR Press.