CUADERNOS.INFO Nº 55
Versión electrónica: ISSN 0719-367x
http://www.cuadernos.info
https://doi.org/10.7764/cdi.55.52631

Received: 08-06-2022 / Accepted: 12-29-2022

Machine learning for cross-platform political communication research: Argentine government and opposition in Facebook, Instagram and Twitter

Aprendizaje automático para el análisis cross-plataforma de la comunicación política: Gobierno y oposición argentinos enFacebook, Instagram y Twitter

Aprendizagem de máquina para análise entre plataforma da comunicação política Governo e oposição argentinos no Facebook, Instagram e Twitter

Federico Albanese, Instituto de Ciencias de la Computación, Universidad de Buenos Aires, Buenos Aires, Argentina (falbanese@dc.uba.ar)

Esteban Feuerstein, Instituto de Ciencias de la Computación, Universidad de Buenos Aires, Buenos Aires, Argentina (efeuerst@dc.uba.ar)

Gabriel Kessler, Universidad Nacional de La Plata, La Platay Universidad Nacional de San Martín, San Martín, Argentina (gkessler@fahce.unlp.edu.ar)

Juan Manuel Ortiz de Zárate, Instituto de Ciencias de la Computación, Universidad de Buenos Aires, Buenos Aires, Argentina (jmoz@dc.uba.ar)

ABSTRACT | This article studies political communication in different platforms, applying data science methods to analyze similarities and differences among Facebook, Instagram, and Twitter posts of 50 Argentinian politicians in 2020. This is a pioneering cross-platform study for our region, and its objectives are heuristic and methodological. Regarding the former, we show that strategies differ among platforms: Twitter is the battlefield for controversy and interpellations among politicians, and toxicity is rewarded, while on Facebook and Instagram politicians expand on the topics in which they seem to consider themselves stronger. The closs-platform study shows that even in a polarized context such as the Argentinean one, there are common and non-controversial topics. Methodologically, we use novel analytical methods and implemented a recent topic-detection algorithm, we apply sentiment analysis techniques to understand if texts have positive or negative intentions, and deep neural networks to detect toxicity in a text, among others. Readers are offered access to the toolbox developed during the research, which can be useful for working large text corpora.

KEYWORDS: social media; Twitter; Instagram; Facebook; Argentina; politics; natural language processing; topic modeling.

HOW TO CITE

Albanese, F., Feuerstein, E., Kessler, G. & Ortiz de Zárate, J.M (2023). Aprendizaje automático para el análisis cross-plataforma de la comunicación política: Gobierno y oposición argentinos en Facebook, Instagram y Twitter. *Cuadernos.info*, (55), 256-280. https://doi.org/10.7764/cdi.55.52631

RESUMEN | *Este artículo indaga acerca de la comunicación política en las distintas plataformas, aplicando métodos de las ciencias de datos para analizar similitudes y diferencias entre las publicaciones en Facebook, Instagram y Twitter de 50 políticos argentinos durante 2020. Es un estudio pionero en la región entre los trabajos cross-plataformas y sus objetivos son heurísticos y metodológicos. En relación a lo primero, se demuestra que hay estrategias diferentes según las plataformas: Twitter es el terreno de controversias e interpelaciones entre los políticos y allí la toxicidad es recompensada, mientras que en Facebook e Instagram los políticos despliegan los tópicos en los que parecen considerarse más fuertes. Así, el estudio cross-plataformas permite observar que aun en un contexto polarizado como el argentino existen temas comunes y sin polémicas entre sectores opuestos. En lo metodológico, utilizamos métodos novedosos e implementamos un reciente algoritmo de detección de tópicos, aplicamos análisis de sentimiento con el objetivo de entender si son textos positivos o negativos, y redes neuronales profundas para medir la toxicidad, entre otros. El artículo pone a disposición la caja de herramientas desarrolladas durante la investigación, las que pueden ser de utilidad para trabajar corpus de texto de gran magnitud.*

PALABRAS CLAVE: *redes sociales; Twitter; Instagram; Facebook; Argentina; política, procesamiento del lenguaje natural; modelado de tópicos.*

RESUMO | O artigo investiga a comunicação política em diferentes plataformas, aplicando métodos de ciências de dados para analisar as semelhanças e diferenças entre as postagens no Facebook, Instagram e Twitter de 50 políticos argentinos durante 2020. Trata-se de um estudo pioneiro na região no trabalho interplataformas e seus objetivos são tanto heurísticos quanto metodológicos. Em relação aos primeiros, o artigo mostra que existem estratégias diferentes segundo as plataformas: o Twitter é terreno de controvérsia e interpelações entre políticos, onde a toxicidade é recompensada, enquanto no Facebook e no Instagram os políticos expõem os tópicos nos quais eles parecem se considerar mais fortes. O estudo interplataformas permite-nos observar que mesmo num contexto polarizado como o da Argentina, existem questões comuns e não controversas entre setores opostos. Metodologicamente, nós usamos novas técnicas e implementamos um algoritmo recente de detecção de tópicos; aplicamos técnicas de análise de sentimentos com o objetivo de entender se os textos são positivos ou negativos, e redes neurais para detectar toxicidade nas mensagens, entre outros. O artigo oferece acesso à caixa de ferramentas desenvolvidas durante a pesquisa, e que podem ser úteis para trabalhar com outros grande corpus de textos.

PALAVRAS-CHAVE: redes sociais; Twitter; Instagram; Facebook; Argentina; Política; Processamento de linguagem natural; Modelagem de tópicos.

INTRODUCTION

How does the existence of multiple platforms affect the communication strategies of politicians, and what are the similarities and differences displayed by their publications on each of them? These questions gathered sociologists and computer scientists, who in this article share their methodological contributions to the incipient cross-platform research. To answer these questions, we focus on the Facebook (FB), Twitter (TW) and Instagram (IG) posts of 50 Argentine politicians from the ruling party and the opposition during 2020.

We are in the phase that Diana Owen (2017) has called New media, new politics 2.0, whose beginning dates back to the 2008-2010 US election campaign (election of Barack Obama and midterm, respectively), and whose novel features would be the widespread and sophisticated use of digital technology, the management of different platforms, and increased interaction with audiences and of users with each other.

There has been an interest in Twitter in the last decade, due to its role as a space for political controversies and its accessibility to collect data. In Latin America there have been comparative works (Cárdenas, 2020; López-López & Vásquez-González, 2018), as well as case studies in Argentina (Calvo, 2015), Brazil (Paulino & Waisbord, 2021), Chile (González-Bustamante, 2015), Colombia (Prada Espinel & Romero Rodríguez, 2018), and Mexico (Salgado Andrade, 2013), among others. However, there is consensus that Twitter is a space restricted to those more interested in politics, more polarized, and with more cultural resources. Thus, caution is recommended when considering it as representative of the entire digital sphere and, even more, of the general political conversation (Stier et al., 2018), and it is suggested to expand the gaze to other platforms, although this is made difficult by restrictions on access to data. Cross-platform analysis considers each user and the different social networks with which he or she interacts on a frequent basis as the object of study and unit of analysis (Rogers, 2017). As said author suggests, a thorough analysis of why more and more users use several platforms at the same time –and even more public personalities– is needed. For now, there is a nascent field of cross-platform political communication studies of electoral campaigns from core countries, such as the United States (Bossetta 2018), Germany (Stier et al., 2018), Norway (Enli & Skogerbø, 2013), and Sweden (Larsson, 2015). Conceptual studies ask how political logic influences the architecture of different interactive media (Chadwick et al., 2015; Owen, 2017), and others compare traditional media with Twitter (Karlsen & Enjolras, 2016).

Methodologically, these works have resorted to metadata analysis (Likes, retweets, etc.) and, to a lesser extent, to qualitative techniques (Spierings & Jacobs, 2019) and discourse analysis with novel approaches (Stier et al., 2018). These show that politicians or parties use different strategies according to each platform and that there is still a predilection for Facebook despite the central place of Twitter in debates and polemics. They also note a growing use of Instagram (and, until a few years ago, Snapchat), although there is still little work on TikTok. They stress the need for methodological innovations to increase the studies' scope and rigor (Hasebrink & Hepp, 2017; Nielsen & Schrøder, 2014; Owen, 2017; Thorhauge & Lomborg, 2016). Stier and colleagues (2018) point out three limitations of most work on political communication in digital media. The first is that they are generally based on campaign periods in core countries, but few account for communication in ordinary times; second, they tend to focus on a single platform, and third, they analyze more metadata than content.

This research, a pioneering study in the region, aims to overcome these limitations. The database is made up of the official and public accounts of 50 very relevant Argentine political figures (in terms of positions, responsibilities, or notoriety) of the national government (Frente de Todos), which we will call ruling party, and of the opposition (Juntos por el Cambio) that rules in some provinces and major cities, during 2020, a year without national elections (although the period considered naturally includes the COVID-19 pandemic). Our theoretical framework articulates theories of agenda setting, sociology of public problems, and framing, as developed in the following section. The methodology focuses on both publications and metadata, since we took each politician's presence in three platforms and studied them using an original computational approach. Indeed, the application of computational methods to the study of social sciences is proving to have enormous potential, so much so that the main research associations in our region have created permanent groups to discuss their use (Arcila Calderón et al., 2021). These approaches have made it possible to work with important corpora for the analysis, among others, of networks and discourses of different types. However, many of the most widespread tools for natural language processing are limited and inefficient, as they are difficult to set up and computationally expensive for large data volumes. Therefore, our main contribution is the use of modern data science tools –natural language processing (Angelov, 2020), machine learning (Bishop, 2006), and toxicity analysis (Fortuna et al., 2020), among others– for the study of discussions in digital platforms, but which can be used for different corpora. The techniques and methods developed in this work form a toolbox that will be available in a public repository[1].

1. https://anonymous.4open.science/r/Aprendizaje-automatico-para-el-analisis-crossplataforma-de-la-comunicacion-politica-B994/README.md

As mentioned, we posed the question of whether politicians communicate in a similar way in the three social networks that have the greatest presence in politics or whether there are differences in the way they do so in each of them. Our initial hypothesis was that Twitter should exhibit distinctive characteristics from the other two, being a space of interaction (Gruzd et al., 2018; Jaidka et al., 2018), of talking to others, while Facebook and Instagram would have more similarities with each other. With this in mind, we conducted tests aimed at finding similarities and differences between platforms and between the ruling party and the opposition in a highly polarized country (Ramírez & Quevedo, 2021). We found that, indeed, the government and the opposition are more interpellative on Twitter than on the other two networks, and that on this network politicians (regardless of the sector they belong to) tend to discuss issues in common. In Twitter, we also found a strong correlation between the degree of toxicity of the messages and their impact. In contrast, in the other two networks, the ruling party and the opposition mainly talk about different topics, about which they have more ownership, and the level of toxicity is low. We also detected shared topics between the ruling party and the opposition in which there is no confrontation.

The paper is organized as follows: first, we present the hypotheses and their foundations, then, the tests performed to prove each hypothesis, with emphasis on the methodological approach, and, finally, the article's conclusions.

THEORETICAL FRAMEWORK AND HYPOTHESIS

As stated, our hypotheses are grounded in different long-standing communication theories that have been applied to political debate; agenda setting studies (Aruguete, 2015), framing theory (Scheufele, 2000; Scheufele & Iyengar, 2012), and work on ownership from the sociology of public issues (Gusfield, 2014). We apply this to cross-platform differences. We assume that messages can be differentiated (1) in relation to the agenda, i.e., talking about different topics on each platform, (2) by framing, i.e., talking about the same topics, but framed differently depending on the network, or (3) in their interpellative or vocative dimension, i.e., with respect to the recipient to whom they would be addressed. Options 1 and 2 would be mutually exclusive; on the other hand, dimension 3 can be combined with either 1 or 2 (for example, the topic and the framing can be maintained, but vary on one platform, and on another to whom it would be addressed)).

In terms of the ruling party and opposition variable, property theory (Kelley & Mirer 1974; Petrocik 1992) asserts that politicians had to refer to issues in which they felt more at ease: traditionally, in the United States, it was convenient for Democrats to talk about racial integration and welfare, and for Republicans to talk about crime and national security. In contrast, others argued that ownership was not a convincing strategy for audiences and that it was necessary to ride the wave (Ansolabehere & Iyengar, 1994), without slipping away from the issues of the moment, at the risk of being considered cynical or out of tune with the public's concerns (Iyengar, 1990). On the other hand, from the works that ask about how the architecture of each platform gravitates, we took what Bossetta (2018) calls network structure, i.e., the technical rules that regulate the relationship between users on each platform. We thus presupposed that Twitter promotes an interpellative to-each-other conversation, since it favors controversy between users with different ideas, because followers are not selected, while in the other two the followers are usually more like-minded people and do not usually engage in controversy; thus, they are more conducive to a past-each-other communication: the emitter chooses on which topics to publish and can guide the agenda with less interference from opponents.

In this regard, based on Kaplan and colleagues (2006), we assume that the ruling party and the opposition are more likely to talk about the same topics (low ownership) on Twitter and about different topics (high ownership) on Facebook and Instagram. In other words, we conjecture that ruling party and opposition choose (or have no choice) one network to debate and the other(s) to promote themselves on the topics in which they consider themselves stronger. Likewise, we assumed that in ordinary times such as the one we studied (not election campaigns) the political space is not only one of confrontation with the opponent and celebration of one's own actions, but that there would be messages in common from both the ruling party and the opposition in which controversy is less plausible.

Based on the above, our hypotheses are as follows:

> *H1a.* Topics: each group chooses Facebook and Instagram to talk about the topics in which it has ownership, while Twitter becomes the platform on which topics without exclusive ownership of one or the other group are discussed.
>
> *H1b.* Topics in common: the topics in common between the ruling party and the opposition do not only include confrontations, but also coincidences or topics of low potential conflict.

Our second hypothesis is related to the differences in the platforms' framing. As we argued, one option would be for the ruling party and opposition to talk about the same issues with a different framing, specifically, with a different and often opposite valuation. In this regard, valuation theory within framing studies (Martin & White, 2005) focuses on the linguistic resources through which texts/speakers come to express, negotiate, and naturalize intersubjective and

–ultimately– ideological positions. This theory pays attention to the valuation, attitude, and emotion embedded in discourses that denote different positions of the enunciator. We conjecture that a difference between the ruling party and the opposition will be the valuation of the main topics on the agenda. Thus, the same topic will have a positive connotation for some and the others will criticize it, which would change their affective valuation (for example, a government action for the ruling party in each jurisdiction). We assume that the negativity will be mainly on Twitter, since it is the network of controversy. Therefore, our second hypothesis is:

> *H2a.* Sentiments: ruling party and opposition tend to enunciate messages with different sentiments (positivity/negativity) depending on the network through which they are expressed.

> *H2b.* Negativity on Twitter: Twitter is the platform where there is the highest proportion of messages expressing negative sentiments due to the higher frequency of confrontational interactions.

Our third hypothesis is related to the fact that on Facebook and Instagram content is displayed mainly based on the accounts the user follows, while on Twitter it is based on topics of interest. This promotes more discussion among users, not only in spaces characterized by homophily (McPherson et al., 2001) but also by people with other points of view. Based on this, our third hypothesis is:

> *H3.* Interpellation: politicians tend to interpellate each other more on Twitter than on Instagram and Facebook.

METHODOLOGIES AND EXPERIMENTS

In this section we will detail the techniques and methods we applied to test our hypotheses. A widely used tool for topic detection is Voyant-Tools[2] (Flores-Márquez & González Reyes, 2021) which uses the Latent Dirichlet Allocation algorithm (Blei et al., 2003). While it is useful on certain data corpora, its performance worsens when dealing with large volumes of unstructured data, such as social networks. Other popular techniques for sentiment analysis, such as SentiStrength (Thelwall, 2017), do not perform well beyon English (Garimella et al., 2018) and, therefore, methods need to be developed to address digital data from our region, in Spanish and Portuguese.

Dataset construction

Argentina has been ruled since 2019 by Alberto Fernández, elected that year with Cristina Fernández de Kirchner as vice-president, heading the Frente de

2. https://voyant-tools.org/

Todos, an alliance between different currents of Peronism that defeated former president Mauricio Macri, who was seeking reelection with the coalition Juntos por el Cambio. This alliance is formed by Propuesta Republicana (PRO), Unión Cívica Radical (UCR), Coalición Cívica ARI, and Peronismo Republicano, which in this paper we call opposition, while we call the former ruling party. To build our corpus, we selected 50 political figures –25 from each political trend–, with characteristics as homogeneous as possible in both groups in terms of positions, responsibilities, or notoriety, making sure that all of them had official accounts on Facebook, Twitter, and Instagram (see appendix for details).

From the ruling party, we selected 12 personalities with positions in the executive branch (main ministers and first line of the national executive branch), and 13 senators and deputies from different provinces and with high public exposure. From the opposition, we selected 11 executive positions, of which seven are current (mayors of the main cities and governors) and four are former (former president, former governor of the province of Buenos Aires, the country's main province, president of Pro and former minister of Security, former governor of the province of Mendoza and president of the UCR), and 14 relevant deputies and senators. Using Twitter[3] and CrowdTangle[4] APIs, we downloaded all the posts they published during 2020 on the three platforms, with a total of 150 accounts (three for each political figure) and 84,435 posts, of which 56,622 are from Twitter, 16,133 from Facebook, and 11,680 from Instagram. Although images are an important component of Instagram (Bast ,2021; Figuereo-Benítez et al., 2021), for this paper we have limited ourselves to analyzing the text of the posts. A first finding is that Argentine politicians make more than twice as many posts on Twitter than on Facebook and Instagram.

How to elucidate the topics discussed? H1. Topics in common and themes

Method

To test H1, we needed to identify the own topics and shared topics of the ruling party and the opposition in each platform. Traditional algorithms for topic detection and modeling, such as LDA (Blei et al., 2003), demand that we provide a priori the number of topics into which we want to divide the corpus; therefore, the coherence of the resulting division depends on matching this parameter with the real one, which requires testing with different parameters until we find the correct value. As our dataset was voluminous, the number of discussed topics was likely to be very high, and we would have needed to perform numerous attempts until we reached the correct value (Röder et al., 2015). We thus resorted to a

3. https://developer.twitter.com/en/docs/twitter-api

4. https://www.crowdtangle.com/

more recent technique, Top2Vec (Angelov, 2020), which estimates the number of topics without first validating the consistency of each possible value, reducing the computation time. According to Top2Vec, 1,028 topics were discussed. This technique, moreover, does not require eliminating stopwords (articles, prepositions, etc.) or normalizing the text, and makes it possible to identify deterministically what topic a given post was about[5]. After identifying the topics, we analyzed which ones belonged to each political sector. We defined the categorization between own and topics in common as:

- Own topic: topic in which 95% or more of the posts come from the same political sector.
- Topic in common: topic on which each group produced between 45% and 55% of the posts.

On which platforms they discuss their own topics and on which ones the common ones?

To test H1a, we measured the proportions of own and shared topics in each network. It was necessary to normalize the number of posts per politician and per social network, since politicians generally make more daily posts on Twitter than on Instagram and Facebook. The following figures show these proportions.

Twitter is the most used network for topics in common, and Facebook and Instagram for owned ones, and we can confirm that there is a different agenda on the platforms, so we were able to verify H1a.

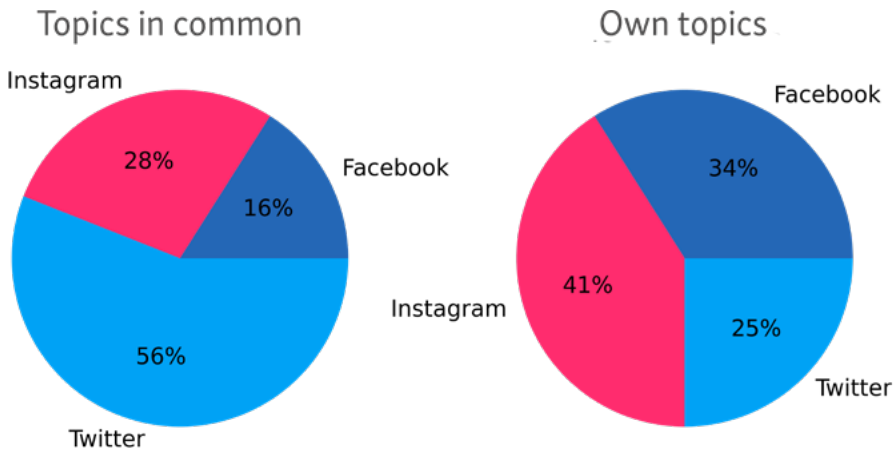


Figure 1. Proportion of posts by social network of topics belonging to a single political sector (own topics) and shared by both (common topics)

Source: Own elaboration.

5. LDA only produces a score for which it is then necessary to define a threshold from which it is possible to say that a post talked about a certain topic.

H1b Results (common topics)

We analyzed the main words and posts of each topic by group:

Ruling party

- Defense of the river: campaign of the national ruling party regarding the discussion on what to do with fiscal lands adjacent to the Río de la Plata.
- Women's rights: campaigns promoted by the ruling party.
- Levantarnos: the government's campaign to get out of the economic crisis and the pandemic.
- Revolución de las Viejas: campaigning for the rights of older women.
- Anti-discrimination campaigns: campaigns to combat xenophobia, sexism, classism, among others.

Opposition

- COVID figures, San Isidro: reports on COVID-19 cases in San Isidro, a district governed by the opposition.
- COVID CABA reports: reports on COVID-19 cases in the Ciudad Autónoma de Buenos Aires (CABA), a district ruled by the opposition.
- Volunteering to care for the elderly: CABA government program to assist the elderly during quarantine.
- Virtual meetings with neighbors: virtual activities with neighbors of the districts governed by the opposition.
- Criticism of Kirchnerism: criticism of the opposition regarding the ruling party represented by Cristina Fernández de Kirchner.

The topics coincide with the agendas of each sector in the media. Indeed, the ruling party focuses on management campaigns and social or rights issues, while the opposition refers to management in their districts and criticizes the ruling party (focusing on the main sector of the ruling coalition represented by vice-president Cristina Fernández de Kirchner).

What are the topics in common?

To test H1b, we set out to see which topics had similar participation by party, which we defined as topics in common, and we detected the following ones (among others):

Non-controversial topics in common

- Greetings and recognition to workers: greetings to firefighters, healthcare workers and other workers on their respective feast days.

- Condolences on death: on the occasion of the death of political figures (e.g., a federal judge, former national senator, or former governor of a province).
- Malvinas War Commemoration: on the occasion of the anniversary of the Malvinas War against the United Kingdom in 1982.
- Retiree care: Messages on the importance of caring for retirees in a pandemic.
- National holidays: messages for national commemorations, such as Independence Day or National Day.

Common controversial topics

- Sputnik vaccine: discussions about this vaccine of Russian origin; the national government posted about its purchase and the opposition denounced that it was of poor effectiveness.
- Mentions about Ginés: mentions concerning Ginés González García, National Minister of Health. While the government announced activities with him, the opposition criticized him for his management.

We thus confirm that hypothesis H1b is met, since most of the topics in common are non-controversial, with the exception of the Sputnik vaccine and the mentions of the Minister of Health.

H2. Sentiment and negativity on Twitter

H2a method (sentiment)

To test hypothesis H2a we sought to characterize the messages' positivity and negativity (Ain et al., 2017) by applying a convolutional neural network[6] (LeCun et al., 1989) for sentiment analysi[7]s on the posts. Then, with a statistical test, we tried to detect if there were significant differences between the proportion of positive and negative posts from each of the networks.

Results

We found no significant disparities. Therefore, our hypothesis that politicians express themselves with different positivity or negativity depending on the social network was not verified with this method.

6. Convolutional neural networks are a deep neural network architecture widely used for image and text analysis.

7. Sentiment analysis library in python: sentiment-spanish (https://pypi.org/ project/sentiment-analysis-spanish/).

H2b method (negativity on Twitter)

We performed new tests with novel developments around toxicity: a message is considered toxic if its rude and disrespectful tenor can cause the interlocutor to leave a conversation (Fortuna, 2020). To measure toxicity, we use the Perspective API (Wulczyn et al., 2017) that uses deep neural networks for natural language processing pre-trained for such a task. This algorithm assigns each text a value between 0 and 1, which represents the probability of the message being toxic. Following the methodology used by other authors (Hua et al., 2020), we define a cutoff value above which we consider a message as toxic.

Results

When quantifying the number of toxic messages in each social network, it was observed that the proportion, although small in the three social networks, was considerably higher in Twitter: 7.6% against 1.2% in Instagram and 0.4% in Facebook. The question arises: Why do politicians have incentives to publish messages with greater toxicity on one network compared to the other two? We found that, on Twitter, the higher toxicity corresponds to a much higher number of Likes, but the same does not happen on the other networks. The following figure shows the results for each social network.

When performing a Spearman test (Zwillinger & Kokoska, 2000), on Twitter we observed a statistically significant ($p<0.05$) and positive correlation between a post's number of Likes and toxicity with a correlation coefficient of 0.15. On the other hand, in Facebook this coefficient is less than half, 0.08. On Instagram there is not even a significant correlation between toxicity and Likes. We conclude that, as H2b states, politicians would have incentives to be toxic on Twitter, but not so much on Facebook and Instagram, since the former rewards toxicity with a much higher number of Likes.

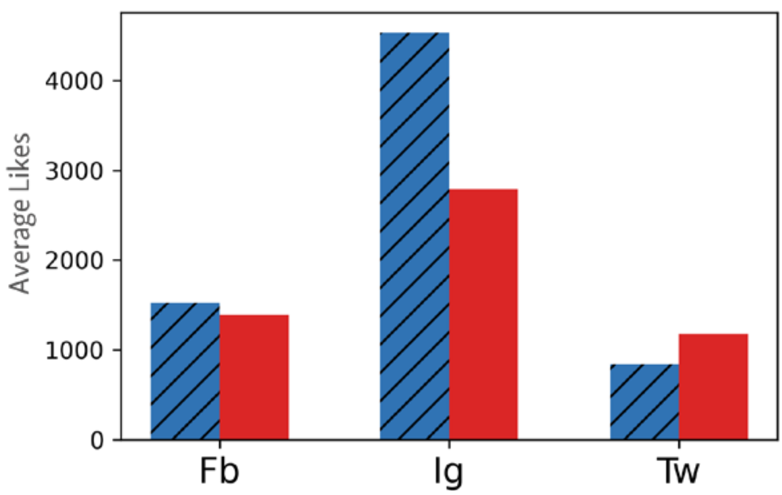


Figure 2. Average number of Likes on toxic (red) and non-toxic (blue) messages on each platform

Source: Own elaboration.

H3. Interpellation

Method 1

To test our third hypothesis, we performed several steps. First, we compared the messages' discursive form to observe differences between platforms. We used natural language processing techniques that measure the similarity of the texts according to the number and significance of shared words: two texts are considered more similar if they share many words that are not very common in the rest of the corpus. To do so, we vectorized the text using the Term frequency - Inverse document frequency (Tf-idf) word frequency counting technique, and measured similarity through cosine similarity (Manning & Schütze,1999). Thus, we found that the accounts of the ruling party and the opposition on Twitter were, overall, very similar to each other, much more so than those of the other two networks. To capture the uniqueness of Twitter, we analyzed word distribution patterns to discover which words were most frequently found together. Using the singular value decomposition (SVD) technique[8], we found the main groups of words (dimensions), and then trained a decision tree (Bishop, 2006) to predict which platform each user belonged to; a decision tree trained to classify texts according to whether they belonged to Twitter, Facebook, or Instagram can detect if there is a group of words that is mainly used in one social network and not in the others. The tree was trained on 75% of the randomly selected accounts, leaving the remaining 25% to calculate its performance (test set). Each account was represented by the concatenation of all its posts.

Results 1

Regarding the model's effectiveness, of the 26 instances of class 1 (FB/IG), 24 were correctly predicted and of the 12 of class 2 (Twitter), 10. Thus, the accuracy of the predictive model is 89.4%, and the area under the ROC curve (Müller & Guido, 2016) is 0.919[9]. We then focused on dimension 50, the one that most significantly separated and classified texts according to social network. We call this dimension interpellative, due to the fact that the main words most used are you, disclaimer, greetings, good day, and back. Although some words related to current issues or slogans such as resignation or back are seen, we found it striking that the main word is you, since this may denote a dialogue of direct interpellation with

8. SVD is an algebraic method to reduce the dimension of a matrix; applied to a matrix of Tf-Idf vectors, as in our case, it results in the detection of words that tend to appear together in a text corpus.

9. This metric is the standard used in machine learning developments for binary classification cases. It produces values between 0 and 1; 1 corresponds to a perfect classification and 0.5, to a totally random one..

another user, presumably another politician. The most important posts within the interpellative dimension are listed below:

- "Usted, sí. https://t.co/zjRiDBvEvE" (FerIglesias) (You, yes).
- "@SolciPlata Usted, en cambio, sí." (FerIglesias) (You, instead, yes).
- "@clarigv1 A usted" (WolffWaldo) (To you).
- "@Damian_Deglauve @WorldGrace saludos!" (gabicerru) (Greetings!).
- "Soy yo la que lo quiere a usted, @caramellocumpa!!! https://t.co/1qZjlEvADi" (fvallejoss) (I am the one who loves you, @caramellocumpa).
- "@shetpwk94 Que tengas un buen día Delfi!!! No salgas de tu casa !!! Cuídate mucho" (alferdez) (Have a nice day Delfi !!!! Don't leave your house !!!! Take care of yourself).
- "Si usted insistía en adjudicar esta compra con sobreprecios, hubiéramos realizado la denuncia al PAMI. Pero entendemos que ha procedido como corresponde." (gracielaocana) (If you insisted on awarding this purchase with overpricing, we would have filed a complaint with PAMI. But we understand that you have proceeded properly).

The list shows that all the posts are direct dialogues, i.e., the person who generates them is questioning another user, not necessarily in a negative tone (which is consistent with what was seen when trying to distinguish the networks through sentiment analysis).

The following figure is a histogram of the users differentiated by social network according to how much they used the words of this interpellative dimension (using the score obtained with SVD).

It can be observed that most of the Twitter accounts are on the right of the X-axis, which means that they have a significant component in this dimension (they frequently used the words associated with it). On the other hand, almost all blue and orange bars (Facebook and Instagram accounts, respectively) are over the left of the X-axis.

It is interesting to note that, in some cases, accounts in different networks belonging to the same person are in opposite places with respect to the 0 value of the X-axis. This tells us that that person had a different way of communicating on Instagram and Twitter concerning the words linked to that dimension.

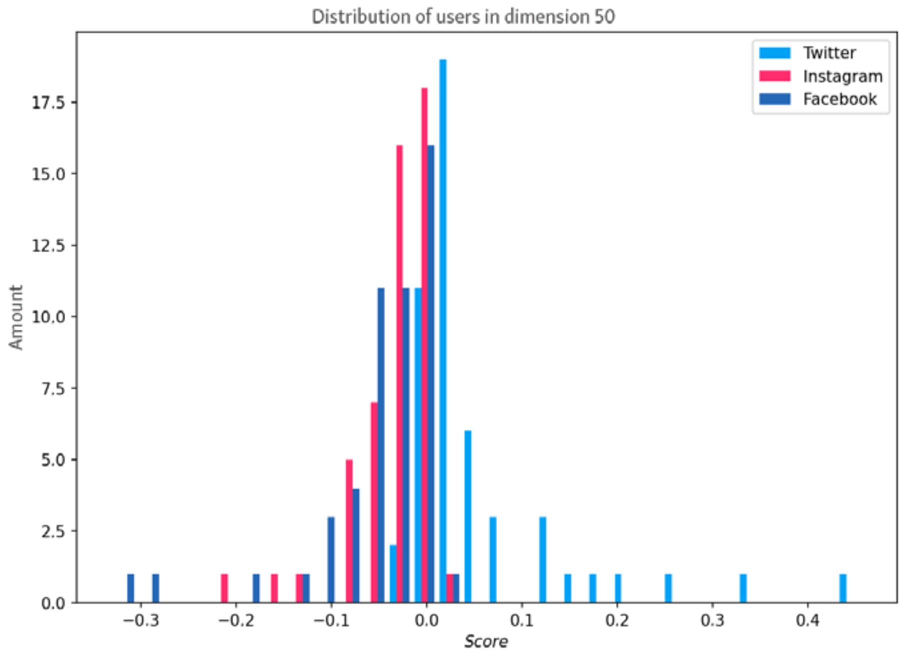


Figure 3. Scores obtained on the interpellative dimension by accounts. The X axis shows the scores and the Y axis shows the number of accounts with this score

Source: Own elaboration.

An outstanding case is that of opposition congressman Fernando Iglesias, an important spokesman against the government, whose Twitter account had a score of 0.25 on the X-axis, while on his Instagram account the score was almost -0.3, occupying opposite ends of the figure.

Method 2

We then measured the relative frequency of each term in each social network. To do so, we plotted the words according to their importance on Twitter and Facebook + Instagram (figure 4), positioning on the X axis the score on Twitter and on the Y axis the score on the other two. We added a red line indicating equivalence, i.e., those terms positioned at or near it have similar frequencies in both cases.

Results 2

It is verified that you and *vos*[10] have a much higher importance on Twitter, with a score higher than 0.4 on that network and lower than 0.2 on Facebook and Instagram. The most important term on Twitter is alferdez, the account of President Alberto Fernández. We also see that terms referring to management, such as works and neighbors (widely used by municipal governments), are very important on Facebook and Instagram and not on Twitter, which would suggest that these platforms are more popular to communicate public management.

10. In Argentina *vos* is used instead of *tú* as second person singular.

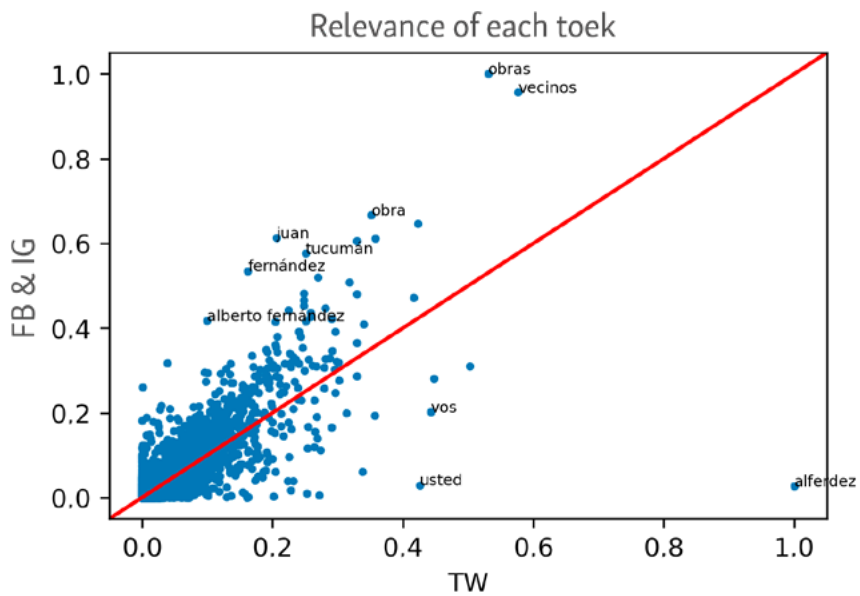


Figure 4. Term frequency by social network

Source: Own elaboration.

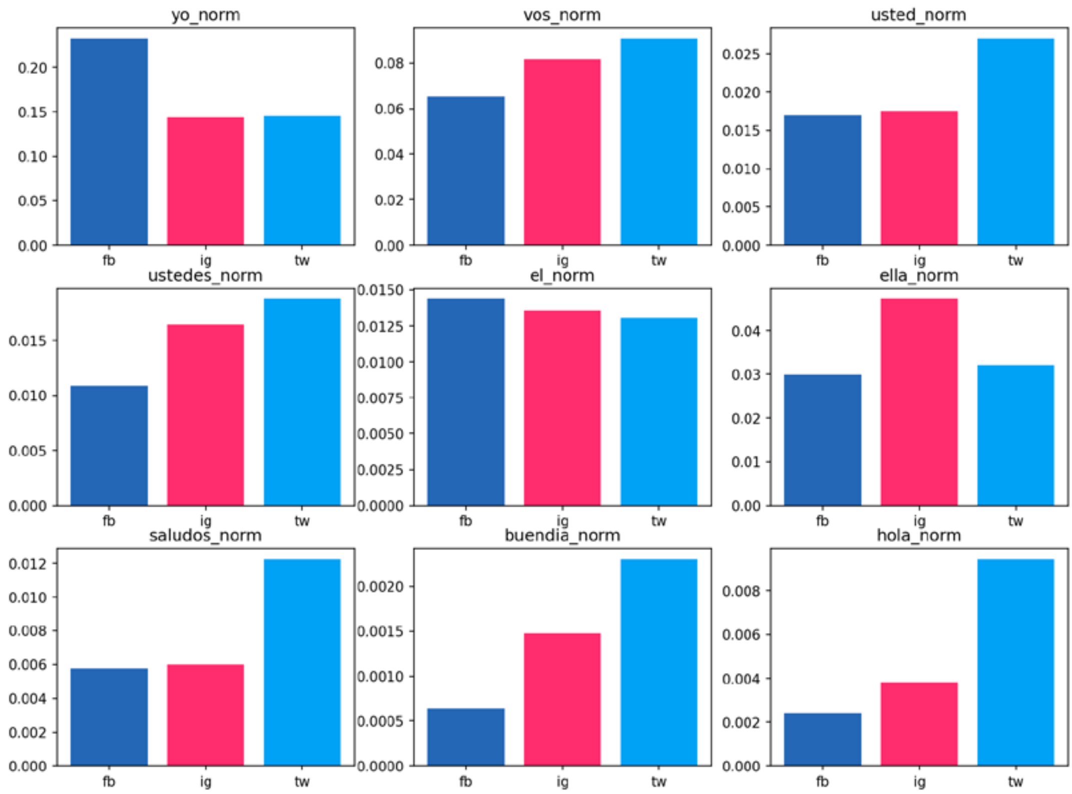
Method 3

To further attempt to test H3 (Twitter is the ground for interpellations), we checked the use of different pronouns across platforms; specifically, the normalized frequency of certain interpellative words (the number of times that word appears divided by the total number of words used in that network).

Results 3

The figure shows that second person pronouns such as *vos* and *usted* (both meaning you, but *usted* implies respect, distance, authority and superiority) are more used on Twitter than on Facebook or Instagram; the same happens with expressions addressed to another interlocutor, such as *buen día, hola* or *saludos* (good morning, hello, greetings). In contrast, the first-person pronoun *yo* (I) appears to a greater extent on Facebook than on Instagram or Twitter. Likewise, third person pronouns are more used on Facebook (he) and Instagram (she) than on Twitter.

These results reinforce the previously confirmed hypothesis 3 that on Twitter there is more dialogue between users, while Facebook and Instagram are less interpellative networks and use first- or third-person personal pronouns more.



Normalized frequency of each term in the three social networks: Facebook (fb), Instagram (ig) and Twitter (tw)

Source: Own elaboration.

CONCLUSIONS

In this paper we set out to elucidate similarities and differences in the communication of 50 Argentine politicians from the ruling party and the opposition, analyzing their posts during 2020 on Facebook, Instagram, and Twitter. We found that politicians talk more about the topics about which they have ownership and in which they presumably feel more comfortable on Facebook and Instagram than on Twitter. The architecture of the latter network enables a field of controversy between like-minded and non-like-minded followers, unlike the first two, where political affinity would predominate. Moreover, the main feature is the general interpellation of other political actors; the controversial nature of this platform is also verified since there is a higher proportion of messages with toxicity, the ones that generate more adhesion. In other words, toxicity is rewarded on Twitter. This is why focusing the gaze and research only on this network reinforces a conflictive image of politics and with little reference to specific political actions, a compelling reason to follow the same emitters on the other platforms. We also observed low conflict shared topics, such as common celebrations but, above all, references to the policies implemented.

In this paper we share novel techniques for comparative text analysis. Although we apply them to publications in social networks, they can be used in a broad spectrum to compare media in relation to toxicity, type of interpellation, and discussion topics. Likewise, they could be focused on other types of public figures and not only in Spanish, since our methods can be adapted to different languages. Therefore, we made all the code available in a public repository to be used in a simple way. Thus, our goal has been to contribute to the development and innovation in the field of communication research and social sciences in general through the use of recent computational techniques of natural language processing. Regarding its contributions to political communication, this paper attempted to show the productivity of the cross-platform approach to better capture the complexity and nuances of current political communication, both in message production and in user interaction and reception. We strongly believe that cross-platform studies will advance both the understanding and the articulated planning of the strategies of politicians and other public figures.

FUNDING

Modelos y herramientas algorítmicas avanzadas para redes y datos masivos-parte 2 (Advanced Algorithmic Models and Tools for Networks and Massive Data-Part 2), UBACyT, Project 20020190100126BA, 2020-2022, Universidad de Buenos Aires, (E. Feuerstein, Universidad de Buenos Aires).

REFERENCES

Ain, Q. T., Ali, M., Riaz, A., Noureen, A., Kamran, M., Hayat, B., & Rehman, A. (2017). Sentiment Analysis Using Deep Learning Techniques: A Review. *International Journal of Advanced Computer Science and Applications*, *8*(6). https://doi.org/10.14569/IJACSA.2017.080657

Angelov, D. (2020). Top2vec: Distributed Representations of Topics. *arXiv preprint arXiv:2008.09470*. https://doi.org/10.48550/arXiv.2008.09470

Ansolabehere, S. & Iyengar, S. (1994). Riding the wave and claiming ownership over issues: The joint effects of advertising and news coverage in campaigns. *Public Opinion Quarterly*, *58*(3), 335-357. https://doi.org/10.1086/269431

Arcila Calderón, C., Van Atteveldt, W., & Trilling, D. (2021). Métodos Computacionales y Big Data en la Investigación en Comunicación: Editorial Dossier (Special Issue Computational Methods and Big Data in Communication Research). *Cuadernos.Info*, (49), I-IV.

Aruguete, N. (2015). *El poder de la agenda: política, medios y público* (The power of the agenda: politics, media, and public). Biblos.

Bast, J. (2021). Managing the Image. The Visual Communication Strategy of European Right-Wing Populist Politicians on Instagram. *Journal of Political Marketing*. https://doi.org/10.1080/15377857.2021.1892901

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer. https://link.springer.com/book/9780387310732

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*, 993-1022.https://jmlr.org/papers/v3/blei03a.html

Bossetta, M. (2018). The Digital Architectures of Social Media: Comparing Political Campaigning on Facebook, Twitter, Instagram, and Snapchat in the 2016 US Election. *Journalism & Mass Communication Quarterly*, *95*(2), 471-496. https://doi.org/10.1177/1077699018763307

Calvo, E. (2015). *Anatomía política de Twitter en Argentina. Tuiteando #Nisman* (Political anatomy of Twitter in Argentina. Tweeting #Nisman). Capital Intelectual.

Calvo, E. & Aruguete, N. (2020). *Fake news, trolls y otros encantos: Cómo funcionan (para bien y para mal) las redes sociales* (Fake news, trolls, and other charms: How do social networks work (for better and for worse)). Siglo XXI Editores.

Cárdenas, A. (2020). Uso de Twitter como estrategia de comunicación política en campañas presidenciales (México, Colombia y Perú 2012-2016). *Razón y Palabra*, 24, (109). https://doi.org/10.26807/rp.v24i109.1716

Chadwick, A., Dennis, J., & Smith, A. P. (2015). Politics in the Age of Hybrid Media: Power, Systems, and Media Logics. In A. Bruns, G. Enli, E. Skogerbø, A. O. Larsson, & C. Christensen, *The Routledge Companion to Social Media and Politics* (pp. 7-22). Routledge.

Enli, G. S. & Skogerbø, E. (2013). Personalized campaigns in party-centred politics. Twitter and Facebook as arenas for political communication. *Information, Communication & Society*, *16*(5), 757-774. https://doi.org/10.1080/1369118X.2013.782330

Figuereo-Benítez, J. C., González-Quiñones, F., & Machin-Mastromatteo, J. D. (2021). Instagram como objeto de estudio en investigaciones recientes. Una revisión de literatura con enfoque en revistas científicas (Instagram as an object of study in recent research. A literature review with a focus on scientific journals). *Ámbitos: Revista internacional de comunicación*, *53*, 9-23. https://doi.org/10.12795/Ambitos.2021.i53.01

Flores-Márquez, D. & González Reyes, R. (2021). En busca de coordenadas metodológicas para estudiar la cultura digital (In search of methodological coordinates to study digital culture). In D. Flores-Márquez & R. González Reyes (Coords), *La imaginación metodológica. Coordenadas, rutas y apuestas para el estudio de la cultura digital* (The methodological imagination. Coordinates, routes, and stakes for the study of digital culture) (pp. 15-23). Tintable.

Fortuna, P., Soler, J., & Wanner, L. (2020, May). Toxic, Hateful, Offensive or Abusive? What Are We Really Classifying? An Empirical Analysis of Hate Speech Datasets. In *Proceedings of the Twelfth Language Resources and Evaluation Conference* (pp. 6786-6794). European Language Resources Association. https://aclanthology.org/2020.lrec-1.838

Garimella, K., Morales, G. D. F., Gionis, A., & Mathioudakis, M. (2018). Quantifying Controversy on Social Media. *ACM Transactions on Social Computing*, *1*(1), 1-27. https://doi.org/10.1145/3140565

González-Bustamante, B. (2015). Evaluando Twitter como indicador de opinión pública: una mirada al arribo de Bachelet a la presidencial chilena 2013 (Assessing Twitter as an indicator of public opinion: a look at Bachelet's arrival to the Chilean presidential election in 2013). *Revista SAAP*, *9*(1), 119-141. https://revista.saap.org.ar/articulos/revista-saap-volumen-9-1.html

Gruzd, A., Lannigan, J., & Quigley, K. (2018). Examining government cross-platform engagement in social media: Instagram vs Twitter and the big lift project. *Government Information Quarterly*, *35*(4), 579-587. https://doi.org/10.1016/j.giq.2018.09.005

Gusfield, J. R. (2014). *La cultura de los problemas públicos: el mito del conductor alcoholizado versus la sociedad inocente* (The culture of public problems: the myth of the drunk driver versus the innocent society). Siglo XXI Editores.

Hasebrink, U. & Hepp, A. (2017). How to research cross-media practices? Investigating media repertoires and media ensembles. *Convergence: The International Journal of Research into New Media Technologies*, *23*(4), 362-377. https://doi.org/10.1177/1354856517700384

Hodges, J. L. (1958). The significance probability of the smirnov two-sample test. *Arkiv för Matematik*, *3*(5), 469-486. https://doi.org/10.1007/bf02589501

Hua, Y., Ristenpart, T., & Naaman, M. (2020). Towards Measuring Adversarial Twitter Interactions against Candidates in the US Midterm Elections. *Proceedings of the International AAAI Conference on Web and Social Media*, *14*(1), 272-282. https://doi.org/10.1609/icwsm.v14i1.7298

Iyengar, S. (1990). The accessibility bias in politics: Television news and public opinion. *International Journal of Public Opinion Research*, *2*(1), 1-15.

Jaidka, K., Guntuku, S., & Ungar, L. (2018, June). Facebook versus Twitter: Differences in self-disclosure and trait prediction. *Proceedings of the International AAAI Conference on Web and Social Media, 12*(1). https://doi.org/10.1609/icwsm.v12i1.15026

Kaplan, N., Park, D. K., & Ridout, T. N. (2006). Dialogue in American Political Campaigns? An Examination of Issue Convergence in Candidate Television Advertising. *American Journal of Political Science*, *50*(3), 724-736. https://doi.org/10.1111/j.1540-5907.2006.00212.x

Karlsen, R. & Enjolras, B. (2016). Styles of Social Media Campaigning and Influence in a Hybrid Political Communication System: Linking Candidate Survey Data with Twitter Data. *The International Journal of Press/Politics*, *21*(3), 338-357. https://doi.org/10.1177/1940161216645335

Kelley, S. & Mirer, T. W. (1974). The Simple Act of Voting. *American Political Science Review*, *68*(2), 572-591. https://doi.org/10.2307/1959506

Larsson, A. O. (2015). Green light for interaction: Party use of social media during the 2014 Swedish election year. *First Monday*, *20*(1). https://doi.org/10.5210/fm.v20i12.5966

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, *1*(4), 541-551. https://doi.org/10.1162/neco.1989.1.4.541

López-López, P. C. & Vásquez-González, J. (2018). Agenda temática y *Twitter*: elecciones presidenciales en América Latina durante el período 2015-2017 (Thematic agenda and *Twitter*: Presidential elections in Latin America during the 2015-2017 period). *Profesional De La Información*, *27*(6), 1204-1214. https://doi.org/10.3145/epi.2018.nov.04

Manning, C. & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press.

Martin, J. R. & White, P. R. R. (2005). *The Language of Evaluation. Appraisal in English*. Palgrave Macmillan.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology, 27,* 415-444. https://doi.org/10.1146/annurev.soc.27.1.415

Müller, A. C. & Guido, S. (2016). *Introduction to Machine Learning with Python. A Guide for Data Scientists*. O'Reilly Media, Inc.

Nielsen, R. K. & Schrøder, K. C. (2014). The Relative Importance of Social Media for Accessing, Finding, and Engaging with News. An eight-country cross-media comparison. *Digital Journalism, 2*(4), 472-489. https://doi.org/10.1080/21670811.2013.872420

Owen, D. (2017). New Media and Political Campaigns. In K. Kensi & K. Hall Jamieson (Eds.), *The Oxford Handbook of Political Communication* (pp. 829-836). https://doi.org/10.1093/oxfordhb/9780199793471.013.016

Paulino, F. O. & Waisbord, S. (2021). Las narrativas del populismo reaccionario: Bolsonaro en Twitter durante la pandemia (The narratives of reactionary populism: Bolsonaro on Twitter during the pandemic). *Mediapolis–Revista de Comunicação, Jornalismo e Espaço Público*, (12), 33-48. https://doi.org/10.14195/2183-6019_12_2

Petrocik, J. R. (1996). Issue Ownership in Presidential Elections, with a 1980 Case Study. *American Journal of Political Science*, *40*(3), 825-850. https://doi.org/10.2307/2111797

Prada Espinel, Ó. A. & Romero Rodríguez, L. M. (2018). Polarización y demonización en la campaña presidencial de Colombia de 2018: análisis del comportamiento comunicacional en el Twitter de Gustavo Petro e Iván Duque (Polarization and Demonization in the 2018 Presidential Campaign of Colombia: Analysis of Twitter's Communication Behavior by Gustavo Petro and Iván Duque). *Revista Humanidades*, *9*(1). https://doi.org/10.15517/h.v9i1.35343

Quevedo, L. A., & Ramírez, I. (2022). Polarizados ¿Por qué preferimos la grieta?(Aunque digamos lo contrario). *Revista SAAP*, *16*(1), 220-222.

Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the Space of Topic Coherence Measures. In X. Cheng & H. Li (Chairs), *WSDM'15: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining* (pp. 399-408). https://doi.org/10.1145/2684822.2685324

Rogers, R. (2017). Digital Methods for Cross-Platform Analysis. In J. Burgess, A. Marwick, & T. Poell, *The SAGE Handbook of Social Media* (pp. 91-110). SAGE. https://doi.org/10.4135/9781473984066

Salgado Andrade, E. (2013). *Twitter* en la campaña electoral de 2012 (Twitter in the Electora Campaign of 2012). *Desacatos*, (42), 217-232. https://doi.org/10.29340/42.78

Scheufele, D. A. (2000). Agenda-Setting, Priming, and Framing Revisited: Another Look at Cognitive Effects of Political Communication. *Mass Communication & Society*, *3*(2-3), 297-316. https://doi.org/10.4324/9781315679402-5

Scheufele, D. A. & Iyengar, S. (2012). The State of Framing Research: A Call for New Directions. In K. Kenski & K. Hall Jamieson (Eds.), *The Oxford Handbook of Political Communication Theories* (pp. 1-26). https://doi.org/10.1093/oxfordhb/9780199793471.013.47

Spierings, N. & Jacobs, K. (2019). Political parties and social media campaigning. *Acta Politica*, *54*(1), 145-173. https://doi.org/10.1057/s41269-018-0079-z

Stier, S., Bleier, A., Lietz, H., & Strohmaier, M. (2018). Election Campaigning on Social Media: Politicians, Audiences, and the Mediation of Political Communication on Facebook and Twitter. *Political Communication*, *35*(1), 50-74. https://doi.org/10.1080/10584609.2017.1334728

Thelwall, M. (2017). The Heart and Soul of the Web? Sentiment Strength Detection in the Social Web with SentiStrength. In J. Holyst (Ed.), *Cyberemotions. Understanding Complex Systems* (pp. 119-134). Springer. https://doi.org/10.1007/978-3-319-43639-5_7

Thorhauge, A. M. & Lomborg, S. (2016). Cross-media communication in context: A mixed-methods approach. *MedieKultur*: *Journal of Media and Communication Research*, *32*(60), 16. https://doi.org/10.7146/mediekultur.v32i60.22090

Wulczyn, E., Thain, N., & Dixon, L. (2017). Ex Machina: Personal Attacks Seen at Scale. In R. Barret & R. Cummings (Chairs), *WWW' 17: Proceedings of the 26th International Conference on World Wide Web* (pp. 1391–1399). International World Wide Web Conferences Steering Committee. https://doi.org/10.1145/3038912.3052591

Zwillinger, D. & Kokoska, S (2000). *CRC Standard Probability and Statistics Tables and Formulae*. Champan & Hall.

APPENDIX

Selected political referents by sector and role.

Ruling party

Name	Role
Alberto Fernández	President
Cristina Fernández de Kirchner	Vice-president
Santiago Cafiero	Chief of staff
Wado de Pedro	National minister
Gabriel Katopodis	National minister
Victoria Donda	INADI director (National Institute against Discrimination)
Axel Kicillof	Governor
Gildo Insfran	Governor
Gustavo Bordet	Governor
Juan Manzur	Governor
Omar Perotti	Governor
Sergio Uñac	Governor
Anabel F. Sagasti	Senator
Oscar Parrilli	Senator
Facundo Moyano	Congressman
Fernanda Vallejos	Congresswoman
Gabriela Cerrutti	Congresswoman
Itai Hagman	Congressman
Jorge Antonio Romero	Congressman
José I de Mendiguren	Congressman
Jose L. Gioja	Congressman
Leonardo Grosso	Congressman
Lucia Corpacci	Congresswoman
Pablo Carro	Congressman
Pablo Yedlin	Congressman

Opposition

Name	Role
Gerardo Morales	Governor
Rodolfo Suarez	Governor
Horacio R. Larreta	Government's head
Diego Santilli	Deputy head of government
Gustavo Posse	Mayor
Jorge Macri	Mayor
Néstor Grindetti	Mayor
Mauricio Macri	Former president
Maria E. Vidal	Former governor
Alfredo Cornejo	Former governor
Alfredo De Angeli	Senator
Humberto Schiavoni	Senator
Luis Naidenoff	Senator
Martin Lousteau	Senator
Alfredo Schiavoni	Congressman
Brenda Austin	Congresswoman
Cristian Ritondo	Congressman
Elisa Carrió	Congresswoman
Fernando Iglesias	Congressman
Graciela Ocaña	Congresswoman
Luis A. Juez	Congressman
Mario R. Negri	Congressman
Maximiliano Ferraro	Congressman
Waldo Wolff	Congressman
Patricia Bullrich	PRO president

ABOUTH THE AUTHORS

FEDERICO ALBANESE, holds a degree in Physics from the Universidad de Buenos Aires, Ph.D. student in Computer Science at the same university, and member of the Institute of Computer Science (CONICET - UBA). He teaches in the master's program in Data Exploitation at the Faculty of Exact and Natural Sciences, UBA. His research is currently focused on the study of machine learning models in graphs and their applications to social networks.

https://orcid.org/0000-0001-7140-2910

ESTEBAN FEUERSTEIN, holds a degree in Computer Science from the Escuela Superior Latinoamericana de Informática (ESLAI) and a Ph.D. in Computer Science from the University of Rome La Sapienza. He has more than 25 years of experience as a researcher and consultant in public and private organizations. He is an associate professor in the Department of Computer Science at the Faculty of Exact and Natural Sciences (UBA). His area of expertise is in the field of algorithms and data structures, and the search and organization of information and big data.

https://orcid.org/0000-0003-2985-810X

GABRIEL KESSLER, Ph.D. in Sociology from EHESS-Paris, CONICET researcher and professor at the Universidad Nacional de La Plata and the Universidad Nacional de San Martín. His research areas are inequality, violence, and political polarization. He is the author of numerous books and academic articles. His latest books are (with Gabriela Benza) Uneven Trajectories. Latin America Societies in the XXI Century (Cambridge University Press, 2020) and La ¿Nueva? estructura social de América Latina (Latin America's New (?) Social Structure) (Siglo XXI, 2021).

https://orcid.org/0000-0003-4204-8961

JUAN MANUEL ORTIZ DE ZÁRATE, holds a degree in Computer Science from the Universidad de Buenos Aires, Ph.D. student in Computer Science at the same university, and member of the Institute of Computer Science (CONICET - UBA). With experience in both the academic and private sectors, his research is currently focused on the study of social networks through natural language processing and graph modeling.

https://orcid.org/0000-0002-0291-1997