



OPEN ACCESS

EDITED BY

Nicolas Tsapatoulis,
Cyprus University of Technology, Cyprus

REVIEWED BY

Pilar Rodriguez Martinez,
University of Almeria, Spain
Alejandra Tirado Garcia,
University of Jaume I, Spain

*CORRESPONDENCE

Javier J. Amores
✉ jamores@uc.cl

†PRESENT ADDRESS

Javier J. Amores,
Department of Journalism, School of
Communications, Pontifical Catholic
University of Chile, Santiago, Chile

RECEIVED 12 May 2025

ACCEPTED 24 June 2025

PUBLISHED 23 July 2025

CITATION

Amores JJ, González-Baquero W and
Oller-Alonso M (2025) Toxicity levels in
Spanish political communication on
Twitter/X: a comparative analysis of major
political parties, 2015–2023.
Front. Polit. Sci. 7:1627474.
doi: 10.3389/fpos.2025.1627474

COPYRIGHT

© 2025 Amores, González-Baquero and
Oller-Alonso. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Toxicity levels in Spanish political communication on Twitter/X: a comparative analysis of major political parties, 2015–2023

Javier J. Amores^{1*†}, William González-Baquero² and
Martín Oller-Alonso²

¹Department of Social and Political Sciences, Faculty of Political, Economic and Social Sciences, University of Milan, Milan, Italy, ²Department of Sociology and Communication, Faculty of Social Sciences, University of Salamanca, Salamanca, Spain

Introduction: Social media, especially Twitter (now X), have long since become integral to the communication strategies of political parties, providing a direct and agile medium for interaction with the public and dissemination of messages. But this digital omnipresence also poses worrying challenges, such as the polarisation of political discourse. Several studies have explored the communication strategies of political parties on social media, as well as the toxic language and hate present in such communication. However, most of the work has looked at specific parties and in electoral contexts.

Methods: The present study focuses on examining in depth the levels of toxicity in political discourse on Twitter, particularly on the accounts of the 10 major Spanish political parties, from 2015 to 2023. Computational methods and the Google Perspective API have been used to identify the levels of toxicity, severe toxicity, insult, profanity, threat, and identity attack present in the contents published on these accounts during that period ($N = 265,122$).

Results: Among the main findings, a generalized temporary increase in the presence of all toxicity indicators since 2015 is highlighted. Comparatively, a significant difference is perceived between most of the parties and Vox, the party that scores highest in virtually all indicators of toxicity. Furthermore, higher levels of toxicity appear to generate greater engagement in terms of interaction.

Discussion: These findings confirm trends noted earlier at the international level, such as the leading role of the populist far-right in the propagation of toxic discourses and the generalised rise of toxicity and polarization in political debate.

KEYWORDS

toxicity, toxic language, polarization, political communication, political parties, social media, Twitter, X

1 Introduction

Social media platforms, particularly Twitter (now known as X), have profoundly reshaped political communication, providing politicians and political parties with direct and instantaneous channels to disseminate messages, interact with the public, and mobilize support. While these digital environments offer substantial advantages in terms of immediacy and reach, they also present significant challenges, notably the escalation of toxic and polarized political discourse. The proliferation of aggressive and harmful communication styles online has been recognized as a pressing issue due to its potential to exacerbate societal divisions,

discourage constructive debate, and undermine democratic processes (Froio and Ganesh, 2018; Guerrero-Solé and Philippe, 2020).

Previous research on political toxicity in social media has consistently highlighted its association with increased polarization, political hostility, and the deterioration of public discourse. Specifically, recent studies in Spain have identified notable trends towards aggressive and confrontational rhetoric within digital political communication, particularly by populist and far-right political actors. For example, Vico and Rey (2020) analysed the Instagram account of the Spanish far-right party Vox, focusing on the toxicity identified in the contents published by that party. Along the same lines, Guerrero-Solé and Philippe (2020) analysed the messages of Spanish parliamentarians on Twitter during the COVID-19 pandemic, concluding that the party that propagated more toxic discourses was Vox, followed by PP. For their part, Galindo-Domínguez et al. (2023) explored political discourse on social platforms regarding immigrants and highlighted the amplification of racist discourse on these platforms, especially led also by Vox (Hutchins and Halikiopoulou, 2019). Similarly, studies by Arcila-Calderón et al. (2020) or Vicente et al. (2021) have demonstrated how toxic language on Twitter effectively mobilizes voters by leveraging negative emotions such as fear and anger, frequently targeting minorities, immigrants, or political opponents.

In this context, the current study addresses a crucial gap by conducting an extensive computational analysis of toxic political discourse across official Twitter accounts of the major Spanish political parties from 2015 to 2023. This research is timely and particularly relevant given the growing concern about the impact of online toxicity on democratic health, public debate quality, and societal cohesion in Spain, a country marked by increased political fragmentation, polarization, and repeated electoral cycles during the past decade. By employing advanced computational methods and Google's Perspective API, the present study aims to identify patterns and temporal trends in the use of toxic language, including insults, threats, profanity, severe toxicity, and identity attacks.

Understanding these dynamics is especially pertinent in the Spanish political context, characterized by the recent emergence and consolidation of populist movements and deepening ideological divides. The analysis presented here seeks not only to quantify and compare the levels of toxicity present in the communication of different Spanish parties, but also to offer insights into the broader implications for political debate, voter engagement, and democratic resilience. Ultimately, this research contributes valuable empirical evidence to inform policy-makers, political actors, and social media, fostering strategies aimed at mitigating online toxicity and promoting healthier democratic discourse.

2 Political communication and social media

As mentioned, the emergence of social media has substantially transformed contemporary political communication, marking a paradigm shift in interactions between political actors and citizens. These platforms not only facilitate the dissemination and virality of political messages but also reshape agenda-setting dynamics, citizen engagement, and public debate, enabling direct interaction, social mobilization, and instant public discourse (Škorić et al., 2016). A

significant impact of social media is the elimination of traditional media intermediaries, such as print, radio, and television, previously responsible for filtering and reinterpreting political messages, thereby influencing public perception of political candidates and parties (Campos-Domínguez, 2017; Ramos Antón, 2021). This direct communication has strategically benefited political actors such as Podemos and Vox in Spain, allowing them to broadcast polarized narratives without mediation, effectively mobilizing their supporters (Ramos Antón, 2021). Additionally, social media has revolutionized political participation by enabling two-way interactions between politicians and the public, contrasting with the predominantly one-directional nature of traditional media (Said-Hung et al., 2017; Lava Santos, 2021). This shift has democratized political information access and increased involvement among traditionally disengaged demographics, notably younger populations (Castro Martínez and Díaz Morilla, 2021). Social media platforms also play a crucial role in political agenda-setting. Through strategic selection and repetition of specific topics, parties guide public attention toward priority issues (Blanco-Herrero et al., 2020). For instance, Vox has effectively amplified discussions around immigration and national identity, positioning these issues centrally in the Spanish political debate (Esteban, 2022; Fernández Romero et al., 2021). This strategic use of social media demonstrates how parties can influence public opinion and mobilize supporters appealing to shared identities.

Given these transformations, extensive research has concentrated on analysing candidate interactions on platforms like Twitter, particularly in highly polarized political contexts. Bustos Díaz and Ruiz del Olmo (2021) emphasize the importance of studying candidates' communication strategies during electoral campaigns, as social media offers an optimal environment for efficient campaign execution and allows researchers to assess discursive and persuasive tactics (Maarek, 2015; Otalora, 2017). Amores et al. (2022) further underline that digital political discourse is crucial due to its rapid dissemination and mass reach, shaping interactive networks that alter the relationship between audiences and political figures. Furthermore, social media has reshaped political mobilization, enabling political actors to efficiently organize and disseminate information about rallies and marches, thereby significantly increasing event visibility and participation (Ruano Ibarra et al., 2018; Lagares Díez et al., 2021). Consequently, political communication has acquired greater immediacy and virality, allowing political messages to reach vast audiences rapidly (Castro Martínez and Díaz Morilla, 2021; Congosto, 2015).

Nevertheless, despite democratizing information access, social media platforms also present notable challenges, such as increasing public debate polarization, simplifying political discourse, and facilitating misinformation proliferation (Mazzoleni, 2018). Scholars such as Mazzoleni and Splendore (2020) indicate that digital platforms have contributed to the spread of unverified content, resulting in a more polarized yet less informed public, exacerbating existing media challenges related to trust and accountability. Another critical issue is the creation of echo chambers, where users predominantly interact with like-minded individuals, reinforcing pre-existing beliefs and limiting exposure to diverse viewpoints (Ramos Antón, 2021). This phenomenon contributes significantly to political polarization by fostering ideological communities resistant to external perspectives (de Borja Navarro and Yeh, 2022). Authors such as Oller-Alonso et al. (2025) specifically notes how social platforms like Facebook and Twitter amplify negative sentiments, toxic narratives and polarized,

discriminatory opinions, further entrenching public division on sensitive issues such as immigration (e.g., Amores and Arcila-Calderón, 2025; Arcila-Calderón et al., 2020; Blanco-Herrero et al., 2024; Latorre and Amores, 2021; Saridou et al., 2023). Lastly, social media's emphasis on populist and emotional strategies has led to simplified, impactful political narratives. Far-right parties like Vox, for instance, successfully leverage direct and emotional language to mobilize supporters through constructing common enemies such as political elites, separatists, or immigrants (Álvarez-Benavides and Jiménez Aguilar, 2021). This approach effectively captures audience attention within digital environments characterized by immediacy, sensationalism, and visual impact amidst information overload (Vicente et al., 2021).

3 Political parties in the Spanish context

The political landscape in Spain has undergone significant changes over the past decade, notably characterized by the decline of traditional bipartisanship between the Popular Party (PP) and the Spanish Socialist Workers' Party (PSOE) and increasing polarization. The 2008 economic crisis and corruption scandals affecting both major parties eroded public trust, facilitating the emergence of new political actors such as Podemos, representing a more radical left (Otalora, 2017), and Ciudadanos, a renewed center-right and anti-independence alternative (Rodríguez-Teruel et al., 2016), now in deep decline. Although these parties initially experienced rapid electoral growth, their popularity has since waned. By 2023, Podemos had integrated into the broader left-wing platform Sumar, while Ciudadanos dramatically lost electoral support, eventually forfeiting representation in the national parliament and the European Parliament, to the point of not running in the 2023 general elections (for this reason it is not included in this analysis). Concurrently, the rise of the far-right party Vox introduced a nationalist, anti-immigrant, and anti-feminist discourse, significantly heightening public polarization (Turnbull-Dugarte, 2019).

This fragmented scenario, compounded by increasing digital platform usage and successive economic, migration, and health crises, further fuelled public distrust in governmental and institutional stability, particularly after 2015, resulting in repeated elections and challenging multi-party coalitions. Pedro Sánchez has maintained power since 2018 through coalitions with Unidas Podemos—a progressive coalition formed by Podemos, Izquierda Unida, and smaller leftist groups since 2016—and parliamentary support from regional parties such as the Basque Nationalist Party (PNV) and Esquerra Republicana de Catalunya (ERC) (Carozzi et al., 2022). Territorial tensions, particularly regarding Catalan independence, have prominently influenced national politics and relationships between the central government and regional administrations. Parties such as ERC and Junts per Catalunya (Junts) have played crucial roles following the unsuccessful 2017 independence referendum. These tensions prompted traditional parties to clarify their positions, with PP adopting a rigid anti-independence stance and PSOE attempting dialogue-oriented approaches. Meanwhile, Vox intensified its explicitly nationalist and conservative discourse, whereas Podemos and allies promoted more conciliatory progressive policies (Turnbull-Dugarte, 2019).

Within this complex environment, the present study aims to determine which major Spanish political parties have used the most toxic and violent language on Twitter since 2015, potentially reflecting sensationalist, hostile and polarizing communication strategies. Similarly, we aim to explore whether there is an evolution in the use of toxic language in general lines. Specifically, the official Twitter accounts of the 10 main political parties contesting the early general elections of July 23, 2023, were analysed. These selected political parties are briefly described below.

3.1 PSOE (@PSOE)

Founded in 1879, the Spanish Socialist Workers' Party (Partido Socialista Obrero Español) advocates social democratic policies focused on equality and mixed economy approaches. Despite experiencing fragmentation and corruption issues, it remains electorally strong, historically promoting progressive social welfare, gender equality, public healthcare, and labour rights. Under Pedro Sánchez's leadership, PSOE has pursued progressive coalitions, implementing significant social and labour reforms, often adopting moderate, consensus-oriented language (Carozzi et al., 2022). The official Twitter account was created in June 2009.

3.2 Partido Popular (@ppopular)

Evolving from conservative post-Francoist origins, the Spanish Popular Party (PP) represents Spain's center-right, promoting conservative, liberal, and pro-European policies, advocating market economy, public spending reductions, traditional social values, and national unity. Although it has historically played a central role in Spanish politics, corruption cases have also affected its credibility (Royo, 2014). Its communication typically emphasizes moderation, national unity, economic stability, and security, avoiding excessively aggressive rhetoric (Royo, 2014). Its Twitter account was created in February 2009.

3.3 Vox (vox_es)

Established in 2013, Vox rapidly gained support through nationalist, anti-immigrant, and anti-feminist rhetoric. Vox promotes a conservative and centralist Spanish vision, opposing separatist movements and progressive gender policies. Known for polarizing discourse, Vox strategically leverages explicit hostility, appealing to fear and indignation, especially regarding immigration (Amores et al., 2024; Turnbull-Dugarte, 2019). Its Twitter account dates from November 2013.

3.4 Podemos (@PODEMOS)

Founded in 2014, emerging from the 15-M movement, Podemos represents the popular radical left, advocating redistributive economics, labour rights, and social justice (Jerez et al., 2015). Initially combative, criticizing corruption and elites with emotional and mobilizing rhetoric, its tone moderated following governmental coalition participation with PSOE from 2019 and subsequent integration into Sumar (Barba and Blanco, 2011; Rodríguez-Teruel et al., 2016). Its Twitter account was established in January 2014.

3.5 Más País (@MasPais_Es)

Formed in 2019 as a pragmatic, moderate left-wing offshoot of Podemos, Más País emphasizes ecological transition and feminism, favouring broad alliances and sustainable solutions. It currently participates in the Sumar coalition. Its discursive style tends to be more restrained. Its Twitter account was launched in September 2019.

3.6 Sumar (@sumar)

Initiated in 2022 by Yolanda Díaz, Minister of Labor and Second Deputy Prime Minister under Sánchez, Sumar seeks to unify progressive leftist forces. Its agenda prioritizes social justice, labour rights, feminism, and ecological transition, employing a conciliatory discourse distinct from traditional combative leftist rhetoric. Its Twitter account was established in June 2022.

3.7 EAJ-PNV (@eajpnv)

Advocating enhanced Basque regional autonomy, PNV is described as social-democratic and Christian-democratic, incorporating both centre-right and centre-left factions. Historically significant in Spanish politics, PNV promotes progressive social policies coupled with conservative, economically liberal regional governance, strategically providing parliamentary support for regional concessions. Its style is more conciliatory than that of other regionalist parties. Its Twitter account was created in June 2009.

3.8 EH Bildu (@ehbildu)

A leftist Basque independence coalition founded in 2012, advocating more radical sovereignty and socialist redistribution than PNV. Despite controversial historical associations with Basque conflict, EH Bildu has regional and national relevance, using a confrontational rhetoric (Lecours, 2021). Its discourse tends to be more hostile and combative, especially in the national context. Its Twitter account was created in June 2012.

3.9 Esquerra Republicana (@Esquerra_ERC)

Key actor in Catalan independence movement, historically advocating socialism alongside independence, progressive healthcare, education, and social rights policies. ERC recently pursued pragmatic governmental dialogue, maintaining relatively moderate yet satirical discourse. Its Twitter account was established in June 2009.

3.10 Junts per Catalunya (@JuntsXCat)

Catalan center-right independence party, evolving from a broader coalition and gaining independent structure in 2020 under Carles Puigdemont's leadership. Junts promotes Catalan nationalism with conservative, economically liberal policies. It often employs

more confrontational rhetoric towards central governance compared to ERC. Its Twitter account was launched in June 2016.

4 Toxic language and political communication

The proliferation of toxic discourse within political communication, particularly on social media, has become increasingly concerning for contemporary democracies (Wulczyn et al., 2017). According to Guerrero-Solé and Philippe (2020) and Wulczyn et al. (2017), toxic language includes insults, offences, derogatory terms, and unacceptable expressions used to attack opponents or demean others publicly. Such messages may either discourage civic participation or conversely foster addictive behaviours, triggering similar conduct from others, thus creating a toxic communication spiral (Guerrero-Solé and Philippe, 2020). In Spain, this phenomenon is especially pronounced among far-right political groups, which frequently employ aggressive rhetoric to mobilise supporters, attack adversaries, and reinforce populist narratives (de Borja Navarro and Yeh, 2022; Vicente et al., 2021; Ferreira, 2019; Lava Santos, 2021; Rodríguez de Mora, 2023). The use of hostile language on platforms such as Twitter raises critical concerns about its impact on social polarisation, public debate quality, and democratic health (Froio and Ganesh, 2018; Vicente et al., 2021).

Social media platforms inherently facilitate the formation of echo chambers, spaces where users predominantly engage with individuals sharing similar perspectives, limiting exposure to divergent opinions and amplifying extreme views. Consequently, political polarisation intensifies as users reinforce their existing beliefs without external critique (Ramos Antón, 2021). In such contexts, toxic discourse strategically serves to discredit opponents, demonise rivals, and consolidate supporters' convictions (Lava Santos, 2021). Furthermore, social media's inherent simplification encourages sensationalist and emotionally charged messages, which quickly become viral, enhancing their appeal (Froio and Ganesh, 2018; Congosto, 2015).

Platforms such as Twitter, with strict character limitations, promote simplified political discourse, favouring emotionally charged over rational messages (Pérez-Curiel, 2020). Toxic, aggressive, and polarising messages therefore gain wider appeal, especially when addressing sensitive topics or leveraging negative emotions like fear or indignation (Froio and Ganesh, 2018). In Spain, messages promoting hatred or fear towards stigmatised social groups or ridiculing political elites resonate strongly among followers of parties like Vox (de Borja Navarro and Yeh, 2022; Fernández Romero et al., 2021). Consequently, toxic rhetoric effectively supports populist strategies, positioning 'the people' against perceived corrupt elites disconnected from public needs (de Borja Navarro and Yeh, 2022). Aggressive narratives thus mobilise voters against political opponents framed as enemies, employing personal attacks and exaggerations to delegitimise rivals and foster a victimisation narrative that reinforces loyalty (Pérez-Curiel, 2020). Vox frequently employs such tactics, associating opponents with perceived threats like a exacerbated globalism or progressivism, frequently using fake contents, thereby fostering distrust and confusion amidst rampant disinformation and limited content verification (Arcila-Calderón et al., 2020; Cano-Orón et al., 2021; Lava Santos, 2021; Castro Martínez and Díaz Morilla, 2021).

A critical risk associated with toxic language on social media is its potential escalation into personal attacks, harassment, and targeted hate speech. This digital aggression is particularly alarming when originating from political leaders or parliamentary parties, normalising verbal violence and intolerance (Guerrero-Solé and Philippe, 2020). Messages dehumanising opponents or vulnerable minorities not only deepen polarisation but may incite real-world discriminatory or violent behaviours (Arcila-Calderón et al., 2020). The reciprocal relationship between toxic discourse and political polarisation perpetuates a damaging cycle, eroding public debate quality and democratic foundations (Guerrero-Solé and Philippe, 2020).

Undoubtedly, aggressive rhetoric, insults, and personal attacks exacerbate ideological divisions, fostering a hostile atmosphere detrimental to constructive dialogue (Castillo Jara et al., 2019). Depicting political adversaries as existential threats rather than mere opponents amplifies tensions, diminishing opportunities for consensus or compromise (Fernández Romero et al., 2021). Parties such as Vox frequently utilise bellicose metaphors and consistent aggression to reinforce this confrontational dynamic. Additionally, social media's segmentation capabilities enhance homogeneous interactions, further reinforcing polarisation and toxic exchanges (de Borja Navarro and Yeh, 2022).

As revised, social media characteristics such as anonymity, immediacy, and global reach facilitate rapid dissemination of toxic messages, enabling a single offensive tweet to achieve disproportionate public impact (Arcila-Calderón et al., 2020). Such dynamics amplify toxicity and accelerate polarisation, rewarding aggressive behaviour with increased visibility (Åkerlund, 2020). Aware of these dynamics, some political actors deliberately use toxic language to evoke emotions like fear or anger, mobilising supporters, and discouraging opposition participation. The rise of populist and radical ideologies in Spain may thus contribute to increased online political toxicity, strategically polarising public opinion and capitalising on social discontent (de Borja Navarro and Yeh, 2022). In addition, such speeches could promote greater user interaction and greater engagement in the followers of such parties.

To address these issues, based on the reviewed literature, this study seeks to answer the following research questions:

RQ1. Which major Spanish political parties exhibit the highest levels of toxicity in their Twitter content?

RQ2. Have toxicity levels increased in the Twitter content of major Spanish political parties since 2015?

RQ3. In which year have the highest toxicity levels been recorded in the Twitter content of major Spanish political parties since 2015?

RQ4. Is there a correlation between the toxicity levels of Twitter content posted by major Spanish political parties and the interaction they generate?

5 Method

5.1 Sample

To examine toxicity levels within the Twitter content posted by Spain's major political parties, we compiled a comprehensive dataset covering the period from 1 January 2015 to 31 May 2023. Tweets were

harvested via Twitter's v2 API (prior to the closure of free research access in 2023) using the Python library Tweepy, in six-month batches. A purpose-built Jupyter notebook retrieved all tweets from each party's official account together with associated metadata. Collected variables included publication date and time, public profile metrics, and interaction statistics for every tweet (retweets, replies, likes, and quotes). Retweets were excluded to ensure the analysis focused exclusively on original content.

The initial database comprised 265,122 tweets. The data were pre-processed by normalising dates and standardising metadata to guarantee analytical consistency and quality. Because the Perspective API offers limited reliability for Catalan, Basque, or Galician, posts were subsequently filtered by language so that only Spanish-language tweets were retained. This step was especially pertinent for accounts such as JuntsXCat and ERC, which produced a substantial volume of Catalan content. After cleaning, the final analytical sample contained 203,554 tweets. Table 1 reports the number of tweets gathered from each party before and after filtering.

5.2 Toxicity analysis

Toxicity within the collected tweets was assessed using Perspective API, a free tool resulting from a collaborative research effort between Jigsaw and Google's Anti-Abuse Technology team, in an initiative called Conversation-AI. This tool applies machine-learning techniques to estimate the probability that a text exhibits different toxic attributes. It is based on a series of machine learning models (initially trained as multilingual BERT-based models and then distilled into single-language Convolutional Neural Networks for each supported language), which were trained on millions of comments from diverse online sources, including Wikipedia and The New York Times, across multiple languages. The training process involves human annotators who evaluates and labels comments for specific toxic attributes, with these annotations forming the basis for the models' scoring. Google continually updates these models and actively works to mitigate biases in the training data.

For this study, we used a dedicated Jupyter notebook leveraging the "google-api-python-client" to access the API. Through the "comments: analyze" endpoint each tweet was submitted to obtain probability scores ($0 \leq p \leq 1$) for the six standard attributes defined

TABLE 1 Tweets collected from each party account.

Party	Original sample	Final sample
PSOE	67,975	67,112
PP	50,257	49,788
Vox	16,087	15,856
Podemos	64,778	57,192
Más País	1,991	1,953
Sumar	542	468
PNV	11,883	2,133
EH Bildu	18,789	8,271
ERC	22,077	673
Junts	10,743	108

Own elaboration. Final counts represent the Spanish-language tweets after the filtering.

by Jigsaw. All requests forced the parameter “languages”: [“es”], were sent in batches of 100 to comply with rate limits, and were executed on the Castilla-y León Supercomputing Centre (SCAYLE) to handle the computational load. Specifically, Perspective assigns a score from 0 to 1 (higher values indicate greater likelihood) across the following six dimensions.

Toxicity: Comments perceived as disrespectful, offensive, or disruptive.

Severe Toxicity: Highly toxic comments with heightened insult, attack, and/or verbal violence.

Identity Attack: Comments targeting a person or group based on identity traits.

Insult: Derogatory or demeaning expressions towards an individual or group.

Profanity: Vulgar or blasphemous language unacceptable in civil discourse.

Threat: Explicit statements of violence or physical harm towards an individual or group.

Perspective and much of the scholarly literature recommend flagging individual messages when $p > 0.70$ – 0.90 for moderation purposes (since a lower threshold could result in many false positives). However, no hard threshold was applied in this case because the objective here was comparative and longitudinal. Retaining the continuous scores preserves variance, enabling detection of subtle yet substantively meaningful differences between party accounts and across years; differences that would be obscured by dichotomisation. Furthermore, this study pursued a large-scale statistical comparison of average toxicity levels across party accounts and over time; we therefore anticipated generally low mean scores across all dimensions within such an extensive dataset, where most tweets are expected to be informational. At the same time and with these considerations, small variations between parties are expected to represent large differences in terms of toxicity.

Following score retrieval, we conducted statistical analyses to explore inter-party differences and temporal variation in toxicity indicators. A composite variable averaging the six Perspective dimensions was created. Because assumptions of homoscedasticity were violated and distributions were skewed with outliers, non-parametric tests were employed. This robust approach enabled comparison of toxicity levels across accounts, detection of significant longitudinal trends, and examination of potential correlations between toxicity and tweet interaction metrics.

6 Findings

To begin, we compared the public metrics of the parties’ Twitter accounts at the point of data extraction, using the original (unfiltered) dataset. As shown in Table 2, PSOE led overall engagement on Twitter with 140,963 posts (second-highest) and the second-largest follower base (857,241). Podemos, although slightly less active (137,591 tweets), recorded the largest community of followers (1,534,441) and the second-highest following count (18,586), indicating a particularly robust presence. PP, with 119,391 tweets, also maintained a substantial follower base (852,779), only marginally below that of PSOE. Notably, both major parties enjoy almost identical follower figures.

ERC exhibited the greatest activity (179,341 tweets) and by far the highest following count (92,421), suggesting an extensive networking

TABLE 2 Public metrics of major Spanish parties’ Twitter accounts.

Party	Tweets	Followers	Following	Listed
PSOE	140,963	857,241	13,207	5,730
PP	119,391	852,779	5,298	5,045
Vox	74,706	513,571	1,637	2,150
Podemos	137,591	1,534,441	18,586	5,999
Más País	6,905	68,243	621	394
Sumar	1,168	96,917	186	258
PNV	64,307	35,819	86	560
EH Bildu	96,756	78,796	1,380	526
ERC	179,341	369,679	92,421	2,356
Junts	62,035	120,508	601	645

Own elaboration.

TABLE 3 Mean interaction metrics by party.

Party	RTs	Replies	Likes	Quotes
PSOE	98.50	59.68	130.25	9.16
PP	102.25	39.27	143.06	7.50
Vox	616.53	89.84	1,254.61	47.50
Podemos	162.97	29.57	241.86	9.18
Más País	48.44	16.94	117.05	6.09
Sumar	87.57	105.16	334.30	19.51
PNV	9.30	2.84	16.14	1.09
EH Bildu	21.88	4.56	36.45	1.95
ERC	69.90	17.86	108.52	3.92
Junts	97.92	10.72	231.81	3.01

Own elaboration.

strategy. Podemos was the most frequently listed account (5,999 lists), followed by PSOE (5,730) and PP (5,045). Predictably, the newest formations, Más País and Sumar, showed the lowest tweet volumes. Nevertheless, in terms of followers and following, these accounts already surpass PNV, which attracts the fewest interactions among the parties considered.

We then examined average interaction metrics for tweets posted between 2015 and 2023 (retweets, replies, likes, and quotes). As shown in Table 3, in this case Vox dominates engagement, recording the highest mean values for retweets (616.53), likes (1,254.61), and quotes (47.50). This indicates exceptionally strong supporter commitment, despite its low activity on the platform compared to other parties. Vox is surpassed only in replies by Sumar (105.16), which have generated substantial interaction and dialogue in a short period of time. This party also records the second-highest mean likes (334.30) and quotes (19.51). Podemos ranks second in mean retweets (162.97) and third in likes (241.86) and quotes (9.18). By contrast, PNV and EH Bildu post the lowest averages across all metrics, confirming limited engagement. On the other hand, PP attracts more likes on average (143.06) than PSOE (130.25), although both parties have similar and relatively high RT figures. It is noteworthy in any case that the PP account is the third in terms of RTs, with an average of 102.25. Among the Catalan pro-independence parties, ERC and Junts reach notable like averages (108.52 and 231.81, respectively), reflecting committed followings.

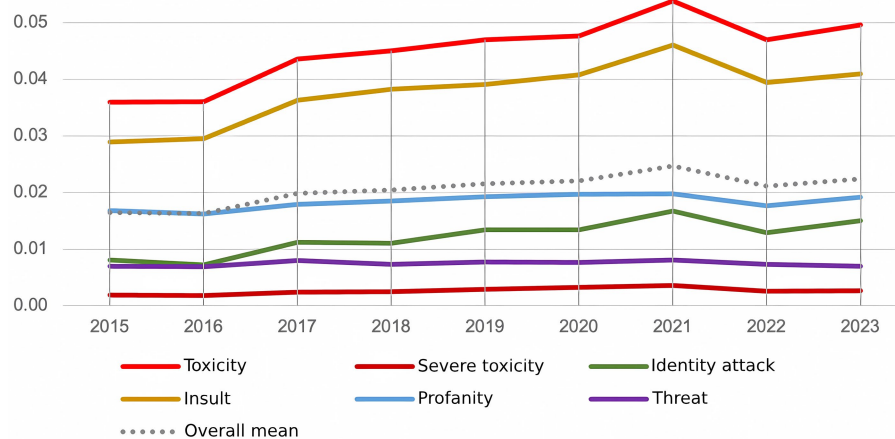


FIGURE 1

Mean values of the various toxicity dimensions on the official Twitter accounts of Spain's major political parties from 2015 to 2023.

6.1 Toxicity levels across major Spanish parties' Twitter accounts

To evaluate the presence of toxicity in tweets disseminated by Spain's major political parties on Twitter, we conducted a robust Welch one-way analysis of variance, a method appropriate for samples exhibiting unequal variances, as confirmed by Levene's test ($p < 0.01$). The dependent variable was a composite toxicity score, computed as the mean of all six Perspective API dimensions, enabling a comparison of overall toxicity levels across party accounts. Exploratory descriptive statistics revealed substantial variation in mean toxicity: Vox's account recorded the highest average toxicity ($M = 0.042$, $SD = 0.061$), followed by ERC ($M = 0.025$, $SD = 0.034$), whereas PNV exhibited the lowest mean toxicity ($M = 0.011$, $SD = 0.020$).

The Welch ANOVA confirmed statistically significant differences in toxicity between accounts, $F(9, 223,576) = 385.308$, $p < 0.001$, $\eta^2 = 0.044$, suggesting that political affiliation is a meaningful predictor of average toxicity in party communications. Subsequent Games–Howell *post hoc* comparisons revealed multiple significant pairwise differences (all $p < 0.001$), underscoring the heterogeneity of political discourse on social media. Vox's toxicity significantly exceeded that of every other party, with mean differences of -0.025 compared to PSOE, -0.026 compared to PP, and -0.022 compared to Podemos (all $p < 0.001$), highlighting Vox's notably aggressive rhetoric. Conversely, PSOE and PP differed only minimally in mean toxicity ($\Delta M = 0.001$, $p < 0.001$), indicating almost identical toxicity profiles. Both parties differed significantly from ERC (ΔM of 0.008 for PSOE–ERC and 0.009 for PP–ERC, both $p < 0.001$). Podemos and Más País maintained comparatively low toxicity relative to Vox and ERC; nonetheless, Podemos's toxicity was still significantly higher than that of PSOE ($\Delta M = 0.003$, $p < 0.001$). Less pronounced yet significant differences emerged among regionalist parties such as PNV and EH Bildu, suggesting more moderate or infrequent use of toxic language. All reported differences achieved significance at the 1% level, indicating high statistical reliability.

When examining which parties formed statistically indistinguishable clusters, we found that PSOE ($M = 0.017$, $SD = 0.027$), PP ($M = 0.016$, $SD = 0.025$), Más País ($M = 0.017$,

$SD = 0.027$), EH Bildu ($M = 0.016$, $SD = 0.027$) and JuntsXCat ($M = 0.017$, $SD = 0.026$) did not differ significantly among themselves; Podemos ($M = 0.020$, $SD = 0.032$) would clustered with ERC, even with lower values; and finally Sumar ($M = 0.011$, $SD = 0.024$) grouped with PNV ($M = 0.011$, $SD = 0.019$), the two parties exhibiting the lowest average toxicity. On this basis, the accounts can be classified into four principal tiers: Vox thus stands alone as the most toxic account, followed by ERC and Podemos as a secondary tier, with a moderate-high toxicity tendency; a moderate group comprising PSOE, PP, Más País, EH Bildu and JuntsXCat, and finally, at the lowest level, Sumar and PNV with minimal toxicity. Notably, PNV's exceptionally low standard deviation suggests a consistently moderate tone across its tweets. Vox not only recorded the highest mean toxicity but also the maximum scores on every individual dimension. These average toxicity levels across all indicators and the total average level for each are depicted in Figure 1.

In addition, given that Vox and ERC exhibited markedly higher mean toxicity scores than all other parties, and to address RQ1 with greater precision, we conducted three planned contrasts to test: first, whether Vox's average toxicity differed significantly from the combined mean of all remaining parties; second, whether ERC's average toxicity differed from all other parties; and third, whether Vox and ERC differed from one another. All three contrasts yielded highly significant results, each with a small but consistent effect size (Cohen's $d \approx 0.032$, corroborated by Hedges's correction). Specifically, the contrast of Vox versus the pooled other parties produced $t(11,076) = -31.870$, $p < 0.001$, $d = -0.032$, unequivocally confirming that Vox's tweets were substantially more toxic than those of any other party. Likewise, ERC versus the remainder yielded $t(46,578) = -3.749$, $p < 0.001$, $d = -0.032$, demonstrating that ERC's toxicity also significantly exceeded the aggregate toxicity of the other accounts. Finally, directly comparing Vox and ERC resulted in $t(867,044) = -12.371$, $p < 0.001$, $d = -0.032$, indicating that Vox's toxicity was significantly greater than that of ERC. Although these effect sizes are small by conventional benchmarks, they consistently reveal a clear hierarchy, Vox at the apex of toxicity, followed by ERC, with all other parties falling significantly below.

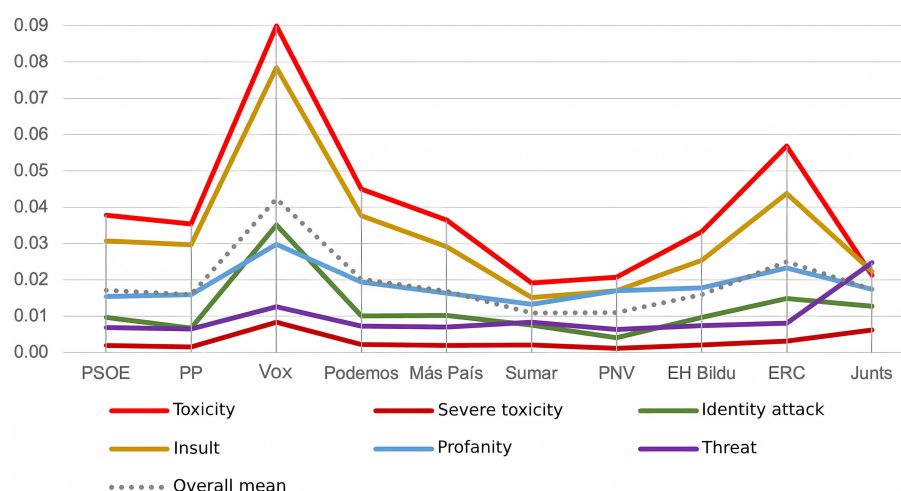


FIGURE 2

Temporal evolution of mean levels of the different toxicity dimensions detected on the official Twitter accounts of Spain's major political parties from 2015 to 2023.

6.2 Temporal evolution of overall toxicity levels

To analyse temporal differences in toxicity levels and thereby address RQ2 and RQ3, we also conducted a Welch one-way ANOVA comparing the annual mean composite toxicity scores from 2015 through 2023. We once again selected the Welch procedure for its robustness when variances across groups are unequal. In this case, descriptive inspection revealed that, while toxicity fluctuated year to year, the peaks occurred in 2021 ($M = 0.0247$) and 2023 ($M = 0.0224$). The Welch ANOVA reported statistically differences in toxicity levels between the analysed years, $F(8, 51266.233) = 200.676$, $p < 0.001$, $\eta^2 = 0.008$, indicating that toxicity levels have varied meaningfully over the study period, albeit with a generally upward trajectory since 2015.

To pinpoint which years differed, we performed Games–Howell post-hoc tests. These revealed significant pairwise contrasts (all $p < 0.001$), underscoring a pronounced evolution in toxicity prevalence over time. Notably, although the data exhibit a steady rise in toxicity from 2015, 2021 ($M = 0.025$, $SD = 0.041$) clearly emerged as the year with the highest mean toxicity, differing significantly from every other year, including 2023 ($M = 0.022$, $SD = 0.035$), which in turn was significantly higher than earlier years. The largest gap occurred between 2021 and 2016 ($M = 0.016$, $SD = 0.026$), the year with the lowest toxicity ($\Delta M = -0.008$, $p < 0.001$), marking 2021 as a clear apex far above prior levels.

On the other hand, no significant differences were found between 2015 and 2016 ($M = 0.016$, $SD = 0.025$), nor among 2017 ($M = 0.020$, $SD = 0.033$), 2018 ($M = 0.020$, $SD = 0.034$), 2019 ($M = 0.022$, $SD = 0.036$), 2020 ($M = 0.022$, $SD = 0.039$) and 2022 ($M = 0.021$, $SD = 0.034$). Thus, in response to RQ2, although raw means suggest an almost uninterrupted escalation of toxicity in official party tweets since 2015, the ANOVA's multiple comparisons show that most year-to-year increases are not statistically significant, implying a relative plateau from 2017 onwards, punctuated by marked peaks in 2021 and

again in 2023. Nevertheless, comparing these later years against the two initial years reveals a distinct upward shift: every year from 2017 forward differs significantly from both 2015 and 2016 (all $p < 0.001$), the two lowest-toxicity years. The full trajectory of annual mean toxicity across all parties is depicted in Figure 2.

Finally, to test RQ3 more rigorously, we carried out planned contrasts for the two peak years. Comparing 2021 to the pooled mean of all other years yielded $t(24\ 804.206) = -15.189$, $p < 0.001$, $d = -0.033$; comparing 2023 to the pooled remainder gave $t(4\ 369.585) = -3.804$, $p < 0.001$, $d = -0.033$; and directly contrasting 2021 with 2023 resulted in $t(8\ 011.302) = 2.644$, $p < 0.001$, $d = -0.032$. These findings confirm that 2021 is the year with the highest toxicity, significantly exceeding the composite average of all other years, and that 2023 maintains the second-highest level. While these effect sizes remain small in absolute magnitude, they are meaningful within the context of a large, heterogeneous dataset and indicate a clear upward trend in toxic language among Spain's leading political parties on Twitter. Despite the variability reflected in measures of dispersion, the consistent rise in mean toxicity points to a troubling intensification of toxic political commentary on the platform over the period studied, which at the same time could be seen as an indicator that polarisation has grown.

6.3 Temporal evolution of toxicity levels by political party

To chart how toxicity evolved for each party from 2015 onwards, we examined the temporal trajectories of all six toxicity dimensions on each party's official Twitter account. Vox stood out unequivocally, registering the highest toxicity levels across nearly every year and in every dimension. Only in 2022 did Podemos slightly exceed Vox in Severe Toxicity; ERC surpassed Vox in Profanity in both 2022 and 2023; and in the Threat dimension for 2023, both ERC and Junts recorded higher scores. Furthermore,

Vox's most pronounced toxicity peaks occurred not in 2023 but in 2020 and 2021, with mean values of 0.136 and 0.134 in Toxicity, 0.015 and 0.016 in Severe Toxicity, 0.056 and 0.064 in Identity Attack, 0.126 and 0.122 in Insult, 0.042 and 0.038 in Profanity, and 0.016 and 0.019 in Threat, respectively. These figures indicate that Vox's account was chiefly responsible for the overall toxicity zenith in 2021, although in 2023 toxicity rose further through contributions from ERC and Junts, and to a lesser extent Podemos. In contrast, ERC and Podemos themselves exhibited similar –but less extreme–temporal patterns, with noticeable spikes especially in 2021 and again in 2023.

Opposite to Vox's profile, the Basque Nationalist Party (PNV) maintained the lowest toxicity levels across all dimensions for every year. Remarkably, PNV's toxicity even dipped to its lowest mean values in 2021 (0.005 in Toxicity, 0.000 in Severe Toxicity, 0.001 in Identity Attack, 0.007 in Insult, 0.011 in Profanity and 0.005 in Threat) underscoring that the general toxicity peak of 2021 definitely cannot be attributed to this party. Indeed, PNV's toxicity was only surpassed in terms of lower levels by Junts in Toxicity (2018, 2022), in Severe Toxicity (2020), in Insult (2018 and 2020), in Profanity (2018, 2022) and in Threat (2018, 2022). Similarly, the newly established Sumar account registered particularly low levels of Profanity and Threat during its 2 years of existence (2022–2023). The two major national parties, PSOE and PP, displayed comparatively moderate toxicity levels, remaining relatively stable over time with only slight fluctuations across all dimensions. EH Bildu and Más País likewise maintained low to moderate, steady toxicity profiles since their respective inceptions. The full party-by-year evolution of each toxicity dimension is illustrated in [Figure 3](#).

6.4 Relationship between toxicity and interaction

Finally, to investigate the relationship between toxicity levels in tweets posted by the analysed political parties and their corresponding interaction metrics and thus answer RQ4, we employed Spearman's rank-order correlation. This statistic allowed us to assess associations between quantitative variables without presuming linearity. As we had no *a priori* hypothesis regarding the direction of these associations, we used a two-tailed significance test. Specifically, we examined the correlation between each tweet's composite toxicity score (the mean across all toxicity indicators) and its number of likes, retweets, and replies.

The analysis yielded positive but moderate and statistically significant correlations in all cases: toxicity versus likes ($\rho = 0.101$, $p < 0.001$), toxicity versus retweets ($\rho = 0.110$, $p < 0.001$) and toxicity versus replies ($\rho = 0.065$, $p < 0.001$). These results suggest that more toxic messages on the official Twitter accounts of the parties under study tend to attract a greater number of likes and replies and are more frequently retweeted. However, the strength of these associations is modest and does not permit reliable prediction of one variable from another. Nonetheless, recognising that multiple factors influence a tweet's reach and engagement, the observed correlations indicate a discernible trend: Twitter users may show a modest preference or heightened engagement with political content that is toxic or controversial in nature.

7 Discussion and conclusion

This study comparatively has analysed the levels of toxicity present in the content published by the main Spanish political parties on Twitter from 2015 to 2023, revealing both the heterogeneity between parties and a temporal dynamic that carries important theoretical and practical implications. Our findings demonstrated that toxicity levels differ significantly by party affiliation, with Vox consistently exhibiting the highest mean toxicity across all six dimensions measured (toxicity, severe toxicity, identity attack, insult, profanity and threat), followed by ERC and Podemos in a secondary tier, and with Sumar and PNV at the lowest end of the spectrum. These inter-party differences extend earlier research on political party communication styles (e.g., [Guerrero-Solé and Philippe, 2020](#); [de Borja Navarro and Yeh, 2022](#)) by systematically comparing the full set of major parties' accounts over an extended period of more than 8 years. The striking contrast between Vox's aggressive rhetoric and the generally more moderate discourse of the other parties underscores the role of radical ideological positioning, particularly far-right populism, in fostering toxic language online. As seen in the initial review, several authors have already pointed out and warned about the toxic and hostile language used as a strategy by Vox in the Spanish political landscape (e.g., [Galindo-Domínguez et al., 2023](#); [Guerrero-Solé and Philippe, 2020](#); [Esteban, 2022](#); [Fernández Romero et al., 2021](#); [Vico and Rey, 2020](#)), something this study empirically confirms. Nevertheless, this is not something that affects only Spaniards, it is a global trend. The international extreme right seems to have adopted for years communication strategies loaded with toxic and hostile language, as well as disinformation and hate speech, something that seems to have done nothing but give them popularity and votes (e.g., [Baum-Baicker, 2020](#); [Caiani et al., 2021](#); [Collins, 2023](#); [Daniel et al., 2025](#); [Neiwer, 2017](#)).

Temporal analyses further revealed that toxicity on party accounts has not remained static but has followed a generally upward trajectory since 2015, punctuated by pronounced peaks in 2021 and again in 2023. The apex in 2021, when mean toxicity reached its highest point, may plausibly be linked to the convergence of significant political events, such as the COVID-19 pandemic's social and economic fallout, regional elections, and high-stakes national debates, that heightened political tensions and incentivised more combative online strategies. The subsequent rise in 2023 may reflect a continuation of this polarised environment, compounded by contentious policy discussions at both national and regional levels. These temporal fluctuations corroborate the notion of toxicity as a dynamic phenomenon ([Wulczyn et al., 2017](#); [Froio and Ganesh, 2018](#)), and point to the necessity of adopting a longitudinal lens when assessing digital political communication. In particular, these findings seem to highlight that the most toxic and populist parties in Spain, such as Vox, could be taking advantage of controversial and affecting episodes to further inflame public opinion with toxic and violent discourses.

Lastly, the moderate yet significant positive correlations between toxicity and interaction metrics (likes, retweets and replies) suggest that more toxic tweets generate slightly higher engagement on average. While these associations are not strong enough to permit predictions at the individual-tweet level, they align with patterns observed in other contexts where emotionally charged or provocative content attracts greater user attention ([Froio and Ganesh, 2018](#); [Vicente et al., 2021](#)). This “engagement premium” for toxic content raises both

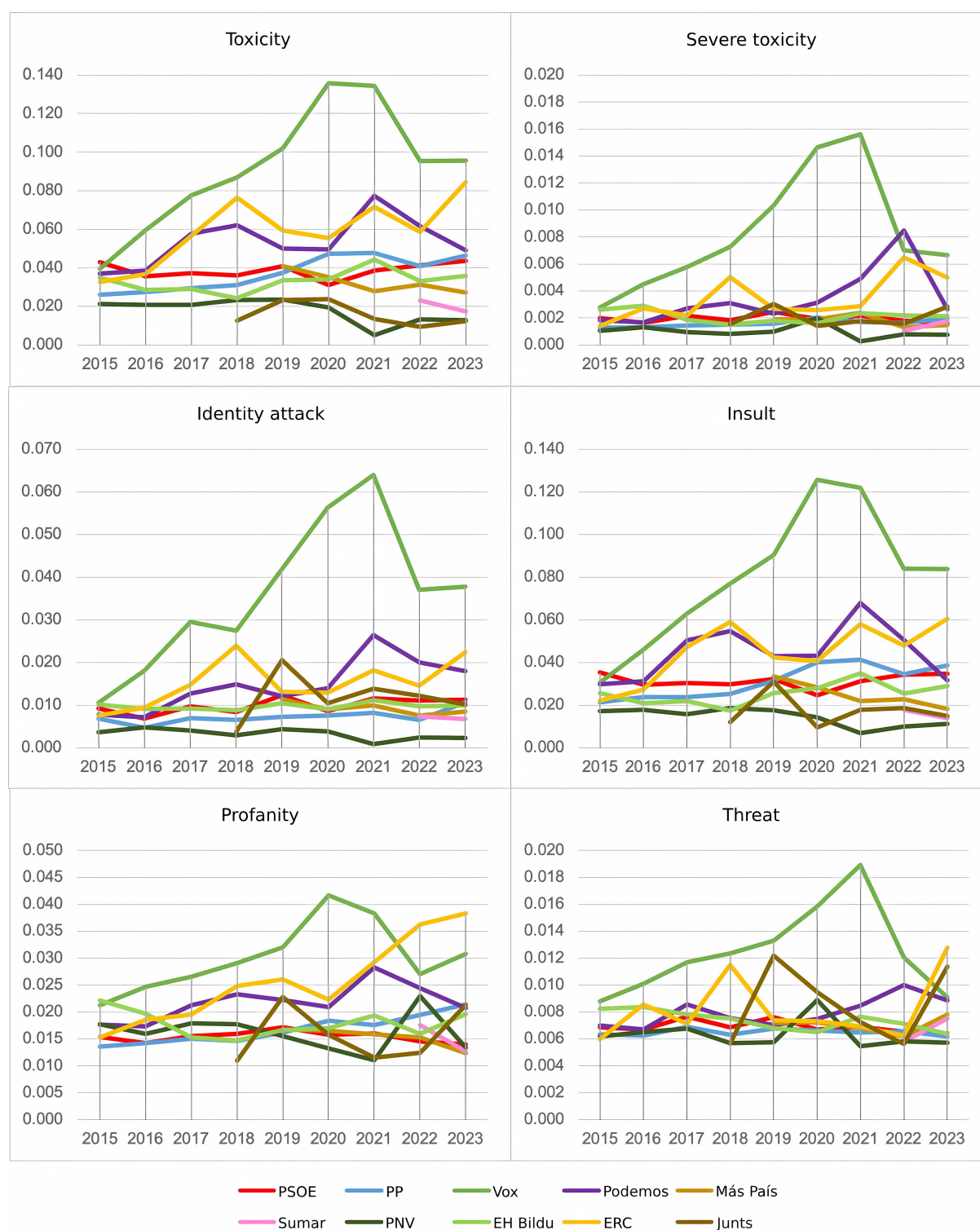


FIGURE 3

Year-on-year evolution of mean toxicity dimension levels on the official Twitter accounts of Spain's major political parties from 2015 to 2023.

theoretical questions about the incentives shaping party communication strategies and practical concerns regarding the amplification of hostile discourse on social media platforms such as Twitter, currently X. Moreover, this could partly explain the digital popularity and relative electoral success of the Spanish far right in

such a short period of time, just as has been seen with the far right around the world. Although it should be noted that the number of followers and engagement on social media does not always translate directly into votes. Furthermore, this positive correlation between toxicity scores and higher interaction volumes does not in itself

establish causality, yet it resonates with the well-documented “negativity bias” of online attention economies.

Inspection of tweets located in the upper 10% of the toxicity distribution mainly reveals three recurrent patterns. First, the bulk of these posts constitute confrontational and aggressive attacks on rival parties. They can be read as part of the broader “culture war” that right-wing actors frequently claim to be waging online against progressive ideas, and they often deploy explicitly hostile and violent language. Second, there seems to be an association of hostile language with unverified or blatantly false claims. In this regard, several authors had already warned of the political use of disinformation on platforms such as Twitter, especially by ultra-conservative groups (e.g. [Cano-Orón et al., 2021](#); [Castro Martínez and Díaz Morilla, 2021](#); [Lava Santos, 2021](#); [Pérez-Curiel et al., 2022](#)). This overlap suggests that toxicity can serve as a stylistic vehicle for political disinformation, or conversely, that political parties use toxic language in the fake content they spread on Twitter, trying to reach more people and/or generate more visceral reactions. Third, a smaller but salient group, mostly led by Vox, contains implicit or explicit hate speech directed at stigmatised minorities, most commonly immigrants or LGBTIQ+ communities, thereby moving beyond inter-party rivalry to target vulnerable populations. Taken together, these patterns imply that toxic rhetoric not only amplifies partisan polarisation but also risks normalising exclusionary narratives and eroding deliberative norms in the Spanish Twittersphere. And this is especially serious if such messages –perhaps precisely because of their confrontational nature, which triggers emotional reactions–, are rewarded with increased interaction.

By integrating these findings with existing literature, our study fills several gaps. First, it offers the most exhaustive cross-party comparison of toxicity levels in Spanish political Twitter to date, extending prior analyses that have often focused on single parties or electoral cycles ([Amores et al., 2022](#); [Guerrero-Solé and Philippe, 2020](#)). Second, it highlights the temporal dimension of toxicity, demonstrating that online hostility is not merely a static attribute of particular actors but a responsive feature of the broader political context. Third, by linking toxicity to user engagement, we contribute to understanding the feedback loop between party communication and audience behaviour. And these contributions carry clear implications. Theoretically, our findings enrich our understanding of digital political communication by integrating toxicity as a central variable across parties, time and engagement outcomes. This underscores the importance of incorporating toxicity and engagement metrics into models of digital political communication, particularly in multiparty systems characterised by fragmentation and polarisation, to account for the interplay between ideology, context and platform incentives. Practically, our findings provide empirical guidance for social media platforms and regulators, they suggest that platform designers and policymakers should prioritise moderation strategies during periods of heightened political tension, such as elections or major legislative debates, when toxicity peaks, as well as promote transparency around algorithmic boosts of hostile content. And at the social level, civil society organisations and party communication teams might also use these insights to calibrate their messaging strategies, aiming to reduce hostile rhetoric without sacrificing reach. Furthermore, strategies should be developed to pressure parties (especially the polarized and far-right ones), to moderate, soften, and professionalize their online communication strategies, and to pressure the platforms themselves so that algorithms do not favour this type of emotional and

sensationalist content and do not benefit from greater impact and user interaction. Nonetheless, parties themselves should reconsider rhetorical strategies that trade civility for short-term engagement, reflecting on the long-term impact of toxicity on democratic norms, as the documented rise in toxic rhetoric carries implications for public and institutional trust, public debate, and civic participation. As hostile discourse becomes more visible, it risks normalising aggression and discouraging moderate voices, thereby fracturing the deliberative fabric of democracy ([Mazzoleni and Splendore, 2020](#)). Civil society organisations, educators and journalists must therefore collaborate to foster digital literacy and promote constructive dialogue, counteracting the polarising effects of toxic language, frequently overlapped, moreover, with disinformation practices ([Blanco-Herrero et al., 2021](#)).

In sum, this study highlights a complex, evolving landscape of toxic political communication on Twitter in Spain. The pronounced inter-party differences, temporal escalation and engagement dynamics collectively underscore the urgency of sustained monitoring, evidence-based policy interventions and proactive communication strategies to safeguard the quality of public debate and the health of democratic discourse.

8 Limitations and future research

First, as mentioned above, the Perspective API, although widely validated, offers very limited reliability when analysing co-official languages in Spain. Consequently, messages authored in Catalan, Basque or Galician (8.4% of the full corpus) were excluded, a decision that disproportionately affects regionalist parties such as ERC, Junts and, to a lesser extent, PNV. Therefore, the reported mean toxicity scores for these accounts may be biased and should be interpreted as lower-bound estimates. A second constraint stems from Perspective’s cross-lingual architecture. Recent research shows that the API can mis-estimate toxicity in non-English texts and exhibits language-specific biases ([Nogara et al., 2025](#)). Although Spanish performance is comparatively strong, dialectal variation, regional slang or code-switching could still induce measurement error.

On the other hand, the uneven temporal coverage of certain parties, due to the recent emergence of some of them, such as Más País and Sumar, in addition to the elimination of content in regional languages from others, such as Junts (which did not offer content in Spanish until 2018), limits the comparability of the entire time series. Also, it should be noted that this study focuses solely on digital political communication that takes place on Twitter, ignoring other online and offline platforms. Furthermore, computational methods, which allow large samples to be analysed, may not achieve the depth, precision, interpretability, and contextualization capabilities of manual analyses.

Future research should address these limitations by incorporating more parties and new multimodal data sources derived from different platforms, which would allow for cross-platform comparisons and more robust generalizations. It would also be convenient to expand the sample and continue longitudinally exploring levels of toxicity in Spanish political communication. Future work should also combine manual validation with explicitly multilingual classifiers. The implementation of advanced linguistic models tailored to Spanish political discourse could improve sensitivity to contextual subtleties and minority languages. Open-source

models such as *toxic-bert*, fine-tuned on the Jigsaw corpus for seven languages including Spanish and Catalan, already achieve competitive F1 scores and could be further adapted to Basque via transfer learning. In addition, new computational techniques such as network analysis or topic modelling could also be integrated, which would shed light on which are the topics and narratives that predominate in toxic content and how this content spreads in broader political ecosystems, potentially revealing the role of parties, influencers and bots. Finally, it would be interesting to expand and complement this type of analysis with mixed methods, seeking to deepen our understanding of the motivations behind toxic messages and their actual effects on audience attitudes and behaviour, as well as to contextually link them to explicit hate speech and disinformation campaigns. These avenues would be relevant for developing comprehensive strategies to mitigate toxicity and foster healthier digital public spheres.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: <https://zenodo.org/records/15353369>.

Ethics statement

Ethical approval was not required for the study involving human data in accordance with the local legislation and institutional requirements. Written informed consent was not required, for either participation in the study or for the publication of potentially/indirectly identifying information, in accordance with the local legislation and institutional requirements. The social media data was accessed and analyzed in accordance with the platform's terms of use and all relevant institutional/national regulations.

Author contributions

JJA: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. WG-B: Conceptualization, Funding

References

- Åkerlund, M. (2020). The importance of influential users in (re)producing Swedish far-right discourse on Twitter. *Eur. J. Commun.* 35, 613–628. doi: 10.1177/0267323120940909
- Álvarez-Benavides, A., and Jiménez Aguilar, F. (2021). La contraprogramación cultural de Vox: secularización, género y antifeminismo [Cultural counter-programming by Vox: Secularisation, gender and antifeminism]. *Polit. Soc.* 58:e74486. doi: 10.5209/POSO.74486
- Amores, J. J., and Arcila-Calderón, C. (2025). What is the nature of the racist and xenophobic hate speech disseminated on social media in Southern Europe? Topic modelling of anti-immigration messages posted on X and YouTube in Spain, Italy and Greece. *Commun. soc.* 38, 163–180. doi: 10.15581/003.38.1.013
- Amores, J. J., Frias-Vázquez, M., Sánchez-Holgado, P., and Blanco-Herrero, D. (2022). Análisis comparativo de la información política publicada en Twitter por los principales medios impresos y digitales en torno a las elecciones generales del 10N de 2019. *AdComunica.* 23, 301–328. doi: 10.6035/adcomunica.6274
- Amores, J. J., Lara, M. C. F., and Cogo, D. (2024). Racismo, discursos de odio y migraciones en el contexto de los medios de comunicación y las plataformas digitales. *Anu. Electrón. Estud. Comun. Soc. "Disertaciones".* 17:13898. doi: 10.12804/revistas.urosario.edu.co/disertaciones/a.13898
- Arcila-Calderón, C., De la Vega, G., and Blanco-Herrero, D. (2020). Topic modeling and characterization of hate speech against immigrants on twitter around the emergence of a far-right party in Spain. *Sociol. Sci.* 9:188. doi: 10.3390/socsci9110188
- Barba, C. H., and Blanco, V. S. (2011). Activismo político en red: del Movimiento por la Vivienda Digna al 15M. *Teknokultura: Revista de Cultura Digital y Movimientos Sociales* 8, 157–175. Available online at: <https://revistas.ucm.es/index.php/TEKN/article/view/48025>
- Baum-Baicker, C. (2020). Not fake news: toxic consequences of the trump stress effect. *J. Humanist. Psychol.* 60, 730–746. doi: 10.1177/0022167820911757
- Blanco-Herrero, D., Amores, J. J., and Arcila-Calderón, C. (2024). "Inequalities and the media: spread and reception of hate speech against migrants and refugees in social media" in Research handbook on inequalities and work. eds. C. Forson, G. Healy, M. B. Öztürk and A. Tatli. (Cheltenham, UK: Edward Elgar Publishing), 211–227.

acquisition, Software, Writing – original draft, Writing – review & editing. MO-A: Conceptualization, Funding acquisition, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research received no financial support from any funding agency, commercial or not-for-profit sectors. The Article Processing Charge was funded by internal publication funds of the University of Salamanca.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpos.2025.1627474/full#supplementary-material>

- Blanco-Herrero, D., Amores, J. J., and Sánchez-Holgado, P. (2021). Citizen perceptions of fake news in Spain: socioeconomic, demographic, and ideological differences. *Publica* 9:35. doi: 10.3390/publications9030035
- Blanco-Herrero, D., Frías-Vázquez, M., Amores, J. J., and Sánchez-Holgado, P. (2020). "Comunicación política en Twitter: Análisis de los perfiles de los principales medios de comunicación españoles durante la campaña del 10N" in *El apagón analógico y el despertar del periodismo digital*. ed. J. Cruz. (Quito, Ecuador: Pontificia Universidad Católica del Ecuador).
- Bustos Díaz, J. B., and Ruiz del Olmo, F. J. R. (2021). Presencia e interacción de los candidatos a la presidencia del gobierno de España en las principales redes sociales Durante la campaña electoral de noviembre de 2019. *Revista OBETS* 16:63. doi: 10.14198/OBETS2021.16.1.04
- Caiani, M., Carloti, B., and Padoan, E. (2021). Online hate speech and the radical right in times of pandemic: the Italian and English cases. *Javn. - Public* 28, 202–218. doi: 10.1080/13183222.2021.1922191
- Campos-Domínguez, E. (2017). Twitter y la comunicación política. *Prof. Inferm.* 26:785. doi: 10.3145/epi.2017.sep.01
- Cano-Orón, L., Calvo, D., García, G. L., and Baviera, T. (2021). Disinformation in Facebook ads in the 2019 Spanish general election campaigns. *Media Commun.* 9, 217–228. doi: 10.17645/mac.v9i1.3335
- Carozzi, F., Cipullo, D., and Repetto, L. (2022). Political fragmentation and government stability: evidence from local governments in Spain. *Am. Econ. J. Appl. Econ.* 14, 23–50. doi: 10.1257/app.20200128
- Castillo Jara, S., Marchena Montalvo, V., and Quiliche Arévalo, S. (2019). El discurso antifeminista como recurso electoral en Brasil y España. *Politai: Revista de Ciencia Política* 10, 35–59. doi: 10.18800/politai.201902.002
- Castro Martínez, A. C., and Díaz Morilla, P. (2021). La comunicación política de la derecha radical en redes sociales: De Instagram a TikTok y Gab, la estrategia digital de Vox. *Rev. Dig.* 7:67. doi: 10.7203/rd.v1i7.2102021
- Collins, J. (2023). Mobilising extremism in times of change: analysing the UK's far-right online content during the pandemic. *Eur. J. Crim. Policy Res.* 29, 355–377. doi: 10.1007/s10610-023-09547-9
- Congosto, M. L. (2015). Elecciones Europeas 2014: Viralidad de los mensajes en Twitter. *Redes. Rev. Hispana Anal. Redes Soc.* 26, 23–52. doi: 10.5565/rev/redes.529
- Daniel, W. T., Frelin, E., Robert, M. V., and Rowley-Abel, L. (2025). Individual drivers of toxicity in radical right-wing populist legislative campaigns. *West Eur. Polit.* 48, 590–617. doi: 10.1080/01402382.2024.2311040
- de Borja Navarro, F., and Yeh, A. (2022). The dangerous discourse of "us" vs. "them": Spain's Vox discursive practices. *J. Natl. Mem. Lang. Polit.* 16, 211–233. doi: 10.2478/jnmlp-2022-0005
- Esteban, A. N. (2022). Debate político en España sobre los menores extranjeros no acompañados en los medios de comunicación digitales. *RIEM Rev. Int. Estud. Migr.* 11:4769. doi: 10.25115/riem.v11i2.4769
- Fernández Romero, C., Prieto-Andrés, A., and Uldemolins Julve, E. (2021). Entre la aceptación de la inmigración y la amenaza identitaria: El reflejo en la prensa de los discursos políticos sobre la inmigración durante las elecciones al parlamento de España en abril de 2019. *Estud. Mensaje Period.* 27, 123–132. doi: 10.5209/esmp.70898
- Ferreira, C. (2019). Vox como representante de la derecha radical en España: un estudio sobre su ideología. *Rev. Esp. Cienc. Polit.* 51, 73–98. doi: 10.21308/recp.51.03
- Froio, C., and Ganesh, B. (2018). The transnationalisation of far right discourse on twitter. *Eur. Soc.* 21, 513–539. doi: 10.1080/14616696.2018.1494295
- Galindo-Domínguez, H., Magariño, S. G., and Iglesias, D. L. (2023). Discurso político en redes sociales sobre el colectivo minoritario inmigrante: Análisis de las narrativas e implicaciones educativas. *OBETS: Revista de Ciencias Sociales* 18:97. doi: 10.14198/obets.22342
- Guerrero-Solé, F., and Philippe, O. (2020). La toxicidad de la política española en Twitter durante la pandemia de la COVID-19. *Hipertext.net Revista Académica Sobre Documentación Digital y Comunicación Interactiva* 21, 133–139. doi: 10.31009/hipertext.net.2020.i21.12
- Hutchins, R. D., and Halikiopoulou, D. (2019). Enemies of liberty? Nationalism, immigration, and the framing of terrorism in the agenda of the front national. *Nations Natl.* 26, 67–84. doi: 10.1111/nana.12555
- Jerez, A., Maceiras, S. D., and Maestu, E. B. (2015). Esferas públicas, crisis política e internet: el surgimiento electoral de Podemos. *Hist. Cienc. Saude-Manguinhos* 22, 1573–1596. doi: 10.1590/S0104-59702015000500003
- Lagares Díez, N., López-López, P. C., Rivera Otero, J. M., and Jaráiz Gullás, E. (2021). La comunidad digital y el uso político de las redes sociales: elecciones generales del año 2019 en España. *Araucaria* 23. doi: 10.12795/araucaria.2021.i48.11
- Latorre, J. P., and Amores, J. J. (2021). Topic modelling of racist and xenophobic YouTube comments. *Analyzing hate speech against migrants and refugees spread through YouTube in Spanish*. In *Ninth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'21)*. 456–460. doi: 10.1145/3486011.3486494
- Lava Santos, D. (2021). El discurso populista de VOX en las Elecciones Generales de 2019: Análisis comparativo de las agendas temáticas en televisión y Twitter. *Rev. Dig.* 7:37. doi: 10.7203/rd.v1i7.200
- Lecours, A. (2021). *Nationalism, secessionism, and autonomy*. Oxford University Press.
- Maarek, P. J. (2015). "Political communication, electronic media and social networks in France" in *Political parties in the digital age*. Eds. G. Lachapelle and P. Maarek (Berlin, Germany: De Gruyter), 165–180.
- Mazzoleni, G. (2018). "Populism and the media" in *The Oxford handbook of populism*. eds. K. A. Hawkins, R. E. Carlin, L. Littvay and C. R. Kaltwasser (Oxford, UK: Oxford University Press), 436–454.
- Mazzoleni, G., and Splendore, S. (2020). Media and populism in Italy: the rise of right-wing populism and the role of media framing. *Journal. Stud.* 21, 305–319. doi: 10.1080/1461670X.2020.1713163
- Neiwert, D. (2017). *Alt-America: The rise of the radical right in the age of trump*. New York, USA: Verso Books.
- Nogara, G., Pierri, F., Cresci, S., Luceri, L., Törnberg, P., and Giordano, S. (2025). Toxic bias: perspective API misreads German as more toxic. *Proc. Int. AAAI Conf. Web Soc. Media* 19, 1346–1357. doi: 10.1609/icwsm.v19i1.35876
- Oller-Alonso, M., Arcila Calderón, C., Amores, J. J., Sánchez-Holgado, P., Frías-Vázquez, M., González-Baquero, W., et al. (2025). Comparative analysis of sentiments towards different stigmatized groups on social media platforms in Spain. *Cuadernos.Info*, 61, 117–141. doi: 10.7764/cdi.61.87210
- Otalora, A. U. (2017). El populismo como vanguardia del desencanto político en Europa: el fenómeno «Podemos» en España. *Revista de Estudios Políticos* 177, 213–255. doi: 10.18042/cepc/rep.177.07
- Pérez-Curiel, C. (2020). Trend towards extreme right-wing populism on twitter: an analysis of the influence on leaders, media and users. *Commun. Soc.* 33, 175–192. doi: 10.15581/003.33.2.175-192
- Pérez-Curiel, C., Marín, G. J., and Domínguez-García, R. (2022). Twitter como herramienta de comunicación y desinformación: Un análisis sobre la transparencia informativa y la corrupción política en España. *Polit. Soc.* 59:e75666. doi: 10.5209/poso.75666
- Ramos Antón, R. (2021). La crisis de la Covid-19 como oportunidad política: El uso de Twitter por Santiago Abascal durante el estado de alarma en España (14 de marzo–20 de junio de 2020). *Doxa Comun.* 33, 75–98. doi: 10.31921/doxacom.n33a993
- Rodríguez de Mora, S. (2023). La metáfora bélica en el discurso político de Vox. *FRASEOLEX Revista Internacional de Fraseología y Lexicología* 2, 122–150. doi: 10.5565/rev/fraseolex.57
- Rodríguez-Teruel, J., Barrio, A., and Barberà, O. (2016). The emergence of Podemos: A new political actor in Spain. *South European Society and Politics*.
- Royo, S. (2014). Institutional degeneration and the economic crisis in Spain. *Am. Behav. Sci.* 58, 1568–1591.
- Ruano Ibarra, L. E., López, J. C., and Mosquera, J. F. (2018). La política y lo político en Twitter: Análisis del discurso de los candidatos presidenciales de Colombia. Bogotá, Colombia: Repositorio Institucional Universidad Cooperativa de Colombia.
- Said-Hung, E. M., Prati, R. C., and Cancino-Borbón, A. (2017). La orientación ideológica de los mensajes publicados en Twitter durante el 24M en España. *Palabra Clave* 20, 213–238. doi: 10.5294/pacla.2017.20.1.10
- Saridou, T., Amores, J. J., Oller, M., Alonso, A. V., and Panagiotou, N. (2023). "Negative perceptions and hate speech toward migrants and refugees in southern Europe through social media" in *Migrants and refugees in southern Europe beyond the news stories: Photographs, hate, and journalists' perceptions*. eds. C. Arcila Calderón and A. Veglis (Lanham, MD, USA: Rowman & Littlefield), 101–150.
- Škorić, M. M., Zhu, Q., Goh, D., and Pang, N. (2016). Social media and citizen engagement: a meta-analytic review. *New Media Soc.* 18, 1817–1839. doi: 10.1177/1461444815616221
- Turnbull-Dugarte, S. J. (2019). Explaining the end of Spanish exceptionalism and electoral support for Vox. *Res. Policy*. 6:2053168019851680.
- Vicente, G., Rodríguez Martínez, G., and González Gómez, Á. (2021). Auge electoral de la extrema derecha española: Análisis de la irrupción y evolución del voto de Vox (2018–2019). *MARCO (Marketing y Comunicación Política)* 7:7891. doi: 10.15304/marco.id7891
- Vico, E. A., and Rey, P. R. (2020). Discurso, estrategias e interacciones de Vox en su cuenta oficial de Instagram en las elecciones del 28-A: Derecha radical y redes sociales. *Rev. Lat. Comun. Soc.* 77, 203–229. doi: 10.4185/rlds-2020-1455
- Wulczyn, E., Thain, N., and Dixon, L. (2017). "Ex machina: personal attacks seen at scale." In *Proceedings of the 26th international conference on world wide web*. pp. 1391–1399.