ORIGINAL ARTICLE

Overview of the Twitter conversation around #14F 2021 Catalonia regional election: an analysis of echo chambers and presence of social bots

Raul Broto Cervera[1] · **Cristina Pérez-Solà**[1] · **Albert Batlle**[1]

Abstract

The omnipresence of the digital ecosystem makes it increasingly important in our societies, which implies that the analysis and study of the digital battlefield in political elections is also becoming more necessary to protect our democracies. Previous literature showed the existence of information operations around the world, designed to manipulate the political perception of citizens, and therefore, the electoral results. This paper examines the Twitter conversation around #14F 2021 Catalonia regional elections, which had special significance due to the pandemic situation and the highly polarized scenario around Catalonia and Spain, using tools and techniques from Big Data Analytics and Artificial Intelligence. The results obtained show that the conversation existed inside robust echo chambers within each political party community, which became even more powerful if parties are unified into political affinity blocks. Also, focusing on the analysis related to the social bot presence, a significant quantity of results showed a higher presence of social bots in VOX party community compared to the rest of communities. This study corroborates other existing studies regarding the Catalan and Spanish scenario on the presence of echo chambers and on the existence of social bots with their tendency to basically amplify content; it also uncovers the lack of existence of cross-conversation between the independentist and unionist political block claimed in other studies.

Keywords Network analysis · Twitter · Catalonia · Spain · Echo chamber · Social bots

1 Introduction

The use of information operations (disinformation, propaganda, astroturfing, and other techniques, as explained in Wardle and Derakhshan (2017)) leads to a poorer public debate, modifying public perception and polarizing societies, which has disastrous consequences (Lazer et al. 2018). There is plenty of literature about it, but one of the most remarkable pieces is "The Global Disinformation Order" report (Bradshaw and Howard 2019), where it is possible to obtain a global look at the rise of propaganda in social media around the World. As noted in the report, this problem is not only affecting countries with authoritarian regimes, it is also affecting consolidated democracies in Western countries.

The present and future of our societies depend on using technology ethically, and from the academic world it is possible to provide the arguments and knowledge needed by the policymakers to understand the digital environment, and if necessary, legislate and control it. This investigation intends to make some contributions to the global analysis on this topic, adding relevant data and insights that reinforce aspects previously studied by other researchers. As has been presented in previous studies, which will be further detailed in the following section, this investigation reinforces the existence of echo chambers within each political community and shows that the echo chamber effect is even more robust when communities are grouped in political affinity blocks. Also, the study investigates the presence of social bots in the global conversation, focusing on each political community.

This research work presents an analysis with a political perspective of the social media conversation that took place during the election campaign period of a Spanish regional election: the #14F 2021 Catalan Parliamentary elections.

✉ Raul Broto Cervera
  rbroto@uoc.edu

  Cristina Pérez-Solà
  cristina.perez@uab.cat

  Albert Batlle
  abatlle@uoc.edu

1   Universitat Oberta de Catalunya, Barcelona, Spain

The data extraction was centered on Twitter, a key platform for political discussions, with approximately 30% of the Spanish population using Twitter and 18% using it for news (Newman et al. 2023). Moreover, Twitter offered numerous tools for analyzing public data, and the platform has been extensively studied, as evidenced by the multiple references on this paper.

This paper focuses on the 14th February 2021 Catalan Parliamentary Elections (#14F 2021), an important election contest for several reasons. To offer a contextual backdrop, preceding elections occurred at the close of 2017, following the suspension of Catalan autonomy. These elections resulted in a fragile majority for the pro-independence block, complicating governance throughout the legislative term (Guerrero-Solé 2022). The opposition primary objective on the #14F 2021 elections was to prevent a repeat of a pro-independence majority. Additionally, the emergence of the far-right VOX party, with a clear upward trend in the Spanish scenario, introduced a new and noteworthy element. The party, with a significant chance of securing parliamentary seats, marked its potential debut in the Catalan Parliament.

The significance of the #14F 2021 elections is highlighted by a convergence of factors: firstly, they were the first elections after the 2017 independence conflict that extremely polarized Catalan society (Guerrero-Solé 2022); secondly, there were still politicians imprisoned for these events; thirdly, the elections were called due to the disqualification of the previous President Joaquim Torra; and finally, they took place during the Covid-19 pandemic and specifically during months of high incidence (the elections were called during the third wave of the Covid-19 pandemic, and were even postponed to May 30th 2021 for a short period of time, but were rescheduled to the original date by the judiciary). Table 1 shows the results of this election, focusing on the parties that obtained parliamentary representation.

The investigation will center on two of the multiple possibilities that arise in this field, concretely on echo chambers and social bots. Echo chambers are commonly referred to as the digital bubble where users participate in a conversation mostly with like-minded others (Terren and Borge 2021). It is usually related to homophily or confirmation bias but can also be directly associated with the effects of social media algorithms. On the other hand, this paper understands social bots (Yang et al. 2019) as a broad bot classification that includes from totally automated bots to accounts with less automation but with highly targeted activity, for example where one person or organization runs hundreds of accounts at the same time. The main objective of using these social bots is to distort the debate by amplifying certain messages (or suppress contrary messages), while artificially boosting the popularity of certain users or political figures. As numerous other academic studies (Yang et al. 2024), we have used the Botometer tool (Davis et al. 2016), which is considered the best option to detect bots on Twitter.

The paper is organized as follows: we first analyze the related work on these topics, with a national scope specifically. Secondly, the research questions that englobe the investigation are formulated. Then, the methodology used to collect and analyze the data is presented, as well as the results obtained from the analysis. Finally, these results are discussed and compared to other studies, and the main conclusions are presented.

2 Related works

In this section, we navigate through the body of literature concerning echo chambers and social bots, following a logical sequence that progresses from abstract concepts to more specific considerations. We begin by exploring the theoretical concepts and influential studies within the political sphere. Later, we place special emphasis on national investigations within both domains. International studies provide valuable methodological approaches, while Spanish investigations facilitate direct comparisons of results, due to the shared political system. Once more, it is evident from the extensive body of literature that both topics have accumulated significant attention, revealing consistent trends

Table 1 #14F election results

Political party	Popular vote	Vote %	Parliamentary seats
Partit socialista de catalunya (PSC)	652.858	23.04%	33
Esquerra republicana de catalunya (ERC)	603.607	21.30%	33
Junts per catalunya (JxCAT)	568.002	20.04%	32
VOX (VOX)	217.883	7.69%	11
Catalunya en comú (Comuns)	194.626	6.87%	8
Candidatura d'unitat popular (CUP)	189.087	6.67%	9
Ciutadans (Cs)	157.903	5.57%	6
Partit popular (PP)	109.067	3.85%	3

in research outcomes, and admitting some occasional divergences. Studies focusing on echo chambers predominantly underscore the polarized nature of interactions, especially noticeable within retweet networks. Simultaneously, investigations into social bots highlight their role in shaping online interactions. Together, these studies emphasize the complex relation among digital communication, political polarization, and the influence of automated entities.

Digital social networks analysis and information operations are fields with a prolific international academic literature in the last decade. As examples in the global literature conversation: Wardle and Derakhshan (2017), Nielsen and Graves (2017), Mele et al. (2017), Tucker et al. (2018), Marchal et al. (2019), or Neudert et al. (2019). These highly respected papers analyze a wide range of topics and platforms, using different strategies and approaches, to study information operations in the digital environment. Similarly, there are studies about how the disinformation issue affects democracy in the USA, UK or China (Bradshaw et al. 2020; Howard et al. 2018; Gorodnichenko et al. 2018; King et al. 2017). Also, the Covid-19 pandemic multiplied the attention of network analytics researchers to this topic (Brennen et al. 2020; Gallotti et al. 2020), a sub-field with similarities in the analysis, even though it is not a full political framework.

As a brief introduction to the national studies on social media analysis, studies exist where the main objective is the analysis of misinformation structures and propagation. Stella et al. (2018) exposes that around the events of October 2017 in Catalonia, the political debate was inflamed with artificial hatred messages. In Aparici et al. (2019), there is an analysis of the disinformation effect on the Catalonia public debate, stating that "with our analysis, we have found that much of the misinformation does not come from the traditional media, but its propagation depends increasingly on its circulation on the Internet and the participation of users sharing and discussing the information".

The concept of echo chambers, extensively studied in recent academic investigations, has received significant attention for its influence on social networks. Two literature reviews on the topic, Terren and Borge (2021) and Ross Arguedas et al. (2022), play a crucial role in this paper. Terren and Borge offers an overview of the literature debate, analyzing the differences obtained between studies depending on the data collected or the platform under study. Ross Arguedas et al. focuses on defining concepts such as echo chambers, filter bubble and polarization. Echo chambers are often described as the situation in which users mostly communicate with and consume content from like-minded others (Terren and Borge 2021). The phenomenon of echo chambers is linked to key concepts, especially homophily, the human tendency to interact and associate with similar others (McPherson et al. 2001). There are also other similar and related concepts, such as the filter bubble, which

place more emphasis on the algorithmic selection of content (Ross Arguedas et al. 2022; Thorson et al. 2019). In order to simplify the concepts discussed in this article, the primary focus will be on echo chambers, with special attention on interactions between users and communities, as highlighted by some authors (Barberá et al. 2015).

Research on echo chambers remains a subject of extensive and ongoing debate, with discussions focusing on their impact and even their existence. This research is centered on examining users and community interactions through digital trace data, and this are the investigations that we will center in our literature review and discussion. Barberá et al. (2015), a highly cited work in this field, presented varied findings concerning echo chambers in social media. They identified strong echo chambers in discussions on political topics, where users predominantly engaged with those who shared similar ideological preferences. Two distinct social network analyses centered on climate change (Williams et al. 2015) and gun control (Merry 2015) demonstrated that a significant portion of Twitter users primarily engaged with individuals who held similar views, avoiding direct confrontation with their ideological opponents. Additionally, Tsai and Chuan (2020) observed that the echo chamber effect was particularly pronounced in the retweet and mention network, while the reply network exhibited a higher occurrence of intergroup communication.

Moreover, in the Spanish context, several specific studies have investigated into this topic. Guerrero-Solé (2017) and Esteve Del Valle and Borge Bravo (2018) have both emphasized the occurrence of highly polarized interactions within echo chambers. Esteve Del Valle and Borge Bravo specifically noted that the follower and retweet networks tend to display higher political homophily. Also, Balcells and Padró-Solanet (2020) conducted a qualitative analysis focused on cross-conversations between polarized communities, with a particular emphasis on examining tweet replies. The results of Balcells et al. showed that users interact with people holding opposing views on the debate on Catalonia Independence. In addition, Aragón et al. (2013) examined the usage of Twitter by major Spanish political parties and their respective communities during the national election in 2011. Their findings supported the concept of a "balkanization" of the online political landscape in Spain, determined through an exploration of diffusion and conversational patterns. Consistent with other studies, this research observed a limited occurrence of retweets between members of different political communities but identified more interactions between communities when analyzing replies. Additionally, the article emphasizes differences in emotional content depending on political community, noting that the most popular party typically adopts a more positive tone in its messages.

The definition of social bots has evolved over time within the academic discourse, reflecting the dynamic nature of

these entities in online environments. Social bots are accounts controlled by software, algorithmically generating content and establishing interactions (Ferrara et al. 2016; Varol et al. 2017), designed to simulate human behavior on social media platforms. This paper understands social bots (Yang et al. 2019) as a broad bot classification that includes from totally automated bots to accounts with less automation but with highly targeted activity, for example where one person or organization runs hundreds of accounts at the same time. The primary goal of social bots is often to influence public opinion, amplify certain narratives, or manipulate online discussions, while artificially boosting the popularity of certain users or political figures.

In relation to social bot investigations, the literature is prolific (Yang et al. 2024). Researchers have explored various aspects of social bots, examining their detection methods, behavioral patterns, and the implications of their presence in online ecosystems. One prominent tool employed in the study of social bots is Botometer (Davis et al. 2016), a platform developed to assess the likelihood of a Twitter account being automated. In a study from 2017, researchers estimated that approximately 9–15% of active Twitter accounts were automated entities using this tool (Varol et al. 2017). Also, other studies deal with the social bot characterization and analysis, focusing on certain political events across a wide range of countries: Ferrara et al. (2020), Bello and Heckel (2019), Beskow and Carley (2018), Brachten et al. (2017), Keller and Klinger (2019), among many others. Lastly, underscore the multitude of studies that examined the presence of social bots during the COVID-19 pandemic, including Zhang et al. (2022) and Chang and Ferrara (2022). Zhang et al. noted a substantial increase in the creation of social bots during the pandemic period. Moreover, certain investigations establish a connection between bot behavior and astroturfing campaigns, noting that bots often engage in extensive retweeting to amplify specific messages (Keller et al. 2020) or attempt to artificially influence Trending Topics (Elmas et al. 2021).

In the Spanish context, studies examining social bots have also contributed to the broader understanding of their quantification, prevalence and impact. The relation between bot behavior and astroturfing campaigns, has been studied with a national focus on the studies García-Orosa (2021), García-Orosa et al. (2021), Arce-Garcia et al. (2022) and Arce-Garcia et al. (2023). Stella et al. (2018) analyzed social bots during the 2017 Catalan referendum, providing evidence that social bots targeted mainly human influencers and promoted violent content aimed at Independentists, ultimately exacerbating social conflict online. Kusen and Strembeck (2018) analyze the propagation of misinformation in four specific cases, and one of them is about the independence process in Catalonia in 2017. The analysis is mainly focused on how bots and humans interact, and how this interaction helps message propagation. It uses sentiment analysis to detect tweet emotions, and the results show that extreme emotions have higher propagation. Also, Gonzalez-Bailon et al. (2020) analyzes the centrality of bots during contentious political events, specifically the interactions between bot, human and media users. An additional study centered on the Spanish context is Pastor-Galindo et al. (2020), which identifies the existence of social bots during the 2019 Spanish General election. This investigation reveals that the far-right party VOX had a considerable number of potential social bots compared to other political communities. Lastly, Martinez Torralba et al. (2023) examined the presence of political bots during the COVID-19 crisis in Spain, discovering that approximately 20% of users engaging in the conversation were bots. In terms of sentiment analysis, the study indicates that government parties (PSOE-Podemos) expressed more positive messages than those in opposition (PP-VOX).

3 Research questions

Based on the extensive literature discussed earlier, our study aims to analyze globally the Twitter conversation around the #14F Catalan Elections, and more specifically quantify and evaluate the effects of echo chambers and social bots. Our research questions, outlined in this section, provide a structured approach to examining the theoretical frameworks and empirical findings discussed in the preceding Related Works section. The section provides instances of relevant prior literature that directly relate to our concrete research questions, establishing a framework for later comparison with our own results in the discussion section.

Firstly, a central point of echo chambers in the literature revolves around their existence and their distinctive characteristics. Previous investigations have studied this phenomenon on Twitter, both international context (Barberá et al. 2015; Tsai and Chuan 2020) and the specific context of Spain (Guerrero-Solé 2017; Esteve Del Valle and Borge Bravo 2018; Balcells Padró-Solanet 2020). Some of these studies have specifically analyzed echo chambers within the framework of electoral campaigns (Aragón et al. 2013). Our study is designed to analyze the interactions within and between political communities during our specific case study, allowing us to draw comparisons with these previous investigations.

Q1 Does the conversation around #14F reflect echo chambers for each political party community? How do these echo-chambered communities interact inside and between them?

Additionally, continuing the preceding question, the academic debate on echo chamber effects is linked with the dynamics of polarized political groups (Barberá et al. 2015; Williams et al. 2015; Merry 2015). Our goal is to replicate these analyses and identify parallels in our own research.

Q2 Can the three main political blocks (independentist, constitutionalist and Spanish government parties) be grouped into three greater echo chambers? Are these supra-communities even more isolated?

In relation to Social Bots, the primary goal of this investigation is to identify and measure their presence. While numerous examples can be found in international literature (Varol et al. 2017; Ferrara et al. 2020), this study will adopt a similar approach to Pastor-Galindo et al. (2020). The characterization and detection of bots actively participating in the conversation will be categorized based on their association with each political community identified in Q1. Furthermore, the obtained results will be compared with those from prior Spanish investigations such as Pastor-Galindo et al. (2020) and Martinez Torralba et al. (2023).

Q3 How extensively were social bots utilized in the digital #14F electoral campaign? When associating these bots with political communities, is there an equitable distribution among them, or does any particular community exhibit a higher prevalence of bots?

In addition to their mere presence, it is essential to explore the interactions of social bots, particularly with human users (Ferrara et al. 2016; Varol et al. 2017; Gonzalez-Bailon et al. 2020). Specifically, certain researchers have associated the predominant retweeting bot behavior with astroturfing campaigns on an international scope (Keller et al. 2020; Elmas et al. 2021) and within the Spanish context (Arce-Garcia et al. 2022; Arce-Garcia et al. 2023). Previous research has also investigated the nature of the messages propagated by these bots (Stella et al. 2018; Kusen and Strembeck 2018).

Q4 In what ways do social bots interact with human users, considering interaction types such as retweets, replies, or quotes? Are there noticeable distinctions in the content and sentiment analysis between social bots and human users?

4 Methodology

In this section, the methodological aspects are presented, focusing on certain decisions that were made to collect and analyze the data of this study. Also, an introductory analysis is presented to reinforce the methodological choices made and to assure the validity of our approach. We have obtained a substantial amount of findings from our study, leading us to highlight the most significant ones within the main text. Additionally, we have included some secondary results, considered less pivotal but still relevant, in Appendix A.

4.1 Data collection

Data collection has been performed through Twitter Streaming API (API v1.1) 2023, using "Tweepy" Python library (Roesslein 2020). The data was captured in real-time from January 1st 2021 to February 21st 2021 (one week after Election Day). The tweets collected contained specific hashtags or keywords, a complete list of which can be found in Appendix B. For analysis purposes, the total dataset has also been divided into 3 sub-periods (Table 2):

The target list of hashtags and keywords was manually updated through all the collection period (following the strategy of in Abilov et al. 2021), especially considering all the hashtags used from the contender political parties, which were selected from the parties list with possibilities of obtaining parliamentary representation on the last "Baròmetre del CEO" before the elections (Centre d'Estudis d'Opinió 2020): Cs, JxCat, ERC, PSC, Comuns, PP, CUP and VOX. Common to all parties was the use of a single hashtag (or group of hashtags) related to their campaign slogan. The exception to that point are Cs and Comuns, which sometimes used daily hashtags for specific topics, but these cases were not considered in this analysis.

Also, other hashtags related directly to the electoral period were summed (#14F, 14F, #eleccions14F, etc.), as well as the ones correspondent to relevant events (i.e., TV debates), or other hashtags that achieved notoriety and were related to the elections (for example #pucherazo or #RompeTuVoto, which were used to criticize the democratic system). The criteria used to add these hashtags to the collection list is that they were added only if the hashtag achieved

Table 2 Tweet distribution over datasets

Dataset	Extraction period	Tweets total	Number of users
Pre-campaign	Jan. 1st to Jan. 28th	460.447 (17.23%)	84.845 (33.66%)
Campaign	Jan. 29th to Feb. 14th	2.069.633 (77.45%)	206.769 (82.04%)
Post-campaign	Feb. 15th to Feb. 21st	142.051 (5.32%)	60.969 (24.19%)
Total	Jan. 1st to Feb. 21st	2.672.131 (100%)	252.034 (100%)

enough notoriety to become Trending Topic in Barcelona or Spain.

Table 19 shows the tweet distribution by type, where it can be observed that most of the tweets by type are retweets (around 80–85% in every dataset), followed by the original tweet type (around 5–10% in every dataset). Quotes and replies represent a small percentage of all tweets collected. There are slight variations in the percentages over the different collection periods, but these are not significant. Also, Fig. 1 represents the temporal analysis over the collection period, grouping the number of tweets collected each day. The graph show that the number of daily tweets was increasing until the Election Day, with peaks that correlate with relevant dates.

4.1.1 RT (ReTweet) coverage

As seen in multiple literature examples (Morstatter et al. 2013; Abilov et al. 2021), Twitter streaming API only allows capturing a sample of all the tweets under the target hashtags or keywords. To evaluate the completeness of the data captured on the Twitter API stream compared to the real tweet amount, a test has been performed to quantify the percentage of tweets covered by our datasets.

Following the process implemented in Abilov et al. (2021), a retweet coverage percentage has been calculated. Every tweet captured has metadata with information related to the tweet, and for this parameter calculation the value of "RT counter" takes special importance in the metadata, a value that indicates the number of previous retweets that a certain tweet had when it was captured. The RT coverage value that we calculate is the result of comparing the count of retweet type tweets captured of a certain tweet, with the maximum value in the RT counter metadata on the last retweet captured. The results (Table 3) show an impressive result on this evaluation, with a value higher than usually

Table 3 RT Coverage table

Dataset	RTs captured	Max RT value captured	RT coverage %
Campaign	1.730.082	1.941.093	89.13%
Total	2.231.744	2.447.108	91.19%

expected working with the Twitter Streaming API (around 60–80%), which means that a considerably high percentage of the tweets under investigation have been captured, and the losses due to Twitter Streaming API functioning have been low. This could probably be related to the fact that the number of tweets captured is considerably lower than in other literature datasets taken as a reference.

4.2 Ethics and limitations

When studying public data from social media, it is important to keep in mind that it contains personal information that could be linked to an individual identity in many cases. We followed Twitter guidelines on responsible data and platform use (Twitter Developer Policy and Terms 2023), and all the data treatment processes were validated by our University Ethics Committee.

As a summary, the investigation anonymized the user ID and tweet ID on every tweet captured, using an algorithm that randomized the identification to a new Globally Unique Identifier (GUID). After the community detection analysis, where the username is used to search for correspondences between the detected communities and the political parties, the username has also been anonymized. For this labeling, only the usernames of public figures and political parties have been used. In addition, in many analysis the data is used in an aggregated way.
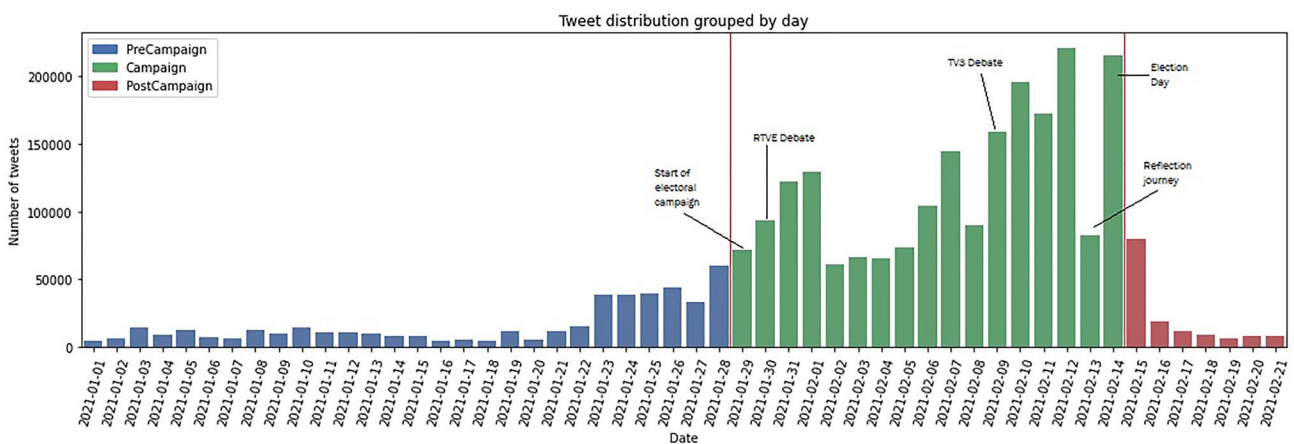
Fig. 1 Tweet distribution grouped by day over collection period

The code used on all this investigation will be uploaded to Github, in a necessary effort of transparency and replicability with other datasets. The URL of the Github repository will be included in the camera ready version of the paper.

Recognizing the limitations inherent in the methodology applied in this study is vital, particularly concerning the tools and methods employed. Firstly, the scope of data collection was confined to a single social network and a specific set of hashtags/keywords, potentially resulting in an incomplete representation of interactions and discussions across the broader Twitter platform.

Secondly, specific methodological decisions made during the study could introduce biases and influence the obtained outcomes, necessitating additional experiments to fortify the reliability of our results. Thirdly, while Botometer is widely used, it is not without potential biases or errors that might impact its functionality. Finally, the analysis of text sentiment faced constraints associated with the inherent limitations of such analyses when applied to Latin languages.

4.3 Community analysis

Once the complete tweet dataset was collected, a community analysis has been performed, using different tools. First, the tweet dataset was transformed to a directed graph using "NetworkX" Python library (Hagberg et al. 2008), containing all the relations between users involved. As we are searching for community affinities to build the directed graph, retweets, quotes, and replies were considered (not mentions inside tweet text, which can be used both to criticize or to praise).

We define $G = (V, E)$ as a directed weighted graph, with V being the set of Twitter users and E the set of directed edges representing interactions between those users through tweets (retweets, quotes and replies). Edges e in E have a weight, which represents the number of repetitions the relation has taken place. This type of graph is widely used on the literature (Stella et al. 2019; Pastor-Galindo et al. 2020), and Table 4 shows a summary of its general metrics, where "average degree" represents the average number of edges that go in/out on graph nodes (considering the weight of edges):

After the graph building process, community detection operations were performed on Gephi software (Bastian et al.

Table 4 Network directed graph metrics

Dataset	Nodes |V|	Edges |E|	Average degree
Pre-campaign	86.704	286.700	6.61
Campaign	210.844	1.024.424	9.71
Post-campaign	64.221	107.876	3.35
Total	257.653	1.301.051	10.10

2009). The data processing consisted in a k-core filter (where $k = 1$) that was applied to facilitate operations and eliminate isolated nodes and, after that, the Louvain method community partition algorithm (Blondel et al. 2008) was executed, splitting groups of users with strong connections between them (the weight of each edge indicates the strength of that relationship). These partitions are not directly related to a specific political community a priori, so the method applied to rename these communities is a manual inspection of the top in-degree users on each of them, considering that the users with more incoming relations inside their community are the users who flag it. As will be presented in the results section, the accounts with higher in-degree usually match the official party account, the party candidate or other important party figures, which simplifies the categorization of each community.

Regarding graph visualization operations, the ForceAtlas2 algorithm (Jacomy et al. 2014) was executed to place spatially each community and their relations of closeness, and each community has been colored according to the color the party uses in their iconography.

4.4 Bot detection

For the bot detection process, the widely accepted Botometer API tool has been used (Davis et al. 2016). This tool analyzes more than 1.000 user account features such as friends and interaction network, metadata, content and sentiment, among others; to provide an associated user bot score in the range [0, 1]. Some users did not have a botscore due to several reasons: the user has a private profile, the user does not have tweets in the timeline, or the user has been banned or eliminated by Twitter. In those cases, these users are not considered for the bot analysis. In other words, Botometer "only" gives a score, but does not return a label indicating whether it is a bot or not.

Once we obtained the score associated with each user involved in the election Twitter conversation, each user was tagged in one of these three categories: Human, Bot or Unclear. To categorize the users, the method presented by Pastor-Galindo et al. (2020) has been used. Considering the distribution of all the botscores on the user dataset, a user will be considered as a bot when its botscore is on the percentile 95 of the sample, which for the Total dataset is 0.78 (0.76 for Campaign dataset). On the other hand, each user with a botscore below percentile 75 (0.32 in both datasets) will be categorized as a human. Finally, all the users between these two values have been labeled as unclear, by a criterion of prudence. Table 5 presents the results obtained.

As an example, a botscore distribution plot is presented for the total dataset. As can be noticed, most users have botscore punctuations close to 0 (probably human), so the percentile 95 threshold seems a conservative limit to ensure

Table 5 Bot detection initial results

Dataset	Total users	#Users (after cleaning)	Human	Bot	Unclear	Limits
Pre-campaign	85.573	83.237	62.808	4.487	15.942	[0.32–0.74]
Campaign	208.226	202.717	152.839	10.823	39.055	[0.32–0.76]
Post-campaign	63.312	61.531	46.394	3.189	11.948	[0.4–0.83]
Total	253.716	245.983	184.955	12.676	48.352	[0.32–0.78]

that users inside it will have high possibilities of being social bots (Fig. 2).

4.5 Text analyisis

To complement the community and bot analysis, a text analysis is performed to search for similarities and differences on the words used by communities or user types. Also, with the sentiment analysis we evaluate the words used in each tweet to give a result of a negative, positive or neutral emotion associated to the tweet text. So the text analysis performed can be separated in two main methodological lines: word analysis and sentiment analysis.

For the word analysis, different texts have been retrieved depending on the unit under analysis. These units are the user bio (where users write a short description of themselves) and the text of the tweets. The methodology behind each case is very similar: once the text is retrieved, the text string is divided for each word and all the non-alphabetical characters are filtered (as well as other elements like web links). Then using the "NLTK" Python library (Bird et al. 2009) another filter is performed on the stopwords for the

English, Spanish and Catalan corpus. Table 20 presents the general overview of the words inside the Total dataset, which will be extended in the results section with an analysis by communities.

The sentiment analysis conducted employed the Polyglot toolkit (Chen and Skiena 2014), selected for its ability to perform analyses in both Spanish and Catalan. We acknowledge that the reliability of lexicons in Latin languages, including Spanish and Catalan, is generally considered to be less robust than those in English. Despite the existence of alternative lexicons, we opted for Polyglot due to its ease of use compared to other available options. Significantly, over 90% of the captured tweets were identified as being in either Spanish or Catalan, highlighting the effectiveness of the Polyglot toolkit in handling these languages.

To process the text of the tweets and ensure the maximum accuracy on the analysis, two important decisions were made to filter the Campaign dataset (2.069.633 tweets):

– Only the tweets in Spanish or Catalan were considered (1.887.122 tweets, 91.18% of Campaign dataset tweets), to simplify the complexity of the analysis and taking into

Fig. 2 Users Botscore values distribution on the Total dataset
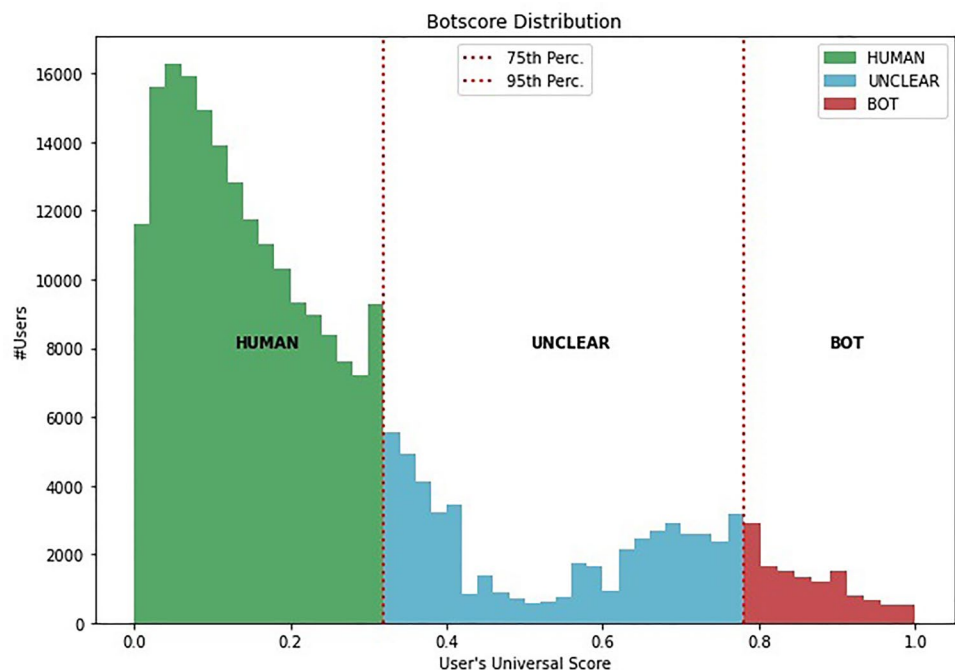
consideration the high percentage that these languages represent.

– Polyglot has also a language detection feature, so every tweet was filtered to ensure the coincidence between the Polyglot detected language and the language 'lang' field on the tweet metadata. The Sentiment Analysis Dataset finally contained 1.589.860 tweets, 76.82% of Campaign dataset tweets.

The analysis with Polyglot toolkit returns a value for each word on the tweet, being + 1 for positive sentiment words, −1 for negative sentiment words, and 0 for neutral sentiment words. After classifying each word with Polyglot, an average calculation is made programmatically in our code for the set of all words on the tweet, whose result is evaluated to determine the tweet sentiment tag: if is minor than zero (NEGATIVE), greater than zero (POSITIVE), or equals zero (NEUTRAL).

5 Results

This section presents the specific results obtained after applying the methodology and operations presented in the previous section. First, global data exploration is offered, to better understand the datasets we work with. Subsequently, community analysis is presented, which focus on the political communities detected and the interactions inside and between them; then, bot analysis, with a presentation of social bot presence in each community and the interactions between them and human users; and finally text analysis is shared, with a global study of the words used prominently in each community, in conjunction with the sentiment analysis in each community, as well as per user type (bot/human).

As an important general statement, and with the main and only objective of highlighting the relevant information,

the results presented will be centered on the "Campaign" dataset (which represents the 77.45% of the Total tweets captured and is the period with higher activity and events). Even though the experiments have been performed for all the datasets, the results do not provide relevant or different contributions than the Campaign dataset.

To start the result analysis, some general aspects will be presented as a general data exploration. This information will be an important base to compare on the further analysis that looks into each community or focuses on bot/human accounts. First, the number of accounts by creation year is presented. The global result seems reasonable, with a high number of accounts created in 2010–2012 (the years where Twitter became mainstream in Spain), and then a slight rebound in 2017 (a politically turbulent year, especially in Catalonia), and finally a relevant quantity in the recent years 2020 and 2021 (considering that 2021 only reached up to February 21st) (Fig. 3).

Also, it is interesting to study the "numberplate" accounts type more concretely, accounts where their username is a name followed by 8 numbers. This is the default user screen name that Twitter assigns, and during the data collection in 2021, it was formerly linked with bot and troll farms, although this connection is no longer prominent because they are easily detectable. The results obtained show an accumulation of this type of account in recent years, a casuistic that will be studied in the next sections (Fig. 4).

5.1 Community analysis

The second step in the analysis performed on this study is a political community analysis. In this section, the results are presented, obtained following the methodology mentioned in Sect. 4.3. Starting with a global community analysis, then some relevant statistics will be presented separating the data for each community.
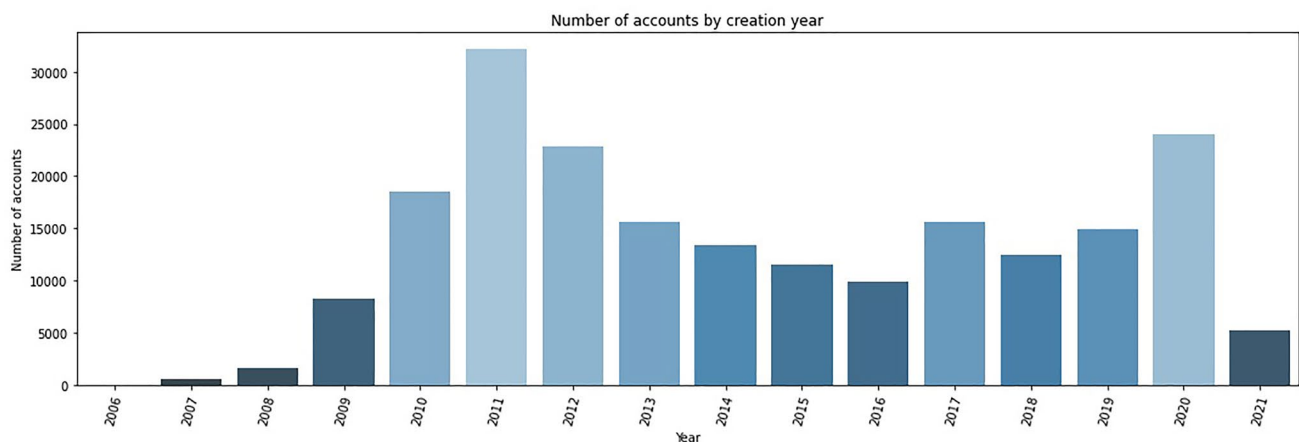
Fig. 3 Number of accounts by creation year (Campaign dataset, 206.769 different users)
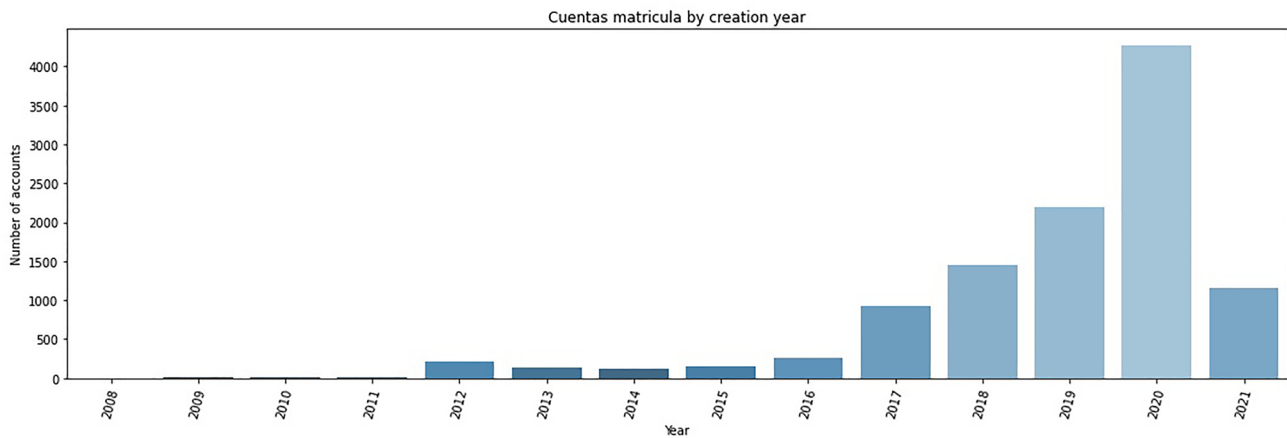
Fig. 4 Number of "numberplate" accounts by creation year (Campaign dataset, 10.886 different users)

5.1.1 Community network graph

As presented in Fig. 5, the "Campaign" graph perfectly shows each community of the 8 political parties that obtained representation in the #14F election.

Tables 6 and 21 present the number of users inside each political community detected, and the top 3 most relevant users in each community that allowed us to tag them without contradictions. As we can see in Table 21, each

community can be related to its analogous main accounts (official party accounts, candidates, significant political figures, etc.).

Another important point is that if we look at the number of users in each political community, we start noticing surprising results. The size of the communities is not correlated with the electoral result, and therefore should be a matter to be considered in the future analysis of the results in this study.

Fig. 5 Graph interaction network of "Campaign" Dataset
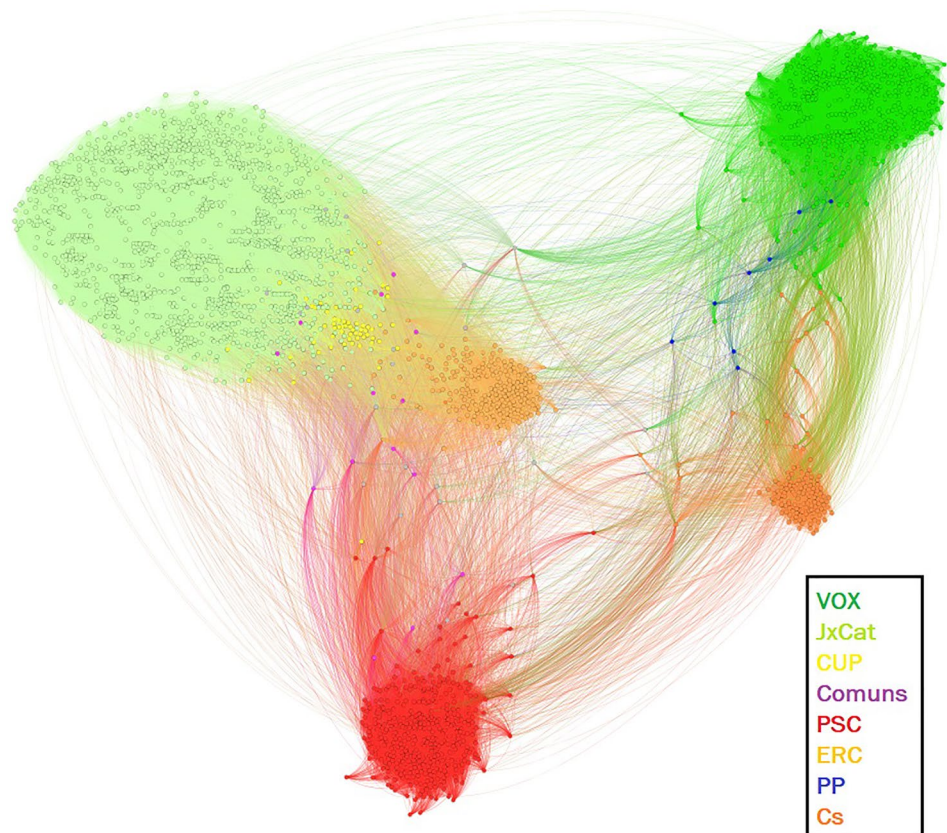
Table 6 Number of users in each political community

Community	# Users
VOX	44.690
JxCat	32.714
CUP	24.935
Comuns	17.787
PSC	12.916
ERC	9.113
PP	8.585
Cs	6.614

Also, focusing on the interactions of the users in each community (Table 22), it is possible to differentiate active interactions (a user that retweets, replies or quotes) and passive interactions (a user is retweeted, replied, or quoted). Additionally, these active and passive interactions can be categorized by communities. Figure 6 represents all the active and passive interactions between all the communities, and Table 23 presents the 15 most numerous relations

between an active user community and their respective passive user community, along with the count and the percentage that these interactions represent on the active interactions of the community. As expected, the most common relations are inside each community:

Also, when analyzing this data, the next step was the grouping of the different parties with their more similar political space considering their closeness in real life:

o   Independentist parties: JxCat, ERC and CUP
o   Constitutionalist parties: VOX, Cs and PP
o   Spanish Government parties: PSC and Comuns

The results obtained after this grouping are presented in Table 7 and Table 24:

In addition, this section presents a complementary data exploration focusing on the differences between each community. It is possible to observe in Table 25 and Table 26 that for most cases the results have similar percentages, with the only exception of VOX, which represents an outlier in

Fig. 6 Heatmap of Communities Active/Passive interactions
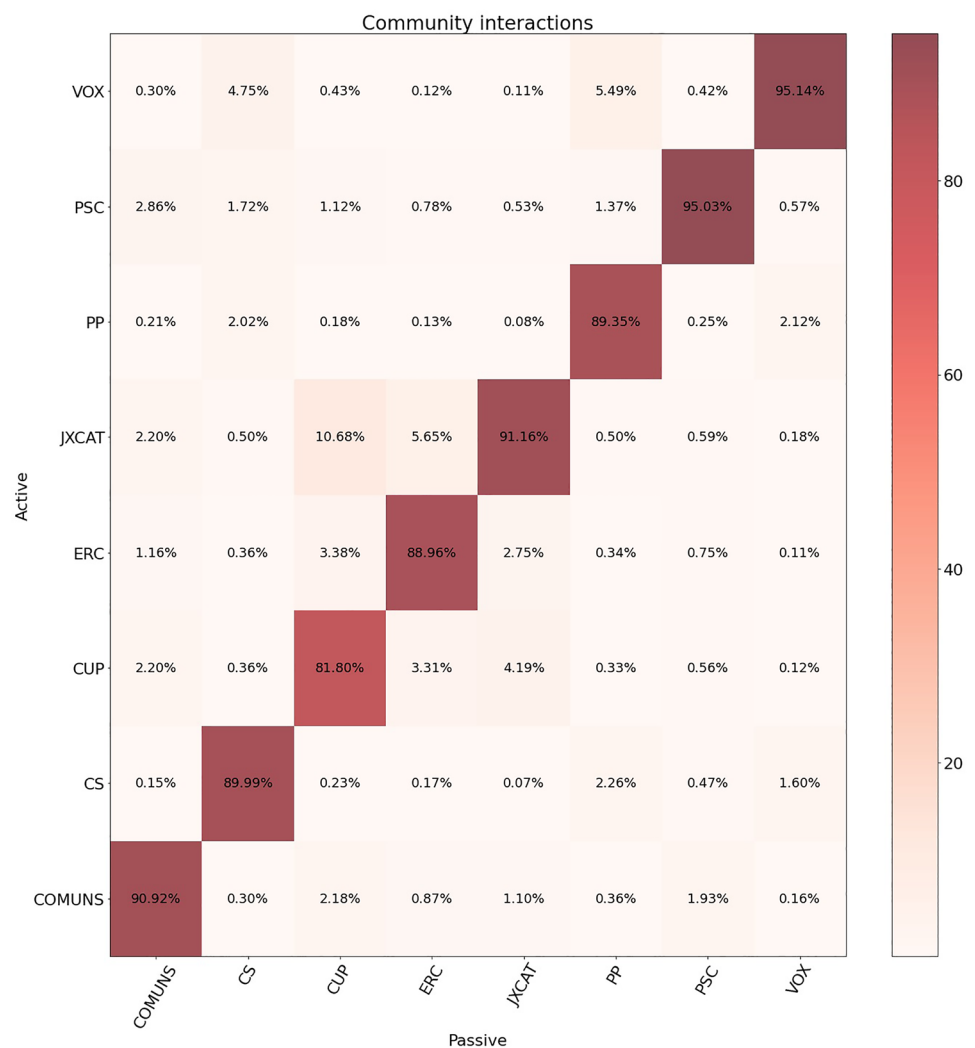
Table 7 Number of interactions between community blocks

Active user community	Passive user community	Count	Percentage in active interactions
Independentist	Independentist	695.137	96.13%
Constitutionalist	Constitutionalist	596.748	97.26%
Spanish Gov	Spanish Gov	366.309	94.60%
Independentist	Spanish Gov	13.723	1.90%
Spanish Gov	Independentist	12.417	3.21%
Constitutionalist	Spanish Gov	6.742	1.10%
Constitutionalist	Independentist	4.054	0.66%
Spanish Gov	Constitutionalist	3.686	0.95%
Independentist	Constitutionalist	2.820	0.39%

the percentage of users created in 2020–2021 and in the numberplate accounts percentage. This could be because it is the newest party and is experiencing a growth in followers in recent years or could be an indicator of artificial enlargement of their community using low quality social bots (considered low-quality due to their short lifespan and for their generalization regarding their "numberplate" username).

Some insights can be gained from the network graph and all the data related that has been presented in this section. First, the official accounts of the parties and relevant political figures have an important centrality on each community (especially considering their in-degree). But only 2 out of 8 candidates on the election are on the top 3 of their community users by in-degree (they are even surpassed by Spanish political figures of their party). Also, the size of the communities does not correlate with the results of the elections (VOX is the largest user community, but it was placed in the 4th position). Neither does the number of interactions. Also, JxCat and PSC communities are prolific, the number of interactions is high if it is compared to the number of users. On the other hand, the CUP community has the lowest average interaction rate per user (Table 8).

Secondly, it is possible to spatially place every party, with 3 main blocks that coincide with the closeness of each party:

independent parties (JxCat, CUP and ERC), constitutionalist parties (VOX, PP and Cs) and a third block with the Spanish government parties, PSC and Comuns. We can observe that relations inside the independentist (and constitutionalist) block represent a significant percentage of the total interactions of the users in those communities (Table 7). On the other hand, we can observe that the communities of VOX and PSC are mostly hermetic, and do not interact normally outside their community (Table 23). When grouping political parties' communities by political affinity, the effects of the echo chambers are even higher and more obvious, with almost every interaction inside this "supra" community. Only the Spanish government parties group has a 3% of active interactions with the independentist block, but on the other hand, the relations between independentists as active with the constitutionalist is as low as 0.39%.

5.1.2 Analysis of the reply interactions graph

In this section, we performed a segmented analysis specifically targeting the graph of reply interactions. When a user chooses to reply to a tweet, various motivations may drive this action, ranging from expressing agreement or support for the original tweet to offering a critique

Table 8 Coefficients between votes, and community users and interactions

Political party	Popular vote	Community #users	Coef. votes/users	Community active interactions	Coef. votes/interactions
PSC	652.858	12.916	50.55	244.097	2.67
ERC	603.607	9.113	66.23	136.627	4.41
JxCAT	568.002	32.714	17.36	460.250	1.23
VOX	217.883	44.690	4.87	409.652	0.53
Comuns	194.626	17.787	10.94	143.104	1.36
CUP	189.087	24.935	7.58	126.228	1.50
Cs	157.903	6.614	23.87	103.342	1.53
PP	109.067	8.585	12.70	100.538	1.08

of the presented statement or engaging in a debate with other users. This study does not enter into these specific motivations, instead focusing on how reply interactions occur between users based on their respective communities. It is reasonable to expect that replies would show a higher degree of cross-community interaction, but we now quantify how this happened in our captured datasets.

Figure 7 displays the network graph for reply tweets, which presents similar results compared to the total network graph. All political communities are clearly identifiable, with the exception of the Cs community, which has integrated into the VOX community. Additionally, there is a higher level of connection between communities.

Despite these graph differences, an examination of Fig. 8 reveals results similar to those observed in the complete dataset. Intra-community interactions are lower than in the complete graph, ranging from approximately 77% to 88%, representing a decrease around 5–10% of interactions inside their own community compared to the graph comprising all tweet types.

5.2 Bot analysis

This section evaluates the presence of social bots in the general conversation of the elections on Twitter. As stated in the methodology section, the users were tagged according to their Botometer score, and now some metrics referring to their interaction between humans and bots or their prevalence in each community are presented.

5.2.1 Bot analysis by community

First, we analyze the proportion of humans and bots in each community. The results are presented in Table 9, where we can gain the following insights: bot quantity follows the size of each complete community closely, but the high percentage must be emphasized on Cs and Vox compared with the other communities, and the low percentage on CUP community; secondly, regarding human percentage, it is remarkable how low the percentage of humans present in VOX community is, but this is due to a considerable amount of users on the boundary limit between human user and unclear user, so this fact does not give relevant information.

When we specifically analyze the distribution of bots created in 2020 and 2021 by their community, it is possible to discover the information presented in Table 10. It is a piece of interesting information due to the short life span that social bots usually have (Ferrara et al. 2020), and could indicate that some social bots were created specifically for this election. The most important observations that can be extracted from data in Table 10 are two. First, 2020 was a year with an increase in social network use allegedly due to the pandemic. Also, the social bot creation had a golden period in their participation in information operations (Zhang et al. 2022). In this case, we can state that in global numbers, the quantity of VOX social bots created are much ahead of other communities. If we look at the percentage inside their set of bots of each community, Comuns has a high percentage of bots created in 2020, but this percentage by itself does not give relevant information. Second, if we analyze 2021, we must take into consideration that only 52 days of 2021 are under the collection period of analysis, which means that in some cases nearly one account per day

Fig. 7 Graph interaction network filtering by Reply type tweets
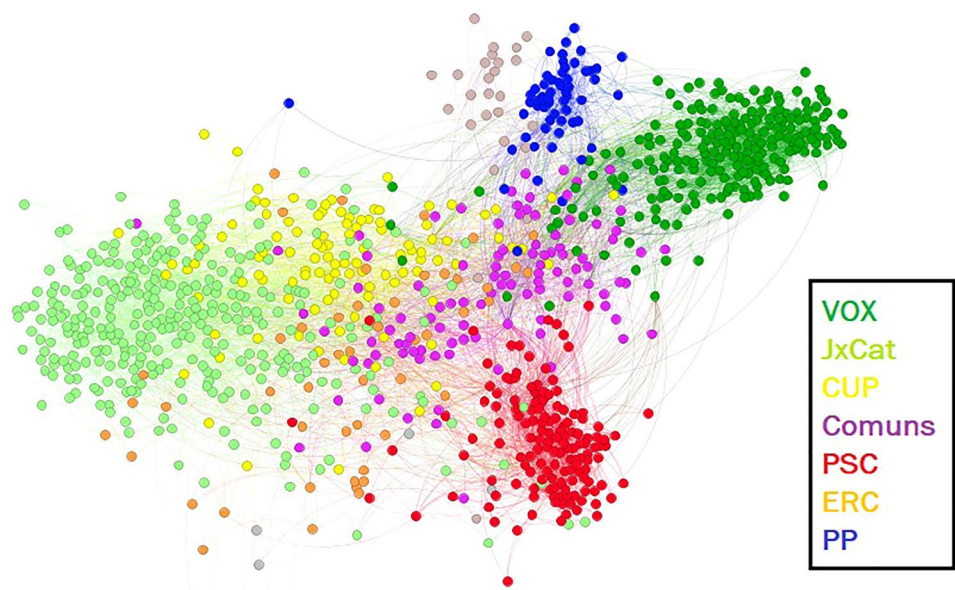
Fig. 8 Heatmap of community interaction filtering by Reply type tweet
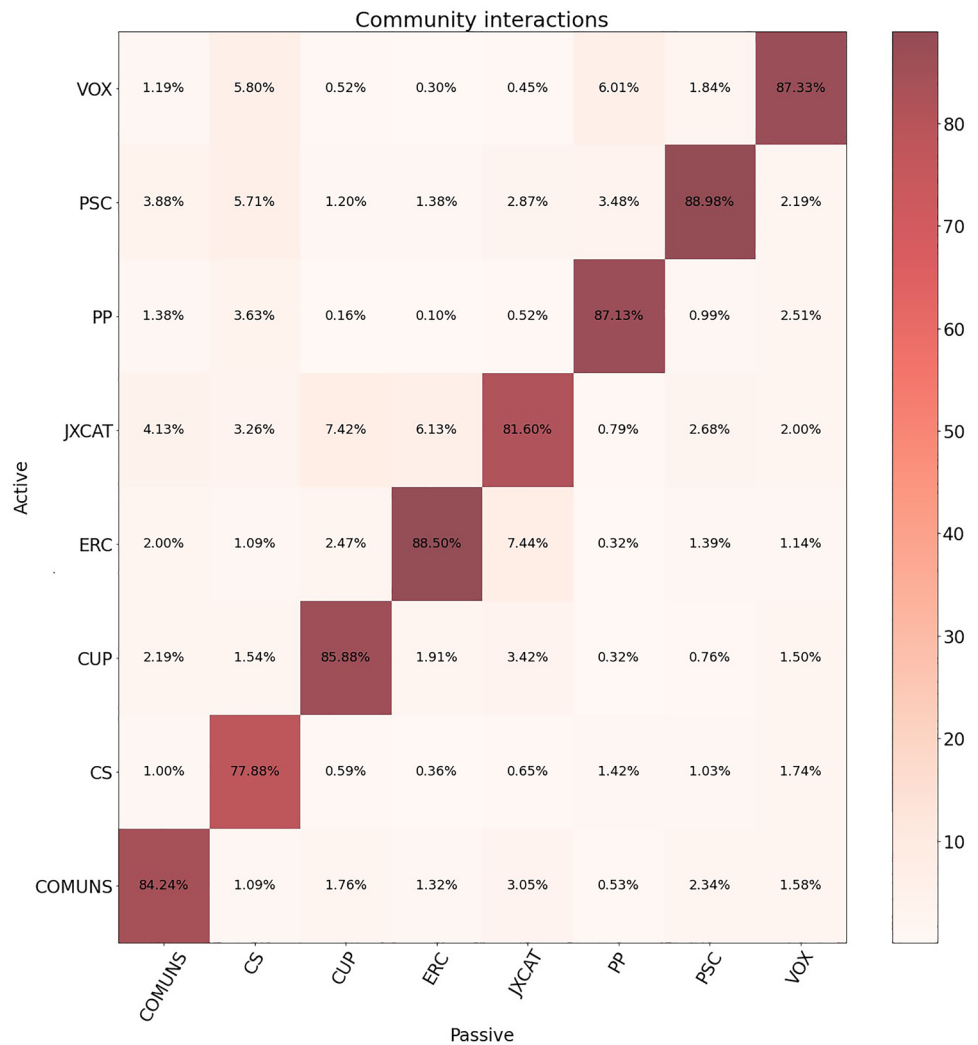
Table 9 Number of bots and human users by community

Community	#Bots	Bot % in community	#Humans	Human % in community
VOX	1.745	4.95%	23.821	67.54%
JxCat	1.102	3.37%	24.835	75.91%
CUP	770	2.57%	20.439	81.97%
Comuns	642	4.32%	12.839	72.18%
PSC	388	3.00%	9.844	76.22%
ERC	308	3.38%	7.079	77.68%
PP	369	4.30%	6.113	71.20%
Cs	435	6.57%	4.770	72.11%

Table 10 Number of bots created in 2020/2021 by community (the percentage represents the fraction of 2020/2021 bots inside the set of bots of each community)

Community	Total bots	Bots created 2020 (%)	Bots created 2021 (%)
VOX	1.745	332 (19.03%)	53 (3.04%)
JxCat	1.102	118 (10.07%)	85 (7.71%)
CUP	770	100 (14.41%)	53 (6.88%)
Comuns	642	137 (21.33%)	48 (7.48%)
PSC	388	71 (18.30%)	39 (10.05%)
ERC	308	44 (14.29%)	30 (9.74%)
PP	369	66 (17.89%)	13 (3.52%)
Cs	435	42 (9.66%)	21 (4.83%)

was created that is considered a bot. In this regard, the 85 social bot accounts on the JxCat community look unnatural. The percentage of bots created in 2021 of PSC and ERC is also high, but as in the previous case, this does not give much information (their communities are small compared to others).

If we center the analysis on the "numberplate" accounts (Table 27), it is possible to detect again a high value on the VOX community, as well as on the PSC and Comuns communities.

5.2.2 Bot-human interactions

When analyzing social bots, one of the most important parts is the analysis of their behavior. In this section we will focus on the interactions activated by bots, especially when they interact with human users, to discover insights into these interactions.

First, we will analyze the global conversation to find out how the human and bot users interact between them and with each other. We can see on Fig. 9 and Table 28 that the results obtained are interesting. Human users present what seems a "human" behavior, they usually interact with themselves and in a very small proportion with bot users, as previously seen in literature (Ferrara et al. 2016 and Varol et al. 2017). This means that bot users do not have relevance in the global conversation and are almost never interacted with by human users. On the other hand, bot users interact half of the time with human users, which can represent an artificial way to increase the impact of those publications. Also, the interaction of bots with other bots is unnaturally high considering the small percentage that they represent.

In addition, in Table 11 we can study the interactions of these users by the type of tweet that they have published.

Table 11 Number and percentage of tweet type by user type

Active user	Tweet type	Count	% inside active user group
Human	RT	882.284	88.73%
Bot	RT	156.382	96.54%
Human	Quote	87.331	8.78%
Human	Reply	24.627	2.47%
Bot	Quote	2.878	1.78%
Bot	Reply	2.712	1.67%

The clearest insight is that the detected bots mostly only retweet, with quotes or replies representing less than 2% of their interactions. Regarding human users, they also retweet in a high percentage, but in their case, they also quote in more than 8% of their interactions, which can be considered a more "humanized" behavior.

Fig. 9 Number of interactions between human and bots users

Bots (Active): 92.507 — Bots (Passive): 46.518 — Humans (Active): 663.880 — Humans (Passive): 709.869
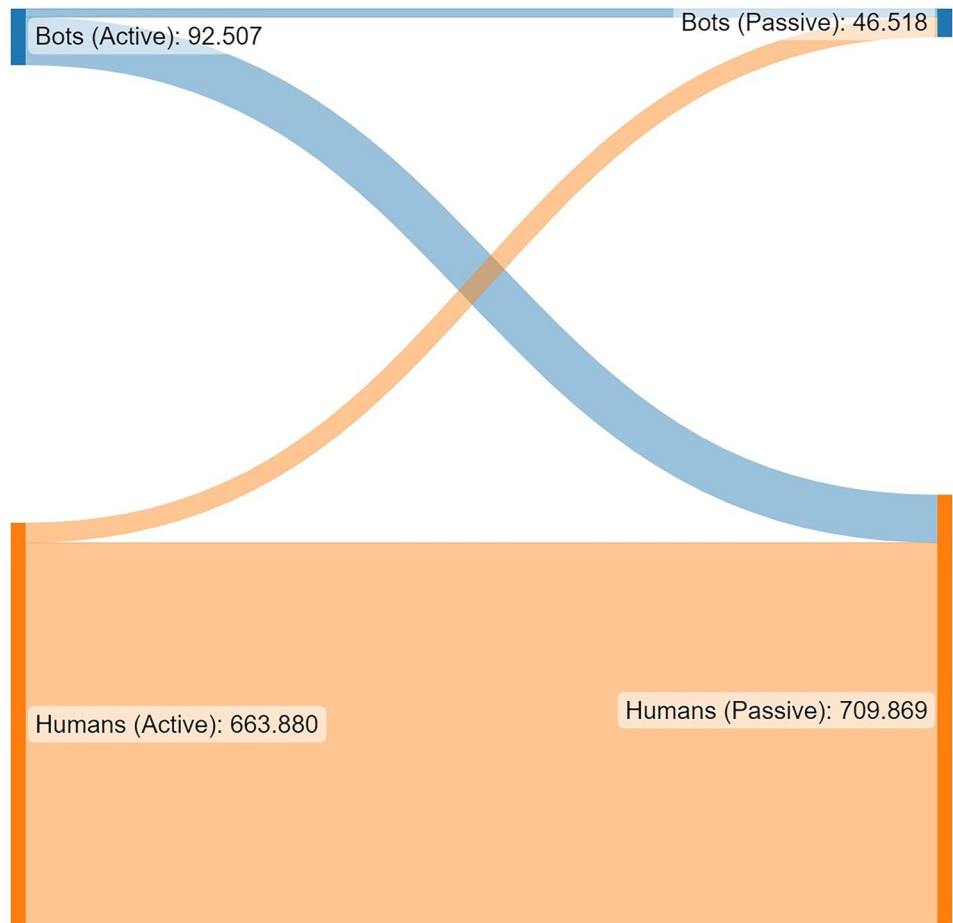
5.2.3 Bot-human interactions by community

In this section, the insights will be particularized by community and type of user to find more detailed information. Important information that can be extracted from Table 12 (which only contains the top 20 relations by interactions count) is that the human–human relations are the most common in every community. The only value that appears slightly different is the percentage inside the VOX active community, which represents only 18% of the total active VOX user relations. Also we can observe human–human relations between different communities (JxCat with ERC and CUP, and vice versa), that fit exactly with the previous section results that showed the strong bonds between those communities.

The bot-human relations also bring interesting results to the table regarding the relations inside each community. We can notice that JxCat has a high number of this type, which also represent nearly a 5% of the total relations of the community. In that case, VOX has high numbers as well but not so elevated as in other aspects (quantity or percentage). On the other hand, we have relevant results regarding Comuns and Cs, in both cases, we can notice a high count of relations compared with their community size (especially in the Cs casuistic). PSC also appears in this table, but with a low percentage that seems irrelevant to this purpose. Regarding the

human-bot relations, as we have seen in previous sections, they represent a small percentage of all interactions, even though the two bigger communities with this type of relation appear on the table. This result does not give any relevant information. Finally, also one case of bot-bot relation appears, which corresponds to the bigger bot community (VOX), and as in the previous case does not give relevant information either.

5.3 Text analysis

Finally, this section analyzes the text used by the users when they tweet or when they present themselves, specifying the community where they belong, and specifying the results for the detected social bot users. Then, an investigation of the text sentiment analysis is presented.

5.3.1 User bio content analysis

In the first place, we are analyzing the user bio texts, a piece of information that each user provides to describe itself. This analysis is performed taking into consideration the community tagging of each user, to notice difference between communities and compare the results to the expectations of their real communities and associated political parties.

Table 12 Number and percentage of interactions between human-bots by community (top 20 by interactions count)

Active user	Passive user	Active community	Passive community	Count	% Inside active community interactions
Human	Human	JxCat	JxCat	161.050	34.99%
Human	Human	VOX	VOX	74.836	18.26%
Human	Human	PSC	PSC	74.524	30.53%
Human	Human	CUP	CUP	52.536	41.62%
Human	Human	ERC	ERC	51.475	37.67%
Human	Human	Comuns	Comuns	38.098	26.62%
Human	Human	Cs	Cs	27.209	26.32%
Human	Human	PP	PP	23.734	23.61%
Bot	Human	JxCat	JxCat	23.290	5.06%
Bot	Human	VOX	VOX	15.924	3.89%
Bot	Human	Cs	Cs	8.906	8.62%
Human	Bot	JxCat	JxCat	8.809	1.91%
Bot	Human	Comuns	Comuns	8.431	5.89%
Human	Human	JxCat	CUP	8.260	1.79%
Human	Human	CUP	JxCat	8.024	5.77%
Human	Bot	VOX	VOX	7.026	1.71%
Bot	Human	PSC	PSC	6.523	2.67%
Bot	Bot	VOX	VOX	5.297	1.29%
Human	Human	JxCat	ERC	4.924	1.06%
Human	Human	ERC	JxCat	3.736	2.73%

The "% inside active community interactions" refers to the percentage of a certain row combination count of user-community and active–passive, i.e. JxCat "Active Community" has a total of 460.250 active interactions, and the 34.99% are between human–human user types and JxCat-JxCat communities

Table 13 presents the top 10 most common words of each community, and the results match the prior expectations that we could have, which at the same time reinforces the suitability of the community partitions. For example, it is possible to notice that the language difference (Spanish vs. Catalan) matches the typology of each party (Independentist vs. all the rest). Also, the top 10 words in each community matches the expected ideology of their party, we can even find specific words in each community that match the prototypical profiling of each community: CUP, Comuns and PSC have "Feminista" in a high position, CUP and JxCat have "Barcelona", and PSC and Comuns have a strong connection with their partners in the Spanish Government ("Podemos", "PSOE", "Madrid"). Finally, a remarkable detail is that there are some words like "Cuenta", "Oficial", "Regidor", "Concejal", "Perfil", "Portavoz", which show the importance that the number of official party accounts have within the community in quantity.

5.3.2 Tweets language by community

Regarding the language analysis of the tweets captured, the next table presents the results separated by community. Only Catalan and Spanish languages are considered, as explained in Sect. 4.5.

Again, the results match expectations as can be seen in Table 29 and its representation on Fig. 10, presenting very clearly the three main political blocks:

– Independentist: JxCat, ERC and CUP have a high predominance of Catalan language.
– Constitutionalist: VOX, PP and Cs have a high predominance of Spanish language, especially in the extreme case of VOX with 99.24% of Spanish tweets.
– Spanish Government parties: PSC and Comuns present an intermediate situation, in line with their policies.

5.3.3 Tweets content analysis

In relation to the tweets content text analysis, the evidence obtained show interesting results that could be related to the campaign strategies of each political party. First, results for the complete set of users are presented in Table 14, and then compared with the set of bot users (Table 15).

Regarding the complete pool of users, it is possible to observe that the language used for each community matches the language expected, as seen in the previous sections. It is also possible to see words related to the campaign slogan, asking for the vote, the usual topics of the party or some of their relevant figures:

– VOX: Libertad, Vamos or Odio (in this last case referred to their victimization strategy)
– JxCat: Junts
– CUP: -
– Comuns: Podemos, Gobierno, Pública
– PSC: Cambio
– ERC: Junqueras (historic party leader)
– PP: Alejandro Fernández (new candidate that needs relevance)
– Cs: Procés (disparaging term commonly used by the party)

In addition to these results, it is necessary to highlight another significant element, the presence of references to other parties or candidates between communities, in what can be considered as cross-community attacks. For example, it is possible to recognize that "VOX" is used in CUP and Comuns communities in considerable portion of their tweets, which probably indicates that their political strategy was centered in criticizing VOX. In parallel, we can notice the same casuistry with "Illa" (referencing PSC candidate Salvador Illa), which has a significant relevance in ERC, PP and Cs tweets, and probably indicates a considerable number of attacks to that candidate. In relation to that, the name

Table 13 Most common words in users bio by community (Bold words with associated insights)

VOX	JxCat	CUP	Comuns	PSC	ERC	PP	Cs
España	Catalunya	Periodista	**Podemos**	Socialista	Catalunya	Popular	Oficial
Español	Catalana	Social	Vida	PSOE	Republicana	Partido	Ciudadanos
Vox	República	Vida	Mundo	Vida	ERC	**Oficial**	Cs
Vida	Català	Barcelona	**Feminista**	Política	**Oficial**	España	**Perfil**
Libertad	Vida	Catalunya	Social	Periodista	Twitter	**Madrid**	Política
Siempre	Independentista	**Feminista**	Siempre	Social	Periodista	Vida	Liberal
Cuenta	País	Política	Política	**Feminista**	Esquerra	PP	Vida
Madrid	Llibertat	Català	Republicano	Mundo	País	Política	**Portavoz**
Mundo	**Barcelona**	Sempre	Periodista	Siempre	Vida	Twitter	**Concejal**
Viva	Lliure	Educació	Antifascista	**Madrid**	**Regidor**	Español	Derecho

Fig. 10 Language used by community after using filtering

Language used by community after user filtering

CATALAN   SPANISH

| | VOX | JxCat | CUP | Comuns | PSC | ERC | PP | Cs |
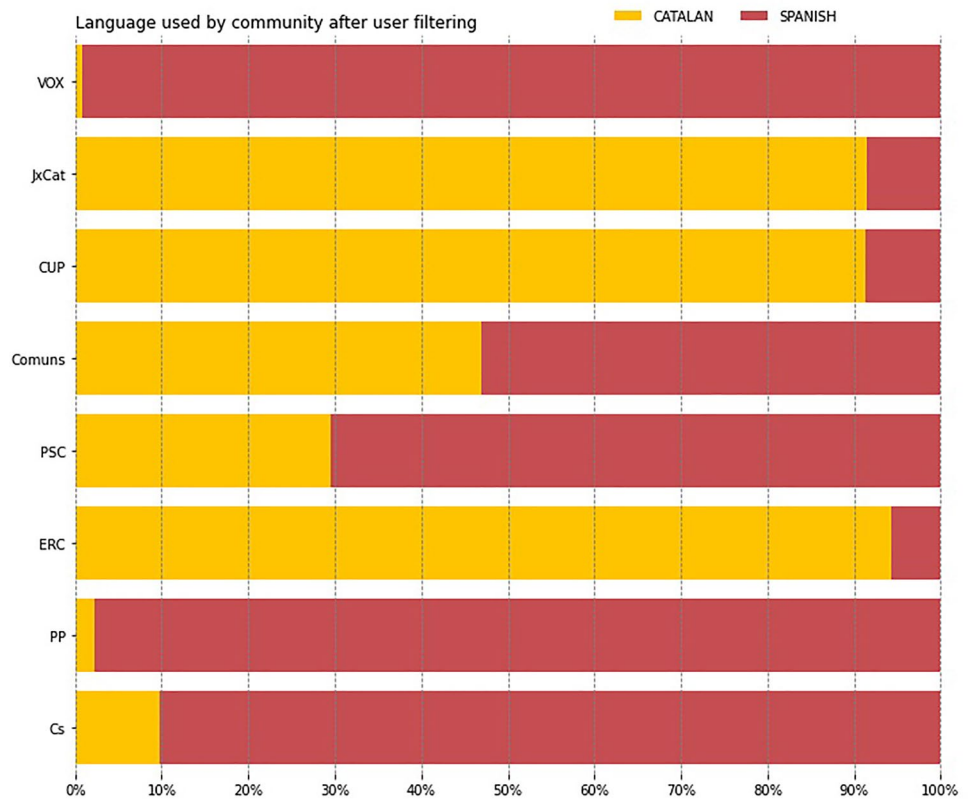
Table 14 Most common words in users tweets by community (Bold words with associated insights)

VOX	JxCat	CUP	Comuns	PSC	ERC	PP	Cs
Cataluña	Votar	Votar	Catalunya	PSOE	**Junqueras**	Cataluña	Cataluña
VOX	Vot	CUP	**Podemos**	Cataluña	Votar	**Illa**	Cs
Hoy	Catalunya	Catalunya	Govern	Catalunya	País	**Alejandro**	Votar
Catalanes	**Junts**	Vot	Votar	Illa	Catalunya	**Fernández**	Voto
Vamos	Eleccions	Electoral	**VOX**	Votar	Avui	Partido	Hoy
Libertad	Campanya	Avui	**Pública**	**Cambio**	Vot	Mejor	**Procés**
Odio	Avui	Eleccions	**Gobierno**	PSC	**Illa**	España	**PSC**
Mañana	**Illa**	Cap	Eleccions	Govern	Campanya	Gobierno	**Illa**
Campaña	Dia	Persones	Campanya	Voto	Cap	Catalanes	Gobierno
España	Cap	**VOX**	Vot	Candidato	**PSC**	Hoy	Vota

Table 15 Most common words in bots tweets by community (Bold words with associated insights)

VOX	JxCat	CUP	Comuns	PSC	ERC	PP	Cs
Cataluña	Votar	CUP	Catalunya	PSOE	**Junqueras**	Cataluña	Cataluña
VOX	Catalunya	Votar	**Podemos**	Cataluña	Votar	**Illa**	Cs
Hoy	Junts	Vot	Govern	Catalunya	Campanya	Partido	**Procés**
Catalanes	Vot	Avui	Votar	Illa	Catalunya	**Alejandro**	Votar
Libertad	Eleccions	Catalunya	Gobierno	Votar	País	Mejor	Hoy
Mañana	Campanya	Eleccions	Avui	**Cambio**	Vot	España	Voto
Odio	**Illa**	Electoral	Vot	Govern	Cap	Gobierno	**PSC**
Vamos	Avui	Cap	Campanya	Voto	Govern	**Fernández**	Gobierno
Campaña	Dia	Campanya	Canvi	Candidato	**PSC**	Votar	**Illa**
España	Cap	Acte	Comú	Gobierno	Avui	Catalanes	Vota

of the party PSC also features prominently in ERC and Cs tweets.

If the focus is among the bot users, we cannot observe substantial differences with the previous analysis, only slight variations. The reason behind this could be that bot users have a high percentage of RT, so they basically amplify the tweet content of human users.

5.3.4 Sentiment analysis

On this section, an analysis of the tweets sentiment is performed. The study will center basically on three important aspects: how was the tweet sentiment for the detected bots, how was the sentiment between the bot-human interactions, and finally how was the sentiment projected by the users in each community.

First, if we look at Table 16, we can observe two important characteristics: there is nearly no difference in the percentage of tweets of each sentiment if we divide by bot/human, and also there are more negative tweets than positive ones:

Also, if we analyze the interactions between types of users in Table 17, it is possible to detect a very similar pattern to the one before: there is not a substantial difference between bots and human users behavior, with a similar percentage of negative tweets when interacting with humans and some minor variations when interacting with bots (the percentages are low in that case, so this is not enough relevant):

Finally, if the corpus is analyzed by community of active user, it is possible to extract some interesting insights present on Table 18: the three independentist parties (ERC, JxCat and CUP) have the higher positive rate (above 38%), a fact that coincides with other studies, where the government ruling parties have a higher positive sentiment rate (Aragón et al. 2013; Martinez Torralba et al. 2023; Chen et al. 2016). On the other hand, the 3 constitutionalist parties (VOX, Cs and PP) have the higher negative rates (more than 50%), which correlates with their aggressive style of opposition to the government.

Table 16 Bot-Human percentage of tweets by sentiment

Active User	Sentiment	% of active user interactions
Bot	Positive	33.38%
	Neutral	15.95%
	Negative	50.67%
Human	Positive	35.39%
	Neutral	15.19%
	Negative	49.42%

Table 17 Bot-Human percentage of interactions by sentiment (percentages does match 100% for every active user type)

Active User	Passive User	Sentiment	% of active user interactions
Bot	Bot	Positive	4.05%
		Neutral	2.12%
		Negative	7.15%
	Human	Positive	30.46%
		Neutral	14.11%
		Negative	42.11%
Human	Bot	Positive	1.58%
		Neutral	0.72%
		Negative	1.58%
	Human	Positive	34.52%
		Neutral	14.82%
		Negative	46.78%

Table 18 Sentiment Analysis by active users community

Active user community	Sentiment	% of active user interactions
Comuns	Positive	35.60%
	Neutral	15.96%
	Negative	48.44%
Cs	Positive	29.76%
	Neutral	14.96%
	Negative	55.28%
CUP	Positive	38.62%
	Neutral	14.55%
	Negative	46.83%
ERC	Positive	43.22%
	Neutral	14.01%
	Negative	42.77%
JxCat	Positive	39.94%
	Neutral	15.19%
	Negative	44.87%
PP	Positive	32.91%
	Neutral	14.50%
	Negative	52.59%
PSC	Positive	33.83%
	Neutral	17.12%
	Negative	49.05%
VOX	Positive	26.85%
	Neutral	15.93%
	Negative	57.22%

6 Discussion

This section discusses the results obtained, and more specifically, the questions and objectives presented in Sect. 3. Also, the results collected are contrasted with those documented in the current academic literature, with a particular emphasis on studies that examine the Catalan and Spanish scenarios.

Q1 Does the conversation around #14F reflect echo chambers for each political party community? How do these echo-chambered communities interact inside and between them?

Derived from the conducted analysis in this study, it can be asserted that echo chambers existed in the interactions around the #14F Catalan Elections. These findings align with other studies on the subject, including Aragón et al. (2013), which affirms the "balkanization" of the digital conversation in a Spanish electoral campaign, as well as studies by Barberá et al. (2015) and Esteve Del Valle and Borge Bravo (2018).

The results obtained in Sect. 5.1.1 indicate that each of the 8 identified communities exhibit more than 80% of interactions occurring within their respective communities. Notably, the VOX community demonstrates the highest proportion, with 94.26% of interactions taking place within its own community. When examining the network of replies (Sect. 5.1.2), the outcomes indicate a higher degree of cross-interaction. However, it is important to emphasize that intra-community interactions remain predominant, consistent with the observations made by Aragón et al. (2013).

Additionally, the analysis of the most frequently used words in tweets by each community (Sect. 5.3.4) reinforces this statement because it is possible to detect that the most used words are related to campaign slogans, asking for the vote, or referring to a relevant political figure of their community. Even though we have detected some words that possibly show inter-community attacks ("Illa", "VOX" and "PSC"), these are used as an internal product to attack the opposition, and not to engage in a real conversation with the opposite communities. This constitutes an innovative analysis that, to the best of our knowledge, has not been previously introduced in the existing literature.

Moreover, Table 8 suggests that the largest communities, which also experienced the highest number of interactions, did not necessarily receive the most votes. One possible explanation could be the echo chambers that exist in social networks, which lead to interactions primarily among individuals with more extreme political positions. However, earlier studies have proposed an alternative interpretation, suggesting that these emerging political parties exhibit greater strength and adaptability in social networks. This adaptability is attributed to their restricted access to traditional mass media channels, which stands in contrast to the traditional political parties (Sampedro 2021).

In conclusion, there is clear evidence of echo chamber presence in the interactions surrounding the Catalan elections on Twitter. These findings align with the knowledge established by previous studies, including Barberá et al. (2015), Aragón et al. (2013), Guerrero-Solé (2017), and Esteve del Valle and Borge Bravo (2018), offering additional insights into the existence of echo chamber effects. Even in scenarios where cross-community interactions were anticipated to be more prevalent, such as reply interactions, the quantity and strength of intra-community interactions remained significantly higher than inter-community interactions. This pattern is consistent with the results reported in the study by Aragón et al. (2013), which also focused on an electoral campaign in the Spanish context. Thus, it can be concluded that the majority of user interactions occur with individuals who share similar ideological views.

Q2 Can the three main political blocks (independentist, constitutionalist and Spanish government parties) be grouped into three greater echo chambers? Are these supra-communities even more isolated?

Absolutely, the results of Sect. 5.1.1 show that echo chambers are even more isolated when we group these 3 political blocks, obtaining an astonishing result of more than 94% of interactions inside each block, which basically means that there is almost no conversation outside these bubbles. These results are consistent with research on echo chambers that explore interactions between polarized political groups on social media, as illustrated by Barberá et al. (2015), Williams et al. (2015), and Merry (2015). Additionally, they support the presence of echo chambers in the Spanish context, as indicated by Guerrero-Solé (2017) and Del Valle and Borge Bravo (2018).

However, these results diverge from the study conducted by Balcells Padró-Solanet (2020), where pro and anti-independence blocks exhibited interaction. We attribute the discrepancies in findings between our study and that of Balcells Padró-Solanet (2020) to methodological differences. Specifically, their study employed a qualitative analysis of the entire replies network for a set of tweets, while our approach involved capturing replies associated with specific hashtags or keywords, which may not cover the full range of replies involved. In our study, the highest inter-block community interaction is the "Spanish government" parties block which has around 3% of active interactions with the independentist block (mainly due to Comuns interactions with them). On the other side, we can observe the deep isolation between

independentist and constitutionalist blocks, with only 0.39% and 0.66% of interactions respectively.

This analysis is reinforced by the sentiment analysis in Sect. 5.3.4, validating the findings from Aragón et al. (2013) and Martinez Torralba et al. (2023) regarding the variations in emotional content depending on the political community. Specifically, our study identified that the independentist block (Catalan government) had more than 38% of positive sentiment tweets, while the opposition (constitutionalist block) exhibited over 50% negative sentiment tweets. This might be associated with the well-known fact that the opposition employed an aggressive campaign style against the independentist.

Q3 How extensively were social bots utilized in the digital #14F electoral campaign? When associating these bots with political communities, is there an equitable distribution among them, or does any particular community exhibit a higher prevalence of bots?

The findings from this investigation confirm the presence of social bots employed in the digital #14F electoral campaign on Twitter. Previous academic investigations by Varol et al. (2017) indicated that between 9 and 15% of active Twitter users were bots. Additionally, Martinez Torralba et al. (2023) reported that during the Covid-19 crisis in Spain, 19% of users engaged in the Twitter conversation were identified as bots. Following the methodology outlined in Pastor-Galindo et al. (2020), our study adopted conservative statistical parameters when utilizing the Botometer classification results. In alignment with the findings of that study, it is highly probable that the actual number and impact of social bots are considerably higher than presented in this study. The reasoning behind the conservative classification is the prioritization of analyzing the distribution by communities over obtaining a global number with less reliability. Therefore, it is crucial to place special emphasis not only on the bot count (and percentage) but also to analyze the number of human users (and their percentage).

In the examination of bot analysis results by communities, an uneven distribution becomes apparent, particularly highlighted by an increased presence within the VOX political community. This community not only exhibited the highest absolute number of bots (second-highest in percentage) but also displayed the lowest percentage of human users by a considerable margin. These findings are consistent with the outcomes presented in Pastor-Galindo et al. (2020), despite the fact that their study was focused on a national electoral campaign. Additionally, alternative studies, like Ferrara et al. (2020), observed a notably higher proportion of bots in right-wing political communities. On the contrary, Martinez-Torralba et al. (2023) reported in their research that government parties (in their case left-wing) exhibited a

greater prevalence of bot activity, a situation that does not coincide with our results.

As previously mentioned, in addition to examining the distribution of social bots across political communities, we conducted supplementary analyses on the bot distribution, to reinforce and increase robustness of the results obtained. Following the insights provided by Ferrara et al. (2020), which highlighted the typically short lifespan of social bots, we specifically focused on the analysis of bots created in the recent previous years (2020 and 2021), considering the pandemic period impact on bot creation (Zhang et al. 2022). Moreover, given that "numberplate" accounts were previously considered potential indicators of automated accounts linked to bot or troll farms, we conducted an analysis to identify such accounts within each community and in recent previous years. All these supplementary analyses consistently highlight the VOX community, as it emerges prominently in each of these evaluations. Once again, it is noteworthy to highlight the alignment of these findings with the results from Pastor-Galindo et al. (2020), where the cumulative count of bots in the VOX community during the analyzed period significantly exceeded other political communities. Additionally, it is important to consider that the lower percentage of human users within the VOX community could potentially lead to a considerable increase in the number of bots with a less conservative analysis.

Q4 In what ways do social bots interact with human users, considering interaction types such as retweets, replies, or quotes? Are there noticeable distinctions in the content and sentiment analysis between social bots and human users?

When analyzing the interaction patterns by type of user, it is possible to observe that human users have what can be qualified as human behavior, as previously seen in literature (Ferrara et al. 2016 and Varol et al. 2017): they usually interact with other human users and in very few cases they interact with bot users. Also, human users have a considerable amount of quote and reply tweets, which fits in normal use of the Twitter application. On the other hand, bot users nearly always retweet (more than 96% of tweets captured), and half of these interactions go to human users.

This allows us to state that the relevance of bots in the general conversation is low, and they are mostly used to amplify the messages thrown by human users. This casuistic has been noticed previously in Ferrara et al. (2016), Gonzalez-Bailon et al. (2020) and Pastor-Galindo et al. (2020). More concretely, this situation can be related to astroturfing campaigns as explained by Keller et al. (2020), when in the flood phase there is a significant increase in tweets with the help of automated bots (Arce-Garcia et al. 2022; Arce-Garcia et al. 2023). Also, Elmas et al. (2021) considered that these astroturfing campaigns could be related to

promote artificially Trending Topics through coordinated and inauthentic activities.

No considerable differences have been noted in the text analysis between bot and human users, as well as in the sentiment analysis of their respective texts. The probable reason behind that is the high rate of RT in bot tweets, which implies that bots are used to amplify the messages of human users, thus the text and sentiment analysis have a high similarity between both types of users. These findings align with the results reported by Kusen and Strembeck (2018), emphasizing the prevalence of retweets among bots. However, they diverge from the outcomes of previous investigations such as Stella et al. (2018), which highlighted bots predominantly disseminating negative and inflammatory content, specifically targeting human users. In our study, while we observe bots targeting human users, we did not identify a predominant negative sentiment among them when compared to the sentiment expressed by human users. This contrasts also with the results of Martinez Torralba et al. (2023), who reported a prevalence of negative sentiment in bot-generated tweets.

7 Conclusions

In conclusion, this study added results and analysis to the ongoing academic conversation regarding the digital political ecosystem during political campaigns, concretely in a very specific case during #14F 2021 Catalonia regional elections.

The existence of echo chambers leads to poor cross-community conversation and debate, which would be the desirable environment in a healthy democracy, in parallel with the debate in the real world. We are experiencing a rise in polarization around the world, and the omnipresence of social media could be behind this trend. The concept of echo chambers and their repercussions on social networks has been a focal point of academic attention in recent years, particularly in the fields of digital communication and social network analysis, and the debate about their existence and effects it is still ongoing. In our study, echo chambers have been detected and have considerable robustness, even higher when users are grouped into polarized supra-communities. These findings are in accordance with the insights provided by previous studies, including Barberá et al. (2015), Williams et al. (2015), Merry (2015), which investigate interactions between polarized political groups on social media. Similarly, they coincide with studies on the Spanish context, such as Aragón et al. (2013), Guerrero-Solé (2017), and Esteve del Valle and Borge Bravo (2018), offering additional insights into the existence of echo chamber effects.

Thus, it can be concluded that the majority of interactions occurred within political communities, reinforcing the idea that users tend to engage with like-minded individuals, even in scenarios where cross-community interactions were anticipated. This trend mirrors the results reported in Aragón et al. (2013), which is also centered in a Spanish electoral campaign. However, these results diverge from Balcells and Padró-Solanet (2020), even though we attribute the discrepancies to methodological differences.

The existence and influence of social bots have been extensively explored in academic literature. The rise of propaganda and information operations in the digital ecosystem suppose a threat to our democracies, with a few actors attempting to deceive the broader society through artificially generated content and support. This investigation detected the use of social bots during the election campaign, and the results obtained show an excess of social bots within the VOX community. These findings are in line with the results in Pastor-Galindo et al. (2020) and Ferrara et al. (2020). Contrarily, Martinez-Torralba et al. (2023) reported that government parties (in their case left-wing) exhibited a greater prevalence of bot activity.

Examining the interactions between bots and human users, the observed behavior aligns with patterns associated with astroturfing campaigns, particularly during the flood phase, a phenomenon investigated in both international and national literature. Thus, these social bots play a crucial role in significantly amplifying and disseminating specific ideas, potentially influencing the beliefs of social media users and, consequently, impacting their voting behavior in political elections. This casuistic has been previously observed in Ferrara et al. (2016), Gonzalez-Bailon et al. (2020) and Pastor-Galindo et al. (2020). Specifically, this situation can be linked to astroturfing campaigns, as explained by Keller et al. (2020), and also identified in the works of Arce-Garcia et al. (2023). Additionally, Elmas et al. (2021) suggested that these astroturfing campaigns might be associated with the artificial promotion of Trending Topics.

This study reinforces the imperative need for in-depth exploration of social media and the digital environment. Integrating a social and ethical perspective, it becomes evident that proactive measures are essential to safeguard democratic integrity in digital spaces. Policymakers must prioritize transparency and accountability on social media platforms, creating an environment that contains well-informed and diverse public discussions. However, it is concerning that the direction of most social media platforms appears to be moving in the opposite direction, with increasing restrictions on public access to information. Twitter, in particular, has notably reduced the availability of publicly accessible data in recent years.

Echo chambers, as identified in our study, have significant social implications. In a healthy democratic environment, citizens should have access to diverse opinions and perspectives, enabling them to make informed decisions. However,

the prevalence of echo chambers restricts the range of viewpoints available to citizens, potentially reinforcing existing biases and limiting the diversity of political dialogue.

Our investigation into the presence and activities of social bots indicate the urgent need for regulation and scrutiny. The use of social bots in astroturfing campaigns represents a clear attempt to manipulate the dissemination of artificial information, potentially disrupting democratic processes. Such activities should be subject to deeper investigation and regulation, as they could be interpreted as deliberate attempts to manipulate democracy in a time when voters heavily rely on social media for obtaining information.

Finally, the global trend of polarization and the rise of radical political ideologies raise significant concerns. In terms of future research directions, we believe that exploring the dynamics between online echo chambers, social bots and political polarization can offer valuable insights. Extreme polarization is a threat to the stability of democratic societies, potentially leading to increased social divisions. Addressing the root causes of polarization and understanding its relationship with digital phenomena is essential to ensuring the protection of democratic values. It is essential to conduct further research in these areas to formulate effective strategies for mitigating polarization and promoting a more inclusive democratic digital environment. Also, additional studies focusing on the detection and mitigation of social bots, are necessary for developing robust strategies to combat digital manipulation and safeguard democratic principles.

Table 20 20 most common words on user bios and in tweets on the Total dataset

Most common words in User Bio (#Users)	Most common words in Tweets (#Repetitions)
Vida (6.764)	Cataluña (160.212)
Periodista (4.068)	Illa (129.659)
España (3.711)	Votar (125.505)
Política (3.451)	Catalunya (104.953)
Catalunya (3.425)	Vox (95.880)
Siempre (3.328)	PSOE (79.129)
Mundo (3.309)	Vot (77.143)
Social (3.193)	Eleccions (74.827)
Twitter (2.947)	Campanya (70.392)
Oficial (2.832)	Govern (59.540)
Cuenta (2.749)	Catalanes (54.485)
Barcelona (2.651)	PSC (50.225)
Español (2.622)	Hoy (50.164)
Madrid (2.423)	President (47.503)
Amante (2.349)	España (47.001)
Libertad (2.338)	Podemos (44.433)
Catalana (2.303)	Avui (43.512)
Gusta (2.297)	Cap (43.171)
País (2.216)	Salvador (42.141)
Feminista (2.185)	Gobierno (41.912)

Appendix A

See Tables 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29.

Table 19 Tweet distribution by type (and percentage) over datasets

Dataset	Total tweets	Original tweets	Retweets	Quote tweets	Reply tweets
Pre-campaign	460.447 (100%)	26.677 (5.79%)	388.039 (84.28%)	35.027 (7.61%)	10.704 (2.32%)
Campaign	2.069.633 (100%)	182.511 (8.82%)	1.730.082 (83.59%)	116.075 (5.61%)	40.965 (1.98%)
Post-campaign	142.051 (100%)	15.278 (10.76%)	113.623 (79.99%)	9.307 (6.55%)	3.843 (2.70%)
Total	2.672.131 (100%)	224.466 (8.40%)	2.231.744 (83.52%)	160.409 (6.00%)	55.512 (2.08%)

Table 21 Most relevant users in each community (higher in-degree)

Community	#1 user (#in-degree)	#2 user (#in-degree)	#3 user (#in-degree)
VOX	@Santi_ABASCAL (37.562)	@vox_es (34.818)	@Macarena_Olona (25.360)
JxCat	@JuntsXcat (22.795)	@oriolmitja (10.648)	@Bernat_Deltell (9.323)
CUP	@cupnacional (21.719)	@CUPNordOriental (2.038)	@CUPTarragona (1.749)
Comuns	@EnComu_Podem (15.872)	@jessicaalbiach (9.770)	@FonsiLoaiza (2.723)
PSC	@socialistes_cat (45.029)	@PSOE (36.291)	@sanchezcastejon (7.162)
ERC	@Esquerra_ERC (26.600)	@perearagones (2.016)	@esquerracamp (1.684)
PP	@populares (24.930)	@PPCatalunya (11.107)	@IDiazAyuso (3.213)
Cs	@CiudadanosCs (15.272)	@CiutadansCs (12.830)	@InesArrimadas (5.372)

Table 22 Number interactions (active and passive) for each community

Community	# Active interactions	# Passive interactions
JxCat	460.250	446.658
VOX	409.652	401.233
PSC	244.097	245.961
Comuns	143.104	145.121
ERC	136.627	143.072
CUP	126.228	133.214
Cs	103.342	103.596
PP	100.538	102.087

Table 23 Number of interactions with certain active/passive user

Active user community	Passive user community	Count	Percentage in active community interactions
JxCat	JxCat	412.664	89.66%
VOX	VOX	386.151	94.26%
PSC	PSC	229.464	94.00%
Comuns	Comuns	128.143	89.54%
ERC	ERC	120.151	87.94%
CUP	CUP	101.395	80.33%
Cs	Cs	91.866	88.89%
PP	PP	88.989	88.51%
JxCat	CUP	18.962	4.12%
CUP	JxCat	13.241	10.49%
JxCat	ERC	12.431	2.70%
VOX	PP	8.606	2.10%
ERC	JxCat	7.631	5.59%
VOX	Cs	6.495	1.58%
PP	VOX	5.472	5.44%

Table 24 Number of interactions grouping by political blocks

Community	# Active interactions	# Passive interactions
Independentist	723.105	722.944
Constitutionalist	613.532	606.916
Spanish Gov	387.201	116.180

Table 25 Number of users created in 2020/2021 by community

Community	Total users	Users created 2020	Users created 2021
VOX	35.268	6.053 (17.16%)	1.089 (3.08%)
JxCat	32.714	1.811 (5.53%)	361 (1.10%)
CUP	24.935	1.661 (6.66%)	323 (1.29%)
Comuns	17.787	1.774 (9.97%)	334 (1.88%)
PSC	13.016	1.162 (8.92%)	278 (2.13%)
ERC	9.113	578 (6.34%)	131 (1.44%)
PP	8.585	967 (11.26%)	150 (1.75%)
Cs	6.614	551 (8.33%)	129 (1.95%)

Table 26 Number of "numberplate" accounts by community

Community	Total users	"Numberplate" accounts	Percentage inside community
VOX	35.268	3.526	9.99%
JxCat	32.714	1.460	4.46%
CUP	24.935	390	1.56%
Comuns	17.787	1.020	5.73%
PSC	13.016	683	5.28%
ERC	9.113	192	2.11%
PP	8.585	484	5.63%
Cs	6.614	322	4.87%

Table 27 Number of "numberplate" bots by community (the percentage represents the fraction of "numberplate" accounts inside the set of bots of each community)

Community	Total bots	"Numberplate" accounts
VOX	1.745	248 (14.21%)
JxCat	1.102	119 (10.79%)
CUP	770	55 (8.56%)
Comuns	642	97 (12.59%)
PSC	388	60 (15.46%)
ERC	308	24 (7.79%)
PP	369	34 (9.21%)
Cs	435	34 (7.81%)

Table 29 Language used by community users after filtering

	Catalan	Spanish
VOX	0.76%	99.24%
JxCat	91.45%	8.55%
CUP	91.29%	8.71%
Comuns	46.99%	53.01%
PSC	29.59%	70.41%
ERC	94.34%	5.66%
PP	2.31%	97.69%
Cs	9.75%	90.25%

Table 28 Number and percentage of interactions between human-bots (percentages sum does not match 100% due to the missing unclear users interactions)

Active user	Passive user	Count	% inside active user group
Human	Human	631.088	63.47%
Bot	Human	78.781	48.64%
Human	Bot	32.792	3.29%
Bot	Bot	13.726	8.47%

Appendix B: Hashtag collection list

See Table 30.

Table 30 Hashtags and keywords captured during the period under study (If stopped value is blank, means it stopped at the end of the extraction period (21/2/2021))

Hashtag/Keyword	Started Tracking	Stopped	Comments
#14F	1/1/2021		General hashtag
#eleccions14F	3/1/2021		General hashtag
#30 M	16/1/2021	27/1/2021	General hashtag used in the impasse where the elections where postponed (ref)
#14-F	23/1/2021		General hashtag
14F	23/1/2021		General keyword
eleccions14F	23/1/2021		General keyword
14-F	23/1/2021		General keyword
#DebateCATrtve	31/1/2021	9/2/2021	TV Debate Hashtag
#DebatTV3	9/2/2021		TV Debate Hashtag
#InfoEleccions2021	14/2/2021		Hashtag used by the election organizers
#VuelveIlla	1/1/2021	27/1/2021	Hashtag used by PSC
#SomPSC	1/1/2021	27/1/2021	Hashtag used by PSC
#TornaCatalunya	1/1/2021	28/1/2021	Hashtag used by PSC
#VuelveCatalunya	1/1/2021	27/1/2021	Hashtag used by PSC
#FemHo	28/1/2021		Hashtag used by PSC
#PresidentIlla	29/1/2021		Campaign Slogan PSC
#PSC	14/2/2021		Hashtag used by PSC
#FemLaViaÀmplia	28/1/2021	29/1/2021	Hashtag used by ERC
#ViaAmplia	28/1/2021	29/1/2021	Hashtag used by ERC
#AragonèsSenseFiltres	28/1/2021	29/1/2021	Hashtag used by ERC
#AlCostatDeLaGent	29/1/2021		Campaign Slogan ERC
#JoERC	14/2/2021		Hashtag used by ERC
#JuntsXCat	1/1/2021	27/1/2021	Hashtag used by JxCat
#FemHoJunts	1/1/2021	27/1/2021	Hashtag used by JxCat
#Junts	3/1/2021	27/1/2021	Hashtag used by JxCat
#FemHistoriaJunts	13/1/2021	27/1/2021	Hashtag used by JxCat
#JuntsperGuanyar	28/1/2021	29/1/2021	Hashtag used by JxCat
#JuntsperFer	28/1/2021		Campaign Slogan JxCat
#JuntsperSer	28/1/2021		Campaign Slogan JxCat
#RecuperemosCataluña	1/1/2021		Campaign Slogan VOX
#StopIslamización	28/1/2021	29/1/2021	Hashtag used by VOX during a daily controversy
#ElCanviQueCatalunyaMereix	28/1/2021		Campaign Slogan Comuns
#14FComençaElCanvi	29/1/2021		Hashtag used by Comuns
#JoEnComuPodem	14/2/2021		Hashtag used by Comuns
#PerGuanyar	1/1/2021		Campaign Slogan CUP
#UnNouCicle	3/1/2021		Campaign Slogan CUP
#UnaCataluñaParaTodos	11/1/2021	28/1/2021	Hashtag used by Cs
#PSCIrresponsable	13/1/2021	27/1/2021	Hashtag used by Cs to criticize Salvador Illa while he was not abandoning Health Ministry
#ParaQueGanemosTodos	29/1/2021		Campaign Slogan Cs
#SumantPerCatalunya	1/1/2021	28/1/2021	Hashtag used by PP
#SumandoPorCataluña	1/1/2021	28/1/2021	Hashtag used by PP
#UnaCataluñaMejor	29/1/2021		Campaign Slogan PP
#ElsDeTotaLaVida	1/1/2021	28/1/2021	Hashtag used by PDeCAT
#CentremElPaís	1/1/2021	28/1/2021	Hashtag used by PDeCAT
#DefensemLoNostre	3/1/2021	28/1/2021	Hashtag used by PDeCAT
#SiThoPensesPDeCAT	7/1/2021		Campaign Slogan PDeCAT
#FemFront	3/1/2021		Hashtag used by FNC
#AvalsFNC	3/1/2021	28/1/2021	Hashtag used by FNC
#AlliberaElTeuPaís	31/1/2021		Hashtag used by FNC
#RompeTuVoto	14/2/2021		Hashtag used during Election Day that reached TT
#Pucherazo	14/2/2021		Hashtag used during Election Day that reached TT

Table 30 (continued)

As can be seen in Table 30, hashtags related to PDeCat and FNC (political parties that did not obtain Parliamentary representation) were also collected and monitored, but the number of tweets captured and their users relations do not represent such a significant quantity as to be considered a relevant community on the global conversation

Declarations

Conflict of interest The author(s) declare that there is no conflict of interest.

References

Abilov A, Hua Y, Matatov H, Amir O, Naaman M (2021) Voter-Fraud2020: a multi-modal dataset of election fraud claims on Twitter. Available at: https://arxiv.org/abs/2101.08210

Aparici R, García-Marín D, Rincón-Manzano L (2019) Noticias falsas, bulos y trending topics. Anatomía y estrategias de la desinformación en el conflicto catalán. Prof Inf 28(3):e280313. https://doi.org/10.3145/epi.2019.may.13

Aragón P, Kappler KE, Kaltenbrunner A, Laniado D, Volkovich Y (2013) Communication dynamics in twitter during political campaigns: the case of the 2011 Spanish national election. Policy Internet 5(2):183–206. https://doi.org/10.1002/1944-2866.POI327

Arce-García S, Said-Hung E, Mottareale D (2022) Astroturfing as a strategy for manipulating public opinion on Twitter during the pandemic in Spain. Prof Inf 31(3):e310310. https://doi.org/10.3145/epi.2022.may.10

Arce-García S, Said-Hung E, Mottareale-Calvanese D (2023) Types of astroturfing campaigns of disinformative and polarised content in times of pandemic in Spain. ICONO14 Rev Comun Tecnol Emerg. https://doi.org/10.7195/ri14.v21i1.1890

Balcells J, Padró-Solanet A (2020) Crossing lines in the twitter debate on catalonia's independence. Int J Press Polit 25(1):28–52. https://doi.org/10.1177/1940161219858687

Barberá P, Jost JT, Nagler J, Tucker JA, Bonneau R (2015) Tweeting from left to right: is online political communication more than an echo chamber? Psychol Sci 26(10):1531–1542. https://doi.org/10.1177/0956797615594620

Bastian M, Heymann S, Jacomy M (2009) Gephi: an open source software for exploring and manipulating networks. In: International AAAI conference on weblogs and social media

Bello BS, Heckel R (2019) Analyzing the behaviour of twitter bots in post brexit politics. In: 6th International conference on social networks analysis, management and security (SNAMS), pp 61–66, https://doi.org/10.1109/SNAMS.2019.8931874

Beskow DM, Carley KM (2018) Bot-hunter: a tiered approach to detecting & characterizing automated activity on Twitter. Carnegie Mellon University. Available: http://www.casos.cs.cmu.edu/publications/papers/LB_5.pdf

Bird S, Klein E, Loper E (2009) Natural language processing with Python: analyzing text with the natural language toolkit. O'Reilly Media, Inc

Blondel V, Guillaume J, Lambiotte R, Mech E (2008) Fast unfolding of communities in large networks. J Stat Mech Theory Exp 2008:P10008

Brachten F, Stieglitz S, Hofeditz L, Kloppenborg K, Reimann A (2017) Strategies and influence of social bots in a 2017 German state election–a case study on Twitter

Bradshaw S, Howard PN, Kollanyi B, Neudert L-M (2020) Sourcing and automation of political news and information over social media in the united states, 2016–2018. Polit Commun 37(2):173–193. https://doi.org/10.1080/10584609.2019.1663322

Bradshaw S, Howard PN (2019) The global disinformation disorder: 2019 global inventory of organised social media manipulation. In: Working paper 2019.2. Project on Computational Propaganda, Oxford, UK

Brennen JS, Simon F, Howard PN, Nielsen RK, Cairns I, Shetty P (2020) Types, sources, and claims of COVID-19 misinformation. Reuters Inst 7:1–3

Centre d'Estudis d'Opinió–CEO (2020). Baròmetre d'opinió pública, 3a Onada 2020. Available at: https://ceo.gencat.cat/ca/barometre/detall/index.html?id=7808

Chang HCH, Ferrara E (2022) Comparative analysis of social bots and humans during the COVID-19 pandemic. J Comput Soc Sc 5:1409–1425. https://doi.org/10.1007/s42001-022-00173-9

Chen H, Franks P, Evans L (2016) Exploring government uses of social media through twitter sentiment analysis. J Digit Inf Manag 14:290–301. https://doi.org/10.6025/jdim/2016/14/5/290-301

Chen Y, Skiena S (2014). Building Sentiment lexicons for all major languages. In: Proceedings of the 52nd Annual meeting of the association for computational linguistics, vol 2. Association for Computational Linguistics, Baltimore, MD, pp 383–389

Davis C, Varol O, Ferrara E, Flammini A, Menczer F (2016) BotOrNot: a system to evaluate social bots. pp 273–274. https://doi.org/10.1145/2872518.2889302

Del Valle ME, Borge Bravo R (2018) Echo chambers in parliamentary twitter networks the catalan case. Int J Commun 12:1715–1735

Elmas T, Overdorf R, Ozkalay A, Aberer K (2021) Ephemeral astroturfing attacks: the case of fake twitter trends. In: 2021 IEEE European symposium on security and privacy (EuroS&P), Vienna, Austria, pp 403–422. https://doi.org/10.1109/EuroSP51992.2021.00035

Ferrara E, Varol O, Davis C, Menczer F, Flammini A (2016) The rise of social bots. Commun ACM 59(7):96–104. https://doi.org/10.1145/2818717

Ferrara E, Chang H, Chen E, Muric G, Patel J (2020) Characterizing social media manipulation in the 2020 U.S. presidential election. First Monday. https://doi.org/10.5210/fm.v25i11.11431

Gallotti R, Valle F, Castaldo N, Sacco P, DeDomenico M (2020) Assessing the risks of "infodemics" in response to COVID-19 epidemics. Nat Hum Behav 4(12):1285–1293

García-Orosa B (2021) Disinformation, social media, bots, and astroturfing: the 4th wave of digital democracy. Prof Inf 30(6):e300603. https://doi.org/10.3145/epi.2021.nov.03

García-Orosa B, Gamallo P, Martín-Rodilla P, Martínez-Castaño R (2021) Hybrid intelligence strategies for identifying, classifying and analyzing political bots. Soc Sci 10:357. https://doi.org/10.3390/socsci10100357

González-Bailón S, De Domenico M (2020) Bots are less central than verified accounts during contentious political events. In: Proceedings of the national academy of sciences. Available at SSRN: https://ssrn.com/abstract=3637121 or https://doi.org/10.2139/ssrn.3637121

Gorodnichenko Y, Pham T, Talavera O (2018) Social media, sentiment and public opinions: evidence from #Brexit and #USElection. Available at: https://www.nber.org/papers/w24631

Guerrero-Solé F (2017) Community detection in political discussions on Twitter: an application of the retweet overlap network method to the Catalan process toward independence. Soc Sci Comput Rev 35(2):244–261. https://doi.org/10.1177/0894439315617254

Guerrero-Solé F (2022) The ideology of media. Measuring the political leaning of Spanish news media through Twitter users' interactions. Commun Soc 35(1):29–43

Hagberg A, Schult D, Swart P (2008) Exploring network structure, dynamics, and function using NetworkX. In: Varoquaux G, Vaught T, Millman J (eds) Proceedings of the 7th python in science conference (SciPy2008), Pasadena, CA USA, pp 11–15

Howard PN, Woolley S, Calo R (2018) Algorithms, bots, and political communication in the US 2016 election: the challenge of automated political communication for election law and administration. J Inf Technol Polit 15(2):81–93. https://doi.org/10.1080/19331681.2018.1448735

Jacomy M, Venturini T, Heymann S, Bastian M (2014) ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. PLoS ONE 9(6):e98679. https://doi.org/10.1371/journal.pone.0098679

Keller TR, Klinger U (2019) Social bots in election campaigns: theoretical, empirical, and methodological implications. Polit Commun 36(1):171–189. https://doi.org/10.1080/10584609.2018.1526238

Keller FG, Schoch D, Stier S, Yang J (2020) Political astroturfing on Twitter: how to coordinate a disinformation campaign. Polit Commun 37(2):256–280. https://doi.org/10.1080/10584609.2019.1661888

King G, Pan J, Roberts ME (2017) How the Chinese government fabricates social media posts for strategic distraction, not engaged argument. Am Polit Sci Rev 111(3):484–501

Kusen E, Strembeck M (2018) Investigation of emotion exchange motifs in bot/human interactions during riot events. In: 2018 5th international conference on social networks analysis, management and security (SNAMS), Valencia 2018, pp 19–26

Lazer DMJ, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F (2018) The science of fake news. Science 359(6380):1094–1096

Marchal N, Kollanyi B, Neudert L-M, Howard PN (2019) Junk news during the EU parliamentary elections: lessons from a seven-language study of Twitter and Facebook. Available at: https://comprop.oii.ox.ac.uk/research/eu-elections-memo/

Martínez Torralba Á, Guevara-Gil JA, Fuente AJ (2023) The presence of political bots on Twitter during the COVID-19 crisis in Spain. Rev Esp Investig Sociol 181:61–80. https://doi.org/10.5477/cis/reis.181.61

McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. Ann Rev Sociol 27(1):415–444. https://doi.org/10.1146/annurev.soc.27.1.415

Mele N, Lazer D, Baum M, Grinberg N, Friedland L and Joseph K, Mattsson C (2017) Combating fake news: AN agenda for research and action. Retrieved on 17 Oct 2018

Merry M (2015) Making friends and enemies on social media: the case of gun policy organizations. Iran J Inf Process Manag 30(2):373–396. https://doi.org/10.1108/EL-01-2014-0022

Morstatter F, Pfeffer J, Liu, H, Carley KM (2013) Is the sample good enough? Comparing data from twitter's streaming API with Twitter's firehose. ArXiv, abs/1306.5204

Neudert LM, Howard P, Kollanyi B (2019) Sourcing and automation of political news and information during three European elections. Soc Media. https://doi.org/10.1177/2056305119863147

Newman N, Fletcher R, Eddy K, Robertson CT, Nielsen RK (2023) Reuters institute digital news report 2023. Reuters Institute for the Study of Journalism, Oxford

Nielsen RK, Graves L (2017) "News you don't believe": Audience perspectives on fake news. Reuters Institute for the Study of Journalism Oxford, UK

Pastor-Galindo J, Zago M, Nespoli P, López Bernal S, Huertas A, Pérez M, Ruipérez-Valiente JA, Martinez Perez G, Gomez Marmol F (2020) Spotting political social bots in Twitter: a use case of the 2019 Spanish general election. IEEE Trans Netw Serv Manag 17(4):2156–2170. https://doi.org/10.1109/TNSM.2020.3031573

Roesslein J (2020) Tweepy: Twitter for Python! URL: Https://Github.Com/Tweepy/Tweepy

Ross Arguedas A, Robertson C, Fletcher R, Nielsen R (2022) Echo chambers, filter bubbles, and polarisation: a literature review. Reuters Institute for the Study of Journalism, Oxford

Sampedro V (2021) Comunicación política digital en españa del «pásalo» a podemos y de podemos a vox. UOCPress, Barcelona

Stella M, Cristoforetti M, De Domenico M (2019) Influence of augmented humans in online interactions during voting events. PLoS ONE 14:e0214210. https://doi.org/10.1371/journal.pone.0214210

Stella M, Ferrara E, Domenico MD (2018) Bots increase exposure to negative and inflammatory content in online social systems. Available at: https://www.pnas.org/content/pnas/115/49/12435.full.pdf

Terren L, Borge R (2021) Echo chambers on social media: a systematic review of the literature. Rev Commun Res 9:99–118. https://doi.org/10.12840/ISSN.2255-4165.028

Thorson K, Cotter K, Medeiros M, Pak C (2019) Algorithmic inference, political interest, and exposure to news and politics on Facebook. Inf Commun Soc 24(2):183–200. https://doi.org/10.1080/1369118x.2019.1642934

Tsai T, Chuan H (2020) Echo chambers and social mediators in public advocacy issue networks. Public Relat Rev 46(1):101882. https://doi.org/10.1016/j.pubrev.2020.101882

Tucker J, Guess A, Barbera P, Vaccari C, Siegel A, Sanovich S, Stukal D, Nyhan B (2018) Social media, political polarization, and political disinformation: a review of the scientific literature. SSRN Electron J. https://doi.org/10.2139/ssrn.3144139

Twitter API (2023). https://developer.twitter.com/en/docs/twitter-api

Twitter Developer Policy and Terms (2023). https://developer.twitter.com/en/developer-terms/policy

Varol O, Ferrara E, Davis CA, Menczer F, Flammini A (2017) Online Human-Bot Interactions: detection, estimation, and characterization. Available at: https://arxiv.org/pdf/1703.03107.pdf

Wardle C, Derakhshan H (2017) Information disorder: toward an interdisciplinary framework for research and policymaking, vol 9. Council of Europe, Strasbourg

Williams HTP, McMurray JR, Kurz T, Hugo Lambert F (2015) Network analysis reveals open forums and echo chambers in social media discussions of climate change. Glob Environ Chang. https://doi.org/10.1016/j.gloenvcha.2015.03.006

Yang K, Varol O, Davis C, Ferrara E, Flammini A, Menczer F (2019) Arming the public with artificial intelligence to counter social

bots. Hum Behav Emerg Technol 1:e115. https://doi.org/10.1002/hbe2.115

Yang K-C, Varol O, Nwala AC, Sayyadiharikandeh M, Ferrara E, Flammini A, Menczer F (2024) Social bots: detection and challenges. In: Yasseri T (ed) Handbook of computational social science. Edward Elgar Publishing Ltd, Cheltenham

Zhang M, Qi X, Chen Z, Liu J (2022) Social bots' involvement in the COVID-19 vaccine discussions on twitter. Int J Environ Res Public Health 19:1651. https://doi.org/10.3390/ijerph19031651