Information Fusion 110 (2024) 102459

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Cyber democracy in the digital age: Characterizing hate networks in the 2022 US midterm elections

Andrés Zapata Rozo [a], Alejandra Campo-Archbold [a], Daniel Díaz-López [a,b], Ian Gray [b], Javier Pastor-Galindo [c,*], Pantaleone Nespoli [c], Félix Gómez Mármol [c], Damon McCoy [b]

[a] School of Engineering, Science and Technology, Universidad del Rosario, Bogotá, Colombia
[b] Department of Computer Science and Engineering, Tandon School of Engineering, New York University, NY 11201, USA
[c] Faculty of Computer Science, University of Murcia, Murcia, Spain

ARTICLE INFO

Keywords:
Cyber democracy
Harassment
NLP
Semantic similarity
NER
Sentiment analysis
US midterm elections

ABSTRACT

Social media has become integral to societal discourse and play a role in shaping public engagement, particularly in democratic electoral processes. This paper addresses the pressing issue of hate speech on social media during the 2022 US midterm elections. Unlike previous research, which often relies on limited datasets and classic methodologies, we leverage Open Source Intelligence (OSINT) and Natural Language Processing (NLP) techniques to analyze Twitter data through advanced models of entity recognition, sentiment analysis, and community extraction, having persistence in Knowledge Graphs for consuming the intelligence efficiently. Results indicate that in the US midterm elections 2022, Arizona was the state that provided more content (507,551 tweets) related to a Chief Electoral Official, with 31.58% of them identified in the most aggressive cluster due to its mean attribute values of "attack on commenter" (0.7), "inflammatory" (~0.3), "attack on author" (~0.2), and "toxicity" (~0.2). The name entity recognition model also identified an association between those aggressive tweets and the previous 2020 US Presidential campaign, characterized by attacks on election officials based on conspiracy theories campaigns. Knowledge graphs contributed to understanding the concentration of attacks and connectivity between topics commonly mentioned in hate speech content. Thus, our results offer detailed insights into the actors and dynamics of online harassment in electoral contexts, illuminating the challenges posed by harassment and proposing preventive mechanisms applicable to diverse electoral processes worldwide.

1. Introduction

Undoubtedly, social media plays a significant role in the cyber democracy of a country. Social media provides an open platform for individuals to freely express their ideas and opinions, which can help shape public discourse and influence decision-making processes, e.g., electoral decisions, in an honest and comprehensive manner. Mostly, it permits the creation and maintenance of social connections to create personal networks, which allow the sharing of personal opinions about certain topics in a democratic way. Also, social media can facilitate information dissemination and communication among individuals, such as debates on the political life of a country. Information sharing can also be particularly important in emergencies or for coordinating collective action, e.g., natural disasters or social protests. Additionally, social media enable greater political participation, especially among younger generations, by providing easy access to information about potential modern issues [1].

Nevertheless, some potential negative impacts of social media on cyber democracy also appear. That is, social media can be misused by ill-intentioned entities to carry on malicious activities. For example, the proliferation of fake news and misleading information on social media can undermine the integrity of a democratic process and undermine people's trust in public figures and institutions. Moreover, the use of social media to deceive political campaigns and advertising strategies is a phenomenon happening more frequently nowadays, which, in many cases, attempts to polarize people toward a specific subject or against it [2].

One example of a strategy to disseminate deceptive messages on social media is to use social bots as amplifiers. That is, these software-controlled accounts strive to mimic the behavior of human users, but

* Corresponding author.
E-mail addresses: andresf.zapata@urosario.edu.co (A. Zapata Rozo), alejandra.campo@urosario.edu.co (A. Campo-Archbold), danielo.diaz@urosario.edu.co (D. Díaz-López), iwg210@nyu.edu (I. Gray), javierpg@um.es (J. Pastor-Galindo), pantaleone.nespoli@um.es (P. Nespoli), felixgm@um.es (F. Gómez Mármol), mccoy@nyu.edu (D. McCoy).

https://doi.org/10.1016/j.inffus.2024.102459
Received 2 October 2023; Received in revised form 16 April 2024; Accepted 5 May 2024
Available online 9 May 2024

with the capacity to operate at a much higher rate and remain hidden. In this sense, studies have shown how these coordinated armies of social bots can be used to manipulate democratic elections, perform malware diffusion campaigns, and boost aggressive advertising actions, among others [3]. Social bots allegedly played a central role in the 2016 US Presidential election, as they spread polarized content and disinformation [4]. In this sense, social media facilitates the spread of such misleading content, but social bots can amplify this content and create manually crafted viral trends to capture the attention of online communities [5].

In this context, one could easily argue that the misuse of social platforms can generate or increase several real issues in a cyber democracy. Among them, it is worth mentioning cyberterrorism. This phenomenon refers to acts that promote terror, committed by individuals or groups in cyberspace, generally using an aggressive narrative based on hate speech or promotion of violence [6]. These acts are typically intended to intimidate or coerce a civilian population or to influence government policy through fear and violence. In particular, cyberterrorism can be carried out by individuals or organizations with a variety of motivations, including ideological, religious, or political. In this alarming scenario, it is easy to spot that social media plays a fundamental role in the promotion or struggle of this phenomenon. For instance, cyberterrorist groups can leverage the amplitude of information spreading given by those platforms to reach as many targets as possible in a fast fashion. However, social media can also be used to counteract extremist narratives and propaganda. In this last sense, communities can create and share content that promotes tolerance, inclusion, and counter-narratives to extremist ideologies. Additionally, those platforms allow for direct engagement with communities affected by or vulnerable to cyberterrorism.

Considering this context, hate speech refers to language that is intended to hurt, intimidate, or harass individuals or groups based on their race, ethnicity, national origin, religion, gender, sexual orientation, or other peculiarities [7,8], possibly resulting in violence among people. Hate speech can take many forms, including verbal attacks, slurs, or other derogatory comments, as well as written or visual materials that promote prejudice or discrimination. In this sense, it is easy to argue that hate speech is extremely harmful because it can lead to emotional distress and psychological harm for those targeted by such an attack [9]. It can also contribute to a climate of fear and division within a community or country, leading to violence and riots in extreme cases.

As previously mentioned, social media represent a powerful tool to spread information and communicate with people, but they are also prone to misuse [10]. In certain ways, they can be seen as an agent for spreading hate speech. Specifically, the ease and speed with which information can be shared on social media platforms permit the fast dissemination of hateful or discriminatory messages against certain individuals or groups. Furthermore, the possibility of being masqueraded behind the anonymous or pseudonymous nature of many social media accounts allows individuals to express hateful views without fear of accountability. In this context, it is clear that the spread of hate speech on social media can have serious consequences. First and foremost, it creates tensions and divisions among the users, resulting, as mentioned before, in violence and riots that could lead to physical damage ultimately [11].

In this context, it is possible to spot several techniques and tactics that could be maliciously leveraged to spread hate speech within the social media ecosystem. Among them, one could mention: (i) direct written or multimedia attacks, (ii) dissemination of fake news, (iii) anonymous harassment and threats, (iv) trolling and provocation, and (v) use of amplifier bots, among others [12]. While each tactic may have unique characteristics and methods of execution, they share common goals of spreading hate speech, inciting division, and causing harm or distress to individuals or groups online. They also contribute to creating a hostile or toxic online environment and perpetuating harmful attitudes or stereotypes.

For those reasons, individuals, social media enterprises, and institutions need to be aware of the potential for social media to spread hate speech in order to immediately take steps to mitigate this potentially dangerous phenomenon. This process includes implementing policies and procedures for addressing hate speech on social media platforms, as well as education and awareness campaigns to promote responsible and respectful online behavior. One example of a policy that a social media platform can enforce to stop or limit hate speech among users is implementing a comprehensive Community Guidelines document, entailing a clear prohibition statement, user-friendly report mechanisms, and broad content review while cooperating with Law Enforcement Agencies (LEAs) when speech might cross legal boundaries, such as threats of violence, and poses a serious threat to public safety.

Nonetheless, in order to execute a rapid response against hate speech, it is essential to detect this conduct and the individuals or groups who perpetrate it. To this extent, several technical challenges associated with detecting hate speech on social media platforms exist [7]. Particularly, it is worth remarking that one of the major challenges relies on the inherent subjectivity of hate speech. That is, different people may have different definitions and interpretations of what constitutes hate speech, making it difficult to develop algorithms that can reliably identify and classify such content, especially in the social media ecosystem. Additionally, due to the inherent dissimilarities among countries and cultures (e.g., language, traditions, etc.), it is clear that models generated and trained in distinct countries could output different results. Moreover, one can easily say that many countries do not possess the required resources to maintain infrastructures that support the process of automatic hate speech detection, e.g., through solutions empowered by Artificial Intelligence (AI), Natural Language Processing (NLP), and machine-assisted methodologies.

Additionally, the vast volume of data that is generated on social media platforms hinders the detection. Precisely, with millions of users posting billions of messages every day, it is challenging to manually review and identify hate speech, making it necessary to develop automated approaches to detect such content, leveraging AI, for example. In this sense, it is noteworthy to highlight the initiatives undertaken by various countries in utilizing NLP to identify irregularities on social media. For example, the Big Data Team at the Central Office for Information Technology in the Security Sector (ZITiS)[1] in Germany, the Artificial Intelligence for Law Enforcement of Community Safety (AiLECS) Lab[2] in Australia, the National Police Agency (NPA)[3] in Japan, and the Roxanne EU project.[4]

From a content perspective, the use of slang, abbreviations, and other informal language obstructs algorithms' ability to accurately understand and interpret the meaning of messages [13]. Such characteristics can lead to false positives, where hate speech is incorrectly identified, or false negatives, where it is missed altogether. Finally, the use of obfuscation, social bots, and other tactics to deliberately mislead or manipulate algorithms presents a demanding task for detecting hate speech on social media [14,15]. This requires the development of sophisticated algorithms that can not only identify hate speech but also identify and mitigate attempts to manipulate such algorithms.

In the paper, we present the analysis of hate speech in the 2022 US midterm elections, reviewing specific actions involving chief election officers. Along with this analysis, we apply Open Source Intelligence (OSINT) techniques to compose an extensive tweet dataset, which is processed through our NLP models to extract knowledge that may be used to understand this phenomenon but also to support its detection

[1] https://www.zitis.bund.de/EN/WhoWeAre/WhoWeAre_node.html
[2] https://ailecs.org/
[3] https://www.npa.go.jp/english/
[4] https://www.roxanne-euproject.org/

and prevention. In contrast to prior research, which frequently depends on constrained datasets and binary classification methodologies, the proposed investigation introduces an innovative framework. Our framework integrates comprehensive real-world data with techniques such as entity recognition, sentiment analysis, and community extraction to offer an exhaustive characterization of hate speech directed toward chief election officials during the 2022 US midterm elections.

Specific contributions delivered in this paper are mentioned next:

- A module to collect and preprocess data from social media.
- The integration of a recognized toxicity classifier, i.e., Perspective, to process social media text and identify toxic content.
- A Name Entity Recognition model that allows the identification of actors and targets of hate speech.
- A module to analyze the similarity between the content and cluster communities around hate speech.
- A module to consume interactively knowledge graphs and resolve queries related to hate speech research.

The remainder of this paper is structured as follows. Section 2 describes the fundamental concepts to understand this research. Section 3 describes some scientific works that have researched the phenomena of hate speech in election processes. Section 4 presents our proposed research methodology to analyze hate speech in elections. Sections 5 and 6 show the applicability of the research methodology in the 2022 US midterm elections. Section 7 presents an analysis of the obtained results, focusing also on the limitations of our proposal. Finally, Section 8 includes the main outcomes of this research and proposes some future research initiatives.

2. Understanding a cyber democracy scenario

In this section, we provide some background on essential concepts related to our research. In particular, a review of cyber democracy and its correlation with social media, consequences of hate speech in elections, and regulation related to hate speech is offered, considering examples from the US and worldwide. Finally, hate speech recognition and its technical challenges are also described.

Cyber democracy refers to the use of digital technologies and online platforms to facilitate civic discourse, citizen engagement, and participation in modern democracies [16]. Otherwise, social media has enabled more inclusive political dialogue, information dissemination, and activism across geographic and demographic boundaries. However, social media has also enabled the mass spread of misinformation and created new platforms that can be abused for hate speech and harassment. These platforms can be misused to deepen social divisions, especially during highly charged election cycles [17]. In this regard, a key challenge facing democratic societies is how to maximize the benefits of a free speech right while also protecting vulnerable groups and individuals from targeted abuse. In this context, the spread of cyber harassment aimed at election officials represents one particularly alarming threat to the integrity of a democratic process.

The Obama campaign's innovative use of platforms like Facebook, Twitter, and YouTube enabled them to engage voters, mobilize volunteers, and raise funds on a scale surpassing his opponents in both the 2008 and 2012 US presidential elections [18]. In 2008, however, Obama demonstrated the transformative power of social media in cyber democracy over previous attempts at internet campaigning. The strategic use of social media for voter targeting, messaging, and expanding the electorate, especially among youth, was credited with helping secure Obama's victory [19]. Also, by the time the election results were cast in 2008, Obama had over 2 million Facebook followers, and at that moment Obama's Campaign team sent election night results via Twitter and advertised it through popular YouTube videos. This watershed moment demonstrated the power of social media for political organizing and reshaping cyber democracy in the US, setting a standard for future campaigns, such as the 2018 midterm one [20].

In subsequent elections, social media has become an indispensable tool for campaigns to interact with voters, promote content, respond rapidly to events, and leverage data analytics. However, its open access and reach have also introduced new concerns around misinformation, foreign influence and interference, and online extremism, which played a role in subsequent presidential and midterm campaigns [21,22].

While serving as a tool for political campaigning and pluralistic debate, social media has also enabled the mass spread of harmful content like hate speech, harassment, and threats aimed at election officials. The relative anonymity of platforms like Twitter and Facebook has created opportunities for the rapid dissemination of racist, sexist, antisemitic, violent, and other abusive messaging, which negatively impact a cyber democracy. Coordinated hashtag campaigns, inauthentic accounts, bots, and other tactics can weaponize social networks to amplify hate and intimidation, which in some cases may even scale up to cyber terrorism. The psychological and reputational impacts on the victims can be significant.

So, why is value to study harassment against election officials? Election officials play a pivotal role in the democratic process, ensuring that elections are conducted fairly, transparently, and efficiently. Their responsibilities span from overseeing the logistical aspects of elections to ensuring the integrity of the vote. Given the centrality of their role, any form of harassment or intimidation directed towards them can have profound implications for the democratic process [23]. It is important to note that following the 2020 US Presidential campaign, cited as the "most secure election in American history", election officials were targeted with misinformation campaigns, resulting in threats and undermining trust in the democratic process. Many officials have faced death threats, online harassment, and political interference, leading to a significant turnover in election officials. A high turnover rate can lead to a loss of institutional knowledge, potentially resulting in administrative errors that further fuel conspiracy theories [24,25].

Later, during the 2022 US midterm elections, election administrators, i.e., election officials, faced a surge of violent threats and hostile messages fueled by false conspiracy theories and divisive rhetoric after the 2020 election [25]. The normalization of online hate speech and abuse poses grave risks to the integrity and security of the democratic process in the digital age.

The proliferation of hate and threats toward election administrators poses grave dangers for the future of cyber democracy. The possibility of intimidation erodes public trust in electoral integrity and the competence of those overseeing the process. Officials consumed by harassment, threats and, security needs may be distracted from ensuring safe and secure elections. Fear of retribution for dissent could affect decision-making and independence. The departure of experienced professionals due to burnout and unsafe work conditions may leave electoral management vulnerable to poor administration or manipulation. The entire democratic process rests on the impartial administration of laws guiding campaigns, voting, and results. Threats to election officials could systematically undermine free and fair elections [26]. This growing trend requires both legal protections and cultural shifts toward civil discourse to reinforce democratic norms. Failure to address these impacts risks a crisis of legitimacy at the ballot box.

Harassment and hate speech laws have been implemented in various jurisdictions worldwide to address and curb expressions that incite violence, discrimination, or hostility towards specific groups based on attributes such as race, religion, ethnic origin, sexual orientation, disability, or gender. These laws aim to strike a balance between protecting individual freedom of expression and ensuring personal and societal safety. The First Amendment of the US Constitution protects freedom of speech, including controversial or offensive speech. However, the US Supreme Court has held that speech that incites "imminent lawless action" can be restricted. Despite this, the expression of hate or prejudice, without incitement to imminent violence, has generally been protected in the US [27].

Laws prohibiting harassment and hate speech have remained contentious within the US. In December 2022, the state of New York (N.Y.) attempted to enact a law, N.Y. Gen. Bus. Law §394-CCC that would mandate social media network policies against "hateful content". This was defined in the law as the use of social media networks to humiliate or incite violence against a group based on attributes like race, religion, and gender. Such a law requires social media platforms to provide mechanisms for users to report hateful conduct and mandates platforms to have clear policies on how to address such reports [28]. Critics argued that the law compelled speech by requiring platforms to adopt specific policies. Some of the definitions in the law, like "humiliate" and "vilify" were ill-defined. Further, there were concerns that the law infringes upon US First Amendment rights by compelling platforms to regulate speech that may be constitutionally protected [29].

Several countries outside of the US have enacted laws criminalizing harassment and hate speech. These laws attempt to prevent threatening, abusive, or hateful speech by enacting strict punishments for violations. However, they have also received some resistance as they attempt to balance policing of harmful content online while also protecting free speech. The European Union's Framework Decision on Racism and Xenophobia criminalizes public incitement to violence or hatred directed against a group or a member of such a group defined by reference to race, color, religion, descent, or natural or ethnic origin. Additionally, many EU member states have developed their own national laws, which can be more restrictive than the framework [30].

The Canadian Criminal Code criminalizes the public incitement of hatred against an identifiable group, which can lead to violence. It also prohibits the willful promotion of hatred, except in private conversation [31]. In Australia, the Racial Discrimination Act makes it unlawful to "offend, insult, or intimidate" someone based on their race or ethnicity [32]. In South Africa, the Promotion of Equality and Prevention of Unfair Discrimination Act prohibits hate speech, defined as any communication that clearly intends to be hurtful, harmful, or promote or propagate hatred [33]. Sections of the Indian Penal Code criminalize speech that promotes enmity between different groups on the grounds of religion, race, place of birth, residence, language, etc., and prejudicial acts to the maintenance of harmony [34].

The examination of harassment propagated through social media can be enhanced by integrating OSINT techniques. While textual analysis enables the categorization and assessment of abusive content, OSINT techniques provide additional layers of metadata to fully understand the origins, reach, and real-world impacts of online siege. There are a number of tools available to further understand this ecosystem, such as link analysis of account creations and interactions, tracking the propagation of disinformation across platforms, and sentiment analysis regarding public response and counter-messaging [35,36]. Additionally, as not all platforms are the same in terms of content and posts, researchers generally collect information from different sources to obtain a different perspective on the same phenomena.

NLP techniques offer useful tools for identifying, classifying, and tracking harassment at scale. It provides benefits in terms of processing, speediness automation, and less subjectivity than manual human reviewers. Thus, NLP models can contribute to an OSINT analysis with a large amount of collected data [37,38]. Particularly, NLP can be used to develop classifiers that detect abusive language across massive datasets quickly and consistently. In a complementary way, sentiment analysis models provide insights into the emotional intensity behind threatening messages, aiding prioritization. Additionally, NLP models can be fine-tuned to account for nuances in different dialects and platforms. Topic modeling surfaces emerging themes facing coordinated attacks. Tracking the spread and evolution of terms, slogans, and coded phrases reveals how harassment adapts to avoid bans. Network analysis of account relationships exposes coordination. NLP empowers continuous monitoring and moderation through automated flagging rather than lower human review. However, biases in training data, coding errors, and edge cases remain challenges, making thoughtful development and oversight required so that well-intentioned tools do not inadvertently penalize marginalized groups [38].

3. Related works

A systematic literature review [39] revealed a growing number of articles working on automatic hate speech detection in text. Practically, hate speech detection is intertwined with social media and machine learning, predominantly treated as a binary classification task. In another review of hate speech detection using NLP [40], researchers also outline its typical framing as a supervised learning challenge. Common features like bag-of-words and embedding consistently perform well, with character-level methods [41] outperforming token-level ones [42]. Complex features rooted in linguistic understanding, or those using additional data such as images, are also effective. Moreover, another review [43] suggests that gathering and annotating data for automatic hate speech detection is challenging due to its subjective definition. While various public datasets exist, mostly from Twitter, their applicability is constrained by a unique style and character limitations. Other platforms offer richer contexts but are rarer for analysis. Furthermore, the datasets' imbalance between hate and non-hate content complicates representative sampling.

Other studies aim to profile authors within social media platforms, such as WhatsApp, helping to identify fake profiles [44]. By focusing on code-mixed Tamil messages, these studies seek to identify sociodemographic features of authors, such as gender and age group, leveraging techniques like stacked Convolutional Networks (CNN) combined with k-max pooling and Bidirectional Long Short Term Memory (BiLSTM) models.

In electoral contexts, several studies have investigated aggressiveness on social media, offering automated techniques to detect hate speech or harassment within organic content. During the 2016 Philippine presidential campaign, a study probed into hate expressions on Twitter [45]. Through word analysis and clustering techniques, the research found that roughly 55% of terms were unique between hate and non-hate tweets. Interestingly, automated clustering of hate subjects did not align with manual annotations, highlighting the potential role of lexical diversity in improving hate speech detection.

Another study tackled the challenge of monitoring hate speech on social media, focusing on Kenya during its election periods [46]. The researchers proposed a supervised machine-learning approach and underscored the significance of meticulous data annotation. A framework anchored on Sternberg's hate theory was created, and its efficacy was evaluated on 5000 tweets, each assessed by three human annotators.

In Spain, a study suggested using Latent Dirichlet Allocation (LDA) as an unsupervised model to delineate hate speech in tweets gathered during the 2018 regional elections [47]. Concentrating on the ascendancy of the far-right party, it was found that out of over 240,000 tweets referencing "Vox", a mere 1% manifested hate speech. Predominant themes included derogatory language, disinformation, and threats.

In another study focusing on the influence of social media on political elections, researchers examined the presence and behavior of social bots on Twitter during the November 2019 Spanish general election [48]. By classifying users as either social bots or humans and analyzing their interactions quantitatively and qualitatively, the study shed light on the impact of automated accounts on political discourse. The findings revealed a concerning trend, with a notable number of social bots actively engaging in the election process and supporting various political parties.

For the 2016 US presidential election, a study explored the interplay between news shared on mainstream social media platforms and voting intentions [49]. By marrying sentiment analysis with topic analysis, correlations were discerned between the frequency of candidate mentions of specific issues (such as the Clinton Foundation scandal and immigration) and polling data. The study deduced that solely gauging the sentiment of news articles is not comprehensive enough to predict poll shifts.

Another project [50] investigated the rise of hate speech on Twitter during and after Donald Trump's 2016 presidential campaign. The research utilized machine learning-enhanced dictionary methods and a new classification method based on data from Reddit alt-right communities. Analyzing over 1.1 billion tweets, the researchers found no consistent increase in hate speech. While specific events momentarily boosted hateful language, these spikes were short-lived. Thus, there is no substantial evidence to link Trump's campaign or election to a sustained rise in Twitter hate speech.

In the same election context, another research endeavor [51] zeroed in on detecting hate-filled content in memes—a blend of text and imagery. Introducing the MultiOFF dataset, packed with election-relevant memes, aided in multimodal offensive meme detection. Leveraging an early fusion methodology to integrate text and visuals, the employed classifier surpassed text-only and image-only benchmarks across Precision, Recall, and F-Score metrics. In the subsequent 2020 U.S. election, a project [52] delved into potential hate speech among Biden and Trump followers. After annotating 3000 tweets based on the stance and offensiveness, it was discerned using a BERT classifier that tweets supporting a candidate were simpler to detect than those opposing. Nonetheless, automating hate speech detection posed challenges.

In the backdrop of the 2017 German elections, a classifier was introduced for detecting hate speech in tweets [53]. The analyzed tweets, often associated with right-wing German ideologies, revolved around political doctrine, immigration, and alleged crimes by refugees. Major targets of these hostile tweets ranged from immigrants, political adversaries, and German voters to feminists and the LGBTQ+ community.

A comparative analysis of related works that have studied hate speech in elections is shown in Table 1. As evidenced in the first and fourth columns of such a table, and as far as we know, there was no study of hate speech focused on election officials as we present in this paper. This is due most of the studies are focused on the detection of hate speech around: (i) active participants of campaigns, like supporters, opponents, or candidates, (ii) groups of interests, like a politician party, a group of voters, or (iii) specific topics, like immigration, open commerce, among others. This situation remarks the importance of researching the pressure and stress that democracy's key actors, like election officials, may support and how this may impact the electoral process in terms of stability. Regarding the techniques used by different related works (second column), we may observe extensive use of classification models focused on detecting hate from a single dimension. However, our research is unique as it considers 14 attributes (dimensions) to understand hate speech from a wider perspective. Our research also analyzes the presence of 18 types of entities and contributes knowledge graphs used to analyze the phenomenon and could support a strategy to contain it. In terms of the analyzed data (third column), our study is one of the few that collect data before the election day but also some days after, which allows us to include in our analysis three main related events associated with Twitter polemics around the electoral process which dispute the job of election officials.

Thus, unlike previous research, which often relies on limited datasets and binary classification approaches, our study presents a novel framework that integrates open real-world data with entity recognition, sentiment analysis, and community extraction techniques to comprehensively characterize hate speech during the 2022 US midterm elections. By offering detailed insights into the dynamics surrounding election-related hate speech, our study contributes significantly to the existing body of knowledge. Moreover, it underscores the importance of considering both technical advancements and regulatory measures in addressing the challenges posed by hate speech while safeguarding the principles of free speech and democratic discourse.

4. Methodology to spot hate speech in US midterm elections

This section describes our research goals and the phases that integrate the research process followed in this current paper. Each phase contributed important functionalities in the methodology followed to analyze and detect hate speech in US midterm elections, as shown in Fig. 1.

Thus, our proposed methodology achieves the following research goals:

- Collect and preprocess open-source data, specifically from Twitter, associated with possible victims of hate speech.
- Make a classification of posts and determine the probability of it including hate speech content.
- Identify communities that promote hate speech.
- Recognize and extract entities relevant in posts associated with identified communities.
- Organize data in a consumable way, like knowledge graphs, making it usable in hate speech research.

Our proposed methodology is composed of 4 main phases: business understanding, gathering, preprocessing, modeling, and data organization.

4.1. Business understanding: interpreting the US cyber democracy

Before collecting and processing any data, it is important to develop the first phase of a data science life cycle [54], i.e., business understanding, that allows one to recognize which data are meaningful to contribute to resolving the research problem.

Each State in the US has some particularities in the way it runs elections. However, generally, at the federal and State levels in the US, there are presidential general, midterm, primary, and special elections [55]. Presidential general elections are every 4 years, and all states are consulted and allowed to vote for president. Meanwhile, midterm elections are every 2 years and allow voting for senators, congress representatives, and governors, among other public officials. Primary elections allow registered voters to vote for political party candidates who will compete later in the midterm elections. Finally, special elections occur when someone resigns, dies, or is removed.

The US midterm elections are important since they allow to elect members of Congress, i.e., the House of Representatives and the Senate, which represent the Legislative branch of the government and are in charge of checking the performance and actions of the Executive branch, which is composed mainly by the President and Vice President. As the US Congress has the power to enact/refuse laws and presidential initiatives, among many other capabilities, midterm elections are quite relevant and bring a lot of attention from citizens [56].

Under the previous context, Chief Election Officers/Officials (CEOs) are in charge of running the voting process properly and are essential to guarantee that elections are open, impartial, and trusted [57]. The Chief Election Officer role may be performed by different clerks or sections according to the State, i.e., a non-board member executive (Illinois), a board-appointed commissioner (Wisconsin), a board secretary (Oklahoma), a board co-chair (New York), a commission chairperson (Delaware), a State election director (Maryland), a board selected commissioner (Virginia), an election board (North and South Carolina), a board selected chair (Hawaii), a Lt. Governor (Utah and Alaska) or a Secretary of State (for other 38 states).

In all cases, the CEO is in charge of running and certifying elections of local, State, and national candidates, i.e., county officials, governors, State legislators, senators, and representatives. Such a role should be nonpartisan to guarantee that election results are trustable. However, in 31 states where the CEO is the secretary of State, they are elected. In the other 19 states, the CEO is appointed by the Governor or by the election board [58].

Table 1
Related works that analyzed hate speech in elections.

Proposal	Purpose	Core technique	Dataset composition	Gathering criteria	Dataset Language
[45]	Analysis of hate expressions in tweets along the 2016 Philippine Presidential Election	– Hate speech detection using binary classification with logistic regression, random forest, SVM and Gradient Boost	1,696,613 tweets from November 2015 to May 2016	–	Tagalog, English
[46]	Create an annotation framework to detect hate speech in Kenyan Elections	– Use of the theory of hate of Sternberg to classify tweets as: hate, offensive, neither.	394,000 tweets from 2012 to 2017	–	Swahili, English
[47]	Characterize hate speech against immigrants along regional (Andalusian) Spain Elections of December 2018	– Use of unsupervised model LDA to detect underlying topics in hateful messages.	240,000 tweets from November 2018 to April 2019	Tweets that includes the name of the far-right party "vox"	Spanish
[48]	Analyze the presence and interactions of social bots in Twitter during the Spain Elections of November 2019	– Classifier to calculate sentiment in tweets published by one party, mentioning any of the other parties	6,000,000 tweets from October 4th to November 11th 2019	Tweets that includes any of a 46 hashtags list composed by name, abbreviation and slogan of 5 political parties, and selected trending topics	Spanish
[49]	Analyze the relation between an intention to vote and the news published in social media along 2016 USA Elections	– Sentiment analysis of news using recursive deep learning model – Use of topic detection algorithm to analyze changes in topics of media outlets	5,175 news articles from July 28th to November 8th of 2016 published in: The New York Times, Fox News, CNN and USA Today	News articles in which at least one of the two candidates involved were mentioned	English
[50]	Analysis of hate speech during Trump's 2016 presidential campaign	– Dictionary method supported by a ML model – Classifier trained with Reddit alt-right data	1.1 billion of tweets	Tweets related to 2016 presidential campaign (68%) and random tweets (31%)	English
[51]	Analysis of hate in memes during Trump's 2016 presidential campaign	– Multimodal offensive detection model using image and text	30,000 political memes with captions. 743 annotated memes	Memes selected as associated to 8 politicians participating in the 2016 presidential campaign	English
[52]	Analysis of hate speech between Biden and Trump followers	– BERT classifier to detect support or opposition to a candidate, likewise hate and offensive speech	382,210 tweets from 6 weeks before and 1 week after the election day. 3000 annotated tweets	Tweets that includes any of 20 items list composed by presidential and vice presidential candidates names or nicknames and campaign slogans	English
[53]	Analyze political discourses along the 2017 German Elections	– Perceptron algorithm to classify as hateful or safe speech	50,000 tweets from August 2017 to April 2018	Tweets related to right-wing German	German
Our	Analyze and detect hate speech in 2022 US midterm elections addressed to election officials	– Toxicity model able to calculate 14 attributes – Clusterization model – Entity recognition model able to identify 18 entities – Knowledge graphs	571,998 tweets from October 24th to November 17th 2022	Tweets mentioning or replying to a tweet from a Secretary of States official account	English

Election officials have been the target of hate speech and harassment on different occasions when elections have been discredited, e.g., when former president Donald Trump claimed the outcome of the 2020 presidential race [59]. Regarding this situation, some of the claims about the integrity and transparency of elections are based on misinformation or conspiracy theories, which may undermine the faith of the electorate. In this regard, President Joe Biden pointed out that candidates who deny the results of the 2020 elections are a threat to cyber democracy and criticized the violence against Democrats, Republicans, and nonpartisan officials who are just doing their jobs but are subjected to intimidation due to false claims from those candidates [60]. In the end, misinformation often generates hate speech and harassment against election officials.

4.2. Gathering: identifying where to get data

From multiple social media companies where citizens express their thoughts, Twitter is particularly important in the US cyber democracy, as most Twitter users are from the US, with a total of 55.1 million users in May 2023 [61]. This implies that a meaningful percentage of the population (19,1%) may be influenced or be an influencer through such social media.

Except for 1 CEO (Delaware), all other CEOs had at least one Twitter account during the 2022 US midterm elections. Some CEOs had personal and professional Twitter accounts, resulting in a total of 75 active Twitter accounts by November 2022, which accumulated thousands of followers. Only 1 personal CEO Twitter account (LChapmanEsq from Pennsylvania) was "private", allowing only reciprocally followed

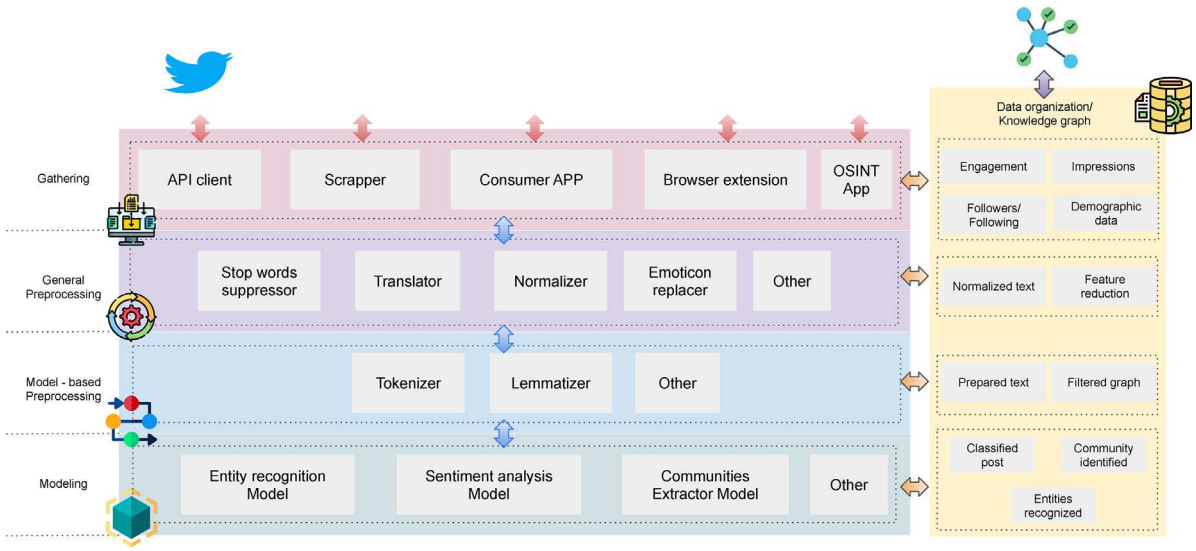

Fig. 1. Phases of the methodology adopted to identify hate speech in US midterm elections.

users could read tweets published from such accounts. Every CEO has a particular behavior on Twitter. Some frequently publish its own content, e.g., katiehobbs from Arizona, and others only retweet, e.g., ltgovmeyer from Alaska. Under this scenario, CEO Twitter accounts are accessible to citizens who want to read their tweets to be informed about their activities or to take a more active position by quoting or replying to their tweets.

There have been some cases where election officials have experienced hate speech and harassment on social media. One instance of harassment was in the 2020 presidential elections when Colorado Secretary of State Jena Griswold received threats through her Facebook, Instagram, and Twitter accounts [62]. This kind of harassment may impact election officials at different levels, not only secretaries of State but also regular election workers who are often temporarily contracted for elections, e.g., Shaye Moss, a Georgia poll worker, who received harassment after some accusations to her appeared for allegedly affecting the 2020 presidential elections results [63]. Part of the actions that social media companies have taken are enacting policies that prohibit threats against elections, including election officials, which allow users to report content that may be inappropriate. A concern about coordinated hate speech and harassment has emerged, but there is not enough monitoring to detect and prevent it.

Gathering data from Twitter may be done in multiple ways. It can be done through requests addressed to Twitter API,[5] a consumer App like TAGS,[6] a browser extension like Twitter Scraper,[7] or even an OSINT application with such capabilities like Maltego.[8] The gathering of content from Twitter depends on different criteria, such as hashtags to be collected, user accounts to consider, types of tweets (original, replies, quote), and the gathering period.

4.3. Preprocessing: preparing data to be consumed

Data preprocessing is an important step in the process of converting the previously collected raw data to the proper format expected by the NLP models and refers to the second phase of a data science life cycle. Different modules are included in the preprocessing layer, and their use will depend on how much transformation should be applied to the raw data.

Between the modules included in this layer, we use a stop word suppressor, which eliminates words in the sentence that are not meaningful. We can use a translator in case some of the collected data is not in the language expected by the NLP models. A normalizer can perform operations to standardize the text to a unique capitalization (generally lowercase) and format (vertical and horizontal writing).

Tweets collected in the previous stage, as part of the analysis of hate speech in the US midterm elections, need preprocessing accordingly to make them usable for the upcoming models in the methodology.

4.4. Modeling: extracting knowledge from data

Monitoring and validating content is one of the concerns to avoid the proliferation of hate speech and harassment messages. Thus, it is important to have mechanisms to review content efficiently but also scalable to be able to follow the high rate of new content that is being generated in social media, e.g., 350,000 tweets/minute are being generated currently. In a production environment, data models can support monitoring, specifically NLP models, which can process large amounts of data quickly and highlight notable anomalies.

4.4.1. Toxicity detection

This layer can include different NLP models in an extensible way, not being limited to a set of predefined models. However, analyzing hate speech in the US midterm elections requires incorporating some specific NLP models. Hate speech is a broad concept affecting a target due to their gender, identity, origin, political position, religious faith, and sexual orientation, among many other factors. Hate speech directed toward CEOs may be initially motivated by some political reason or by other factors e.g., sexual preference, and appears in social media in the form of apparently inoffensive toxic messages or even as threatening ones. Thus, having a module that supports the detection of toxicity in content is essential in our research.

In this regard, the Perspective API[9] is one of the most relevant classifiers that is able to validate a bunch of attributes that are meaningful in the detection of hate speech: toxicity, severe toxicity, identity attack, insult, profanity, and threat. In addition, the Perspective API calculates other attributes that may contribute to understanding the data: attack on the author, attack on commenter, incoherent, inflammatory, likely to reject, obscene, spam, and unsubstantial.

[5] https://developer.Twitter.com/en/docs/Twitter-api
[6] https://tags.hawksey.info/
[7] https://chrome.google.com/webstore/detail/Twitter-scraper/cedomiiokkcmbeoekchahgmfcppnclal
[8] https://www.maltego.com/
[9] https://perspectiveapi.com/

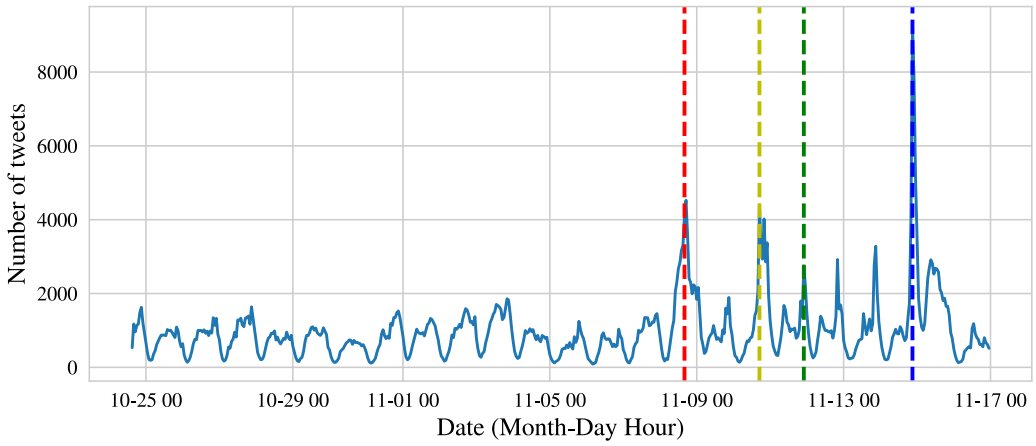

Fig. 2. Histogram of collected tweets registered per hour from October 24th to November 17th, 2022. Election day was on November 8th.

4.4.2. Entity recognition

Extracting the meaning of a Tweet may be complex as it depends on the grammar structure, the context, and the meaning of every word. Researching hate speech cases, some tweet pieces (or entities) may be more relevant than others as they may indicate aspects, such as (i) physical location where the hate speech is happening, (ii) subjects involved (stalker or hater), (iii) time variables (date, time or moment of the day) that inform when the situation happened, (iv) social events correlated, (v) organizations implicated (generating or receiving hate speech), among many others. Thus, it is quite relevant to have a module able to identify these kinds of entities in tweets to execute an eventual query over a big dataset of tweets, including some entities as criteria.

In this regard, the entity recognition model employed is obtained from spaCy.[10] This model is able to detect the following entities: Person (real or fiction), NORP (Nationalities or religious or political groups), FAC (Buildings, airports, highways, bridges, etc.), ORG (Companies, agencies, institutions, etc.), GPE (Countries, cities, states), LOC (Non-GPE locations, mountain ranges, bodies of water), PRODUCT (Objects, vehicles, foods, etc.) not including services, EVENT (Named hurricanes, battles, wars, sports events, etc.), WORK_OF_ART (Titles of books, songs, etc.), LAW (Named documents made into laws), LANGUAGE (Any named language), DATE (Absolute or relative dates or periods), TIME (Times smaller than a day), PERCENT (Percentage, including "%"), MONEY (Monetary values, including unit), QUANTITY (Measurements, as of weight or distance), ORDINAL ("first", "second", etc.) and CARDINAL (Numerals that do not fall under another type).

4.4.3. Similarity

Hate speech may start from a single person generating offensive content. However, many individuals may also coordinate it using tactics to intimidate a common target at the same time. The former situation may conduce toward meaningful damage to the individual being harassed, as they are exposed to a hostile scenario in a massive and continuous way. Thus, a module able to analyze a set of nodes generating offensive content and identify common criteria between them is important to understand the kind of coordination.

Tweets may undergo processing using K-means to obtain a set of clusters using the tweets embedding representation as criteria to do the clustering.

4.5. Data organization: setting up a consumable solution

Supplies generated by the gathering, preprocessing, and modeling layers are stored in the data organization layer. This layer is particularly important as it allows the consumption of the outcomes of the other layers. The storage is done through a knowledge graph that contains different objects (user accounts, tweets, entities), each one with its own attributes and connected to the others through specific relations, such as User A following User B, User B followed by User A, Entity A mentioned in Tweet X, Tweet *N* posted by User K, etc.

5. Detecting hate speech in 2022 US midterm elections

This section describes a set of experiments that use our methodology described in Section 4 to analyze the presence of hate speech in the 2022 US midterm elections.

5.1. Gathering data of elections

Our data collection included tweets from before and after US election day on November 8th, 2022. Thus, our period of data collection spans from October 24th to November 17th, 2022. Our data collection was mainly done using an academic API access granted by Twitter. Our collection was mainly based on two criteria: (i) replies to tweets published by Twitter accounts affiliated with the Secretaries of State, and (ii) tweets that mention Twitter accounts affiliated with the Secretaries of State.

We collected a total of 571,998 tweets, each one including the following 11 attributes: type of tweet (quote, reply, original), language (mainly English), tweet ID, date and time of tweet's creation, tweet author ID, tweet text, ID of the account that is being responded, conversation ID (i.e., ID of the original tweet being responded), entities in the tweet that were identified by Twitter, and hashtags included in the tweet.

Fig. 2 longitudinally shows the number of tweets gathered during our period of collection with hour granularity. We show that a large number of tweets (4526) were generated on election day (November 8th, 2022), specifically between 15:01 and 16:00, marked with a red dashed line. However, there are also peaks on other days, which may be explained by the occurrences of some events related to election officials: (i) November 10th, 2022 (yellow dashed line): when some election projections and results in Arizona and Nevada start generating some polemic and tension between political parties, social media and electors in Twitter [64]. (ii) November 11th, 2022 (green dashed line): reaction to the announcement on Twitter of the Republican Party of Arizona asking for transparency, certainty, and efficiency in the results in Maricopa County [65], replied by the chairman of the Maricopa County Board of Supervisors Bill Gates indicating this was offensive for elections workers [66], (iii), November 14th, 2022 (blue dashed line): reaction to the announcement of Democrat Katie Hobbs (who was also Arizona's CEO) won the gubernatorial race against Trump-endorsed Kari Lake [67], which produced many announcements and reactions

[10] https://spacy.io/api/entityrecognizer

Table 2
Tweets mentioning an account of a CEO by State.

State	Count
Alabama	212
Alaska	4
Arizona	507 551
Arkansas	38
California	385
Colorado	3902
Connecticut	452
District of Columbia	25
Florida	683
Georgia	3936
Idaho	50
Illinois	75
Indiana	48
Iowa	810
Kansas	269
Kentucky	1686
Louisiana	213
Maine	469
Maryland	11
Massachusetts	225
Michigan	24 318
Minnesota	3707
Mississippi	286
Missouri	3983
Montana	101
Nebraska	35
Nevada	455
New Hampshire	308
New Jersey	162
New Mexico	1322
New York	48
North Carolina	59
Ohio	3876
Oklahoma	391
Oregon	2545
Pennsylvania	916
Rhode Island	419
South Carolina	265
South Dakota	13
Tennessee	2915
Texas	782
Utah	2559
Vermont	128
Virginia	130
Washington	783
West Virginia	197
Wisconsin	243
Wyoming	8

on Twitter. Kari Lake had previously announced that she would not accept the election results if she lost.

The number of tweets per state is represented in Table 2, where we can see that the state of Arizona is the one with the highest number of tweets collected, representing 88.73% of all collected tweets.

5.2. Preprocessing data

First, collected tweets are filtered to keep only tweets in English. Then, the tweets are cleaned to remove unwanted characters and symbols, followed by the extraction of hashtags per tweet. Entities mentioned in tweets are converted to a dictionary format. The data types are adjusted, and the name of the State, associated with each secretary of State mentioned in the tweet, is attached. Tweet IDs are extracted, and irrelevant columns in the dataset, such as "withheld", are removed. Duplicate tweets are removed, and the index is reset before saving the preprocessed data as a pkl file. This preprocessing prepares the tweet dataset for subsequent analysis or modeling tasks. The most mentioned account was "katiehobbs" with 496,716 mentions regarding the Arizona Secretary of State, followed by "jocelynbenson" with 22,297 mentions about the Michigan Secretary of State. Also, Arizona is the State with the most amount of mentions (almost 508k).

5.3. Identifying hate speech in all states

We used the Perspective API[11] to classify six different attributes that are meaningful in the detection of hate speech: toxicity, severe toxicity, identity attack, insult, profanity, and threat. These attributes, known as production, were requested from the Perspective API, as these are indicated by Perspective as widely tested in multiple domains and trained using significant amounts of human-annotated inputs. Results from this execution are represented in Fig. 3, and a brief description of each one of these attributes is presented next[12]:

- Toxicity: A rude, disrespectful, or unreasonable comment that is likely to make people leave a discussion.
- Severe Toxicity: A very hateful, aggressive, disrespectful comment or otherwise very likely to make a user leave a discussion or give up on sharing their perspective. This attribute is much less sensitive to mild forms of toxicity, such as comments that include positive uses of curse words.
- Identity Attack: Negative or hateful comments targeting someone because of their identity.
- Insult: Insulting, inflammatory, or negative comment towards a person or a group of people.
- Profanity: Swear words, curse words, or other obscene or profane language.
- Threat: Describes an intention to inflict pain, injury, or violence against an individual or group.

In addition to the six attributes that were previously calculated, Perspective offers eight other attributes, known as New York Times (NYT) attributes, that may contribute to understanding the data: attack on author, attack on commenter, incoherent, inflammatory, likely to reject, obscene, spam, and unsubstantial. These eight attributes are considered experimental as the machine learning models used to calculate them were trained with a single source of comments, i.e., the New York Times, tagged by a moderation team associated with them. The values in the third quartile Q3 of these attributes per state can be seen in Fig. 4, and a brief description of each one of these attributes is presented next:

- Attack on author: Attack on the author of an article or post.
- Attack on commenter: Attack on fellow commenter.
- Incoherent: Difficult to understand, nonsensical
- Inflammatory: Intending to provoke or inflame.
- Likely to reject: Overall measure of the likelihood for the comment to be rejected according to the NYT's moderation
- Obscene: Obscene or vulgar language such as cursing.
- SPAM: Irrelevant and unsolicited commercial content.
- Unsubstantial: Trivial or short comments

Regarding Fig. 3, it is important to realize that the highest mean attribute value in all states, except Alaska and California, is "attack on commenter" followed by "inflammatory". On the other hand, the attribute with the lowest mean attribute value in all states is "severe toxicity". Mean may be used in a first approach to the data however, an analysis based on quartiles may be more adequate to analyze the distribution of the data, which is done in Fig. 4. Such a Figure contains the values for the third quartile of each attribute per state. In this way, we may interpret that the 25% of all collected tweets per state are above the value indicated in the heatmap. Additionally, Fig. 4

[11] https://perspectiveapi.com/

[12] https://developers.perspectiveapi.com/s/about-the-api-attributes-and-languages?language=en_US

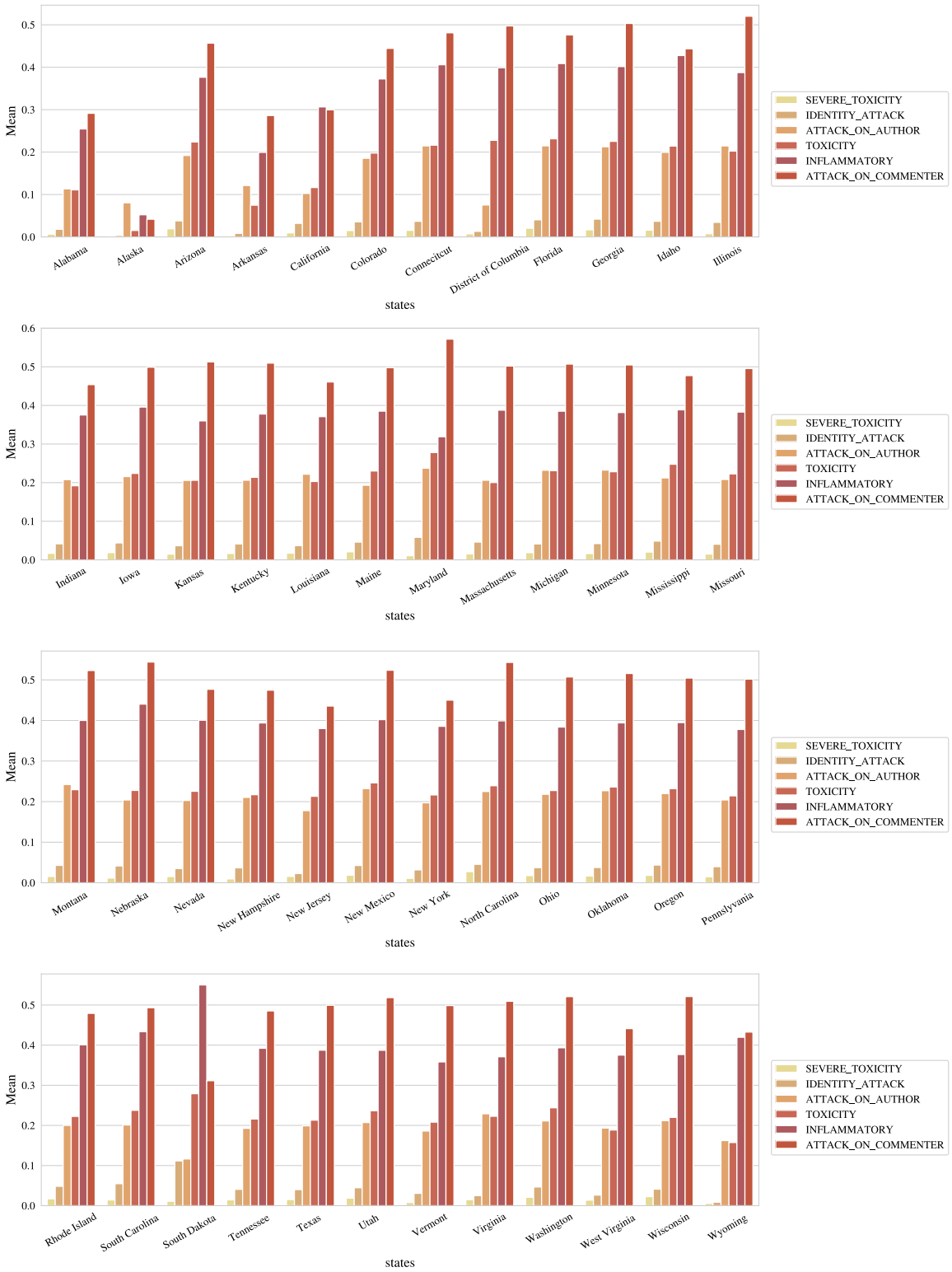

Fig. 3. Perspective results for 6 production attributes in all the States.

also shows a scale of color that indicates the number of tweets per state, being intensive dark orange with the highest number of tweets (507,551) and light orange with the lowest one (4). We may realize here that tweets that mentioned the Arizona CEO are between the ones with the highest Q3 levels of toxicity (0,4), insults (0,3), attack on commenter (0,8), likely to reject (0,9), and unsubstantial (0,8), additionally, it contains the highest amount of tweets (507,551). Other states may have similar Perspective attribute values, like Michigan, but the amount of tweets is considerably lower (24,318). Thus, high levels of Perspective attributes and the highest amount of tweets led us to choose Arizona as an interesting state to conduct a deeper review, as presented in the next section.

6. Use case: Analysis of hate speech in arizona

This section analyzes the existence of hate speech in the State of Arizona, considering the topics identified in all the collected tweets, the internal clusters that may be identified, and the entities identified for

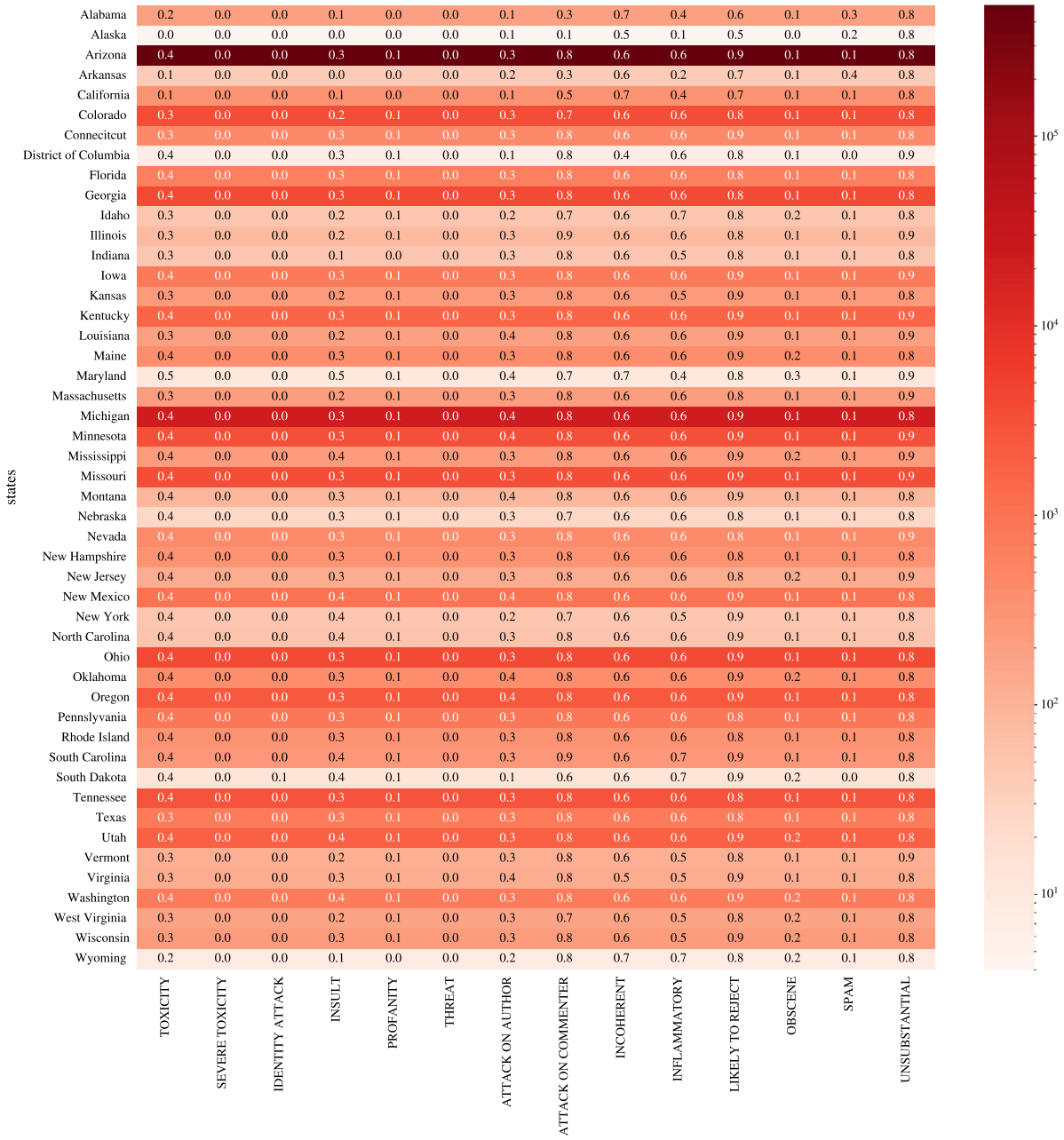

Fig. 4. Heat map of Perspective results in Q3 for production and NYT attributes in all the States.

the cluster that show the most toxic attributes. A knowledge graph for this specific scenario is also built to show the potential of the proposal in terms of providing intelligence information.

6.1. Clustering of tweets according to similarity

All of the 507,551 tweets collected for Arizona were processed using K-means to obtain a set of 5 clusters, as illustrated in Fig. 6. K-means used the embedding representation of each tweet as a criterion to perform the clustering. From Fig. 6, we may realize that clusters 0, 1, and 2 are close, making them difficult to distinguish, suggesting that they have some semantic similarity. The code snippet performs text data clustering on tweets. It first loads data and pre-trained word embeddings, converting tweets into 300-dimensional vectors. After data preprocessing and PCA dimensionality reduction, K-Means clustering with four clusters is applied. The Elbow method is utilized to determine the optimal cluster count. This is a graphical method that helps to search an optimal number of clusters, made by choosing the minor number of clusters in which the sum of the squared distance between each point of a cluster and the center of the cluster does not change significantly if we increase the number of clusters. We can see this graphically when the optimal number of clusters forms an elbow in the graphic. In our case, in Fig. 5, we can observe this behavior for 5 clusters. The code also computes the Calinski–Harabasz score for clustering quality evaluation. It visualizes results with a scatter plot, showing clusters in a reduced 2D space. Finally, the clustered data is saved for future analysis.

As seen in Table 3, cluster 4 has the highest number of tweets (308,003), corresponding also with the highest number of Twitter user accounts (101,710). The following are clusters 2 and 0, with 159,473 and 26,521 tweets, respectively. Finally, clusters with the least amount of tweets are clusters 3 and 1, with 8648 and 2342 tweets, respectively. Table 3 also includes the number of Twitter accounts associated with each cluster.

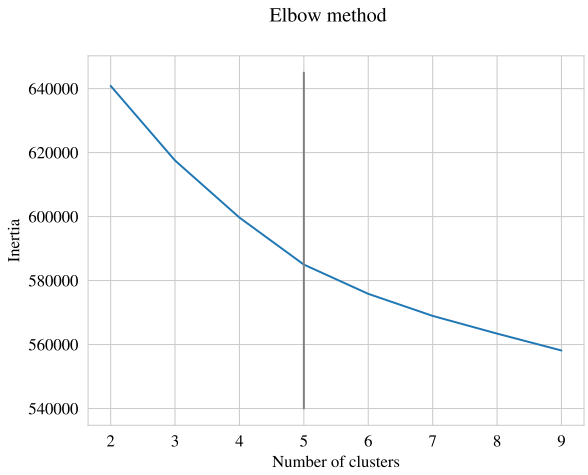

Fig. 5. Elbow method between 2 and 9 clusters.

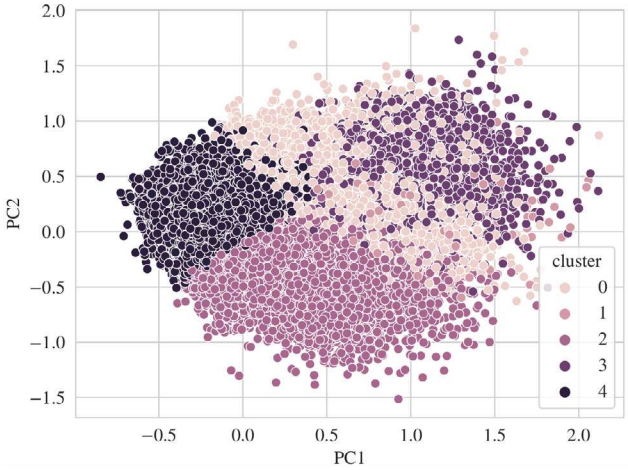

Fig. 6. Clusters of tweets collected for Arizona.

Table 3
Number of tweets and user accounts per cluster.

Cluster	Number of tweets	Percent of tweets	Number of user accounts	Percent of users accounts
0	26 521	5.25%	17 979	9.26%
1	2342	0.46%	2028	1.04%
2	159 473	31.58%	67 884	34.96%
3	8648	1.71%	4588	2.36%
4	308 003	60.99%	101 710	52.38%

6.2. Hate speech detection per cluster

Tweets belonging to each cluster were evaluated using Perspective API in terms of attributes that characterize specifically the hostility in the content of the tweet, and attributes that identify the target of the hostility. Thus, 8 content-associated attributes (toxicity, severe toxicity, identity attack, inflammatory, threat, insult, obscene, and profanity), and 2 target-associated attributes (attack on author and attack on commenter) were considered in our analysis. Fig. 7 shows the boxplot of these 10 attributes for each one of the 5 clusters.

The boxplot analysis will start on clusters 0, 3, and 1, which jointly represent only 7.43% of all the collected tweets for Arizona. Clusters 0 and 3 have some low-medium (<0.3) negative content-associated attributes in their tweets in terms of toxicity, inflammatory, and insult. Also, these tweets are mainly addressed to commenters (~0.3) and not so much to authors (<0.1). On the other hand, cluster 1 is unique as it has the second highest mean value of attack on commenter (~0.4), even if it only has inflammatory (~0.1) as a negative content-associated attribute.

Cluster 4 is the biggest cluster, with 60.99% of all collected tweets (308,003) for Arizona. According to Perspective attributes, it contains inflammatory (~0.4) tweets addressed mainly to commenters (~0.3), i.e., attacks on users who replied to the Arizona Secretary of State's posts.

Finally, cluster 2 is the one that is the most aggressive and also contains a representative amount of tweets (31.58%) of all collected tweets (159,473) for Arizona. The aggressiveness of this cluster is evidenced by the Perspective attributes that qualify the content and the target. Regarding the target, cluster 2 has the highest mean values for "attack on commenter" (0.7) and for attack on author (~0.2). Regarding the content, cluster 2 is consistent in terms of 2 detected negative attributes: toxicity (~0.2) and inflammatory (~0.3). Thus, cluster 2 was chosen for the next analysis in this section.

6.3. Recognition of entities for cluster 2

The entity recognition model employed is accessible through spaCy library.[13] Fig. 8 shows the results of applying the spaCY entity recognition model over tweets belonging to cluster 2. The 12 most relevant entities shown in Fig. 8 are Person (real or fiction), DATE (Absolute or relative dates or periods), GPE (Countries, cities, states), NORP (Nationalities or religious or political groups), ORG (Companies, agencies, institutions, etc.), PRODUCT (Objects, vehicles, foods, etc.) not including services, LOC (Non-GPE locations, mountain ranges, bodies of water), TIME (Times smaller than a day), FAC (Buildings, airports, highways, bridges, etc.), WORK_OF_ART (Titles of books, songs, etc.), LAW (Named documents made into laws) and EVENT (Named hurricanes, battles, wars, sports events, etc.).

In Fig. 8, we may observe that the results for the entity "person" are mainly around the two gubernatorial candidates for Arizona (Katie Hobbs and Kari Lake), which is not surprising as Katie Hobbs was also Arizona's CEO during the 2022 elections, so, many of the most aggressive tweets were addressed toward her criticizing her double role as a candidate and as elections official. Likewise, it is not surprising that the most found entities of type "GPE" and "NORP" are Arizona and Democrats/Republicans as these ones are associated to the State where the elections run and the two main political parties which Katie Hobbs and Kari Lake were affiliated with.

In the results for the entity "DATE", shown in Fig. 8, it is surprising that at the top of the results we find "2020", in reference to the 2020 US Presidential campaign, which feeds some conspiracy theories and misinformation campaigns against election officials. Another surprising insight comes from results for the entity "Event", where "watergate" and "zero" were at the top in reference to one announcement from candidate Kari Lake on Sunday, November 6th, 2022, where she indicated "We are at ground zero when it comes to the border, we're at ground zero when it comes to the fentanyl crisis, we're at ground zero when it comes to election integrity, crime — you name it.".

6.4. Data organization

A complete knowledge graph for cluster 2 would be large and difficult to understand, which is why we built a graph focused specifically on the most aggressive Twitter user accounts, i.e., Twitter user accounts that post tweets mentioning Arizona's CEO with a score of "Attack on commenter" and "Attack on author" >0.9 (Fig. 9). Please note that Blue nodes are Tweets, Orange nodes are Twitter user accounts, and Green nodes are Entities.

[13] https://spacy.io/api/entityrecognizer

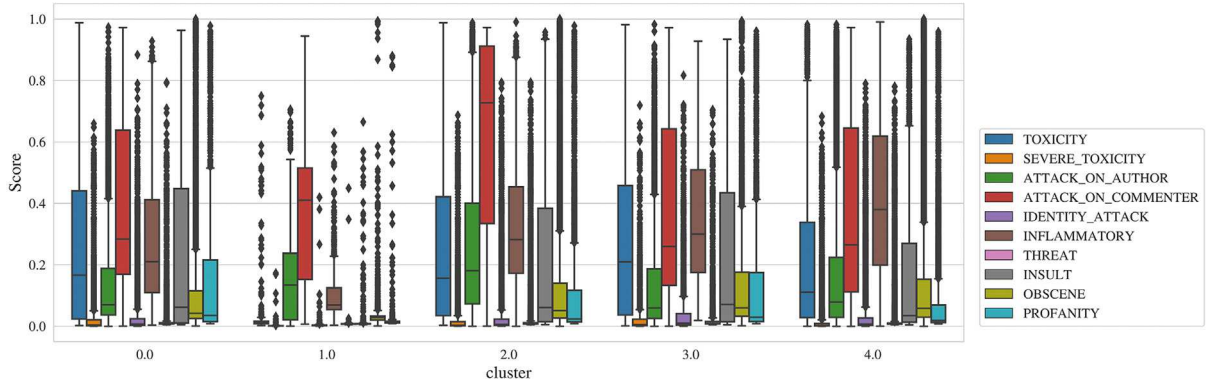

Fig. 7. Boxplot of 10 selected Perspective attributes for Arizona.

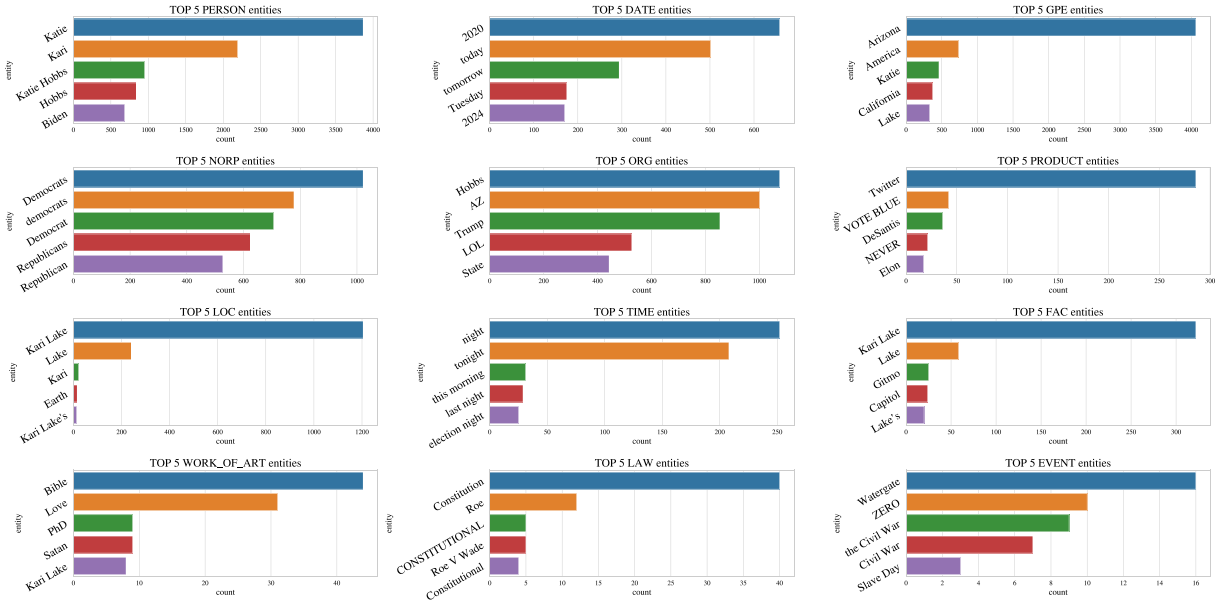

Fig. 8. Top 5 entities by category for cluster 2.

Fig. 9 contains entities identified by the entity recognition module (Section 6.3) for the 271 Twitter user accounts (cluster 2) obtained by the clustering module (Section 6.1) that has a score of "Attack on commenter" and "Attack on author" >0.9, obtained from the hate speech module (Section 6.2). As we can see in Fig. 9, there is some connectivity between these Twitter user accounts, which could be the input for a deeper analysis from different perspectives like network (to identify the connectivity of nodes promoting hate speech according to graph order, size, and density), robustness (to validate the vulnerability of the graph in case some node becomes disconnected), influence (to identify nodes that generate hate speech information and reach a group of consumers or spreader nodes), structure (to identify data paths between nodes reaching other nodes that are not even in cluster 2), temporality (to analyze the timeline behind a spread of a hate speech rhetoric), and virality (to identify nodes contributing to create trend around hate speech).

7. Further discussion

Our research suggests that it is possible to create a data collection and processing pipeline to detect and analyze harassment directed toward election officials. By further leveraging OSINT and NLP, we have demonstrated that it is possible to detect and analyze election-related hate speech on social media platforms. To the best of our knowledge, ours is the first large-scale data pipeline and quantitative study of US election official harassment.

Results indicate that from all Perspective API attributes, the highest mean score values were obtained for "attack on commenter" and "inflammatory". However, an analysis per state indicates that Arizona and Michigan contain tweets that are between the highest Q3 levels of toxicity attributes, i.e. attack on commenters, likely to reject and unsubstantial. These results are consistent between these two states despite Arizona representing a quite bigger amount of collected tweets (507,551) than Michigan (24,318). For the specific case of Arizona, 5 clusters of tweets were identified, cluster #2 (31.58% of tweets) the one with the higher mean values of aggressiveness against a Secretary of State ("attack on author" equal to 0.2) and against commenters of tweets ("attack on commenter" equal to 0.7), and characterized also by toxic (0.2) and inflammatory (0.3) content. It is important to note the relation, discovered by the name entity recognition model, between tweets belonging to cluster 2 and the 2020 US Presidential campaign, which was another scenario where election officials were victims of attacks due to conspiracy theories campaigns. The knowledge graph built for most aggressive tweets belonging to cluster 2 also shows the concentration in attacks using entities like the bible, constitution, or Watergate.

Our findings suggest that US election officials in key battleground states, such as Arizona and Michigan, appear to be receiving copious amounts of harassing messages. These elevated levels of harassment

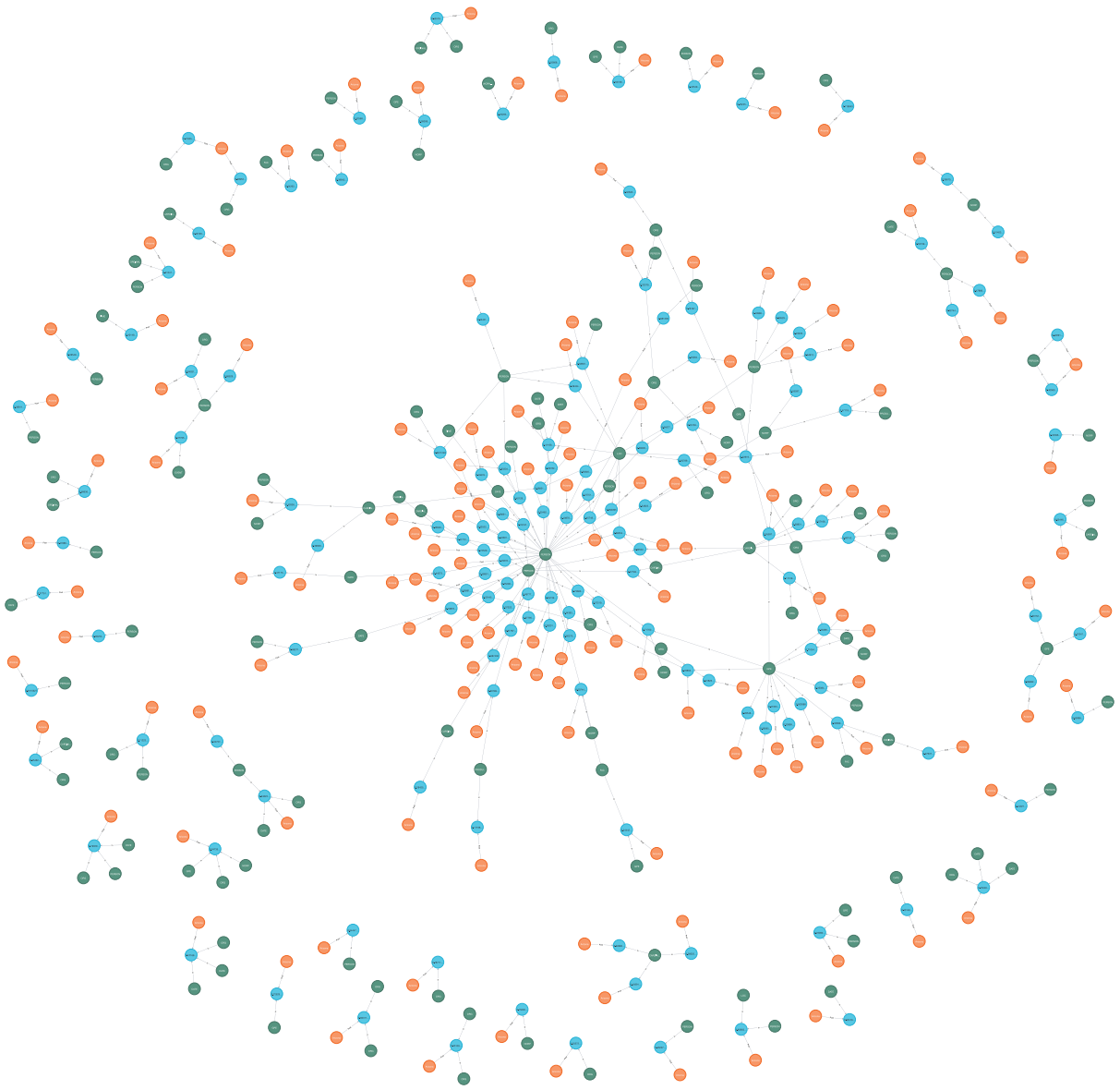

Fig. 9. Knowledge graph with entities and user accounts that post tweets mentioning Arizona's Chief Election Official with a score of "Attack on commenter" and "Attack on author" >0.9. Blue nodes (Tweets). Orange nodes (Twitter user accounts). Green nodes (Entities).

might place stress on election officials. It is important to understand this emerging phenomenon of election official harassment which might further erode cyber democracy.

Regarding the limitations of our research, the first one is related to the fact that there is no universally accepted definition of hate speech. As previously noted, different countries have various laws and policies that may interpret hate speech differently. Also, the subjective nature of hate speech may vary across different cultural and linguistic contexts and may introduce bias or potential errors in the analysis. This limitation is present in our research, as the achieved results are in some way conditioned by the evaluation of hate speech content existing in the datasets employed to build the models used in this research. Further, our study may have been limited by the accuracy and reliability of the NLP techniques used to detect and analyze the hate speech.

We also considered the ethics of conducting this field of research, such as privacy concerns of individuals mentioned in the data. Additionally, we assessed the unintended consequence of amplifying hate speech by drawing attention it through our research.

Our work could be extended in different ways, for example analyzing the long-term impacts of our findings, which include implications of the research on policy-making. This may include research into different demographics or groups, ideally to inform policy decisions. Additionally, exploring the limitations or biases in the current research may prove valuable for future studies focused on a better understanding of the dynamics of election official harassment in the US and other countries. Another promising direction is exploring innovative methods to refrain from election official harassment or to avoid social network users from spreading such type of content.

8. Conclusions and future work

The equilibrium of a democracy can be impacted by different endogenous and exogenous elements. Nowadays, we know that social media is a key influencer in most of the democratic processes, where converge politician proposals and population perceptions, but also fake news, conspiracy theories, cyberbullying, and harassment, among others.

In this paper, we showed an analysis of hate speech in social media during the 2022 US midterm elections. This analysis was performed using a proposed research methodology that integrates a gatherer, preprocessing components, NLP models, and a knowledge database to allow the consumption of the generated outputs. The analysis allows us to pose some insights into the existence of hate speech around election officials, particularly Secretaries of State. Hopefully, this kind of framework and analysis could help to establish mechanisms that may be used in the future by law enforcement agencies or social media companies to eventually regulate and prevent it.

In future work, we plan to study new NLP models that can be integrated into the proposed framework and also define new ways to correlate the outputs of each model to generate new descriptive information that may be used in a preventive and detective way.

Funding

This work has been supported by Universidad del Rosario (Colombia) through a "Beca de Estancia de Docencia e Investigación - EDI 2022-1", National Science Foundation grant 2016061, and by Democracy Fund (US). This study is also funded by the strategic project "Development of Professionals and Researchers in Cybersecurity, Cyberdefense and Data Science (CDL-TALENTUM)" from the Spanish National Institute of Cybersecurity (INCIBE) and by the Recovery, Transformation and Resilience Plan, Next Generation EU.

CRediT authorship contribution statement

Andrés Zapata Rozo: Conceptualization, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. **Alejandra Campo-Archbold:** Conceptualization, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. **Daniel Díaz-López:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing – original draft. **Ian Gray:** Investigation, Writing – original draft. **Javier Pastor-Galindo:** Conceptualization, Investigation, Writing – original draft. **Pantaleone Nespoli:** Conceptualization, Investigation, Supervision. **Félix Gómez Mármol:** Funding acquisition, Supervision. **Damon McCoy:** Funding acquisition, Project administration, Supervision, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] L. Schirch, Digital information, conflict and democracy, in: Social Media Impacts on Conflict and Democracy, Routledge, 2021, pp. 21–42.
[2] L. Luceri, A. Deb, A. Badawy, E. Ferrara, Red bots do it better:Comparative analysis of social bot partisan behavior, in: Companion Proceedings of the 2019 World Wide Web Conference, WWW '19, Association for Computing Machinery, New York, NY, USA, 2019, pp. 1007–1012, http://dx.doi.org/10.1145/3308560.3316735.
[3] M. Zago, P. Nespoli, D. Papamartzivanos, M.G. Perez, F.G. Marmol, G. Kambourakis, G.M. Perez, Screening out social bots interference: Are there any silver bullets? IEEE Commun. Mag. 57 (8) (2019) 98–104, http://dx.doi.org/10.1109/MCOM.2019.1800520.
[4] S. Cresci, A decade of social bot detection, Commun. ACM 63 (10) (2020) 72–83, http://dx.doi.org/10.1145/3409116.
[5] A.J. Patil, A. Deshpande, A comprehensive review on social botnet detection techniques, in: 2022 International Conference on Augmented Intelligence and Sustainable Systems, ICAISS, 2022, pp. 950–957, http://dx.doi.org/10.1109/ICAISS55157.2022.10010877.
[6] A. Zapata, D. Díaz-López, J. Pastor-Galindo, F. Gómez, FCTNLP: An architecture to fight cyberterrorism with natural language processing, in: VII Jornadas Nacionales de Investigación en Ciberseguridad, JNIC, 01, Bilbao, Spain, 2022, pp. 42–49, URL https://2022.jnic.es/Actas_JNIC_2022_v11.pdf.
[7] S. Kumar, A. Nagar, A. Kumar, A. Singh, Hate speech detection: A survey, in: 2022 4th International Conference on Advances in Computing, Communication Control and Networking, (ICAC3N), 2022, pp. 171–176, http://dx.doi.org/10.1109/ICAC3N56670.2022.10074044.
[8] P. Fortuna, S. Nunes, A survey on automatic detection of hate speech in text, ACM Comput. Surv. 51 (4) (2018) http://dx.doi.org/10.1145/3232676.
[9] J. Kansok-Dusche, C. Ballaschk, N. Krause, A. Zeißig, L. Seemann-Herz, S. Wachs, L. Bilz, A systematic review on hate speech among children and adolescents: Definitions, prevalence, and overlap with related phenomena, Trauma, Violence, Abuse 24 (4) (2023) 2598–2615, http://dx.doi.org/10.1177/15248380221108070.
[10] J. Pastor-Galindo, P. Nespoli, J.A. Ruipérez-Valiente, Large-language-model-powered agent-based framework for misinformation and disinformation research: Opportunities and open challenges, IEEE Secur. Privacy (2024) 2–14, http://dx.doi.org/10.1109/MSEC.2024.3380511.
[11] G.A. Marchellim, Y. Ruldeviyani, Sentiment analysis of hate speech as an information tool to prevent riots and environmental damage, in: IOP Conference Series: Earth and Environmental Science, vol. 700, (1) IOP Publishing, 2021, 012024, http://dx.doi.org/10.1088/1755-1315/700/1/012024.
[12] A.A. Siegel, Online hate speech, in: J. Tucker, N. Persily (Eds.), Social Media and Democracy: The State of the Field, Cambridge University Press, Cambridge, 2020.
[13] S. Kusal, S. Patil, J. Choudrie, K. Kotecha, D. Vora, I. Pappas, A systematic review of applications of natural language processing and future challenges with special emphasis in text-based emotion detection, Artif. Intell. Rev. (2023) http://dx.doi.org/10.1007/s10462-023-10509-0.
[14] J. Pastor-Galindo, F.G. Mármol, G.M. Pérez, Nothing to hide? On the security and privacy threats beyond open data, IEEE Internet Comput. 25 (4) (2021) 58–66, http://dx.doi.org/10.1109/MIC.2021.3088335.
[15] E. Aïmeur, S. Amri, G. Brassard, Fake news, disinformation and misinformation in social media: A review, Soc. Netw. Anal. Min. 13 (1) (2023) 30, http://dx.doi.org/10.1007/s13278-023-01028-5.
[16] Politics |, https://web.archive.org/web/20210517031610/https://socialreporter.com/category/politics/.
[17] Public participation guide: Electronic democracy | US EPA, https://www.epa.gov/international-cooperation/public-participation-guide-electronic-democracy.
[18] How the Presidential candidates use the web and social media | Pew research center, https://www.pewresearch.org/journalism/2012/08/15/how-presidential-candidates-use-web-and-social-media/.
[19] Obama and the power of social media and technology | Stanford graduate school of business, https://www.gsb.stanford.edu/faculty-research/case-studies/obama-power-social-media-technology.
[20] Y. Hua, M. Naaman, T. Ristenpart, Characterizing twitter users who engage in adversarial interactions against political candidates, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, in: CHI '20, Association for Computing Machinery, New York, NY, USA, 2020, pp. 1–13, http://dx.doi.org/10.1145/3313831.3376548.
[21] J. Pastor-Galindo, F. Gómez Mármol, G. Martínez Pérez, Profiling users and bots in Twitter through social media analysis, Inform. Sci. 613 (2022) 161–183, http://dx.doi.org/10.1016/j.ins.2022.09.046, URL https://www.sciencedirect.com/science/article/pii/S0020025522010933.
[22] Y. Hua, T. Ristenpart, M. Naaman, Towards measuring adversarial twitter interactions against candidates in the us midterm elections, Proceedings of the International AAAI Conference on Web and Social Media 14 (1) (2020) 272–282, http://dx.doi.org/10.1609/icwsm.v14i1.7298, https://ojs.aaai.org/index.php/ICWSM/article/view/7298.
[23] One in five U.S. Election workers may quit amid threats, politics | Reuters, https://www.reuters.com/world/us/one-five-us-election-workers-may-quit-amid-threats-politics-survey-2022-03-10/.
[24] Information gaps and misinformation in the 2022 elections | Brennan center for justice, https://www.brennancenter.org/our-work/research-reports/information-gaps-and-misinformation-2022-elections.
[25] In the face of threats, election workers say they feel unsafe doing their jobs : NPR, https://www.npr.org/2023/07/02/1185684663/in-the-face-of-threats-election-workers-say-they-feel-unsafe-doing-their-jobs.
[26] Election integrity | Brennan center for justice, https://www.brennancenter.org/issues/defend-our-elections/election-integrity.
[27] Brandenburg test | Wex | US law | LII / Legal information institute, https://www.law.cornell.edu/wex/brandenburg_test.
[28] Section 394-CCC - Social media networks; Hateful conduct prohibited, N.Y. Gen. Bus. Law § 394-CCC | Casetext search + Citator, https://casetext.com/statute/consolidated-laws-of-new-york/chapter-general-business/article-26-miscellaneous/section-394-ccc-social-media-networks-hateful-conduct-prohibited.
[29] NY's 'Hateful conduct' social media law blocked as unconstitutional | Techdirt, https://www.techdirt.com/2023/02/23/nys-hateful-conduct-social-media-law-blocked-as-unconstitutional/.

[30] EUR-Lex - 32008F0913 - EN - EUR-Lex, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32008F0913.

[31] Criminal code, https://laws-lois.justice.gc.ca/eng/acts/c-46/section-319.html.

[32] Racial discrimination act 1975, https://www.legislation.gov.au/Details/C2016C00089.

[33] http://juta/nxt/print.asp?NXTScript=nxt/gateway.dll&NXTHost=jut.

[34] 1501764525-THE-INDIAN-PENAL-CODE-1860.

[35] Online vigilantism in the age of OSINT — Yale cyber leadership forum, https://www.cyber.forum.yale.edu/blog/2021/7/20/online-vigilantism-in-the-age-of-osint.

[36] OSINT for anti-racism 101: An introduction to open source intelligence in the war on hate. | by DeAndre Jones | Medium, https://medium.com/@inteldeandre/osint-for-anti-racism-101-a-cursory-introduction-to-open-source-intelligence-in-the-war-on-hate-69b7b147f878.

[37] NLP as an essential ingredient of effective OSINT frameworks | request PDF, https://www.researchgate.net/publication/261051447_NLP_as_an_essential_ingredient_of_effective_OSINT_frameworks.

[38] Why natural language processing is crucial for open-source intelligence analysts | Flashpoint, https://flashpoint.io/blog/natural-language-processing-for-osint/.

[39] P. Fortuna, S. Nunes, A survey on automatic detection of hate speech in text, ACM Comput. Surv. 51 (4) (2018) http://dx.doi.org/10.1145/3232676.

[40] A. Schmidt, M. Wiegand, A survey on hate speech detection using natural language processing, in: Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media, Association for Computational Linguistics, Valencia, Spain, 2017, pp. 1–10, http://dx.doi.org/10.18653/v1/W17-1101.

[41] S.D. V., S. Kannimuthu, R. G., A.K. M., Kce_dalab@maponsms-FIRE2018: Effective word and character-based features for multilingual author profiling, in: P. Mehta, P. Rosso, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2018 - Forum for Information Retrieval Evaluation, Gandhinagar, India, December 6-9, 2018, in: CEUR Workshop Proceedings, vol. 2266, CEUR-WS.org, 2018, pp. 213–222, URL https://ceur-ws.org/Vol-2266/T4-2.pdf.

[42] V. SharmilaDevi, S. Kannimuthu, G. Safeeq, M.A. Kumar, Kce_dalab@eventxtract-il-fire2017: Event extraction using support vector machines, FIRE (Working Notes) (2017) 144–146.

[43] S. MacAvaney, H.-R. Yao, E. Yang, K. Russell, N. Goharian, O. Frieder, Hate speech detection: Challenges and solutions, PLoS One 14 (8) (2019) 1–16, http://dx.doi.org/10.1371/journal.pone.0221152.

[44] V.S. Devi, S. Kannimuthu, Author profiling in code-mixed WhatsApp messages using stacked convolution networks and contextualized embedding based text augmentation, Neural Process. Lett. 55 (1) (2023) 589–614.

[45] N.T. Solitana, C.K. Cheng, Analyses of hate and non-hate expressions during election using NLP, in: 2021 International Conference on Asian Language Processing, IALP, 2021, pp. 385–390, http://dx.doi.org/10.1109/IALP54817.2021.9675186.

[46] E. Ombui, M. Karani, L. Muchemi, Annotation framework for hate speech identification in tweets: Case study of tweets during Kenyan elections, in: 2019 IST-Africa Week Conference, (IST-Africa), 2019, pp. 1–9, http://dx.doi.org/10.23919/ISTAFRICA.2019.8764868.

[47] C.A. Calderón, G. de la Vega, D.B. Herrero, Topic modeling and characterization of hate speech against immigrants on Twitter around the emergence of a far-right party in Spain, Soc. Sci. 9 (11) (2020) http://dx.doi.org/10.3390/socsci9110188, URL https://www.mdpi.com/2076-0760/9/11/188.

[48] J. Pastor-Galindo, M. Zago, P. Nespoli, S.L. Bernal, A.H. Celdrán, M.G. Pérez, J.A. Ruipérez-Valiente, G.M. Pérez, F.G. Mármol, Spotting political social bots in Twitter: A use case of the 2019 spanish general election, IEEE Trans. Netw. Serv. Manag. 17 (4) (2020) 2156–2170, http://dx.doi.org/10.1109/TNSM.2020.3031573.

[49] F. Albanese, S. Pinto, V. Semeshenko, P. Balenzuela, Analyzing mass media influence using natural language processing and time series analysis, J. Phys.: Complex. 1 (2) (2020) 025005, http://dx.doi.org/10.1088/2632-072x/ab8784.

[50] A.A. Siegel, E. Nikitin, P. Barberá, J. Sterling, B. Pullen, R. Bonneau, J. Nagler, J.A. Tucker, Trumping hate on Twitter? Online hate speech in the 2016 U.S. election campaign and its aftermath, Q. J. Political Sci. 16 (1) (2021) 71–104, http://dx.doi.org/10.1561/100.00019045.

[51] S. Suryawanshi, B.R. Chakravarthi, M. Arcan, P. Buitelaar, Multimodal meme dataset (multiOFF) for identifying offensive content in image and text, in: Proceedings of the Second Workshop on Trolling, Aggression and Cyberbullying, European Language Resources Association (ELRA), Marseille, France, 2020, pp. 32–41, URL https://aclanthology.org/2020.trac-1.6.

[52] L. Grimminger, R. Klinger, Hate towards the political opponent: A Twitter corpus study of the 2020 US elections on the basis of offensive speech and stance detection, in: Proceedings of the Eleventh Workshop on Computational Approaches To Subjectivity, Sentiment and Social Media Analysis, Association for Computational Linguistics, Online, 2021, pp. 171–180, URL https://aclanthology.org/2021.wassa-1.18.

[53] S. Jaki, T. De Smedt, Right-wing german hate speech on Twitter: Analysis and automatic detection, 2019, http://dx.doi.org/10.48550/ARXIV.1910.07518, URL https://arxiv.org/abs/1910.07518.

[54] F. Martínez-Plumed, L. Contreras-Ochando, C. Ferri, J. Hernández-Orallo, M. Kull, N. Lachiche, M.J. Ramírez-Quintana, P. Flach, CRISP-DM twenty years later: From data mining processes to data science trajectories, IEEE Trans. Knowl. Data Eng. 33 (8) (2021) 3048–3061, http://dx.doi.org/10.1109/TKDE.2019.2962680.

[55] D. Kalb, Guide to U.S. Elections, SAGE Publications, 2015.

[56] T. Sisco, J. Lucas, C. Galdieri, The Unforeseen Impacts of the 2018 US Midterms, Springer International Publishing, 2020.

[57] How your secretary of state affects elections and why you should care | PBS News hours, https://www.pbs.org/newshour/politics/why-should-you-care-about-your-secretary-of-state.

[58] Who runs elections in your state? Use our map to find out | PBS news hours, https://www.pbs.org/newshour/politics/who-runs-elections-in-your-state-use-our-map-to-find-out.

[59] Fact checking Trump's claims about 'election integrity', April 12, 2024 | ABC news, https://abcnews.go.com/Politics/fact-checking-trumps-claims-election-integrity/story?id=109171415.

[60] Remarks by president biden on standing up for democracy, Columbus Club, Union Station, Washington, D.C., November 2, 2022 | The white house, https://www.whitehouse.gov/briefing-room/speeches-remarks/2022/11/03/remarks-by-president-biden-on-standing-up-for-democracy/.

[61] US social network users by platform 2019 - 2026 | Insider intelligence eMarketer, https://www.insiderintelligence.com/chart/260841/us-social-network-users-by-platform-2019-2026-millions.

[62] Man gets prison for threatening Colorado secretary of state jena griswold | CPR news, https://www.cpr.org/2022/10/07/nebraska-man-sentenced-threatening-secretary-of-state-jena-griswold/.

[63] Two Georgia election workers cleared of wrongdoing in 2020 elections | The guardian, https://www.theguardian.com/us-news/2023/jun/23/georgia-election-worker-cleared-trump-giuliani-vote-2020.

[64] The latest on the 2022 midterm election | CNN, https://edition.cnn.com/politics/live-news/election-results-congress-senate-house-11-10-2022/index.html.

[65] RNC, RPAZ statement on Maricopa county | GOP Twitter account, https://twitter.com/GOP/status/1591242050887745536.

[66] Nov. 11, 2022 US election coverage | CNN, https://edition.cnn.com/politics/live-news/election-results-congress-senate-house-11-11-2022/index.html?tab=Arizona.

[67] Nov. 11, 2022 US election coverage | CNN, https://www.cnbc.com/2022/11/15/arizona-governor-election-katie-hobbs-defeats-kari-lake-nbc-news-projects.html.