I love to hate! The Racist hate Speech in Social Media

Sandra Miranda[1], Fábio Malini[2], Branco Di Fatima[1] and Jorge Cruz[3]
[1]Escola Superior de Comunicação Social- IPL; CIES – ISCTE, Lisboa, Portugal
[2]Universidade Federal do Espírito Santo, Vitoria, Brasil
[3]Pontificia Universidad Católica del Equador, Quito, Equador
smiranda@escs.ipl.pt
fabiomalini@gmail.com
brancodifatima@gmail.com
JACRUZ@puce.edu.ec

Abstract: It is undeniable that, nowadays, hate speeches have flourished, they have become every day, banal and available to everyone. The interaction allowed by the use of devices, the potential of ICT and Social Media, formed a new participatory culture and contributed to rethinking social dynamics. The supposed illusion of anonymity and the rapid dissemination of narratives and images opened space for the proliferation of p hateful discourses against minority groups, such as those of a racist nature. This research intends to study, from the Digital Social Networks (DSN), the communicational flows of racist hate speeches in Portugal. Specifically, we propose to diagnose racist hate speech in the DSN; understand and characterize the narratives that support the spread of racist hate speech in DSN; and analyze which content and hate narratives generate more engagement in DSN; In this study, a mixed convergent methodology will be adopted. The quantitative approach will combine the use of digital methods with the analysis of social networks and graph theories. There will be 2 panels of 2 social networks (Facebook and Instagram), totaling 24 months of data collection. The qualitative approach will resort to the content analysis of the comments. In terms of results, we intend to strengthen scientific production in the area and develop a barometer on racist hate speech in Portugal.

Keywords: Hate speech, racism, digital social networks, engagement

1. Introduction

This research aims to determine how the terminology used in hate speech is configured in digital environments; changes in human relationships (with their agreements and confrontations) and the way they are presented in DSN constitute a fertile field of analysis. The research proposes several objectives aimed at this line of study: to generate a diagnosis of how these discourses are exchanged and impact the different actors and relationships in a social network; monitor and visualize how terms are related and build semantic networks; and, finally, to indicate how the contents generate different and particular reactions (engagement).

In the first part of the document, the review of the state of the issue shows how the concepts and terms of hate speech become elements that are still complex to identify and classify due to the magnitude of the issue. On the other hand, it is highlighted how the concepts of public sphere, social acceptance and resonance are evidenced by the media's ability to reproduce content and expand the scope of the narratives' intention.

In the description of the method, it is possible to identify how the qualitative and quantitative approaches generate an ideal environment for a discursive analysis that also comes from the accounting of digital content on social platforms. As can be seen, the analysis period (24 months) offers enough time to determine behaviors and patterns; furthermore, the affectation of Covid-19 or the lack of its impact in this regard is brought to the fore. The results offer several elements to consider, on the one hand there will be the reach and impact that the publications and the discourse -content in the terms- will be able to mark in the two years of analysis. On the other hand, both the hating actors and the subjects of their hatred appear represented both in the semantic networks and in the terms found.

Finally, it should be noted that the research seeks to be a turning point and helps in the search for effective answers against a hater culture that recruits new followers; data are offered that show that hate speech must be reflected from the political and attitudinal position of its managers.

2. Literature Review

The popularization of Information and Communication Technologies (ICT) has driven remarkable transformations in the way citizens participate in public life. In this new scenario, the network society and the internet, reorganized the world space and knocked down borders, giving us the power and the possibility to

consume, create, communicate and distribute content in a space of participation and free personal expression. But if, on the one hand, interaction on the internet is a determining factor of organization and social structuring, interfering in the exercise of citizenship, in political, social and economic relations, it is also fertile ground for expanding conflicting aspects of reality and relationships. social, such as hatred and all its manifestations (Pezzella and Borba, 2012; Santos, 2015).

For Timofeeva (2003) the internet created the right conditions and provided exclusive resources to expand the verbosity of hate. It is a relatively inexpensive and highly effective tool for racist individuals or groups to spread hateful ideas to an audience. Dias (2007) adds that the production, legitimation and reproduction of these discourses can be seen on the internet. For the researcher, the content of hate speech eliminates or minimizes the communicative character, since messages, when expressed, are no longer received as messages and are interpreted and felt as attitudes and behaviors.

Revisiting the literature, we find that there is no universal definition of hate speech, with different multidisciplinary understandings and approaches to the phenomenon coexisting between them (Silva, 2018; Costa, 2020; Fernandes, 2020). According to Brugger (2007, p. 21), the term originates from the English hate speech and is one that includes words that tend to insult, intimidate, or harass people because of their race, color, ethnicity, nationality, sex. or religion, among other attributes. It is an attitude of systematic hatred and aggression towards the way of being, lifestyle, beliefs, and convictions of an individual or group of individuals having "the capacity to instigate violence, hatred or discrimination against such persons". Meyer-Plufg (2009) expands the debate by extending intolerance and discrimination by gender, sexual orientation and identity, stating that discrimination and externalization are the two elements that configure the hate speech. In his words, the problem arises when thought goes beyond limits and hatred takes shape through words.

Wieviorka (2007) and Bourdieu (2012) also addressed the issue, for whom hate speech perpetuates a violence, above all, symbolic that is perceived through language and discourse and whose effects can be maintained in this context or extravasated it, moving on to physical violence.

But the topic of hatred and its exultation in the public sphere is not new! It is enough to remember Mandevillhe (1715) and his famous work *Fable of the bees*, pointing out that some private vices, such as hate, are acquiring public acceptability. What is certain is that in the last twenty years, this vice of public demonstration of hatred will have increased (Waldron, 2010), favored by the explosion of Digital Social Networks that work as a resonance chamber, propagating and amplifying its effects; and radicalizing the conflicts of social reality (Daniels, 2008). The supposed anonymity of the haters, the absence of a face-to-face interlocutor and the isolation at the time of constructing argumentative reasoning favor the distillation of hate in the posts, in the comment boxes or in the memes, pregnant with irony. With just a few clicks, a situation of non-recognition on a vast scale is installed, which offends an uncontrollable number of people and calls for countless other internet users to perpetuate this asymmetry (Boyd, 2010).

Busso (2011) when commenting on the unbridled increase in manifestations of hate on DSN states that the effect of hate speech narratives is all the more the greater the diffusing power of the medium. Baurin (2017) argues that haters use social networks to promote their hateful cause and recruit new members, using the most diverse means such as free music downloads with hateful messages, racist games, cartoon characters, images of people maimed for racial or nationalist reasons and messages aimed primarily at young people and children.

It is certain that, despite being one of the most pulsating themes today, in Portugal this narrative source is little studied (Costa, 2020). For this reason, recently, several national institutions (FCT; ACM), in the wake of the UN recommendations (2020), launched a series of initiatives, aimed at action research, recommending the reinforcement of knowledge in the collection of information and analysis of trends about of hate narratives on digital platforms to support effective responses. All these reasons emphasize the relevance of carrying out the present investigation, a multifaceted area of study with an immense space for progression. The objectives are: to diagnose racist hate speech in digital social networks; monitor, understand and characterize the narratives that support the propagation of racist hate speech on digital social networks; indicate which content and hate narratives generate more engagement in digital social networks.

3. Methods

The methods are based on a quantitative and qualitative approach organized in three stages: data extraction, mining, and data visualization from social media platforms. Our team developed a racist speech barometer, consisting of two panels - Facebook and Instagram -, to answer the goals.

From the literature review and cases reported by the media, our team created a list with 35 most frequent terms associated with racism. The intention was to map the vocabulary universe of hate speech against minority groups, based on their origin, religion, or ethnicity. The list includes words directly associated with racism in Portuguese (chink, nigger, monhe, dark-skinned, kike, Jewish quarter, etc.) and compound terms (shit gypsy, bastard nigger, naughty nigger, etc.), but also expressions that favour a more detailed analysis of this phenomenon, such as gringo (slur for foreigner), black, monkey, catinga (slur for nigger) and mulatto. Some of these slurs are difficult to translate because they lose their original meaning.

Data were extracted with CrowdTangle, a tool from Meta company, configured in the Search module. The advantage of this approach is to have access to official Facebook and Instagram data via Application Programming Interface (API). Each term of the list was introduced individually in CrowdTangle, following the tool parameters. For Facebook, the search was delimited by the Portuguese language and country of the pages manager, in our case Portugal. For Instagram, the search was determined only by the language, once the tool does not allow the configuration by place. The result of each term was stored in an individual CSV file.

The sampling period covers 24 months between March 2020 and March 2022, totaling 660 messages from one hundred of social media accounts. These social data are very diverse. It includes the day and time of the publications, URL address of the posts, pages IDs, the potential category of the content, the total number of reactions (likes, sad, love, angry), etc.

Data mining considered two procedures. In the first, the several CSV files were integrated into single databases, according to the platform. In the second, our team read the messages and eliminated information that does not add value to the goals of this study, such as duplicate posts and in other linguistic variants, such as Brazilian and Angolan Portuguese.

The databases were processed with filters and algorithms in open-source software, in an aggregated approach of the social media pages. Voyant Tools was used to determine the frequency of the terms and the association between words by co-occurrences. An analysis tool called Gephi was used to design the network narratives. Our team has also used the Modularity algorithm to delimit the network clusters, the Force Atlas 2 layout to approximate the terms according to their relations in the speech, and the Weighted Degree metric to highlight the most frequent words.

Alldata collected will be stored on a Solid-State Disk (SSD), protected by password, and its access and use must be previously approved by the PI. Whenever necessary, the dataset will also be submitted to an irreversible anonymization process, in accordance with the recommendations of the General Data Protection Regulation (2016/679), in force since May 2018 in the European Union.

4. Results

Our data reveals at least 213 racist messages on Facebook, published between March 2020 and March 2022 (2 years), by 74 accounts that potentially spread cyberhate in Portugal. These fanpages have a very diverse reach, gathering around 15 million supporters. Among them are accounts with more than a million followers, such as *Humor Tuga*, and others with only a few thousand, such as *Ergue-te Coimbra*, *Templarios Portugueses* or *Reconquista 1143*. Not all these pages have as their primary goal the propagation of hate speech. However, at some point, their messages have contributed to the construction of cyberhate, by boosting slurs or reporting cases of discrimination. Generally, the smaller accounts, with a few thousand supporters, tend to be openly hateful.

These fanpages are classified into 15 categories, according to the platform standards. The most representative cases, with five or more occurrences, are activity general (16.2%), media news company (8.1%), news site (8.1%) and entertainment website (6.8%). The high number of categories shows that the hate speech phenomenon mobilizes several social segments. The categories with a lower frequency in the sample – sports (4.1%), political

party (2.7%) or artist (1.4%) – also contributed to the conception of hate imaginary or to the propagation of slurs.

Almost 70% of these fanpages started operating from 2013, in the aftermath of the austerity policies adopted by the Portuguese government in facing the economic crisis. The most significant years are 2014 (17.6%) and 2015 (14.9%). Older fanpages, with ten years or more, represent 39.2%. On the other hand, the effects of the Covid-19 pandemic do not seem to have generated a significant number of hate accounts. On the national level, only 5% of the pages in our sample have been created since 2019. In a digital landscape marked by the crossing of languages, it is interesting to know which content formats are posted. Table 1 shows the most frequent types of messages. Generally, these publications reveal, among others, the visual impact of a narrative, sometimes being more textual or imagetic.

Table 1: Messages posted by fanpages (n = 213)

Contents	Posts (n)	Posts (%)
Photo	133	62.4
Link	60	28.2
Native Video	15	7.0
Status	3	1.4
YouTube	2	0.9
Total	213	100.0

Source: Facebook Graph API via CrowdTangle

A significant fraction of hate messages are images in different formats, such as drawings, photos, and memes (62.4%). In CrowdTangle's metrics, these images are grouped by the same category, even though they deserve specific research. Shared links occupy a considerable space in our sample (28.2%). These URLs tend to take fans to diverse sites, such as news portals, personal blogs, football sites and online shops. Audiovisuals, grouped in the Native Video and YouTube categories, are not very representative (7.9%). Status changes, with information about the pages, occurred punctually (1.4%). A case of these descriptions, highly nationalistic, can be read on page *Reconquista 1143*: "Reality seen as never before, through the beacon of truth. Sentinel at the vanguard of information stripped of political correctness and single-mindedness. Omnia pro Patria (All for the Homeland, in Latin)".

The message, whether photo or video, establishes the bridge between the fanpages manager and its followers. This virtual relationship is created by the fans' interactions. The publications in the sample attracted almost 110 thousand reactions, such as likes, comments or shares. In an equal calculation among the fanpages, the engagement rate is relatively low (0.69%). However, the potential reach of the pages and their number of messages are diverse. This makes it difficult to calculate the exact engagement for the whole sample. Table 2 presents the most frequent reactions by followers. The data also shows what kind of content has mobilized the network audience.

Table 2: Frequency of reactions on fanpages (n = 109,211)

Reaction	Nº	%
Likes	41,964	38.4
Haha	19,842	18.2
Comments	18,005	16.5
Shares	14,265	13.1
Love	7,571	6.9
Angry	3,937	3.6
Sad	2,451	2.2
Wow	794	0.7
Care	382	0.3
Total	109,211	100.0

Source: Facebook Graph API via CrowdTangle

The Like is the most common reaction among users of pages spreading racist posts (38.4%), followed by Haha (18.8%). In some cases, these reactions are associated with recreational racism, irony and debauchery against victims of slurs. In a considerable distance from the first place, Comments (16.5%) and Shares (13.1%) tend to indicate the deepest engagement with the messages. They require more mental effort or desire to propagate a

cosmology. Reactions associated with emotions appear in less than a fifth of our sample, such as Love (6.9%), Angry (3.6%) and Sad (2.2%). It is curious to note that the reaction Care (0.3%), related to the idea of caring for someone or asking someone to be careful, is the least frequent. Thus, many of these fanpages are not supporting someone, but attacking something.

Reactions suggest a pattern of effort. The less dedication required by a reaction creates a more frequent behaviour. This would explain why the Like comes first, to the detriment of Comments or Shares. On the other hand, each message deserves a personalized analysis, and can reveal the values, discords and worldviews of the follower. In any case, the word in these messages helps map the narratives about racism. Figure 1 shows the graph of the most frequent terms in Facebook posts, based on the co-occurrence of the words. This means that the graph allows identifying which words appear together in the same narrative, based on a statistical approach: Modularity algorithm and Force Atlas 2 layout.
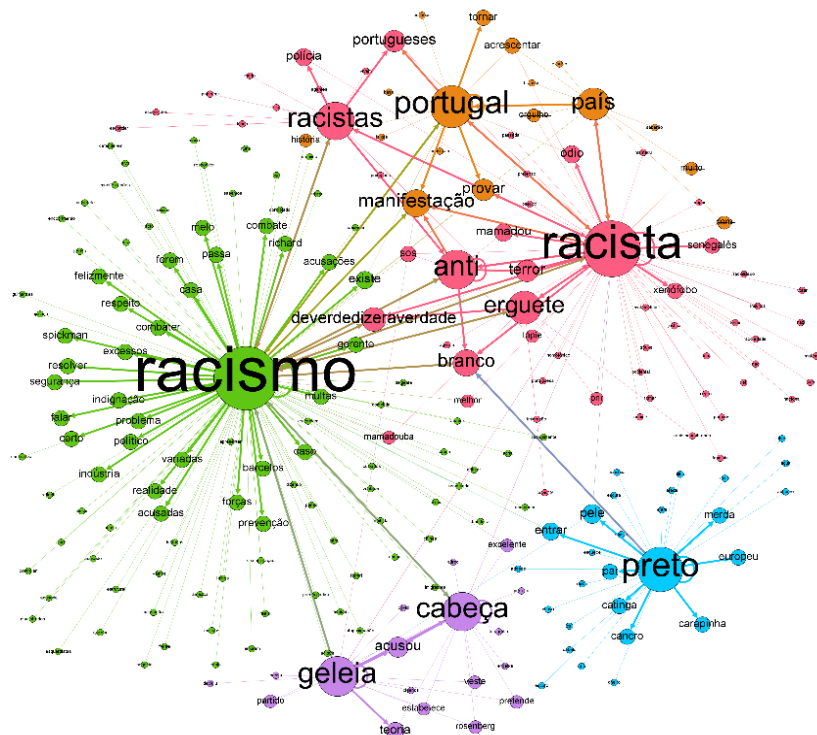

Figure 1: Graph by co-occurrence of posts words

Source: Facebook Graph API via CrowdTangle

The graph shows the emergence of five clusters organized, by their respective colours, into narrative communities. The nodes or points indicate the words, while the edges point out the relations between them. On the one hand, the size of a node indicates how often a word is used. On the other hand, the thickness of an edge reveals the proximity between terms. Each of the clusters tends to suggest a reading, sometimes diverse and opposite, of how racism is discussed.

The green cluster (45.5%) represents a view of racism from profiles with a nationalist background, attachment to traditions and the anti-immigration discourse directed at people from African and Asian. There is a relationship between terms like "respect", "to criminalise", "injustice", "to condemn" and "to fight" against anti-racist activists. The speech is characterized by compound expressions such as "stop immigration" or "anti-white racism". Some of these messages say: "the industry of racism does not stop criminalising white people", "the best way to fight against racism is (...) to ignore that it exists" or "another affront to our glories, mainly of commerce, turned into slave victimhood".

Attacks against Portuguese public personalities of African origin are more notorious in the pink cluster (26.3%). However, this discourse is directed at state institutions in an attempt to link immigrants to welfare aid and thereby the spending of taxpayers' money. Words like "national", "live", "xenophobia", "vote" and "truth" are

associated. Harsh criticism of left-wing activists and the press contrasts with praise of far-right parties such as *Ergue-te* and *Chega*. Some of these messages say: "European governments, Kings of Multiculturalism and of freedoms(?), what are they afraid of?", "the supposed anti-racist associations that live off the public funds" or "where is the sold and whore media?".

The green and pink clusters group more than 70% of the racist terms in our sample. In common, these messages attack public figures of African origin, such as Joacine Katar Moreira, Mamadou Ba and Marcelino da Mata. In the green cluster, the focus is on people. In the pink, the attack also targets government policies and anti-racism NGOs. The other three clusters represent 28.1% of the cyberhate, and they mimic hate arguments from the green and pink. The notable difference is in the blue cluster (12.5%). The narrative has an association with physical attributes, such as hair, skin and sweat. The discourse originated from the words of a participant in the reality show *Big Brother*: "black people are the ones who smell catinga" (a very bad smell). From there, the debate is woven from support to repudiation.

The nationalism shaping messages on Facebook is more visible on many of the 26 Instagram pages in our sample. These accounts posted 447 potentially hate speech messages, and may have reached over 80,000 followers – it is difficult to calculate that impact. Despite their diverse reach – from hundreds of followers, such as *Spam_Politico*, to thousands, such as *Super_Indignado*, these pages share a visual identity. Homeland signs are used as a profile image. Figure 2 shows that the preferred illustration is the national flag, from different eras, emphasising the Cross of the Order of Christ, an important symbol to the Portuguese.

Figure 2: Instagram profile image of the sample (n = 26)

Source: Instagram Graph API via CrowdTangle

In some cases, the images refer to the monarchic periods, to the Portuguese colonization and to the *Estado Novo* – period of fascist dictatorship in the country. Texts, superimposed on the images, allude to the values of the page managers – "I am patriotic" – and to the mythical space of a people in the world – "Portugal for Portuguese" –, as opposed to migratory flows and to the multicultural idea of immigration. Table 3 reveals the most posted formats by these accounts since their launch.

Table 3: Message published on Instagram

Contents	Posts (n)	Posts (%)
Photo	8,534	80.3
Video	1,188	11.2
Album	900	8.5
Total	10,662	100.0

Source: Instagram Graph API via CrowdTangle

Another similarity is the method of content management. Images in various formats, such as cartoon, photo and meme, are the main material shared (80.3%), but their fraction is higher than on Facebook. Videos are also not very representative (11.2%), followed by albums (8.5%) – a compilation of images released to build a narrative.

Not all of these publications are directly related to racism, although they are interspersed with other cyberhates and attacks against minorities (homophobia, sexism, ageism, etc). The search for the most frequent terms associated with racism in the literature review points out that this phenomenon is multidimensional and difficult to map. Therefore, our team also wanted to know what discourse is created and mimicked by these Instagram accounts.

Figure 3 shows the graph of the most frequent words in the messages of the sample (n = 447), based on co-occurrence. Two clusters dominate the narratives about racism on Instagram. Altogether, they represent 83.7% of the conversations. Each of these narrative communities (blue and red) has at least one cosmology of the phenomenon, sometimes marked by dependency and opposition. Although several expressions and terms link these clusters, such as "victims", "kick someone out", "SOS" and "live", many in an allusion to anti-racism NGOs, "white" is the strongest connecting word.
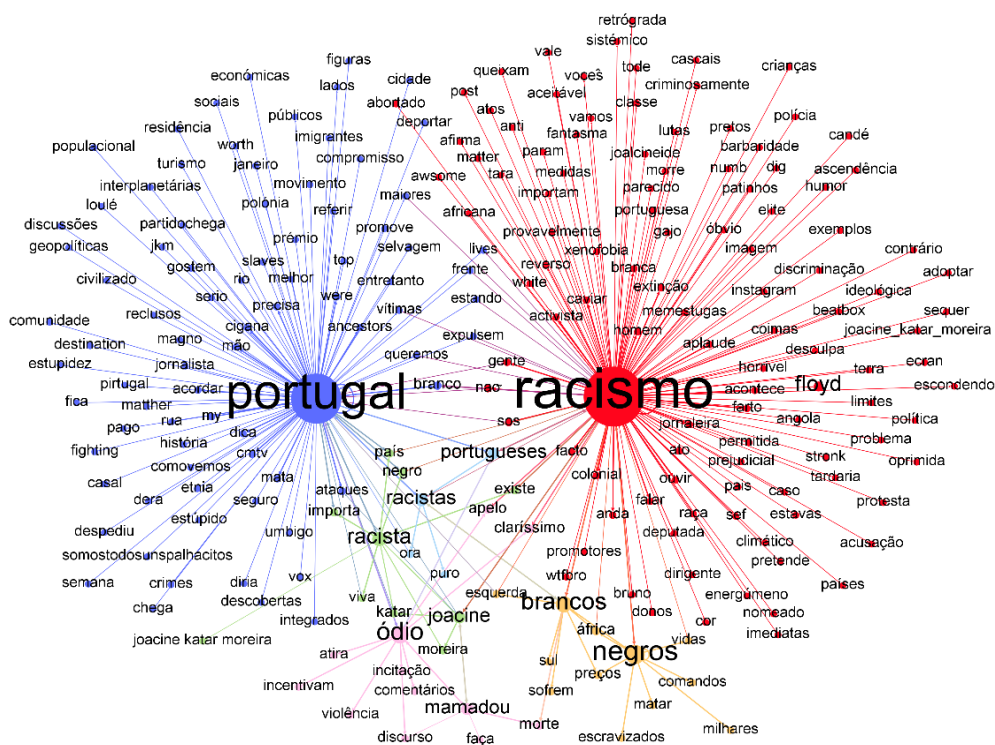

Figure 3: Graph by co-occurrence of posts words

Source: Instagram Graph API via CrowdTangle

The five most recurrent words in the graph are racism (46), Portugal (36), racist (26), white (23) and black (19), in a clear ethnic conflict spread by these pages. Other terms, such as Joacine (16), Mamadou (16) or hate (16), suggest that anti-racism activists from African origin amplify anger with their actions. In that case, interviews and lectures by these personalities tend to be decontextualised.

The red cluster (48.1%) systematically denies the existence of racism in Portugal, in a crossover with other conservative agendas and criticism of left-wing groups. There is a relationship between words like "die", "abortion", "colour", "race", "anti" and "caviar", in an attempt to ensure that white people also experience discrimination. Some messages point to South Africa as an example. Colonialism is another common topic. It is sometimes characterised by criticism of the former Portuguese colonies in Africa, sometimes in a combination of nostalgia and recreational racism. Some of these posts say: "we don't want African immigrants in Portugal, however integrated (…)", "this was the racism of the Portuguese in colonial times. The famous racism!😁" and

"if the Portuguese had stayed in Angola, it would be better than Dubai". There is also the use of English phrases in an effort to connect with transnational far-right movements.

The blue cluster (35.6%) confirms the international approach, also with English phrases. Posts tend to associate immigrants from various ethnic origins with deviant behaviour. There is a relationship between terms like "ethnicity", "gypsy", "Africans", "Islamism", "deport" and "crimes". It is publicised that Portuguese immigration policies are soft, and that legal punishment does not exist against immigrants. There is a clamour for the emergence of right-wing political parties: "I want a VOX in Portugal!". Other messages say: "a group of Roma people causes chaos on the highway", "to deport that savage was too little" and "since when is Islam a part of Portuguese culture?".

5. Conclusions

Findings of this research invite us to reflect on some issues: the scope of the publications, the type of behavior of the audience, the purposes of these actions, and the future of this line of analysis.

First, although the selected platforms showed a considerable scope and significant numbers of reactions, when the depth and impact of hate speech is reviewed, it becomes clear that this type of content coexists -as it happens in digital environments- with more information than diminshes strengh to the effects desired by the producers of this discourse. Also on the platforms discussion, it is interesting to observe how Facebook and Instagram behave in a similar way regarding the majority of the content. It will be vital, after the results of Instagram, to map how the use of photographs is promoted in social networks -such as this one-, which appeals to the most graphic content and that narrative alternative can generate greater impact on the public.

Regarding the behavior of the audience, the idea of the ephemeral is fundamental; users prefer to place "likes" than to comment on posts or even share them with others. The interaction with the content is then marked from a first -very basic- level. This behavior could generate content power, but explain that the final reach does not become exponentially viral. This does not mean that hate speech should be dismissed, based on this variable, but rather it allows us to contextualize how this narrative is reproduced and how strategies can be generated to counteract it. Another element on how the audience and the producers themselves relate to the content is in the linguistic characteristics. On the one hand, it is clear that the grouping of terms is a complex task since it is developed from a folksonomic perspetive more than from a taxonomic standpoint. To this can be added to the interest of studying this phenomenon from the appropriation of anglicisms, a posible line of research in this field.

Finally, regarding the intention of the generators of this discourse, it can be seen that much of the narrative seeks to attack the defenders of migratory and equitable movements. In a -perhaps simpler- vision, one could speak of right or left discourses colliding; This line of discussion also offers spaces for additional and deeper inquiry, particularly because hate speech and its behavior show that the dynamic is developed to seek followers. In order to reduce its impact it must be understood, modeled and thus proposals and responses must be built to combat hate speech, the fundamental basis for all the purposes related to this research.

Funding

References

Abbott, A. (2016). *Processual Sociology*. Chicago: University of Chicago Press.

Arendt, H. (2014). *Homens em tempos sombrios*. Trad. Denise Bottmann. São Paulo: Companhia das Letras.

Alkiviadou, N. (2019). Hate speech on social media networks: towards a regulatory framework? *Information & Communications Technology Law*, *28*(1), 19-35. DOI: https://doi.org/10.1080/13600834.2018.1494417

Barab si, A. (2017). *Network science*. Cambridge: Cambridge University Press.

boyd, d. & Crawford, K. (2012). Critical questions for big data. *Information, Communication &Society*, 15(5), 662-679.

Cabral, A., & Lima, N. (2018). Interações conflituosas e violência verbal nas redes sociais. Lisboa: Polémica.

Castells, M. (2007). *A sociedade em rede. A era da informação: Economia, Sociedade e Cultura*. Lisboa: Fundação Calouste Gulbenkian.

Costa, P. (2020). Uma cartografia do ódio. *Comunicação Pública*, 15(29). 22-34.

Fernandes, P. (2019, 14 de outubro). A cultura de dio. Observador. Disponível em:https://observador.pt/opiniao/a-cultura-de-odio/ [Acedido em outubro, 30, 2019].

Jenkins, H. (2009). *Cultura da convergência*. São Paulo: Aleph.

Jünger, J. & Keyling, T. (2019). *Facepager. An application for automated data retrieval on the web*. [Software]. Github.

Levallois, C. & Totet, M. (2017). *Twitter Streaming Importer*. [Software]. Github.

Mandeville, B. (1715). *A fábula das abelhas, vícios privados, benefícios públicos*. Brasilia: UNESP.

Pereira-Kohatsu, J.C. et al. (2019). Detecting and Monitoring Hate Speech in Twitter. Sensors, 19(21), 1-37. DOI: https://doi.org/10.3390/s19214654

Rieder, B. (2015). *YouTube Data Tools* (Version 1.11). [Software]. Digital Methods Initiative.

Rogers, R. (2013). *Digital methods*. Cambridge: The MIT Press.

Santos, M. (2016). *O discurso de ódio nas redes sociais*. São Paulo: Lura Editorial.

Silva, R., Nichel, A., Martins, A. C. L., & Borchardt, K (2011). Discursos de ódio em redes sociais: Jurisprudência brasileira. *Revista Direito GV*, 7(2), 445-468.

van Dijk, Teun A. (2012). The Role of the Press in the Reproduction of Racism. In Messer M., Schroeder R., Wodak R. (eds) *Migrations: Interdisciplinary Perspectives*. Vienna: Springer.

Vigna, F.D. et al. (2017). Hate me, hate me not: Hate speech detection on Facebook. In: Proceedings of the First Italian Conference on Cybersecurity (ITASEC17), pp. 86-95.