Inflaming public debate: a methodology to determine origin and characteristics of hate speech about sexual and gender diversity on *Twitter*

Sergio Arce-García; María-Isabel Menéndez-Menéndez

Nota: Este artículo se puede leer en español en:
https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/86993

Recommended citation:

Arce-García, Sergio; **Menéndez-Menéndez, María-Isabel** (2022). "Inflaming public debate: a methodology to determine origin and characteristics of hate speech about sexual and gender diversity on *Twitter*". *Profesional de la información*, v. 32, n. 1, e320106.

https://doi.org/10.3145/epi.2023.ene.06

Article received on May 10th 2022
Approved on September 06th 2022

Sergio Arce-García
https://orcid.org/0000-0003-0578-9787

Universidad Internacional de La Rioja
Escuela Superior de Ingeniería y Tecnología (ESIT)
Avda. de la Paz, 137
26006 Logroño, Spain
sergio.arce@unir.net

María-Isabel Menéndez-Menéndez
https://orcid.org/0000-0001-7373-6885
Universidad de Burgos
Facultad de Humanidades y Comunicación
Paseo de los Comendadores, s/n
09001 Burgos, Spain
mimenendez@ubu.es

Abstract

This article is focused on the reproduction of ideologically charged messages whose origins or interests remain hidden from public opinion. There is an urgent need for transparency regarding polarised debates that deform, impede or distort the critical approach that any society should be able to construct concerning issues of great social interest, especially on social media platforms and networks. Research has shown that hostility has colonised digital communication through misogynist, homophobic, transphobic or xenophobic messages, among others, and that, for the most part, these are not spontaneous or individual interactions. In the virtual space, there are forces that, although invisible outside it, construct narratives, generate disinformation and feed generally regressive ideological approaches. Thus, in the name of transparency and social justice, there is an urgent need to investigate these types of messages, as well as their possible destabilising interests at a time of special presence and reputation of discourses such as the feminist one, which is currently experiencing a significant reactionary response. This paper investigates the origin and characteristics of the conversation on the social network *Twitter* concerning gender and sexual identities. To this end, we studied a significant sample of tweets (>1 million) related to women's rights, the LGBTIQ+ collective and trans people, for a full year. Computerised methodologies by means of machine learning techniques, natural language processing (NLP), determination of bots, geolocation, and the application of network theories were used to carry out the study. The results include the highly interrelated presence of groups without clear referents, as well as the existence of what appear to be coordinated networks aimed at causing harm and provoking confrontation.

Keywords

Hate speech; Misogyny; Social media; Social networks; Social networks analysis; Transphobia; Feminism; LGBTIQ+; Gender diversity; Sexual diversity; Transgender; Transsexuality; Transgender; *Twitter*.

1. Introduction

The Internet today is the public square, the place where debate, community and reputation are built. However, the third decade of the twenty-first century is characterised by the presence in the public discourse of ideas that are openly problematic because of their misogynistic, xenophobic, transphobic or homophobic profile. Virtual environments have become the frame of reference for many people, but they have also become generators of dependencies from which human weaknesses are exploited (**Moravčíková**, 2022). Likewise, hate speech has colonised digital platforms and social media and, according to some studies, can act as a mechanism for the propagation of hate crimes by enabling the dissemination of extreme views (**Muller**; **Schwarz**, 2021). In this context, research has demonstrated the persistence of messages as an enabler of influence in an ecosystem where the effect of silencing dissenting views is also important (**Neubaum**, 2022), as well as perceived contingency and its relationship to interactivity (**Jiang**, 2022; **Lew**; **Stohl**, 2022).

With regard to gender issues, the Internet continues to be a space of inequality for women and girls as well as for minority and/or minoritised groups, such as the LGBTIQ+ (lesbian, gay, bisexual, trans, intersex, queer and + for all those not represented by the above terms) community or, following **Fraser**'s term, "subaltern counterpublics" (1990, p. 67). These gender gaps, crossed by **Crenshaw's** (1989) concept of intersectionality, include online harm, an effect far removed from the widespread idea that interprets the digital world as a playful and communicative space with no more impact than the sharing of images or videos (**Berners-Lee**, 2021).

The interest of this article lies in the need to reveal the reproduction of messages with a strong ideological charge whose origins or interests remain hidden from public opinion. It is relevant to do so because a large part of the public believes that participation in social networks is only spontaneous and/or inconsequential. On social media platforms and networks in particular, there is an urgent need for transparency about polarised debates that distort or impede the critical approach that any society should be able to construct on issues of great social interest. Increasingly, citizens are demanding this transparency from political action and the media, but these demands have less of a chance in the world of social platforms and networks, which, on the other hand, remain in private hands, operating with hardly any governmental controls. It is the networks themselves that decide which discourses are acceptable and which are not, without offering the framework of freedom of opinion on which their existence is presupposed. The literature has already demonstrated their contribution to disinformation or the dissemination of content that collides with democratic principles (**Acosta-Quiroz**; **Iglesias-Osores**, 2020; **Alonso-González**, 2019; **Llorca-Asensi**; **Fabregat-Cabrera**; **Ruiz-Callado**, 2021; **Rodríguez-Fernández**, 2019).

Digital networks and platforms are not only a space for entertainment. They are now a space for political and social militancy as well as a place to satisfy information needs, especially when they affect major issues (**Casero-Ripollés**, 2016; **Jiang**, 2022; **Pérez-Zúñiga**; **Camacho-Castillo**; **Arroyo-Cervantes**, 2014; **Sánchez-Duarte**, 2016; **Vite-Hernández**; **Cornelio-Lander**; **Suárez-Ovando**, 2020). They have also proved useful for feminist action in what has been called hashtag feminism, with examples such as #MeToo and #NiUnaMenos (**Dixon**, 2014). In this ecosystem, it is necessary to provide transparency by visibilising the nature of some ideological proposals that are disseminated on social networks, proposals that contribute to the climate of tension, the proliferation of strongly polarised ideas and the dissemination of potentially harmful messages. The investigation of these variables allows us to discover the existence of forces in the virtual space that, invisible outside it, construct narratives, generate disinformation and feed generally regressive ideological approaches (**Berners-Lee**, 2021; **Frischlich**, 2022; **Pérez-Curiel**; **Rúas-Araújo**; **Rivas-de-Roca**, 2022).

Thus, this text seeks to determine the origin and characteristics of the conversation on the social network *Twitter* about gender and sexual identities, aspects on which there is hardly any literature, and none in Spanish. In particular, and although there are already works on hate speech related to LGBTIQ+ groups as well as others interested in misogyny, no research has been found specifically interested in aggressive and/or hate speech, on *Twitter* and in Spanish, among and towards trans people and feminist groups. To this end, an analysis of a significant sample of tweets (over a million, collected over a year) will be carried out to shed light on the origin and possible ideological interest of certain messages that are polarising public opinion in an unprecedented way concerning issues such as women's rights, the LGBTIQ+ collective and trans people. Both the methodology and the results can open up future lines of research for other investigations that either wish to delve deeper into the same object of study or can replicate the methodology in the analysis of a different social problem. There is a surprising scarcity of empirical work on the debate that is covered in this article, despite its prevalence and even its repercussions outside the networks themselves, including in the mass media.

According to **Arcila-Calderón**, **Blanco-Herrero** and **Valdez-Apolo** (2020), the social network *Twitter* was chosen because of its capacity to viralise content quickly, its popularity and its speed of communication, although it is not a medium that represents all citizens. It should also be borne in mind that this platform offers

> "an open register of feelings and opinions about issues of all kinds, including hate speech or other signs of rejection that are expressed freely and without the barriers that are often present in offline spaces" (**Arcila-Calderón**; **Blanco-Herrero**; **Valdez-Apolo**, 2020, p. 23).

In Spain, the *Tuitosfera* is one of the main actors in the channelling of public opinion in the virtual space and the main generator of populist disinformation, building identity communities based on the generation of hashtags that generate both resonance chambers and the very agile organisation of mobilisation actions (**Llorca-Asensi**; **Fabregat-Cabrera**; **Ruiz-Callado**, 2021, p. 129; **Rodríguez-Fernández**, 2021).

2. Theoretical framework and context

2.1. Hate speech and gender on the Internet

It has been documented that, since the early days of the Internet, interaction in virtual spaces has been characterised by violent practices, even when it involves irrelevant topics (**Noblía**, 2015, p. 16). This online participation is deeply damaged by the existence of trolls, campaigns organised for spurious purposes and the influence of systems interested in favouring climates of opinion, such as bot farms, cyberlobbies or the proliferation of influencers who become opinion leaders even when their posts have no veracity whatsoever (**Sánchez-Carballido**, 2008, p. 72). All these practices distort conversations and therefore public opinion, which is exposed to hostile discourses of varying intensity, highly polarised discourses and conspiracy theories (**Frischlich**, 2022; **Neubaum**, 2022).

Hate speech does not enjoy a unanimous definition, mainly because its social meaning is not the same as its legal meaning, but also because it is a concept with a high degree of subjectivity. **Cabo-Isasi** and **García-Juanatey** (2016) explain that the concept has been generalised to refer to a non-homogeneous set of manifestations ranging from threats to the expression of anger, including incitement and apology for terrorism. In any case, it is a practice that is directly contrary to freedom of expression and is usually related to racist, xenophobic, sexist and anti-minority offences. The literature has identified these discourses with fascist groups, extreme right-wing politics or ultras groups (**Cabo-Isasi**; **García-Juanatey**, 2016, p. 9). From this terminological difficulty stems the proposal to define its manifestations in social networks as dangerous speech, understanding that it is a speech based on ideological hatred and likely to generate violence (**Amores** *et al.*, 2021).

In the case of conversations related to the feminist movement or agenda, academic research has recognised the multiplication of anti-feminist reactions (**Alabao**, 2021; **Etura-Hernández**; **Gutiérrez-Sanz**; **Martín-Jiménez**, 2017; **Gutiérrez-Almazor**; **Pando-Canteli**; **Congosto**, 2020), at a time when feminist activism has entered into competition "for discursive space in social networks" (**Núñez-Puente**; **Fernández-Romero**, 2019, p. 386). The advances in rights and opportunities that women have been gaining

> "have often been counteracted by reactive movements, not necessarily explicit, which have become obstacles to further progress" (**Menéndez-Menéndez**; **Amigot-Leache**; **Iturbide-Rodrigo**, 2021, p. 10).

Feminism, as well as trans, homosexual, queer or intersex groups, among others, have been targets established by Breitbart in the American alt-right as "leftist" and "elite" excesses, which must be attacked within their "culture wars" as a way to achieve a future political change in society (**Davis**, 2019). This strategy of using culture as a battlefield to change politics was put into practice by Steve Bannon, owner of *Breitbart News*, through his other company, *Cambridge Analytica*, in the social networks, participating in numerous campaigns in various countries around the world, the best known being his participation in Brexit in the United Kingdom and the election of Donald Trump in the United States (**Wylie**, 2020).

Following **Granovetter**'s (1973) sociological theory of the strength of weak links, social relationships of low intensity and greater social distance are very effective in distributing information. This would have a strong influence on already formed groups, where it would not be necessary to convince a whole group, but only a few users would introduce such messages to the rest. The use of account connections with small links to members of a group can be decisive, a fact that is well known and used in political marketing campaigns on social networks (**Ribera**, 2014).

In places such as *Twitter*, the

> "symbolic resistance of certain social sectors to the advance of feminism" (**Villar-Aguilés**; **Pecourt-García**, 2021, p. 34)

is reinforced through

> "corrective attitudes that show the struggle to construct the true narrative" (**Menéndez-Menéndez** *et al.*, 2021, p. 15).

This violence is not random or casual, but obeys some specific patterns that have increased, especially on social networks such as *Twitter* (**Piñeiro-Otero**; **Martínez-Rolán**, 2021, p. 3).

Online abuse, according to the *World Wide Web Foundation* (**Berners-Lee**, 2021), particularly affects women, who are on the receiving end of a wave of attacks that reveal worrying coordination in executing these attacks and result in self-censorship to avoid harm. The *Foundation*'s research also shows that the moderation of content on networks and platforms does not take into account their experiences and needs, especially if they belong to marginalised communities. Concepts such as misogyny and sexism are often not used to moderate discourse, tolerating in practice abuses against women activists and leaders. In July 2021, at the *Generation Equality Forum* in Paris, the CEOs of *Facebook*, *Google*, *TikTok* and *Twitter* were encouraged to put in place measures to address abuse against women on their platforms as it is estimated that around 45% of

> It is a continuous discourse where hatred (mainly associated with disgust), polarisation and associations of ideas are introduced and embedded in society, even if they are false or minority ideas

women have experienced violence online, with abuse particularly severe for women of colour, LGBTIQ+ women and women from other marginalised communities.

In terms of diversity, the growing visibility of the LGBTIQ+ collective does not guarantee positive treatment in networks that have been marked by the denial of "what is different". **Olveira-Araujo** (2022) explains that we are currently witnessing unprecedented visibility of the trans collective, partly due to the

> "hegemony of the queer perspective [which] has favoured, among others, the analysis of transsexuality through inclusive concepts such as trans(gender)"

in a commitment to the joint analysis of different facts of sexual diversity, which, however, "potentially induces biases" (**Olveira-Araujo**, 2022, p. 3). The author refers to the fact that the concept of "transgender" (as well as "trans", sometimes followed by an asterisk) experienced a rapid evolution in the nineties of the twentieth century in the Anglo-Saxon sphere, but is still "under construction" (**Stryker**, 2017, p. 27). Its original meaning evolved into the current one, an umbrella term that encompasses "all forms of non-normative gender expression and identity" (**Stryker**, 2017, p. 206), referring to the "diversity of people who fall outside the standards" (**Serano**, 2020, p. 11). However, in the Spanish state, it experiences "a more politicised use" (**Platero**, 2017, p. 9), and even as an umbrella term it is problematic because it "drowns out the voices" of transgender people in particular (**Serano**, 2020, p. 408). For Serano, the term is politically useful but too vague to expose the diversity of identities and experiences it supposedly encompasses (**Serano**, 2020, p. 46).

In any case, anti-gender campaigns coincide in criticising what has been called "gender ideology" (**Mulió**, 2020, p. 198) and are usually aimed at combating same-sex marriage or affective-sex education in schools (**Carratalá**, 2021, p. 76). The discourse in social networks constructs a narrative of self-victimisation that problematises the advances of the LGBTIQ+ collective in terms of indoctrination of children, curtailment of rights and discrimination of majorities (**Carratalá**, 2021, p. 90). **Carratalá** points out a not insignificant issue: messages that avoid expressly defamatory rhetoric are disseminated online, despite their profoundly homophobic content. As a discursive novelty, **Alabao** points out that the concept of "gender ideology" has served to link diverse struggles and

> "has shown itself to be a powerful social mobiliser, a good activator of the *moral panics* of cultural conservatism" (**Alabao**, 2021, p. 400).

The attacks on feminism and feminists are neither isolated nor spontaneous. They are related to

> "a current of networked misogyny connected to an international anti-feminist and anti-gender movement driven through organisations and political parties of extreme right-wing ideology in charge of spreading and amplifying this type of attacks" (**Villar-Aguilés**; **Pecourt-García**, 2021, p. 43).

The democratisation of opinion that was predicted by online platforms has shifted to a space that, instead of amplifying gender equality and diversity, has become the public square from which to elaborate messages that harm women and marginalised and/or minoritised communities.

2.2. Feminism and identity politics after *#MeToo*

As already discussed, feminism has acquired an undeniable presence and reputation since 2017, with the *#MeToo* phenomenon crystallising the demands for justice that had begun in 2015 with

> "huge mobilisations of women that took over cities in different corners of the planet" (**Gil**, 2020, p. 289).

March 8th, 2018 would lead to what has been called the violet explosion when a women's strike for work and care had an undeniable global impact (**Varela**, 2019). The success would be repeated in the 2019 event.

March 8th, 2020, however, can be established as the formal beginning of the reactive response. In Spain, conservative political parties applied themselves to blaming the feminist movement for the crisis of the Covid-19 pandemic, making visible the unease at the advance of feminism that was already detectable in social networks and even in mainstream media (**Alabao**, 2021, p. 414). Resistance in the digital environment had shot up precisely because feminist activism had increased significantly, revealing the misogynistic nature of digital culture (**Villar-Aguilés**; **Pecourt-García**, 2021, p. 33).

At the same time, social networks will polarise a series of issues that had recently been causing conflict in the feminist movement and were already the subject of some debate in the academy. The feminist agenda has always been diverse and, at times, has experienced splits between theoretical and activist positions whose focus or priority was not the same. Issues such as prostitution, third-party gestation or even termination of pregnancy have not been problematised in the same way within feminism itself. The new focus of tension will be produced by the collision of theoretical and practical points of view, as well as strategic objectives presented by the feminist agenda in dialogue/dispute with that of trans activism and transfeminism. For different authors, a proposal on "risking the subject of feminism", in an event at the *Popular University of Podemos* in 2018, would trigger the conflict (**Reguero**, 2020, p. 233; **Romero**, 2020, p. 19). From that moment on, the debates between the trans movement and the feminist movement with more presence and belligerence in the social sphere would be articulated around "'whether trans women are women' or whether 'sex work is work'" (**Alabao**, 2020, p. 131). From a philosophical and political point of view, the debate "shifts from equality to diversity" (**Miyares**, 2021, p. 184).

The fact is that, until 2018, the participation of trans women in feminism had been the norm, including in the Spanish state, where it was: "well received, in general, although not appropriately" (**Platero**, 2020, p. 44), giving rise to "transfeminism" and the inclusion of trans men (**Alabao**, 2020, p. 147; **Missé**, 2021, p. 149). However, from the aforementioned climate of confrontation would emerge the accusation of transphobia, synthesised in the use of the term TERF (acronym for *trans exclusionary radical feminists*). Some authors point out that "TERF is a minority position within feminism", although "it is having great social and political influence" (**Mulió**, 2020, p. 198).

The term appears strongly in social networks, specifically *Twitter*: "I didn't know what a TERF was [...] until I came to *Twitter*" (**Ayuso**, 2020, p. 219), although it was not new. **Stryker** (2017, p. 295) locates its origin in the United States and in 2008, when the blogger *TigTog* coined it to describe, neutrally, positions that did not contemplate the participation of trans women in activist groups. It was therefore originally used as a way of defining those who considered transgender people to be part of the feminist movement and those who did not. Today it has become a disqualification, an insult (irrespective of radical feminist affiliation) and even a hate speech that is exercised against people who are considered transgendered (**Miyares**, 2021). The *Real Academia Española* (*RAE*), in a tweet published on April 14th, 2021, defined "terfa(s)" as a derogatory expression.

The trans movement denounces the transgendered discourse that inhabits social networks, identifying some feminists in particular (especially those most closely linked to the *Socialist Party* or the *PSC* and also second-wave feminist academics) as the source of the conversations (**Duval**, 2021, p. 83; **Platero**, 2020, p. 43). In this sense, feminists with a historical presence in Spain, both in political and academic action, are refusing to be "displaced" by new perspectives and protagonists (**Platero**, 2020, p. 63; **Reguero**, 2020, p. 240). Some texts lament the fact that these positions are incorporated into discrimination and "indirectly fertilise and legitimise it" (**Alabao**, 2020, p. 130). In this conflict, it is important to contemplate the difference in approaches between the two parties that share the government (Socialist Party and Podemos), an issue that would favour the presence of feminism in the political agenda but also a discussion about hegemony (**Duval**, 2021, p. 259; **Reguero**, 2020, p. 238).

Miquel Missé, a trans author, refers to these debates and recognises that

> "feminism is more open than ever to listen to the proposals of trans activism"

but, paradoxically,

> "trans activism is more hostile than ever to feminism" (**Missé**, 2018, p. 155).

In this context, very marked by identity politics, the term "transphobia" appears to discredit anyone who thinks differently:

> "When someone says something that is outside the pattern of what can be said on the trans issue, they are directly branded transphobic" (**Missé**, 2018, p. 154).

Thus, critical thinking is seen as an attack on human rights, so the possibility of expressing an opinion is disallowed: the accusation of transphobia closes the debate (**Errasti**; **López**, 2022, p. 19), becoming a stigma (**Soto**, 2021, p. 13). **Julia Serano**, in the same vein, argues that

> "most of the transphobia I have had to face as a trans woman could be described as misogyny" (**Serano**, 2020, p. 21),

hence she proposes that the union between the feminist and trans movements should focus on the struggle against the devaluation of women and femininity (**Serano**, 2020, p. 355).

2.3. Research objectives and hypotheses

To address hostile discourse on *Twitter*, an analysis of conversations around sexual and gender diversity was carried out using a mixed methodology to interpret interactions related to women, feminism and trans activism. A one-year time frame was used, allowing for a meaningful sample to be collected from March 2nd, 2020, given that, as noted in the previous section, March 8th, 2020 had previously been identified as the beginning of the reactive response to egalitarian discourse.

Firstly, we sought a general objective:

> (O1), characterised by determining the existence of hostile or hate speech in *Twitter* conversations about the object of study, based on an initial hypothesis that hostility would manifest itself through interactions with a significant sexist and/or misogynist component. In addition to finding out about this incidence, the secondary objectives were:
>
> (O2) to identify the most frequently used offensive terms and their characteristics from the point of view of the level of hatred, the associated emotions and the positive-negative polarity;
>
> (O3) to investigate the origin of these messages to try to illuminate organised interests along the lines of populist disinformation, the involvement of systems to favour climates of opinion and/or the discussion about hegemony in the feminist movement;
>
> (O4) to observe the possible existence of organised anti-gender campaigns;
>
> (O5) to analyse the conflict between feminism and the trans movement to elucidate whether the accusations raised by both movements against the other are sustainable;

(O6) to determine the role of concepts such as "terf" in the public sphere.

To this end, the following starting hypotheses are established:

(H1) There is a widespread polarised and hateful discourse in the debate on the social network *Twitter* between trans and feminist groups; and

(H2) It is estimated that the debate is partly artificially provoked and promoted to generate conflict.

This is research with methodologies that are hardly applied at present in the field of social sciences, and even less so in the Spanish language in the whole of Latin America. Their novelty also implies that they are development methods and that on the one hand they need to evolve, but at the same time they already allow for the analysis of volumes of data that would be unmanageable from a manual point of view.

3. Methodology

The analysis of large amounts of messages, such as those extracted from social networks, is a task almost impossible to tackle without the help of new technologies, the use of machine learning techniques, natural language processing (NLP) and the application of network theories. Studies such as **Mangold** and **Scharkow** (2022) have already looked at methods for measuring audience polarisation. The use of computerised tools for use in social science has started to develop in academia, business and elections, which allow large amounts of data to be analysed in an objective way, such as sentiment analysis (**Arcila-Calderón** *et al.*, 2017). Taking advantage of these techniques, this research takes as a reference the debate on the social network *Twitter* around and between the groups identified as "trans", "terf" and "feminists".

The sample was collected using the *RTweet* library (**Kearney**, 2019) in *R*, through connection with the *Twitter* application programming interface (API) in its version 1.1. For this purpose, all messages (including retweets and replies) were collected that fulfilled the query "trans OR terf AND mujer" in Spanish at a global level, therefore including the word "mujer" and one of the following: "trans" or "terf". The choice of these words stems from the main debate between the different collectives, located in the definition of what it is to be a woman to define a trans person and the corresponding derogatory attack of the word "terf" towards certain feminist sectors. No further words were added as AND filters to collect the maximum volume of messages around this phenomenon.

The collection period was a full year, between March 2nd, 2020 and March 1st, 2021, reaching a total of 1,079,998 tweets. As *Twitter*'s API version 1.1 only allows for the collection of the last seven days, tweets were collected for a full year, week by week, and all of them were finally merged into a single data set. From this corpus, the following processes were carried out, aimed at achieving the objectives of this research:

- Network analysis applied to the area of communication (**Barabási**, 2016), using *Gephi* software version 0.9.4. With this program, it was possible to develop the assignment of each account to a group or cluster and its area of influence (**Chen** *et al.*, 2020) using the Louvain algorithm (**Blondel** *et al.*, 2008) on the retweets made. The graphical representation of the network was based on the algorithm of **Hu** (2006), designed for networks with around, or more than, one million nodes (accounts) analysed.
- Study and quantification of sentiment or polarity (positive-negative), as well as basic emotions through natural language processing (NLP). For this purpose, the lexicon of the *National Research Council of Canada* (*NRC*) in its version 0.92 in Spanish with more than 14,000 words (**Mohammad**; **Turney**, 2010, 2013) was used through *R* and with the help of the *Syuzhet* library (**Jockers**, 2017). In this way, positive and negative values of polarity and emotions are associated with each message by the words it contains, enhanced or diminished by those around it (**Swati**; **Pranali**; **Pragati**, 2015). The basic emotions described by **Plutchik** (1980) and later **Ekman** (2003) are also labelled and quantified, and are as follows: disgust, anticipation, fear, joy, sadness, surprise and confidence applied to each tweet's discourse (**Sauter** *et al.*, 2010). Levels higher than 3 in any emotion or polarity can be considered significant (**Fitzgerald**, 2017).
- Analysis and quantification of hate using PLN techniques, analogous to the previous point, but this time using the *Hurtlex* lexicon in its 2018 Spanish version with about 5,000 associated words (**Bassignana**; **Basile**; **Patti**, 2018). The presence of words from the hate lexicon will give a level of presence value, which will be enhanced by the words around them. As in the previous point, the values detected with a value of 3 or more will be considered as clear hate, thus avoiding the inclusion of ironic or everyday use of some words in the messages.

  The *Hurtlex* lexicon, according to its authors, has had good detection results in Spanish and is used for its use by machine learning; they are words that have been categorised as hate speech by people, and the algorithm used, *Syuzhet*. It is a methodology similar to, but different from, *Naive Bayes* algorithms, in which hateful phrases are identified, and it is machine learning that searches for the words and relationships between them as a way of training and learning to then be applied to new phrases, as used by **Arcila-Calderón** *et al.* (2021) in Spanish. While there are methodological advances in multilingual dictionaries (**Maier** *et al.*, 2022), hate detection in Spanish is still in its infancy and needs further development.
- Probabilistic determination of bot behaviour through **Kearney**'s (2018) *Tweetbotornot* algorithm. Based on the nature, history and behaviour of each account, a percentage of the probability of being a bot is determined. This method uses the analysis of biography, followers, friends, activity, mentions, photos, location and the last 100 messages issued,

among others, to determine the probability of bot behaviour. Its probability of success, according to the author, is 93.53% for classifying bots and 95.32% for non-bots. This algorithm is considered among the best and most widely used in social sciences (**Martini** *et al.*, 2021).

- Text mining study using a word cloud and dendrogram between the words with the highest presence in the discourse with the highest hate load (levels above a value of 4 according to the analysis and quantification of hate). The aim is to show graphically not only the most frequently used words, but also their relationship. In the dendrogram, a new cluster analysis was carried out to determine the top ten word groups using the *k-means* algorithm. To determine their fit, Dunn's index is calculated, with the highest values being the most desirable (**Gil-Pascual**, 2021). In the analysis, words that do not contribute meaning (stopwords) and symbols (such as #) or punctuation marks have been eliminated.
- Analysis of the statistical relationship between variables employing Pearson correlation in *R*, in this case hate, concerning both the situation in the network and its probability of being a bot or the emotions and feelings they give off.
- A regression tree study, using machine learning techniques, allows the behaviour of a variable to be predicted (**Lantz**, 2019). The variables of the previous point are used, but considering hate as a dependent variable of the others, which would be independent. For this purpose, the *CART* algorithm (**Breiman** *et al.*, 1984) is used through the rpart package in *R*:
  https://cran.r-project.org/web/packages/rpart/index.html

  The reliability of the prediction is assessed through the correlation between actual and predicted values, in addition to the mean absolute error (MAE).
- Parallel to the main message collection in a geolocated manner in the last analysis period, each of the more than 70 countries identified by **Bradshaw**, **Bailey** and **Howard** (2021) as issuers of disinformation on social networks. The search was conducted by entering longitude, latitude and radius of influence coordinates in the query. This technique has a global average reliability of 77.84%, exceeding 90% in some Western countries due to the quality of their network (**Van-der-Veen** *et al.*, 2015).

Hate and bot probability data are represented by box plots, which designate the median value of the value in the middle, while a box representing the first quartile is at the bottom and the third quartile at the top. An asterisk represents the median value, while above and below there is a line representing a value 1.5 above or below the interquartile value, while single dots express statistical outliers.

4. Results

After collecting data between March 2nd, 2020 and March 1st, 2021, a full year, with a total of 1,079,998 tweets, each message was assigned to a group by clustering through the retweets issued. This process resulted in the formation of 5,060 clusters (cluster numbers are offered among all detected clusters), with a modularity value of 0.770, a fairly good assignment value for social sciences. Among all these clusters, the research focused on the top 12 clusters, which account for 72.47% of the total number of messages, with a total of 782,722 tweets, which can be seen in Table 1 and are represented in Figure 1.

Table 1. Top 12 groups detected

Group no.	Trend	Percentage of traffic generated from RT
958	*Queer group*: although varied, especially with origins in Spain and with a lot of attack on feminism for its supposed definition of women.	11.99%
131	*LGBTi Mexico*: support group for transgender people in the country.	10.57%
1163	*Trans group*: trans advocacy group and attack on feminism or authors like JK Rowling.	10.07%
593	*Argentina group*: similar group to Mexico, but from Argentina.	8.63%
577	*Trans* (with numerous accounts suspended or deleted after a random tasting and among the accounts with the highest own-vector value): group attacking feminism and figures such as JK Rowling.	7.29%
464	*Trans group*: support and advocacy group for the definition of women without lacking certain aspects.	6.75%
91	*Trans PSOE*: group supporting the Spanish political party's policies on trans issues.	5.73%
0	*Feminists*: a group that wants to make it clear that they are not attacking the trans group.	5.52%
45	*LGBTi activists*: fight for LGBT rights and support for transgender people.	4.56%
1,717	*Far-right* and suspended accounts (after random tasting and top accounts by highest own-vector value): attack on feminist groups using trans discourse.	3.24%
680	*Latin Influencers*: follower groups of influencers from various Latin American countries, especially Central America, who mainly support attacks on feminism through displacement of trans people.	2.97%
1,429	*Far-right*: attack on feminism and trans groups.	2.68%

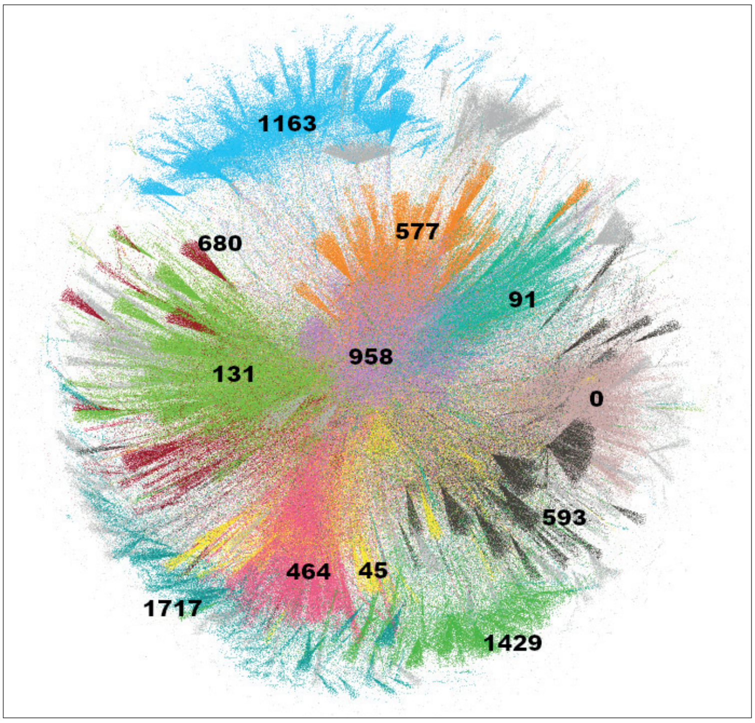

Figure 1. Network relationship graph

The selection among the 12 main clusters yields a total of 146,163 direct messages and 636,559 retweets (RTs). The use of the algorithm for detecting and quantifying hate among these messages determines a total absence in only 8,043 messages (1.03%), with a mean value of 2.73 intensity, a median of 2.00 and a maximum intensity of 32. The presence of hate can be detected in all the groups, but it is particularly present in some of them, namely 593, 577, 464 and 45, supposedly trans groups but without major influencers, and the presence of numerous accounts that have subsequently been suspended. The origin of the accounts is mainly Spanish, although there are important groups from Argentina and Mexico, which are concentrated in their clusters, and to a lesser extent from Colombia, Chile, Venezuela and Peru.

It can be seen in the graph that many of the groups are interconnected with each other, closely intermingled around the main group, 958, the group led by queer influencers. The most distant groups would be group 1163, the trans group, and groups 1429 and 1717, which are situated around far-right political users. This is a circular network structure around a group that is the backbone of the messages and where no separate groups can be seen, except for the most distant ones mentioned above. Thus, from the central queer group onwards, almost all the groups intermingle with each other, with no well-defined groups and no clear opposing influencers. This interrelated structure in most of the groups defines a degree of interconnectedness that does not correlate with the degree of hatred studied in the rest of the article, which should theoretically be represented in more differentiated groups.

Although the average and median hate presence are appreciable and almost 99% of the messages have some hate component, as can be seen in Figure 2, to avoid problems of double-meaning, ironic or commonly employed uses, the study will focus on the analysis of hate messages that are located in the third quartile, which is located in hate values with a value of 4 or higher. This subsample represents a total of 204,700 messages, with 29,985 direct messages and 174,715 RTs. The 29,985 direct messages with the highest hate values were posted by 13,160 accounts, mainly created between 2019 and 2020 and to a lesser extent 2011, of which 9,405 (71.47%) were found to have been deleted by the user or suspended for violating *Twitter*’s rules in February 2022, with a small portion protected as inactive.

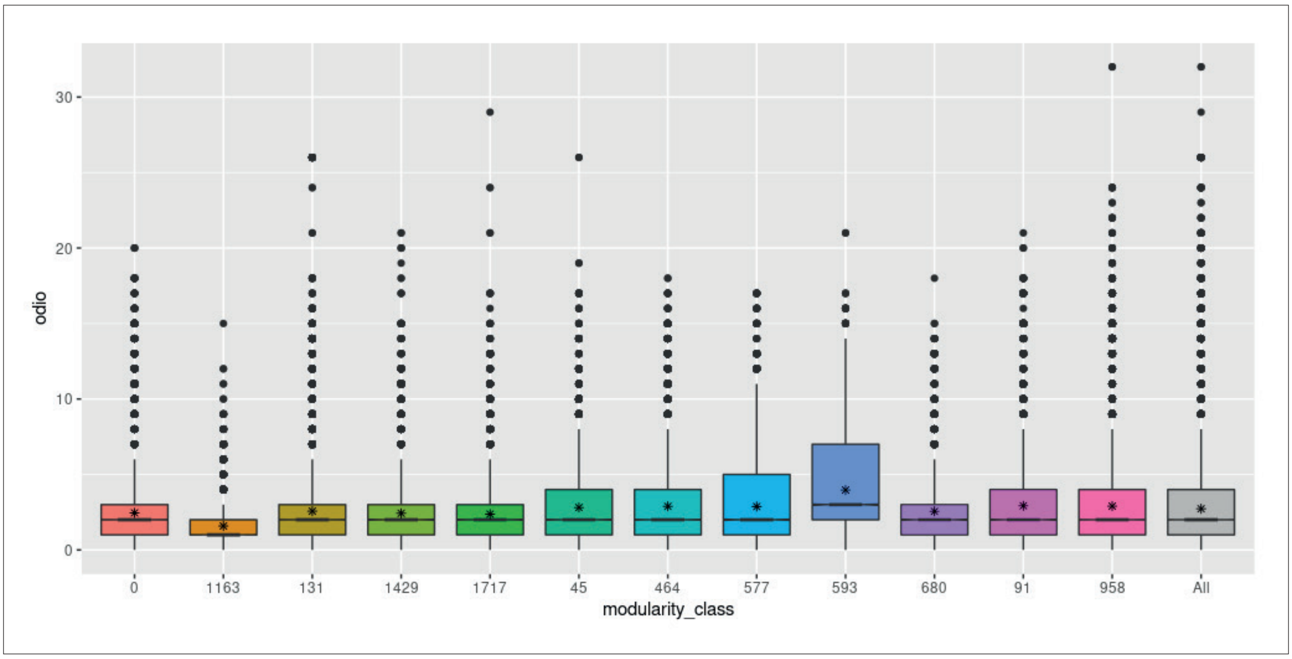

Figure 2. Presence of hate by each main cluster

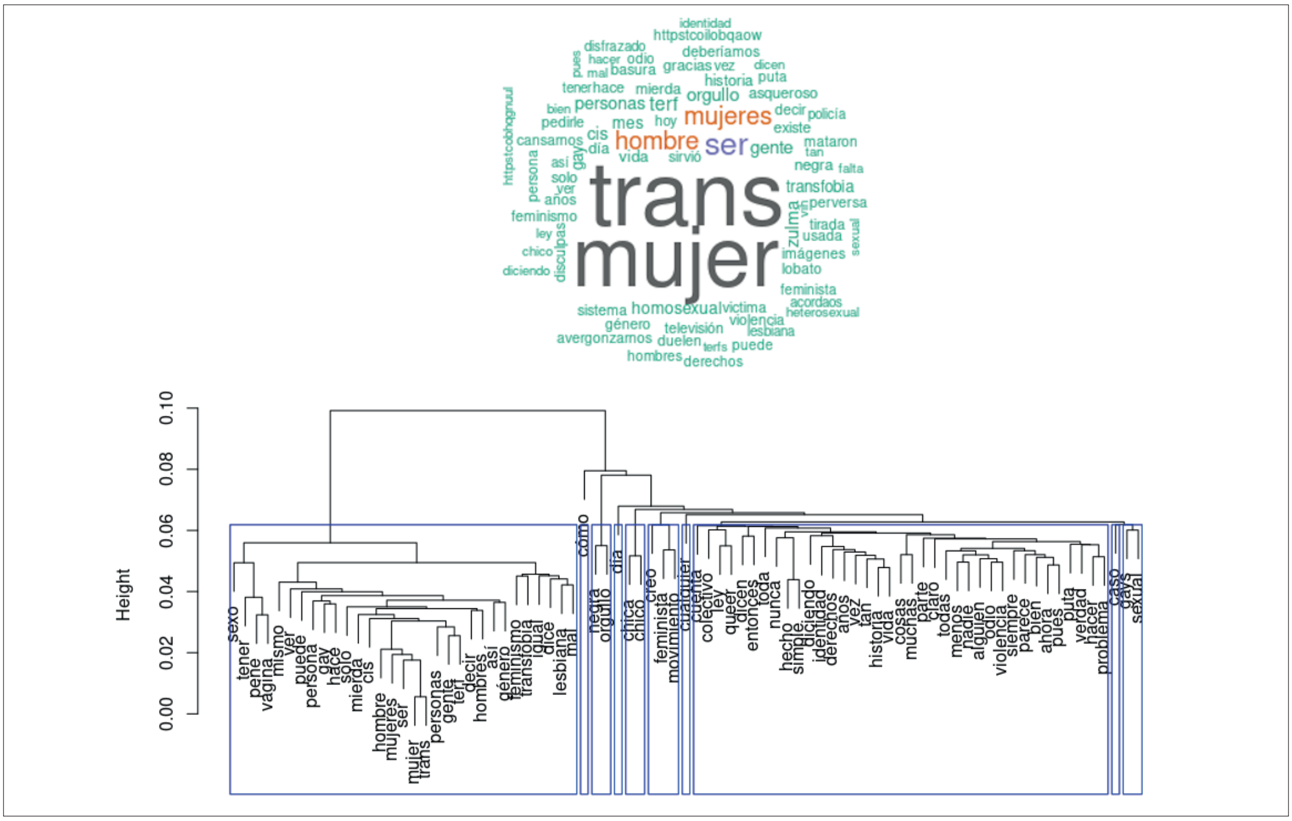

Figure 3. Word cloud and dendrogram separated by clusters of the main words used among the most hateful messages detected

From this subsample, the main words most frequently used were determined by text mining through a representation of a cloud, as well as a dendrogram with the main topics dealt with among the messages with the highest level of hate (Figure 3). As specified in the methodology section, the messages containing the highest level of hate (level 3 and above) were determined. The dendrogram study presents a good fit for social sciences, with a Dunn index value of 0.559 (as a percentage of one), which means that the themes have been well detected according to the process carried out.

It can be seen how the most frequently used words, references to the relationship between feminism, trans, women, pride, rights and various insults, can be identified. The dendrogram shows more precisely the main associations between hate messages. Thus, on the left in the first cluster, one can see the theme that associates feminism and transphobia by the definition of what is said to be female and male according to their sexual organs. There is also another important cluster around the *Queer group* defending their rights in the face of hatred and violence. There are no differences in messages or debates depending on each country, but they are limited to general facts about the group and debates about specific events, such as those produced in response to statements about women and menstruation by the writer JK Rowling in June 2020, after which she was accused of being a "radical feminist who excludes trans people" (*EFE*, 2020).

Correspondence analysis on selected words gives clear association results (together with the percentage of occurrence in terms of the number of times it is linked above 0.10):

- feminism: radical (0.18), subject (0.15), chicxs (0.13), liberal (0.13), women (0.13), colonial (0.12), Marxist (0.12), negrar (0.12), pirado (0.12), political (0.12), putísimo (0.12), transfeminism (0.12), transmisogynist (0.12), victimizo (0.12), lesbophobic (0.11), movement (0.11), chirrian (0.11), dogma (0.10), sect (0.10), exclusionary (0.10).
- feminist: radical (0.21), movement (0.20), nigga (0.15), accompanied (0.15), berf (0.15), qualified (0.15), dominecabra (0.15), elected (0.15), shit (0.15), bird (0.15), pedoliteral (0.15).
- trans: cis (0.19), women (0.19), men (0.17), antidepressants (0.14), boy (0.11), autogynephilia (0.10), questioned (0.10), spanking (0.10), transforms (0.10).
- terf: radical (0.20), feminist (0.18), empowering (0.16), shit (0.16), applauded (0.15), corporately (0.15), descendants (0.15), people (0.15), townies (0.15), wards (0.15), transphobia (0.15), exclusionary (0.14), cisheteras (0.13), people (0.13), speech (0.11), fastidia (0.11), movement (0.11), putito (0.11), supuestoantifa (0.11), transfobas (0.11).
- female: male (0.27), kindness (0.17), lesbian (0.14), black (0.13), personified (0.13), cis (0.12), penis (0.12), cock (0.12), desperate (0.11), vagina (0.10).
- queer: theory (0.48), carnation (0.20), creed (0.20), equivaldrá (0.20), feminismoradical (0.20), implantation (0.20), injerencista (0.20), inquisitorial (0.20), ladinos (0.20), negationism (0.20), pergeñado (0.20), servilismo (0.20), populismo (0.18).

> All of the above exposes an international discourse, sustained over time, which is neither isolated nor spontaneous

The appearance of words linked between the associated words may be for two reasons: the speed of writing in this type of social network, or the possibility of creating words *specifically* to identify and conceptualise a type of association. The very common presence of some of them would give rise to the second reasoning, with the formation of words such as "feminismoradical", "supuestoantifa", "pedoliteral" and "dominecabra". In this way, a series of vocabulary created for the association of certain ideas would be established.

The most hateful tweets detected are, by way of example, those described in Table 2.

Table 2. Messages with the highest hate load detected

Text	Favourites	RT No.	Cluster	Probability of bot
A trans person with data, with their own experiences, etc... says something. TERF's answer: KILL YOURSELF. YOU ARE DUMB. YOU ARE NOT RIGHT CUIR DE MIERDA. FUCK YOU. YOU ARE AN IDIOT. I can tell you have a lot of weighty arguments.	1	0	958	3.96%
@dyJulieh This shit is so ridiculous you don't even know what to say anymore. Now it's all about being black, trans, poor, ugly, feminazi woman, gay. What a BITCH to what level of stupidity this trashy progre agenda is taking us. (the account @dyJulieh no longer exists)	0	0	1,717	44.10%
There is a very simple method to know if the humour is acceptable or not. Ask yourself where the joke is funny, if the answer is "that it's black, female, gay, bi, trans, dumb etc..." the joke is not acceptable. You can also think about whether it attacks an oppressed group, but I don't think you'll get it.... https://t.co/w7A2YWqqzq	0	0	958	81.27%
I thought so too, how fucking ashamed I am now. We owe a lot of our rights to a trans hiv+ woman of African descent. So to make you dumb and discriminate just because you think you're "less gay" is fucked up. Heteronormative gays don't know how to fuck FACT https://t.co/o0pQxPEkqg	505		131	6.73%
@uwuk1ra @satogarca Did you know that trans women have brains shaped in a similar way to biological women? But that's up your ass, you're just a fucking idiot. (@uwuk1ra is suspended for non-compliance with rules, @satogarca no longer exists)	0	0	131	72.36%
@guirimadri @abixquertg @DieBatsuDie We are indeed from the animal kingdom, but treating a human being like a wild animal and giving that classification, especially if you are going to touch the heartstrings of a trans person because of the shit that you are, seems to me to be a fault. I'm a biology student by the way. (@abixquertg no longer exists)	0	0	958	13.52%
It's also anti-transphobia day, and what a fucking nerve to say "this asshole" to a trans woman who is just showing her support for the movement. I've had it up to here with your transphobic, retrograde exclusionary bullshit. https://t.co/otYORArzPY	5		131	73.96%
Man: Miss, you are very beautiful. Woman: FUCKING BITCHY, DISGRACEFUL, MYSOGINIST HARASSER, FUCKING RAPIST!!!! Gay/Trans: Ay perra k culote te cargas y esas pinches chichotas, te pasas maldita, te odio! mmm!!!! *Woman: Ay amiga jiji gracias, eres bien linda!		1	1,717	0.2%
@cxsmic__girl @_AlexClockwork @__erosgarcia @Dalma2206 THEN YOU WILL ASK FOR RESPECT,TRUTH,FUCKING FUCKING TERFS,YOU ARE ESKORIA.YOU TRANS?.YOUR FUCKING MOTHER,FUCKING NUTS.YOU KNOW WHAT YOU ARE GOING TO FACE,AS LONG AS YOU KEEP GIVING X ASS WITH TRANS.NAZIS DAIS ASKO,BUT THERE IS NOT SOMETHING THAT IS MORE ASKO K A FUCKING RETARDED TERF. 👊😠😡	0	0	91	36.92%
How sad to see gays, who enjoy rights and some freedom thanks to trans people who have fought for them in previous years, throwing them under the bus and even questioning who is a man or a woman because of their genitals or something as silly as menstruation, rubbish.		0	593	4.36%

The messages, a significant number of which have hate associated with them, that are collected both among the text mining analyses and among those with the highest intensity of hatred come to expose a clear message against feminism, which is accused of being intransigent and of limiting the definition of a woman. Thus, it is classified as a sect and various insults, especially linked to the derogatory term "terf" (*trans exclusionary radical feminists*). In contrast, the term "trans" is associated with the terms "cis", "women" and "men" and the need for antidepressants. A very visceral discourse is thus expressed, full of hate, but where local debates are not appreciated, but between groups regardless of where they are located. It is noteworthy that many accounts that participate or are challenged in the messages have been deleted by users or suspended one year later for violating the rules of the social network *Twitter*.

Although hate is intense in almost all the groups detected, a very high percentage of bot probability is detected, especially in certain clusters, as shown in Figure 3. There is no statistical correspondence between the position of the account in the network, although there is some correspondence with the emotional discourse of these accounts. Thus, the Pearson correlation values between

> Many accounts, months later, have been deleted or suspended for violating the social network's rules

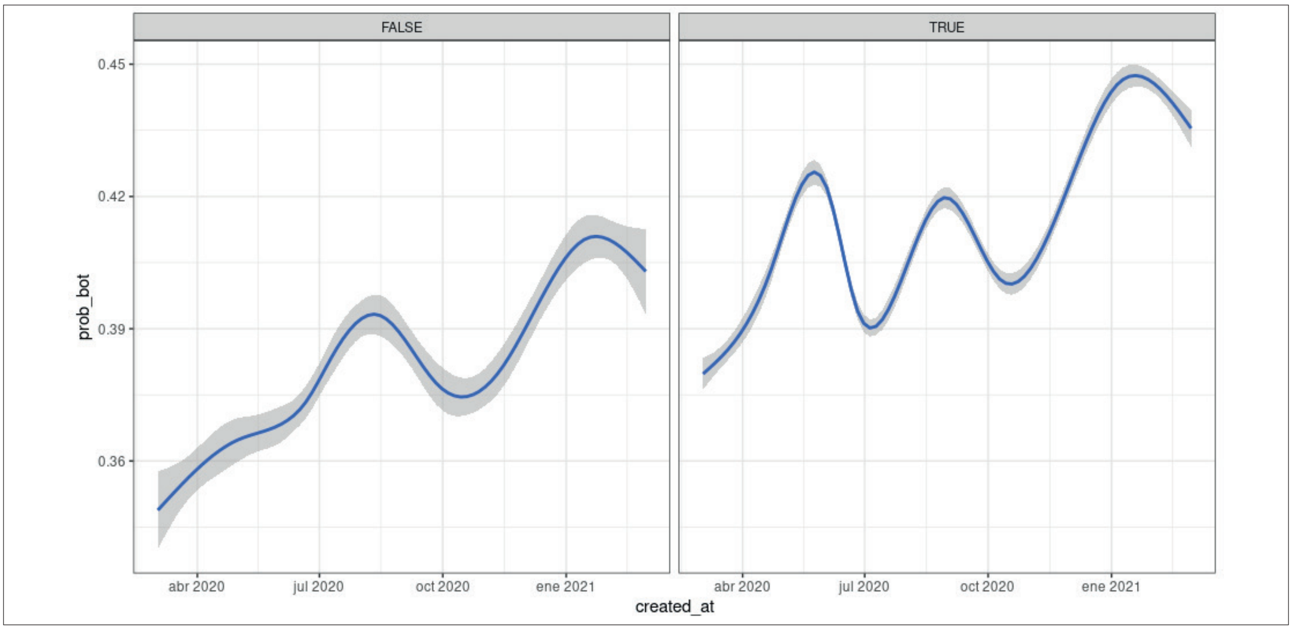

Figure 4. Average probability of being bot over time for accounts that direct messages (False) and those that RT (True)

the level of hatred and other variables of nature and position in the network are completely negligible: -0.016 Bot probability, 0.0002 Favourite, -0.081 RT, 0.152 intermediation centrality, -0.008 with eigenvector, -0.003 with degree and -0.005 with closeness centrality.

Therefore, hate is not directly related to being an influencer and having a larger following, being retweeted more, being favourited more or being a bot. As can also be seen in Table 2, the tweets characterised by the algorithm show messages with a lot of hate regardless of their percentage of bot likelihood, the number of favourites or RTs. For this reason, it cannot be characterised as bots that generate hate, but rather as a user profile that may or may not be closely linked to their influencers (proximity centrality), are the backbone of a network (intermediation) or are influencers of their network (own vector).

According to Figure 4, it can be seen how throughout the year analysed, the accounts that make RTs distribute hate, regardless of their intensity, are more prevalent than the accounts that launch their messages. This corroborates what **Williams** (2021) said, where bots do not create hate, but “inflate” it. The fact that, as mentioned above, there is no correlation between bot probability and intensity also confirms **Vidgen** (2021): the accounts do not exhibit high intensities in their discourse so as not to be easily detected and to remain on the network for much longer. This aspect, in which bots would be in charge of disseminating the messages of others, rather than generating them, would link with what **Keller** (2019) exposed in social networks during the South Korean elections or the Covid-19 pandemic in Spain (**Arce-García**; **Said-Hung**; **Mottareale-Calvanese**, 2022).

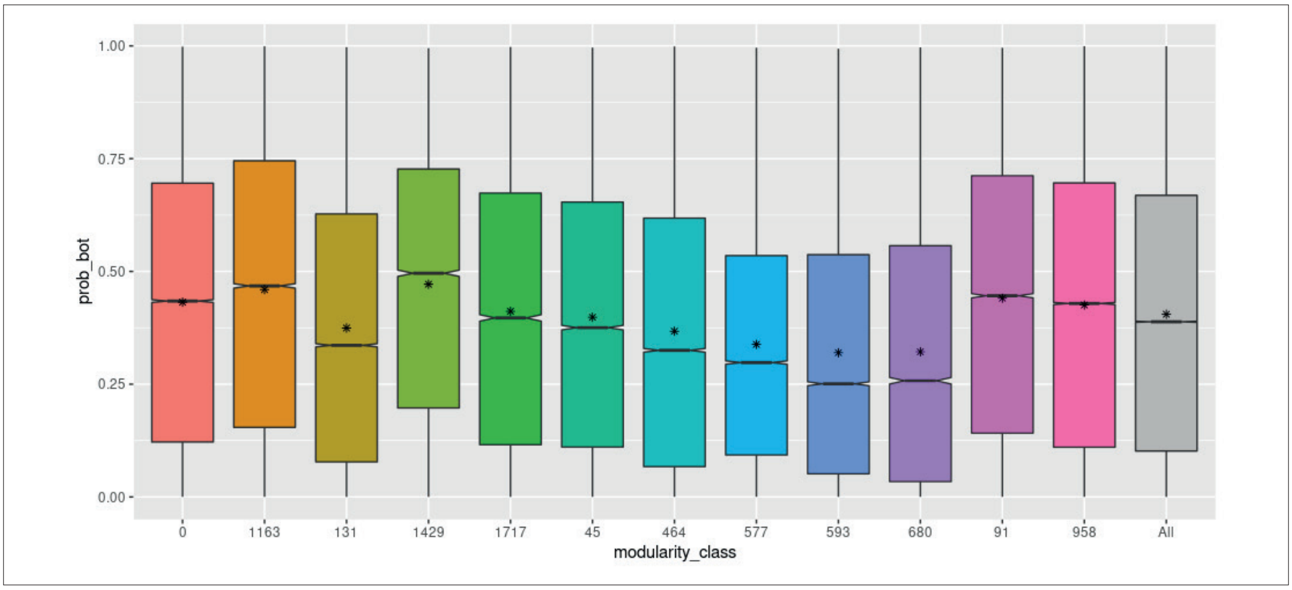

Figure 5. Probability distribution of bots presence in each cluster

The presence of bots according to the cluster or group initially identified, as shown in Figure 5, generally shows an average behaviour, in many cases close to 50%, with several groups in the third quartile reaching a 75% possibility of being bots. This would demonstrate the artificiality of many of the accounts involved, especially in some of the trans and far-right groups.

This could be explained through two possible interpretations: the presence of users who go it alone and have an intense and highly polarised discourse, and/or the use of troll-like accounts designed to establish controversy, set various issues on the agenda-setting or run disinformation and polemic campaigns with the help of confrontations between "false flag" groups (*Russian Bot*, 2022), i.e. accounts posing as opposing groups.

On the other hand, in regard to the basic emotions and feelings or polarity, hatred shows a slight statistical correspondence with sadness and, above all, with disgust or aversion. Thus the Pearson correlations are -0.273 with polarity or vector, 0.068 with confidence, 0.080 with joy, 0.086 with anticipation or rational thinking, 0.097 with surprise, 0.192 with fear, 0.244 with anger, 0.284 with sadness and 0.405 with disgust or aversion. Therefore, it can be seen that the emotions, in this case, that most lead to hatred come mainly from disgust and sadness.

The establishment of a predictive regression tree on the case concerning all messages with detected hate, shown in Figure 6 (the boxes above show the average level of hate intensity and below the number of messages and what it represents over the total), corroborates the main source of hate through disgust or aversion and fear, which reaches its highest values when the feeling of disgust is greater than 2 and fear is equal to or greater than 2. The main weight of hatred comes from disgust at 50%, with 13% for negative polarity and 12% for fear, with all other emotions falling below 10% in importance. Therefore, a mixture of medium to high levels of disgust and fear are those that generate the most hate content. Hate messages that intermingle with sadness have lower values of intensity. The correlation between the predicted and the actual value is 0.498 and the MAE is 1.24, so an acceptably good prediction of the algorithm can be estimated.

According to **Plutchik**'s (1980) wheel of basic emotions, hatred is the increase in intensity of the basic emotion of aversion or disgust, but it can also be achieved from "dyads" (combinations) of other basic emotions, such as sadness and anger or mixtures of fear with anger or sadness. For combinations of emotions, the further apart they are in the wheel (disgust and fear are three steps apart, or tertiary dyads, according to their author), the more difficult they are to maintain over time. Hatred, moreover, activates a secondary behaviour, rejection. For this reason, hatred generated from disgust would be more intense and enduring over time than hatred generated from a combination of other emotions. In the case analysed, hatred is primarily based on disgust or aversion, accompanied by high levels of fear.

For the last collection period, from 25/01/2021 to 01/03/2021, 83,986 messages were obtained, but a parallel collection was also carried out for certain countries specified in the methodology section through geolocation. It was thus possible to obtain the location of 33,795 messages (40.24% of the total), and it was possible to determine that, although it was a conversation in Spanish, the distribution was worldwide. It was thus possible to determine that in that period, of the geolocated messages, 40.21% came from Spain, 15.86% from the United States, 8.26% from Mexico, 6.76% from Colombia, 6.11% from Brazil, 5.66% from Venezuela, 4.27% from the Philippines, 3.83% from Argentina, 2.99% from Russia (Moscow area), 2.93% from Turkey, 0.65% from the Dominican Republic, 0.43% from England (London area), 0.35% from Bolivia, 0.31% from Italy (Rome area), 0.24% from Thailand and 0.23% from Albania. The remaining countries have lower percentages. There are therefore several non-Spanish-speaking countries with non-negligible percentages in the conversation, in which recent studies characterise propaganda and disinformation industries directed, among others, towards Spain from other countries (**Arce-García**; **Said-Hung**; **Mottareale-Calvanese**, 2022).

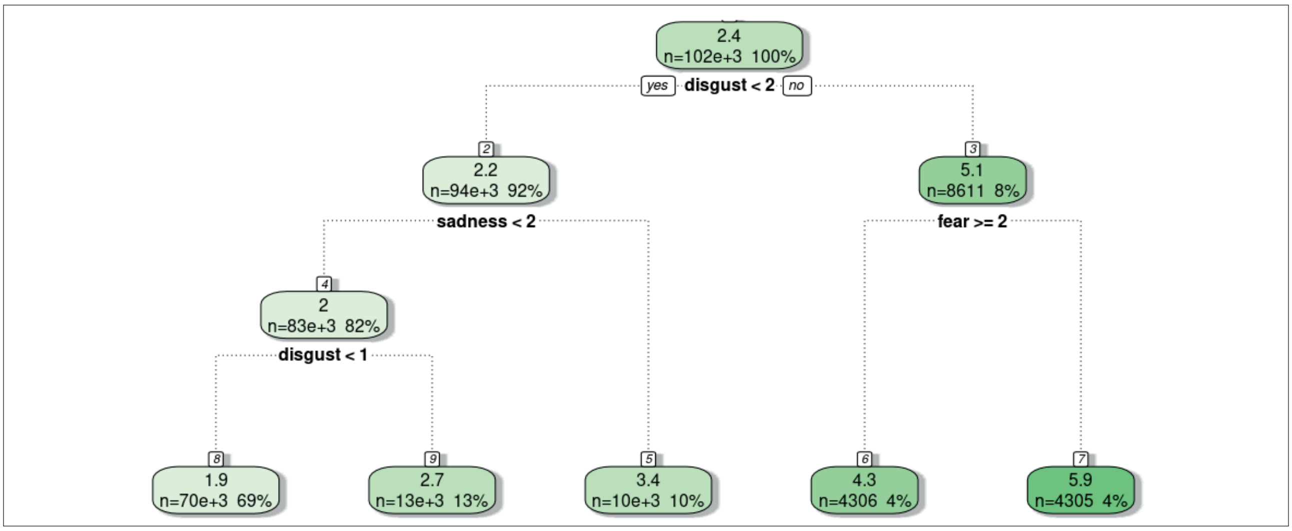

Figure 6. Regression tree on emotions regarding hate speech

5. Conclusions

A year-long analysis of the social network *Twitter* around the conversation between different groups, mainly feminists and transgender people, shows a large debate with a high hate component. Even if one follows the clear formation of groups through cluster analysis, in the face of such a polarising topic one should see a separation of well-marked bubble filters (**Pariser**, 2011), but in many cases there is a great lack of definition between different groups and no marked differences and separations as shown in Figure 1, in contrast to other highly polarised cases. Nevertheless, the high levels of polarity and hatred demonstrate the first hypothesis put forward in this paper between trans and feminist groups.

The network detected, in addition to being highly interrelated, does not have clear referents or major influencers. There are only a few groups far removed from the rest, one with trans issues and two with far-right tendencies, although they still have extensive relations and interconnections with the other clusters. There is a large amount of hate speech in all the groups detected, although the extreme level is much scarcer, with a broad attack on feminism, which is accused of being radical and exclusionary, associated with the term "terf", as opposed to a trans group that suffers (it needs antidepressants), coinciding with the story of self-victimisation as outlined by **Carratalá** (2021, p. 90). There is a very clear profile of word association over the course of a whole year, so these are ideas and discourses maintained over time, as well as created words such as "feminismoradical" or "pedoliteral". Similarly, there is an attempt to equate feminism with the exclusionary term "terf", a derogatory word that, although a minority, is widely used in networks (**Ayuso**, 2020, p. 219).

Other data could suggest that the debate is artificial: many accounts have been deleted or suspended months later for violating the rules of the social network. Similarly, bot-like behaviour was detected in many of the groups, although no significant statistical relationship was found between a greater probability of being a bot and greater dissemination of hate or negative feelings. A greater bot presence was observed among those accounts, which could be a typical troll-type behaviour accompanied by bots of support in a possible false flag campaign. The fact that the bots are especially dedicated to disseminating hate rather than establishing their messages would fit with the studies by **Keller** (2019), which point to organised strategies for disseminating disinformation campaigns or attacks on certain groups typical of the technique known as *astroturfing*.

The presence of far-right groups and the number of allegedly trans accounts suspended or deleted around the debates would contribute to this idea of contaminating the debate of the rest of the clusters. Likewise, the existence of not inconsiderable percentages from various countries, some known for their disinformation industries, of non-native Spanish speakers adds to the suspicion. Therefore, everything points to the idea of a debate that is being fed and fomented by groups that present strategies that have already been studied and cited.

The determination of which emotions mark hate, in this case, comes primarily from disgust or aversion, coupled primarily with fear. As **Wylie** (2020) argues, provoking hatred, especially based on disgust or aversion, is the main component for introducing non-rational thinking that is sustained over time as opposed to that generated solely by fear.

All of the above exposes an international discourse, maintained over time, which is neither isolated nor spontaneous (this study covers the entire Spanish-speaking world in Spanish for one year), as **Villar-Aguilés** and **Pecourt-García** (2021, p. 43) also explain, which attempts to introduce a series of associations of ideas through accounts of little relevance on the network, taking advantage of the sociological theory of **Granovetter** (1973). This theory suggests that it is not necessary to influence all users, but only a few, who will then transfer their new ideas to the rest of the group. The fact that 20% of the messages of the 12 main groups emit intense hate messages is a significant percentage that could end up influencing the entire network. This phenomenon has been contemplated in other academic research (**Alabao**, 2021; **Etura-Hernández**; **Gutiérrez-Sanz**; **Martín-Jiménez**, 2017; **Gutiérrez-Almazor**; **Pando-Canteli**; **Congosto**, 2020), but it could equally fit within the "culture war" towards, among others, the feminist collective as **Davis** (2019) exposed, confirming the second hypothesis put forward.

We could therefore speak of the presence of fairly intense signs of an unnatural structure or movements in social networks attacking feminism, aimed at provoking confrontations between different groups, just when trans authors recognise that when feminism is more open, the environment is more hostile (**Missé**, 2019, p. 155). It is a continuous discourse in which hate is introduced (mainly associated with disgust, which is the basic emotion that makes it last longer in time and intensity), and associations of ideas that are inserted into society, even if they are false or minority ones. The fact that hatred is basically generated from the basic emotion itself, disgust, and therefore generates rejection, ensures that this hatred survives and is not eliminated, as it displaces rational thought. This research also serves to show the scheme of possible influences on society through social networks by provoking confrontation and spreading hatred between groups.

With all these data and confirmation of the two hypotheses put forward, it should be noted that there is a strong polarisation that provokes attacks between trans and feminist groups, in which groups of a very suspicious nature can be seen (use of bot accounts, suspended or deleted accounts, relations with outside groups and

> Attempts are made to equate feminism with the exclusionary term "terf", a derogatory word that, although a minority word, is widely used in networks

not inconsiderable origins from non-Spanish-speaking countries), with methodologies seen in other organised campaigns. The use of these techniques would come to define an environment in which the aim is to generate feelings of hatred that last over time, since repetitive discourse and the provocation of strong basic emotions generate the main basis of social hatred, which is currently seen with the new currents of study in social networks of gaslighting (**Shane**; **Willaert**; **Tuters**, 2022). According to this theory, the generation of strong basic emotions in communication causes the recipients to manipulate and raise doubts about their perception of reality.

> There are several non-Spanish-speaking countries with not insignificant percentages in the conversation

The possible use of the propaganda and disinformation industry, supported by profiles and origins known to be suspicious, would be doing its job to provoke disaffection and mistrust towards trans and feminist groups, exposing them as opponents and disseminators of hate, as they would be the victims of the attack. The use of bots mainly for rebroadcasting and other behaviour deepens the suspicion of the presence of this type of industry, which is very difficult to detect.

The generation of hatred, raising doubts and high polarisation makes society see both groups as conflictive, generating hatred and contempt, especially the feminist group as intransigent and exclusive. All of this would then suggest a possible long-running campaign, in which feminism is seen as an unattractive element in the face of a displacement of the Overton window, that is, the range of ideas that public opinion assumes as acceptable (**Williams**, 2021).

The limitations of this work come mainly from how hate is detected, since it is a very new methodology in Spanish based on one of the first existing lexicons in this language, *Hurtlex*, which has been validated. Its main drawback is the lack of intensity rating among the lexicons separately, but an algorithm has been used that does take into account the conjunction of several words together to increase or decrease their level, in addition to its success in Spanish according to the authors of the lexicon used. There are several projects in Spain to make training samples using neural network methods, but none of them have been completed to date. Similarly, the emotional lexicons used are almost non-existent in Latin languages, which, although they have lower reliability than in English (over 80%), are at least 70% (**Mohammad**, 2016). The sample is considered reliable as it covers a full year of collection.

The presence of texts that may have a double or ironic meaning is also a problem for these methodologies, but given the volume of messages analysed, it is believed that the overall results are not altered, or only minimally so. In future lines, it is proposed to study in greater depth the scheme of attack and confrontation provoked in social networks, not only in this area but also in other areas of society where an attempt could be made to change public opinion in the medium term.

> This phenomenon could be seen as part of the "culture war" against, among others, the feminist collective

6. References

Acosta-Quiroz, Johana; **Iglesias-Osores, Sebastián** (2020). "Covid-19: desinformación en redes sociales". *Revista cuerpo médico HNAA*, v. 13, n. 2, pp. 217-218.
http://doi.org/10.35434/rcmhnaaa.2020.132.678

Alabao, Nuria (2020). "El fantasma de la teoría queer sobrevuela el feminismo". En: VV. AA. (eds.). *Transfeminismo o barbarie*. Málaga: Kaótica Libros, pp. 129-152. ISBN: 978 84 12212921

Alabao, Nuria (2021). "Las guerras de género: La extrema derecha contra el feminismo". En: Ramos, Miquel (ed.). *De los neocon a los neonazis: La derecha radical en el estado español*. Madrid: Fundación Rosa Luxemburgo, pp. 397-423.
https://www.rosalux.eu/es/article/1954.las-guerras-de-g%C3%A9nero.html

Alonso-González, Marián (2019). "Fake news: disinformation in the information society". *Ámbitos. Revista internacional de comunicación*, v. 45, pp. 29-52.
https://doi.org/10.12795/Ambitos.2019.i45.03

Amores, Javier J.; **Blanco-Herrero, David**; **Sánchez-Holgado, Patricia**; **Frías-Vázquez, Maximiliano** (2021). "Detectando el odio ideológico en Twitter. Desarrollo y evaluación de. un detector de discurso de odio por ideología política en tuits en español". *Cuadernos.info*, v. 49, pp. 98-124.
https://doi.org/10.7764/cdi.49.27817

Arce-García, Sergio; **Said-Hung, Elías**; **Mottareale-Calvanese, Daría** (2022). "Astroturfing as a strategy for manipulating public opinion on *Twitter* during the pandemic in Spain". *Profesional de la información*, v. 31, n. 3, e310310.
https://doi.org/10.3145/epi.2022.may.10

Arcila-Calderón, Carlos; **Amores, Javier J.**; **Sánchez-Holgado, Patricia**; **Blanco-Herrero, David** (2021). "Using shallow and deep learning to automatically detect hate motivated by gender and sexual orientation on Twitter in Spanish". *Multimodal technologies and interaction*, v. 5, n. 63.
https://doi.org/10.3390/mti5100063

Arcila-Calderón, Carlos; **Blanco-Herrero, David**; **Valdez-Apolo, María-Belén** (2020). “Rechazo y discurso de odio en Twitter: análisis de contenido de los tuits sobre migrantes y refugiados en español”. *Reis*, v. 172, pp. 21-40.
https://doi.org/10.5477/cis/reis.172.21

Arcila-Calderón, Carlos; **Ortega-Mohedano, Félix**; **Jiménez-Amores, Javier**; **Trullenque, Sofía** (2017). “Análisis supervisado de sentimientos políticos en español: clasificación en tiempo real de tweets basada en aprendizaje automático”. *El profesional de la información*, v. 26, n. 5, pp. 973-982.
https://doi.org/10.3145/epi.2017.sep.18

Ayuso, Olga (2020). “No queremos tu carnet”. En: VV. AA. *Transfeminismo o barbarie*, pp. 217-224. Málaga: Kaótica Libros. ISBN: 978 84 12212921

Barabási, Albert-László (2016). *Network science*. Cambridge: Cambridge University Press. ISBN: 978 1 1070762 6 6
https://doi.org/10.1098/rsta.2012.0375

Bassignana, Elisa; **Basile, Valerio**; **Patti, Viviana** (2018). “Hurtlex: A multilingual lexicon of words to hurt”. In: *5th Italian conference on computational linguistics. CEUR workshop proceedings*, v. 2253.
https://doi.org/10.4000/books.aaccademia.3085

Berners-Lee, Tim (2021). “Why the web needs to work for women and girls”. *World Wide Web Foundation*.
https://webfoundation.org/our-work

Blondel, Vicent; **Guillaume, Jean-Loup**; **Lambiotte, Renaud**; **Lefebvre, Etienne** (2008). “Fast unfolding of communities in large networks”. *Journal of statistical mechanics: theory and experiment*, v. 10.
https://doi.org/10.48550/arXiv.0803.0476

Bot Ruso (2022). *Confesiones de un bot ruso.* Barcelona: Debate. ISBN: 978 84 18619151

Bradshaw, Samantha; **Bailey, Hannah**; **Howard, Philip N.** (2021). *Industrialized disinformation. 2020 global inventory of organized social media manipulation*. The computational propaganda project at Oxford Internet Institute.
https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/127/2021/01/CyberTroop-Report-2020-v.2.pdf

Breiman, Leo; **Friedman, Jerome H.**; **Olshen, Richard A.**; **Stone, Charles J.** (1984). *Classification and regression trees*. Belmont: Wadsworth.
https://doi.org/10.1201/9781315139470

Cabo-Isasi, Alex; **García-Juanatey, Ana** (2016). *El discurso de odio en las redes sociales: un estado de la cuestión*. Barcelona: Ajuntament de Barcelona.

Carratalá, Adolfo (2021). “Invertir la vulnerabilidad: el discurso en Twitter de organizaciones neocón y Vox contra las personas LGBTI”. *Quaderns de filología: estudis lingüístics*, v. XXVI, pp. 75-94.
https://doi.org/10.7203/QF.26.21979

Casero-Ripollés, Andreu (2016). “Estrategias y prácticas comunicativas del activismo político en las redes sociales en España”. *Historia y comunicación social*, v. 20, n. 2, pp. 533-548.
https://doi.org/10.5209/rev_hics.2015.v20.n2.51399

Chen, Jundong; **Li, He**; **Wu, Zeju**; **Hossain, Md-Shafaeat** (2020). “Analyzing the sentiment correlation between regular tweets and retweets”. *2017 IEEE 16th International symposium on network computing and applications (NCA)*.
https://doi.org/10.1109/NCA.2017.8171354

Crenshaw, Kimberlé (1989). “Demarginalizing the intersection of race and sex: a black feminist critique of antidiscrimination doctrine”. *Feminist theory and antiracist politics*, v. 140, pp. 139-167.

Davis, Mark (2019). “A new, online culture war? The communication world at Breitbart.com”. *Communication research and practice*, v. 5, n. 3, pp. 241-254.
https://doi.org/10.1080/22041451.2018.1558790

Dixon, Kitsy (2014). “Feminist online identity: analyzing the presence of hashtag feminism”. *Journal of arts and humanities*, v. 3, n. 7, pp. 34-40.
https://doi.org/10.18533/journal.v3i7.509

Duval, Elizabeth (2021). *Después de lo trans*. Valencia: La Caja Books. ISBN: 978 84 17496524.

EFE (2020). “J.K. Rowling acusada de transfobia en Twitter por comentario sobre menstruación”. *La vanguardia*, 7 de junio.
https://www.lavanguardia.com/gente/20200607/481654614948/jk-rowling-transfobia-twitter-menstruacion.html

Ekman, Paul (2003). “Darwin, deception, and facial expression”. *Annals of the New York Academy of Sciences*, v. 1000, n. 1, pp. 205-221.
http://doi.org/10.1196/annals.1280.010

Errasti, José; **Pérez-Álvarez, Marino** (2022). *Nadie nace en un cuerpo equivocado*. Barcelona: Deusto. ISBN: 978 84 23433322.

Etura-Hernández, Dunia; **Gutiérrez-Sanz, Víctor**; **Martín-Jiménez, Virginia** (2017). "La cultura mediática y el discurso posmachista: análisis retórico de Facebook ante la violencia de género". *Investigaciones feministas*, v. 8, n. 2, pp. 369-384.
http://doi.org/10.5209/INFE.55034

Fitzgerald, Jonathan D. (2017). "Sentiment analysis of (you guessed it!) Donald Trump's tweets". *Storybench*. Northeastern University School of Journalism.
https://www.storybench.org/sentiment-analysis-of-you-guessed-it-donald-trumps-tweets

Fraser, Nancy (1990). "Rethinking the public sphere: a contribution to the critique of actually existing democracy". *Social text*, v. 25/26, pp. 56-80.
https://doi.org/10.2307/466240

Frischlich, Lena (2022). "'Resistance!': collective action cues in conspiracy theory-endorsing Facebook groups. Impact of social media on social cohesion". *Media and communication*, v. 10, n. 2, pp. 130-143.
https://doi.org/10.17645/mac.v10i2.5182

Gil-Pascual, Juan-Antonio (2021). *Minería de textos con R*. Madrid: Universidad Nacional de Educación a Distancia. ISBN: 978 84 362 7711 1

Granovetter, Mark S. (1973). "The strength of weak ties". *American journal of sociology*, v. 78, pp. 1360-1380.
https://www.jstor.org/stable/2776392

Gutiérrez-Almazor, Miren; **Pando-Canteli, María-Jesús**; **Congosto, Mariluz** (2020). "New approaches to the propagation of the antifeminist backlash on Twitter". *Investigaciones feministas*, v. 11, n. 2, pp. 221-237.
https://doi.org/10.5209/infe.66089

Hu, Yifan (2006). "Efficient, high-quality force-directed graph drawing". *The mathematica journal*, v. 10, n. 1, pp. 37-71.
http://yifanhu.net/PUB/graph_draw.pdf

Jiang, Shaohai (2022). "Does social media promote or hinder health learning? The roles of media attention, information discussion, information elaboration, and information seeking experience". *Mass communication and society*.
https://doi.org/10.1080/15205436.2022.2090961

Jockers, Matthew (2017). "Syuzhet, extracts sentiment and sentiment-derived plot arcs from text".
https://rdrr.io/cran/syuzhet

Kearney, Michael W. (2018). "Tweetbotornot: an R package for classifying Twitter accounts as bot or not".
https://github.com/mkearney/tweetbotornot

Kearney, Michael W. (2019). "Rtweet: collecting and analyzing Twitter data". *Journal of open source software*, v. 4, n. 42, p. 1829.
https://doi.org/10.21105/joss.01829

Keller, Franziska G.; **Schoch, David**; **Stier, Sebastian**; **Yang, Junghwan** (2020). "Political astroturfing on Twitter: how to coordinate a disinformation campaign". *Political communication*, v. 37, n. 2, pp. 256-280.
https://doi.org/10.1080/10584609.2019.1661888

Lew, Zijian; **Stohl, Cynthia** (2022). "What makes people willing to comment on social media posts? The roles of interactivity and perceived contingency in online corporate social responsibility communication". *Communication monograph*s.
https://doi.org/10.1080/03637751.2022.2032230

Llorca-Asensi, Elena; **Fabregat-Cabrera, María-Elena**; **Ruiz-Callado, Raúl** (2021). "Desinformación populista en redes sociales: la tuitosfera del juicio del Procés". *Observatorio OBS*, v. 15, n. 3, pp. 124-146.
https://doi.org/10.15847/obsOBS15320211835

Maier, Daniel; **Baden, Christian**; **Stoltenberg, Daniela**; **De-Vries-Kedem, Maya**; **Waldherr, Annie** (2022). "Machine translation vs. multilingual dictionaries assessing two strategies for the topic modeling of multilingual text collections". *Communication methods and measures*, v. 16, n. 1.
https://doi.org/10.1080/19312458.2021.1955845

Mangold, Frank; **Scharkow, Michael** (2022). "Metrics of news audience polarization: same or different?". *Communication methods and measures*, v. 16, n. 2.
https://doi.org/10.1080/19312458.2022.2085249

Martini, Franziska; **Samula, Paul**; **Keller, Tobias R.**; **Klinger, Ulrike** (2021). "Bot, or not? Comparing three methods for detecting social bots in five political discourses". *Big data & society*, v. 8, n. 2.
https://doi.org/10.1177/20539517211033566

Menéndez-Menéndez, María-Isabel; **Amigot-Leache, Patricia**; **Iturbide-Rodrigo, Ruth** (2021). "Narrativas sexistas y hostilidad en foros de prensa digital: análisis en diarios de ámbito local". *Investigaciones feministas*, v. 12, n. 1, pp. 5-17.
https://doi.org/10.5209/infe.68665

Missé, Miquel (2018). *A la conquista del cuerpo equivocado*. Madrid: Egales. ISBN: 978 84 17319 36 6.

Missé, Miquel (2021). "No necesitamos aliados". En: Serra, Clara; Garaizábal, Cristina; Macaya, Laura. *Alianzas rebeldes*, pp. 147-157. Barcelona: Edicions Bellaterra. ISBN: 978 84 18684111

Miyares, Alicia (2021). *Distopías patriarcales. Análisis feminista del "generismo queer".* Madrid: Cátedra. ISBN: 978 84 376420 1 7

Mohammad, Saif (2016). "Sentiment analysis: detecting valence, emotions, and other affectual states from text". *Emotion measurement*, v. 2016, pp. 201-237.
https://doi.org/10.1016/B978-0-08-100508-8.00009-6

Mohammad, Saif; **Turney, Peter-David** (2010). "Emotions evoked by common words and phrases: using mechanical turk to create an emotion lexicon". In: Inkpen, Diana; Strapparava, Carlo (eds.). *Proceedings of the NAACL-HLT 2010 workshop on computational approaches to analysis and generation of emotion in text.* Los Angeles: Association for Computational Linguistics, pp. 26-34.
https://aclanthology.org/W10-02

Mohammad, Saif; **Turney, Peter-David** (2013). "Crowdsourcing a word-emotion association lexicon". *Computational intelligence*, v. 29, n. 3, pp. 436-465.
https://doi.org/10.1111/j.1467-8640.2012.00460.x

Moravčíková, Erika (2022). "Human downgrading. The concept of human degradation on social media". *Communication today*, v. 13, pp. 28-45.

Mulió, Leo (2020). "El debate que no debería serlo". En: VV. AA. (eds.). *Transfeminismo o barbarie.* Málaga: Kaótica Libros, pp. 197-214.

Muller, Karsten; **Schwarz, Carlo** (2021). "Fanning the flames of hate: social media and hate crime". *SSRN Electronic journal*, v. 19, n. 4, pp. 2131-2167.
https://doi.org/10.1093/jeea/jvaa045

Neubaum, German (2022). "'It's going to be out there for a long time': the influence of message persistence on users' political opinion expression in social media". *Communication research*, v. 49, n. 3, pp. 426-450.
https://doi.org/10.1177/0093650221995314

Noblía, María-Valentina (2015). "Un pacto de mutua agresión: la negociación de la imagen y el rol de la audiencia en los diarios digitales". *Textos en proceso*, v. 1, pp. 16-49.
https://doi.org/10.17710/tep.2015.1.1.2nob

Núñez-Puente, Sonia; **Fernández-Romero, Diana** (2019). "Posverdad y victimización en Twitter ante el caso de La Manada: propuesta de un marco analítico a partir del testimonio ético". *Investigaciones feministas*, v. 10, n. 2, pp. 385-398.
https://doi.org/10.5209/infe.66501

Olveira-Araujo, Rubén (2022). "La transexualidad en los cibermedios españoles. Presencia, preeminencia y temas (2000-2020)". *Profesional de la información*, v. 31, n. 1, e310102.
https://doi.org/10.3145/epi.2022.ene.02

Pariser, Eli (2011). *The filter bubble*. New York: The Penguin Press. ISBN: 978 1 59420 300 8

Pérez-Curiel, Concha; **Rúas-Araújo, José**; **Rivas-de-Roca, Rubén** (2022). "When politicians meet experts: disinformation on Twitter about Covid-19 vaccination". *Media and communication*, v. 10, n. 2, pp. 130-143.
https://doi.org/10.17645/mac.v10i2.4955

Pérez-Zúñiga, Ricardo; **Camacho-Castillo, Osvaldo**; **Arroyo-Cervantes, Gloria** (2014). "Las redes sociales y el activismo". *Paakat. Revista de tecnología y sociedad*, v. 4, n. 7.
http://www.udgvirtual.udg.mx/paakat/index.php/paakat/article/view/226

Piñeiro-Otero, Teresa; **Martínez-Rolán, Xabier** (2021). "Say it to my face: analysing hate speech against women on Twitter". *Profesional de la información*, v. 30, n. 5, e300502.
https://doi.org/10.3145/epi.2021.sep.02

Platero, Lucas (2017). "Prólogo". En: Stryker, Susan (eds.). *Historia de lo trans.* Madrid: Continta Me Tienes, pp. 7-15. ISBN: 978 84 947938 0 6

Platero, Lucas (2020). "Conocer nuestras genealogías". En: VV. AA. (eds.). *Transfeminismo o barbarie*. Málaga: Kaótica Libros, pp. 41-68. ISBN: 978 84 12212945

Plutchik, Robert (1980). "A general psychoevolutionary theory of emotion". In: Plutchik, Robert; Kellerman, Henry. (eds.). *Emotion. Theory, research, and experience: V. 1. Theories of emotion.* New York: Academic Press, pp. 3-33.
https://doi.org/10.1016/C2013-0-11313-X

Reguero, Patricia (2020). "Medio siglo de feminismo y transfobia". En: VV. AA. (eds.). *Transfeminismo o barbarie*. Málaga: Kaótica Libros, pp. 227-240. ISBN: 978 84 12212945

Ribera, Carles-Salom (2014). "Estrategia en redes sociales basada en la teoría de los vínculos débiles". *Más poder local*, v. 19, pp. 23-25.
https://dialnet.unirioja.es/servlet/articulo?codigo=4753468

Rodríguez-Fernández, Leticia (2019). "Desinformación: retos profesionales para el sector de la comunicación". *El profesional de la información*, v. 28, n. 3, e280306.
https://doi.org/10.3145/epi.2019.may.06

Rodríguez-Fernández, Leticia (2021). *Propaganda digital. Comunicación en tiempos de desinformación*. Barcelona: Editorial UOC. ISBN: 978 84 9180 792 6

Romero, Carmen (2020). "¿Quién teme al transfeminismo?". En: VV. AA. *Transfeminismo o barbarie*, pp. 17-38. Málaga: Kaótica Libros. ISBN: 978 84 12212945

Sánchez-Carballido, Juan-Ramón (2008). "Perspectivas de la información en Internet". *Zer*, v. 13, n. 25, pp. 61-81.
https://ojs.ehu.eus/index.php/Zer/article/view/3574

Sánchez-Duarte, José-Manuel (2016). "La red como espacio para la militancia política". *Comunicación y sociedad*, v. 29, n. 3, pp. 33-47.
https://doi.org/10.15581/003.29.3.33-47

Sauter, Disa A.; **Eisner, Frank**; **Ekman, Paul**; **Scott, Sophie K.** (2010). "Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations: correction". *PNAS. Proceedings of the National Academy of Sciences*, v. 107, n. 6, pp. 2408-2412.
http://doi.org/10.1073/pnas.0908239106

Serano, Julia (2020). *Whipping girl.* Madrid: Ménades. ISBN: 978 84 121285 4 3

Shane, Tommy; **Willaert, Tom**; **Tuters, Marc** (2022). "The rise of 'gaslighting': debates about desinformation on Twitter and 4chan, and the possibility of a 'good echo chamber'". *Popular communication*, v. 20, n. 5, pp. 178-192.
https://doi.org/10.1080/15405702.2022.2044042

Soto-Ivars, Juan (2021). "Prólogo". En: Shrier, Abigail (eds.). *Un daño irreversible*. Barcelona: Deusto, pp. 13-20. ISBN: 978 84 23432981

Stryker, Susan (2017). *Historia de lo trans.* Madrid: Continta Me Tienes. ISBN: 978 84 947938 0 6

Swati, Ubale; **Pranali, Chilekar**; **Pragati, Sonkamble** (2015). "Sentiment analysis of news articles using machine learning approach". *International journal of advances in electronics and computer science*, v. 2, n. 4, pp. 114-116.
http://www.iraj.in/journal/journal_file/journal_pdf/12-127-1430132488114-116.pdf

Van-der-Veen, Han; **Hiemstra, Djoerd**; **Van-den-Broek, Tijs**; **Ehrenhard, Michel**; **Need, Ariana** (2015). "Determine the user country of a tweet". *Social and information networks*.
https://arxiv.org/abs/1508.02483

Varela, Nuria (2019). *Feminismo 4.0. La cuarta ola*. Barcelona: Ediciones B. ISBN: 978 84 666644 3 1

Vidgen, Bertram (2019). "Tweeting islamophobia" [Doctoral thesis, University of Oxford]. British Library Ethos, e-theses online service.
https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.786187

Villar-Aguilés, Alícia; **Pecourt-Gracia, Juan** (2021). "Antifeminismo y troleo de género en Twitter. Estudio de la subcultura trol a través de #STOPfeminazis". *Teknokultura*, v. 18, n. 1, pp. 33-44.
https://doi.org/10.5209/tekn.70225

Vite-Hernández, Yara; **Cornelio-Landero, Rosa**; **Suárez-Ovando, Asbinia** (2020). "Activismo y violencia de género en las redes sociales en la actualidad". *Perfiles de las ciencias sociales*, v. 8, n. 15, pp. 111-137.
https://revistas.ujat.mx/index.php/perfiles/article/view/3903

Williams, Matthew (2021). *The science of hate*. London: Faber & Faber. ISBN: 978 0 571 35706 2

Wylie, Christopher (2020). *Mindf*ck Cambridge Analytica. La trama para desestabilizar el mundo*. Barcelona: Roca Editorial de Libros. ISBN: 978 84 18014 24 6