

Inflamando el debate público: metodología para determinar origen y características de discursos de odio sobre diversidad sexual y de género en *Twitter*

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

Note: This article can be read in its English original version on:

<https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/86993>

Cómo citar este artículo.

Este artículo es una traducción. Por favor cite el original inglés:

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>

Artículo recibido el 10-05-2022
Aceptación definitiva: 06-09-2022



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, España
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, España
mimenendez@ubu.es

Resumen

Este artículo estudia la reproducción de mensajes de fuerte carga ideológica cuyos orígenes o intereses permanecen ocultos para la opinión pública. Especialmente en las plataformas y redes sociales, se necesita transparencia sobre debates polarizados que deforman, impiden o distorsionan el enfoque crítico que toda sociedad debe poder construir respecto a cuestiones de gran interés social. La investigación ha demostrado que la hostilidad ha colonizado la comunicación digital a través de mensajes misóginos, homófobos, transfóbicos o xenófobos entre otros y que, con frecuencia, no son interacciones espontáneas o individuales. En el espacio virtual existen fuerzas que, invisibles fuera de él, construyen relato, generan desinformación y alimentan enfoques ideológicos generalmente regresivos. En nombre de la transparencia y la justicia social, urge investigar ese tipo de mensajes, así como sus posibles intereses de desestabilización en un momento de especial presencia y reputación de discursos como el feminista que, sin embargo, están experimentando una importante respuesta reactiva desde posiciones reaccionarias. El artículo investiga el origen y características de la conversación presente en la red social *Twitter* en relación con las identidades de género y sexuales. Para ello, se estudia una muestra significativa de tweets, más de un millón y durante un año completo, relacionados con los derechos de las mujeres, del colectivo LGBTIQ+ y las personas trans. Para abordarlo se emplean metodologías computerizadas mediante técnicas de aprendizaje automatizado (*machine learning*), procesamiento de lenguaje natural (PLN), determinación de bots, geolocalización y de aplicación de teorías de redes. Entre los resultados hay que destacar la presencia muy interrelacionada de grupos sin claros referentes, así como la existencia de lo que parecen redes coordinadas para dañar y provocar confrontación.



Palabras clave

Discurso de odio; Misoginia; Redes sociales; Medios sociales; Análisis de redes sociales; Transfobia; Feminismo; LGBTQI+; Diversidad de género; Diversidad sexual; Transexualidad; Transgénero; *Twitter*.

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 LGBTQI+ 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; LGBTQI+; Gender diversity; Sexual diversity; Transgender; Transsexuality; Transgender; *Twitter*.

1. Introducción

Internet es hoy la plaza pública, el lugar donde se construye debate, comunidad y reputación. Sin embargo, la segunda década del siglo XXI se está caracterizando por la presencia en el discurso público de ideas abiertamente problemáticas por su perfil misógino, xenófobo, transfobo u homofóbico. Los entornos virtuales se han convertido en el marco referencial de muchas personas, pero también se han convertido en generadores de dependencias desde las que se explotan las debilidades humanas (Moravčíková, 2022). Asimismo, el discurso de odio ha colonizado las plataformas digitales y las redes sociales y, según algunos estudios, estas pueden actuar como un mecanismo de propagación de delitos de odio al permitir la difusión de puntos de vista extremos (Muller; Schwarz, 2021). En este contexto, las investigaciones han demostrado la persistencia de los mensajes como elemento facilitador de influencia, en un ecosistema donde también es importante el efecto del silenciamiento de opiniones disidentes (Neubaum, 2022) así como la contingencia percibida y su relación con la interactividad (Lew; Stohl, 2022; Jiang, 2022).

En relación con las cuestiones de género, Internet sigue siendo un espacio de desigualdad tanto para las mujeres y las niñas como para grupos minoritarios y/o minorizados, como es el caso del colectivo LGBTQI+ (Lesbiana, Gay, Bisexual, Transgénero, Transexual, Travesti, Intersexual, Queer y + para todos los no representados con las siglas anteriores) o, siguiendo la denominación de Fraser, "contrapúblicos subalternos" (1990, p. 67). Estas brechas de género, atravesadas por el concepto de interseccionalidad de Crenshaw (1989), incluyen el daño online, efecto lejano a la idea generalizada que interpreta el mundo digital como un espacio lúdico y de comunicación sin más repercusión que la puesta en común de imágenes o vídeos (Berners-Lee, 2021).

El interés de este artículo reside en la necesidad de desvelar la reproducción de mensajes de fuerte carga ideológica cuyos orígenes o intereses permanecen ocultos para la opinión pública. Es relevante hacerlo porque gran parte de la ciudadanía considera que la participación en redes sociales es únicamente espontánea y/o intrascendente. Especialmente en las plataformas y redes sociales, urge la transparencia sobre debates polarizados que deforman, impiden o distorsionan el enfoque crítico que toda sociedad debe poder construir respecto a cuestiones de gran interés social. Cada vez más, la ciudadanía exige esta transparencia a la acción política y a los medios de comunicación, pero estas reclamaciones tienen menos recorrido en el mundo de las plataformas y redes sociales que, por otro lado, permanecen en manos privadas, funcionando sin apenas controles gubernamentales. En las redes, son ellas mismas las que deciden qué discursos son aceptables y cuáles no, sin ofrecer el marco de libertad de opinión sobre el que se presupone su existencia. La bibliografía ya ha demostrado su contribución a la desinformación o la divulgación de contenidos que colisionan con los principios democráticos (Llorca-Asensi; Fabregat-Cabrera; Ruiz-Callado, 2021; Acosta-Quiroz; Iglesias-Osores, 2020; Alonso-González, 2019; Rodríguez-Fernández, 2019).

Las redes y plataformas digitales no son únicamente un espacio de entretenimiento. En la actualidad son un espacio de militancia política y social y también un lugar en el que satisfacer necesidades informativas, sobre todo cuando

afectan a problemas de calado (Jiang, 2022; Vite-Hernández; Cornelio-Lander; Suárez-Ovando, 2020; Casero-Ripollés, 2016; Sánchez-Duarte, 2016; Pérez-Zúñiga; Camacho-Castillo; Arroyo-Cervantes, 2014). También han resultado útiles para la acción feminista en lo que se ha denominado *hashtag feminism*, con ejemplos como #MeToo o #NiUnaMenos (Dixon, 2014). En este ecosistema se hace necesario aportar transparencia mediante la visibilización de la naturaleza de algunas propuestas ideológicas que se difunden en redes sociales, propuestas que contribuyen al clima de tensión, la proliferación de ideas fuertemente polarizadas y la divulgación de mensajes potencialmente lesivos. La investigación de estas variables permite descubrir la existencia de fuerzas en el espacio virtual que, invisibles fuera de él, construyen relato, generan desinformación y alimentan enfoques ideológicos generalmente regresivos (Frischlich, 2022; Pérez-Curiel; Rúas-Araújo; Rivas-de-Roca, 2022; Berners-Lee, 2021).

Existe un discurso continuo donde se introduce odio (asociado principalmente al asco), polarización y asociaciones de ideas que se insertan en la sociedad aunque sean falsas o minoritarias

Así, el presente texto busca determinar el origen y características de la conversación presente en la red social *Twitter* en relación con las identidades de género y sexuales, aspectos sobre los que apenas existe bibliografía. En particular, y aunque ya existen trabajos sobre discursos de odio relacionados con los grupos LGBTIQ+ así como otros interesados por la misoginia, no se han encontrado investigaciones específicamente interesadas en discursos agresivos y/o de odio, en *Twitter* y en español, entre y hacia personas trans y grupos feministas. Para ello, se ha llevado a cabo un análisis de una muestra significativa de tweets (superior al millón, recopilados a lo largo de un año) para iluminar la procedencia y posible interés ideológico de determinados mensajes que están polarizando la opinión pública de una manera inédita en relación con temáticas como los derechos de las mujeres, del colectivo LGBTIQ+ y las personas trans. Tanto la metodología como los resultados pueden abrir líneas de investigación futuras para otras investigaciones que, o bien deseen profundizar en el mismo objeto de estudio o bien puedan replicar la metodología en el análisis de una problemática social diferente. Hay que insistir en la escasez de trabajos empíricos que se hayan interesado por el debate que se recoge en estas páginas, a pesar de su prevalencia e incluso de su repercusión fuera de las propias redes, incluyendo medios de comunicación de masas.

La elección de la red social *Twitter* se ha realizado, de acuerdo con Arcila-Calderón, Blanco-Herrero y Valdez-Apolo (2020), por su capacidad para viralizar contenidos rápidamente, su popularidad y su rapidez de comunicación, a pesar de que no es un medio que represente a toda la ciudadanía. Asimismo, hay que tener en cuenta que esta plataforma ofrece

“un registro abierto de sentimientos y opiniones acerca de asuntos de todo tipo, lo que incluye discursos de odio u otras muestras de rechazo que son expresados libremente y sin las barreras que muchas veces están presentes en espacios offline” (Arcila-Calderón; Blanco-Herrero; Valdez-Apolo, 2020, p. 23).

En España, la *twitteresfera* es uno de los actores primordiales en la canalización de la opinión pública en el espacio virtual y el principal generador de desinformación populista, construyendo comunidades identitarias a partir de la generación de hashtags que generan tanto cámaras de resonancia como la organización muy ágil de acciones de movilización (Llorca-Asensi; Fabregat-Cabrera; Ruiz-Callado, 2021, p. 129; Rodríguez-Fernández, 2021).

2. Marco teórico y contexto

2.1. Discurso de odio y género en Internet

Está documentado que, desde los inicios de Internet, la interacción en los espacios virtuales se caracteriza por la existencia de prácticas violentas, incluso cuando afecta a temáticas poco relevantes (Noblía, 2015, p. 16). La participación online está profundamente dañada por la existencia de trolls, campañas organizadas con fines espurios e influencia de sistemas interesados en favorecer climas de opinión, como las granjas de bots, los *cyberlobbies* o la proliferación de influencers que se convierten en líderes de opinión aun cuando sus inserciones no tengan veracidad alguna (Sánchez-Carballido, 2008, p. 72). Todas estas prácticas distorsionan las conversaciones y por tanto la opinión pública, expuesta a discursos hostiles de diferente intensidad, discursos muy polarizados y teorías de conspiración (Frischlich, 2022; Neubaum, 2022).

El discurso de odio no goza de una definición unánime, sobre todo porque no es lo mismo su significado social que el jurídico, pero, también, porque se trata de un concepto con un alto grado de subjetividad. Cabo-Isasi y García-Juanatey (2016) explican que se ha generalizado el concepto para referirse a un conjunto no homogéneo de manifestaciones que van desde las amenazas hasta la expresión de ira, incluyendo también la incitación o apología del terrorismo. En todo caso, se trata de una práctica directamente contraria a la libertad de expresión y que suele estar relacionada con ofensas racistas, xenófobas, sexistas y contra las minorías. La bibliografía ha identificado estos discursos con colectivos fascistas, la política de extrema derecha o grupos ultras (Cabo-Isasi; García-Juanatey, 2016, p. 9). De esta dificultad terminológica se deriva la propuesta de definir sus manifestaciones en redes sociales como *dangerous speech*, entendiendo que se trata de discursos basados en odio ideológico y con probabilidad de generar violencia (Amores et al., 2021).

En el caso de conversaciones relacionadas con el movimiento o la agenda feminista, la investigación académica ha reconocido la multiplicación de reacciones antifeministas (Alabao, 2021; Gutiérrez-Almazor; Pando-Canteli; Congosto, 2020; Etura-Hernández; Gutiérrez-Sanz; Martín-Jiménez, 2017), en un momento en que el activismo feminista ha entrado en la competencia “por el espacio discursivo en redes sociales” (Núñez-Puente; Fernández-Romero, 2019, p. 386).

Los avances en derechos y oportunidades que han ido conquistando las mujeres

“han sido contrarrestados, con frecuencia, por movimientos reactivos, no necesariamente explícitos, que se han convertido en frenos a la posibilidad de seguir avanzando” (Menéndez-Menéndez; Amigot-Leache; Iturbide-Rodrigo, 2021, p. 10).

Tanto el feminismo como los grupos trans, homosexual, queer o intersexuales, entre otros, han sido objetivos establecidos por *Breitbart* en la *alt-right* norteamericana como excesos “izquierdosos” y de la “élite”, que deben ser atacados dentro de sus “guerras culturales”, como vía para alcanzar un cambio político futuro en la sociedad (Davis, 2019). Dicha estrategia para colocar a la cultura como campo de batalla para cambiar la política, fue llevada a la práctica por parte de Steve Bannon, propietario de *Breitbart News*, a través de su otra empresa *Cambridge Analytica* en las redes sociales, participando en numerosas campañas en diversos países del mundo, siendo las más conocidas sus participaciones en el Brexit en Reino Unido o la elección de Donald Trump en Estados Unidos (Wylie, 2020).

Siguiendo la teoría sociológica de la fortaleza de los enlaces débiles de Granovetter (1973), las relaciones sociales de baja intensidad y de mayor distancia social son muy efectivas a la hora de distribuir información. Esto hecho tendría una gran influencia sobre grupos ya conformados, donde no haría falta convencer a un grupo completo, sino solo a unos pocos usuarios que introducirán dichos mensajes al resto. La utilización de conexiones de cuentas con pequeños vínculos con miembros de un grupo puede ser determinante, siendo este hecho conocido y empleado en campañas de marketing político en redes sociales (Ribera, 2014).

En lugares como *Twitter*, se refuerza la

“resistencia simbólica de ciertos sectores sociales ante el avance del feminismo” (Villar-Aguilés; Pecourt-García, 2021, p. 34),

mediante

“actitudes correctivas que muestran la pugna por construir el relato verdadero” (Menéndez-Menéndez; Amigot-Leache; Iturbide-Rodrigo, 2021, p. 15).

No se trata de una violencia aleatoria ni casual, sino que obedece a algunas pautas específicas que se han incrementado especialmente precisamente en redes sociales como *Twitter* (Piñeiro-Otero; Martínez-Rolán, 2021, p. 3).

El abuso online, según la *World Wide Web Foundation* (Berners-Lee, 2021), afecta especialmente a las mujeres, quienes reciben una ola de ataques que revelan una preocupante coordinación para ejecutar esas agresiones y que tienen como consecuencia la autocensura para evitar el daño. Asimismo, los estudios de esta *Fundación* demuestran que la moderación de contenido en las redes y plataformas no tiene en cuenta sus experiencias y necesidades, especialmente si pertenecen a comunidades marginadas. Así, conceptos como misoginia o sexismo no suelen utilizarse para moderar el discurso, tolerando en la práctica los abusos contra activistas y líderes femeninas. En julio de 2021, en el *Foro Generación Igualdad* en París,

<https://forum.generationequality.org/es/foro-generacion-igualdad-paris>

se animó a los CEOs de *Facebook*, *Google*, *TikTok* y *Twitter* a establecer medidas para abordar el abuso contra las mujeres en sus plataformas pues se estima que en torno al 45% de ellas han recibido violencia online, un abuso especialmente grave para las mujeres de color, las de la comunidad LGBTIQ+ y las de otras comunidades marginadas.

En cuanto a la diversidad, la creciente visibilización del colectivo LGBTIQ+ no garantiza un tratamiento positivo en unas redes recorridas por la negación de “lo diferente”. Oliveira-Araujo (2022) explica que estamos asistiendo a una visibilidad sin precedentes del colectivo trans, en parte por la

“hegemonía de la perspectiva queer [que] ha favorecido, entre otros, el análisis de la transexualidad a través de conceptos inclusivos como trans(género)”

en una apuesta por el análisis conjunto de diferentes hechos de diversidad sexual que, sin embargo, “induce potencialmente a sesgos” (Oliveira-Araujo, 2022, p. 3). Se refiere el autor a que el concepto “transgénero” (así como “trans”, a veces seguido de un asterisco), experimentó una rápida evolución en los años noventa del siglo XX en el ámbito anglosajón, pero todavía “está en construcción” (Stryker, 2017, p. 27). Su significado original evolucionó hacia el actual, un término paraguas que engloba “todas las formas de expresión de género e identidad no normativas” (Stryker, 2017, p. 206), haciendo referencia a la “diversidad de personas que caen fuera de los estándares” (Serano, 2020, p. 11). Sin embargo, en el estado español experimenta “un uso más politizado” (Platero, 2017, p. 9) e incluso como término paraguas es problemático porque “ahoga las voces” de las personas transexuales en particular (Serano, 2020, p. 408). Para Serano, el término es políticamente útil pero demasiado vago para exponer la diversidad de identidades y experiencias que supuestamente engloba (Serano, 2020, p. 46).

En todo caso, las campañas antigénero coinciden en criticar lo que se ha denominado “ideología de género” (Mulió, 2020, p. 198) y suelen destinarse a combatir el matrimonio entre personas del mismo sexo o la educación afectivo-sexual en las escuelas (Carratalá, 2021, p. 76). El discurso en las redes sociales construye un relato de autovictimización que problematiza los avances del colectivo LGBTIQ+ en términos de adoctrinamiento de la infancia, recorte de derechos y

discriminación de mayorías (Carratalá, 2021, p. 90). Señala Carratalá una cuestión no menor: en la red se difunden mensajes que sortean la retórica expresamente difamatoria, a pesar de su contenido profundamente homofóbico. Como novedad discursiva, destaca Alabao que el concepto de “ideología de género” ha servido para enlazar luchas diversas y “se ha mostrado como un potente movilizador social, un buen activador de los *pánicos morales* del conservadurismo cultural” (Alabao, 2021, p. 400).

Los ataques al feminismo y a las feministas no son hechos aislados ni espontáneos. Se relacionan con “una corriente de misoginia en red conectada con un movimiento antifeminista y antigénero internacional impulsado a través de organizaciones y partidos políticos de ideología de extrema derecha encargados de extender y amplificar este tipo de ataques” (Villar-Aguilés; Pecourt-García, 2021, p. 43).

La democratización de la opinión que auguraban las plataformas online ha mudado hacia un espacio que, en lugar de amplificar la igualdad de género y la diversidad, se ha convertido en la plaza pública desde la que elaborar mensajes que dañan a mujeres y comunidades marginadas y/o minorizadas.

2.2. Feminismo y políticas identitarias tras el #MeToo

Como ya se ha expuesto, el feminismo ha adquirido una presencia y reputación innegables desde 2017 con el fenómeno #MeToo en el que cristalizan las reclamaciones de justicia que habían comenzado en 2015 con

“enormes movilizaciones de mujeres que tomaron las ciudades en distintos rincones del planeta” (Gil, 2020, p. 289).

El 8 de marzo de 2018 se llegaría a lo que se ha denominado *explosión violeta*, cuando una huelga femenina de trabajos y cuidados tuvo un innegable impacto mundial (Varela, 2019). El éxito se repetiría en la cita de 2019.

El 8 de marzo de 2020, sin embargo, puede establecerse como el inicio formal de la respuesta reactiva. En España, los partidos políticos conservadores se aplicaron a culpar al movimiento feminista de la crisis de la pandemia de Covid-19, haciendo visible el malestar ante el avance del feminismo que ya se detectaba en las redes sociales e incluso en medios de comunicación convencionales (Alabao, 2021, p. 414). Las resistencias en el medio digital se habían disparado precisamente porque había aumentado significativamente el activismo feminista, revelando el carácter misógino de la cultura digital (Villar-Aguilés; Pecourt-García, 2021, p. 33).

Al mismo tiempo, en las redes sociales se polarizó una serie de cuestiones que desde hacía poco estaban produciendo conflicto en el movimiento feminista y que ya eran objeto de cierto debate en la Academia. La agenda feminista siempre ha sido diversa y en ocasiones ha experimentado escisiones entre posiciones teóricas o activistas cuyo enfoque o prioridad no eran los mismos. Temas como la prostitución, la gestación para terceros o incluso la interrupción del embarazo no han sido problematizados de la misma manera dentro del propio feminismo. El nuevo foco de tensión se produjo por la colisión de puntos de vista teóricos y prácticos, así como de objetivos estratégicos que presenta la agenda feminista en diálogo/disputa con la del activismo trans y el transfeminismo. Para varias autoras, una propuesta sobre “arriesgar el sujeto del feminismo”, en un evento de la *Universidad Popular de Podemos* de 2018, fue el detonante del conflicto (Reguero, 2020, p. 233; Romero, 2020, p. 19). A partir de ese momento, los debates entre el movimiento trans y el feminista con más presencia y beligerancia en la esfera social se articularon en torno a “si las mujeres trans son mujeres” o si “el trabajo sexual es trabajo” (Alabao, 2020, p. 131). Desde un punto de vista filosófico y político, el debate “se desplaza de la igualdad a la diversidad” (Miyares, 2021, p. 184).

El hecho es que, hasta 2018, la participación de las mujeres trans en el feminismo había sido la norma, también en el estado español: “bien acogidas, en general, aunque no aproblemáticamente” (Platero, 2020, p. 44), dando lugar al “transfeminismo” y a la inclusión de los hombres trans (Alabao, 2020, p. 147; Missé, 2021, p. 149). Sin embargo, del citado clima de confrontación surgió la acusación de transfobia, sintetizada en el uso del término TERF (acrónimo de *trans exclusionary radical feminists* o feministas radicales trans excluyentes). Algunos autores destacan que “la TERF es una postura minoritaria dentro del feminismo” aunque “está teniendo gran influencia social y política” (Mulió, 2020, p. 198).

El término apareció con fuerza en las redes sociales, específicamente *Twitter*: “no sabía que era una TERF [...] hasta que llegué a *Twitter*” (Ayuso, 2020, p. 219), aunque no era nuevo. Stryker (2017, p. 295) sitúa su origen en Estados Unidos y en 2008, cuando la bloguera *TigTog* lo acuñó para describir, neutralmente, las posiciones que no contemplaban la participación de mujeres trans en los grupos activistas. Por tanto, en su origen se utilizaba como forma de definir a quienes consideraban parte del movimiento feminista a las personas transexuales y a quienes no. Hoy en día se ha convertido en una descalificación, un insulto (independientemente de la adscripción al feminismo radical) e incluso un discurso de odio que se ejerce contra personas consideradas transexcluyentes (Miyares, 2021). La *Real Academia Española (RAE)*, en un tweet publicado el 14 de abril de 2021, definía “terfa(s)” como una expresión despectiva.

Desde el movimiento trans se denuncia el discurso transexcluyente que habita en las redes sociales, identificando a algunas feministas en particular (especialmente las más vinculadas al *PSOE* o al *PSC* y también las académicas del feminismo de segunda ola) como el origen de las conversaciones (Duval, 2021, p. 83; Platero, 2020, p. 43). En este sentido, feministas con presencia histórica en España, tanto en la acción política como en la académica, se estarían negando a

ser “desplazadas” por nuevas perspectivas y protagonistas (Platero, 2020, p. 63; Reguero, 2020, p. 240). Algunos textos se lamentan de que estas posiciones se incorporan a la discriminación y “la abonan y la legitiman indirectamente” (Alabao, 2020, p. 130). En este conflicto es importante contemplar la diferencia de enfoques entre los dos partidos que comparten el Gobierno (*Partido Socialista y Podemos*), cuestión que estaría favoreciendo la presencia del feminismo en la agenda política pero también una discusión por la hegemonía (Duval, 2021, p. 259; Reguero, 2020, p. 238).

Miquel Missé, autor trans, hace referencia a estos debates y reconoce que

“el feminismo está más abierto que nunca a escuchar las propuestas del activismo trans”

pero, paradójicamente,

“el activismo trans es más hostil que nunca con el feminismo” (Missé, 2018, p. 155).

En este contexto, muy marcado por las políticas identitarias, aparece el término “transfobia” para desacreditar a quien opina de manera distinta:

“Cuando alguien dice algo que se sale del patrón de lo que se puede decir sobre el tema trans es directamente tachado de transfobo” (Missé, 2018, p. 154).

Así, el pensamiento crítico se considera un ataque a los derechos humanos, de forma que se desautoriza la posibilidad de opinar: la acusación de transfobia zanja el debate (Errasti; Pérez-Álvarez, 2022, p. 19), convirtiéndose en un estigma (Soto-Ivars, 2021, p. 13). Julia Serano, en la misma línea, argumenta que

“la mayor parte de la transfobia a la que he tenido que enfrentarme como mujer trans se podría describir más bien como misoginia” (Serano, 2020, p. 21),

de ahí que proponga que la unión entre el movimiento feminista y el trans debe centrarse en la lucha contra la devaluación de las mujeres y la femineidad (Serano, 2020, p. 355).

2.3. Objetivos e hipótesis de la investigación

Para abordar el discurso hostil en *Twitter*, se llevó a cabo un análisis de las conversaciones en torno a la diversidad sexual y de género, mediante una metodología mixta que permitiera interpretar las interacciones relacionadas con mujeres, feminismo y activismo trans. Se realizó una acotación temporal de un año, que permitió recopilar una muestra significativa a partir del 2 de marzo de 2020, dado que, como se ha señalado en el epígrafe anterior, se había identificado previamente el 8 de marzo de 2020 como el principio de la respuesta reactiva al discurso igualitario.

En primer lugar se buscaba un objetivo general:

(O1), caracterizado por determinar la existencia de discursos hostiles o de odio en las conversaciones en *Twitter* sobre el objeto de estudio, partiendo de una hipótesis inicial que planteaba que la hostilidad se manifestaría mediante interacciones con un importante componente sexista y/o misógino.

Además de conocer esta incidencia, como objetivos secundarios se planteaba:

(O2) identificar los términos ofensivos más utilizados y sus características desde el punto de vista del nivel de odio, las emociones asociadas y la polaridad positiva-negativa;

(O3) investigar el origen de estos mensajes para intentar iluminar intereses organizados en la línea de la desinformación populista, la participación de sistemas para favorecer climas de opinión y/o la discusión por la hegemonía en el movimiento feminista;

(O4) observar la posible existencia de campañas organizadas antigénero;

(O5) analizar el conflicto entre feminismo y movimiento trans con el objeto de dilucidar si son sostenibles las acusaciones planteadas desde ambos movimientos contra el contrario;

(O6) determinar el rol de conceptos como “terf” en la esfera pública.

Para ello se establecen las siguientes hipótesis de partida:

(H1) Existe un amplio discurso polarizado y de odio en el debate en la red social *Twitter* entre los grupos trans y feministas; y

(H2) se estima que el debate estaría en parte provocado y promovido de manera artificial para generar conflicto.

Se trata de una investigación con metodologías apenas aplicadas en la actualidad en ciencias sociales, y mucho menos en el idioma español en todo el ámbito latinoamericano. Su novedad implica también que son métodos en desarrollo y que por un lado necesitan evolución, pero a la par ya permiten analizar unos volúmenes de datos que serían inabarcables desde el punto de vista manual.

3. Metodología

El análisis de grandes cantidades de mensajes, como los extraídos a partir de redes sociales, es una tarea casi imposible de abarcar sin la ayuda de las nuevas tecnologías, el uso de técnicas de aprendizaje automatizado (*machine learning*), procesamiento de lenguaje natural (PLN) y de aplicación de teorías de redes. Estudios como el de Mangold y Scharkow

(2022) ya se han interesado por métodos para medir la polarización de las audiencias. La utilización de softwares ha empezado a desarrollarse en los entornos académico, empresarial y electoral para analizar grandes cantidades de datos de manera objetiva, como el análisis de sentimientos (**Arcila-Calderón et al.**, 2017). Aprovechando estas técnicas, esta investigación tiene como referencia el debate en la red social *Twitter* en torno y entre los grupos identificados como “trans”, “terf” y “feministas”.

“ Todo lo expuesto revela un discurso internacional, mantenido en el tiempo, que no sería aislado ni espontáneo ”

La recogida de la muestra se realizó mediante la librería *RTweet* (**Kearney**, 2019) en *R*, a través de conexión con el interfaz de programación de aplicaciones de *Twitter* en su versión 1.1. Para ello se recogieron todos los mensajes (incluidos retweets y respuestas) que cumplieran la consulta “trans OR terf AND mujer” en castellano a nivel global (por lo tanto que incluyera la palabra “mujer” y una de las siguientes: “trans” o “terf”). La elección de estas palabras proviene del debate principal entre los distintos colectivos, ubicado en la definición de qué es ser mujer para definir a una persona trans y el correspondiente ataque despectivo de la palabra terf hacia ciertos sectores feministas. No se añadieron más palabras como filtros AND para recoger el máximo volumen de mensajes alrededor de este fenómeno.

El periodo de recogida fue de un año completo, entre el 2 de marzo de 2020 y el 1 de marzo de 2021, alcanzando la cifra de 1.079.998 tweets. Como la versión de la API 1.1 de *Twitter* permite únicamente recogidas de los últimos 7 días, se fueron recogiendo durante un año completo, semana a semana, uniendo al final todos en un único conjunto. A partir de ese corpus se realizaron los siguientes procesos, encaminados a alcanzar los objetivos de esta investigación:

- Análisis de redes aplicado al área de comunicación (**Barabási**, 2016), mediante la aplicación del software *Gephi* en su versión 0.9.4. Con este programa se pudo realizar la asignación de cada cuenta a un grupo o clúster y su zona de influencia (**Chen et al.**, 2020) mediante el algoritmo de Louvain (**Blondel et al.**, 2008) sobre los retweets realizados. En la representación gráfica de la red se aplicó el algoritmo de **Hu** (2006), pensado para redes que rondan o superan el millón de nodos (cuentas) analizados.
- Estudio y cuantificación del sentimiento o polaridad (positiva-negativa), así como de las emociones básicas a través de procesamiento de lenguaje natural (PLN). Para ello se utilizó, mediante *R* y con la ayuda de la librería *Syuzhet* (**Jockers**, 2017), el léxico del *National Research Council of Canada (NRC)* en su versión 0.92 en español con más de 14.000 palabras (**Mohammad; Turney**, 2010; 2013). De esta manera se asocia a cada mensaje valores positivos y negativos de polaridad y emociones por las palabras que contiene, potenciadas o disminuidas por las que están a su alrededor (**Swati; Pranali; Pragati**, 2015). Igualmente se etiquetan y cuantifican las emociones básicas descritas por **Plutchik** (1980) y más tarde **Ekman** (2003), siendo las siguientes: asco, anticipación, disgusto, miedo, alegría, tristeza, sorpresa y confianza, aplicadas a cada discurso de cada tweet (**Sauter et al.**, 2010). La superación de niveles por encima de 3 en cualquier emoción o polaridad puede considerarse significativa (**Fitzgerald**, 2017).
- Análisis y cuantificación del odio mediante técnicas de PLN, de forma análoga al punto anterior, pero empleando en esta ocasión el léxico *Hurtlex* en su versión en español de 2018 con unas 5.000 palabras asociadas (**Bassignana; Basile; Patti**, 2018). La presencia de palabras del léxico de odio da un valor de presencia de nivel, que se ve potenciado por las palabras que están a su alrededor. Al igual que en el punto anterior, los valores detectados con valor a partir de 3 son considerados de odio claro, evitando así la inclusión de uso irónico o cotidiano de algunas palabras en los mensajes.
El léxico *Hurtlex*, según sus autores, ha tenido un buen resultado de detección en español y se emplea en el aprendizaje automático. Son palabras que han sido categorizadas como de discurso de odio por parte de personas, y el algoritmo empleado, *Syuzhet*, posteriormente las analiza. Es una metodología semejante pero distinta a los algoritmos *Naive Bayes*, en los que se identifican frases con odio, y es el aprendizaje automático el que busca las palabras y relaciones que hay entre ellas como modo de entrenamiento y aprendizaje para luego ser aplicado ante nuevas frases, tal como emplean **Arcila-Calderón et al.** en español (2021). Si bien existen avances metodológicos sobre diccionarios multilingües (**Maier et al.**, 2022), la detección de odio en español es todavía muy incipiente y necesita aun desarrollo futuro.
- Determinación probabilística de comportamiento bot a través del algoritmo *Tweetbotornot* de **Kearney** (2018). En función de la naturaleza, historial y comportamiento de cada cuenta se determina un porcentaje de la probabilidad de ser un bot. Este método emplea el análisis de la biografía, seguidores, amigos, actividad, menciones, fotografía, localización y los últimos 100 mensajes emitidos, entre otros, para determinar su probabilidad de comportamiento bot. Su probabilidad de acierto, según el autor, es de 93,53% para clasificar bots y de un 95,32% en no-bots. Este algoritmo está considerado entre los mejores y más empleados en ciencias sociales (**Martini et al.**, 2021).
- Estudio de minería de texto mediante uso de nube de palabras y dendrograma entre las palabras con mayor presencia entre el discurso con mayor carga de odio (niveles por encima de valor 4 según el análisis y cuantificación de odio). Con ello se pretende exponer de manera gráfica no solo las palabras más empleadas, sino también su relación. En el dendrograma se realizó un nuevo análisis de clusters para determinar los diez grupos principales de palabras mediante algoritmo *k-means*. Para determinar su ajuste se calcula su índice de Dunn, siendo los valores más elevados los más deseables (**Gil-Pascual**, 2021). En el análisis se han eliminado las palabras que no aportan significado (*stopwords*) y símbolos (como por ejemplo #) o signos de puntuación.

- Análisis de relación estadística entre variables mediante correlación de Pearson en *R*, en este caso el odio, tanto por la situación en la red como su probabilidad de ser un bot o las emociones y sentimientos que desprenden.
- Estudio de árbol de regresión a través de técnicas de aprendizaje automático (*machine learning*), que permite predecir el comportamiento de una variable (Lantz, 2019). Se utilizan las variables del punto anterior, pero considerando al odio como una variable dependiente de las demás, que serían independientes. Con este fin se emplea el algoritmo CART (Breiman et al., 1984) a través del paquete *rpart* en *R*:
<https://cran.r-project.org/web/packages/rpart/index.html>

La fiabilidad de la predicción se realiza a través de la correlación entre valores reales y previstos, además del error absoluto medio (MAE).

- Recogida de mensajes paralela a la principal de manera geolocalizada en el último periodo de análisis, sobre cada uno de los más de 70 países identificados por Bradshaw, Bailey y Howard (2021) como emisores de desinformación en redes sociales. La búsqueda se realizó mediante la introducción de coordenadas de longitud, latitud y radio de influencia en la consulta. Esta técnica presenta una fiabilidad media mundial del 77,84%, superando el 90% en algunos países occidentales debido a la calidad de su red (Van-der-Veen et al., 2015).

Los datos de odio y probabilidad bot se representan mediante diagramas de caja, que representan en su mitad el valor de la mediana, una caja que representa en su inferior el primer cuartil y en el superior el tercer cuartil. Un asterisco representa el valor medio, mientras que por encima y debajo aparece una línea que representa un valor de 1,5 por encima o debajo del valor intercuartílico, mientras que los puntos sueltos expresan datos atípicos desde un punto de vista estadístico.

4. Resultados

Tras la recogida de los datos entre el 2 de marzo de 2020 y el 1 de marzo de 2021, un año completo, con un total de 1.079.998 tweets se procedió a la asignación de cada mensaje a un grupo o clúster, a través de los retweets emitidos. Este proceso arrojó la formación de 5.060 grupos, con un valor de modularidad de 0,770, bastante bueno de asignación para ciencias sociales. De entre todos esos grupos, la investigación se centró en los 12 principales, que suponen el 72,47% de los mensajes, con un total de 782.722 tweets que pueden observarse en la tabla 1 y representados en la figura 1.

Tabla 1. Principales 12 grupos detectados

Nº grupo	Tendencia	Porcentaje tráfico generado de RTs
958	<i>Grupo Queer</i> : aunque variado, especialmente con origen en España y con bastante ataque hacia el feminismo por su supuesta definición de mujer.	11,99%
131	<i>LGBTi México</i> : grupo de apoyo al grupo trans en el país.	10,57%
1163	<i>Grupo trans</i> : grupo de defensa trans y ataque a feminismo o autoras como JK Rowling.	10,07%
593	<i>Grupo Argentina</i> : semejante al de México, pero desde Argentina.	8,63%
577	<i>Trans</i> (con numerosas cuentas suspendidas o borradas tras una cata aleatoria y entre las cuentas de mayor valor de vector propio): grupo de ataque a feminismo y a figuras como JK Rowling.	7,29%
464	<i>Grupo trans</i> : grupo de apoyo y defensa de la definición de mujer sin falta de determinados aspectos.	6,75%
91	<i>Trans PSOE</i> : grupo de apoyo a las políticas del partido político español en temáticas trans.	5,73%
0	<i>Feministas</i> : grupo que pretende aclarar que no atacan al grupo trans.	5,52%
45	<i>Activistas LGBTi</i> : lucha de derechos del colectivo y apoyo al grupo trans.	4,56%
1.717	<i>Ultraderecha y cuentas suspendidas</i> (tras cata aleatoria y de las principales cuentas por mayor valor de vector propio): ataque a los grupos feministas aprovechando el discurso trans.	3,24%
680	<i>Influencers Latinos</i> : grupos seguidores de influencers de varios países latinoamericanos, especialmente centroamericanos, que apoyan principalmente los ataques contra el feminismo por desplazamiento a los trans.	2,97%
1.429	<i>Ultraderecha</i> : ataque al feminismo y grupos trans.	2,68%

La selección entre los doce clusters principales arroja un total de 146.163 mensajes directos y 636.559 retweets (RTs). El empleo del algoritmo de detección y de cuantificación de odio entre estos mensajes determina una ausencia total únicamente en 8.043 mensajes (1,03%), con un valor medio de 2,73 de intensidad, una mediana de 2,00 y un máximo de intensidad en 32. La presencia de odio puede detectarse en todos los grupos, pero tiene especialmente presencia en algunos de ellos: 593, 577, 464 y 45, grupos supuestamente trans pero sin grandes influencers, y presencia de numerosas cuentas suspendidas posteriormente. La procedencia de las cuentas es principalmente española, aunque existen importantes grupos procedentes de Argentina y México, que se concentran en sus clusters, y en menor medida de Colombia, Chile, Venezuela y Perú.

Puede observarse en el grafo que muchos de los grupos están interconectados entre sí, muy entremezclados alrededor del grupo principal, el 958, el grupo llevado por influencers queer. Los grupos más alejados serían el 1163, grupo trans, y 1429 y 1717 que se sitúan en torno a la ultraderecha política. Se trata de una estructura de red circular alrededor de un grupo que vertebraría los mensajes y donde no se aprecian grupos separados a excepción de los más alejados anteriormente citados. De esta manera, a partir del grupo queer central casi todos los grupos se entremezclan entre sí, no apreciándose grupos bien definidos y seguidores de unos claros influencers opuestos. Esta estructura tan interrelacionada en la mayor parte de los grupos define un grado de interconexión que no se correlaciona con el grado de odio que se estudia en el resto del artículo, que debería representarse, teóricamente, en grupos mejor diferenciados.

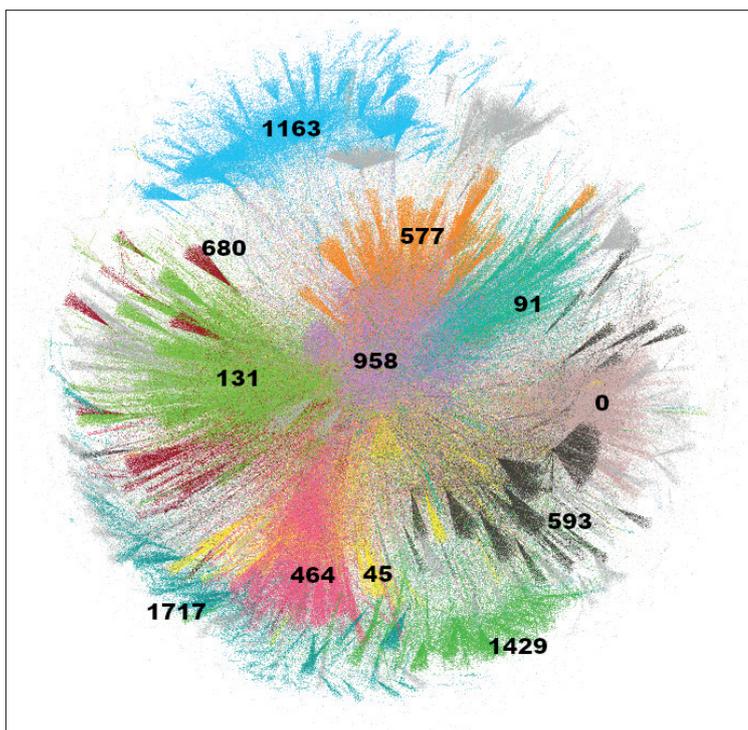


Figura 1. Grafo de relaciones de redes

Aunque la presencia de odio en la media y la mediana son apreciables (figura 2) y casi el 99% de los mensajes tienen algún componente de odio, para evitar problemas de usos con dobles intenciones, irónicos o de uso común, el estudio se centró en los mensajes de odio que situados en el tercer cuartil, con valores de odio 4 o superior. Esta submuestra es de 204.700 mensajes, con 29.985 mensajes directos y 174.715 RTs. Los 29.985 mensajes directos de mayor odio fueron emitidos por parte de 13.160 cuentas, creadas principalmente entre los años 2019 y 2020 y en menor medida 2011, de las cuales en febrero de 2022 se comprobó que 9.405 (71,47%) habían sido borradas por el usuario o suspendidas por infringir las normas de *Twitter*, y una pequeña parte protegidas sin actividad.

A partir de esta submuestra, por minería de texto se realizó la representación de una nube con las principales palabras más empleadas, así como un dendrograma con los principales temas tratados en los mensajes con mayor odio (figura 3). Como se especificó en la metodología, se determina sobre los mensajes que contenían mayor nivel de odio (nivel a partir de 3). El estudio del dendrograma presenta un buen ajuste para ciencias sociales, con un valor del índice de Dunn de 0,559 (en tanto por uno), con lo que los temas han sido bien detectados según el proceso realizado.

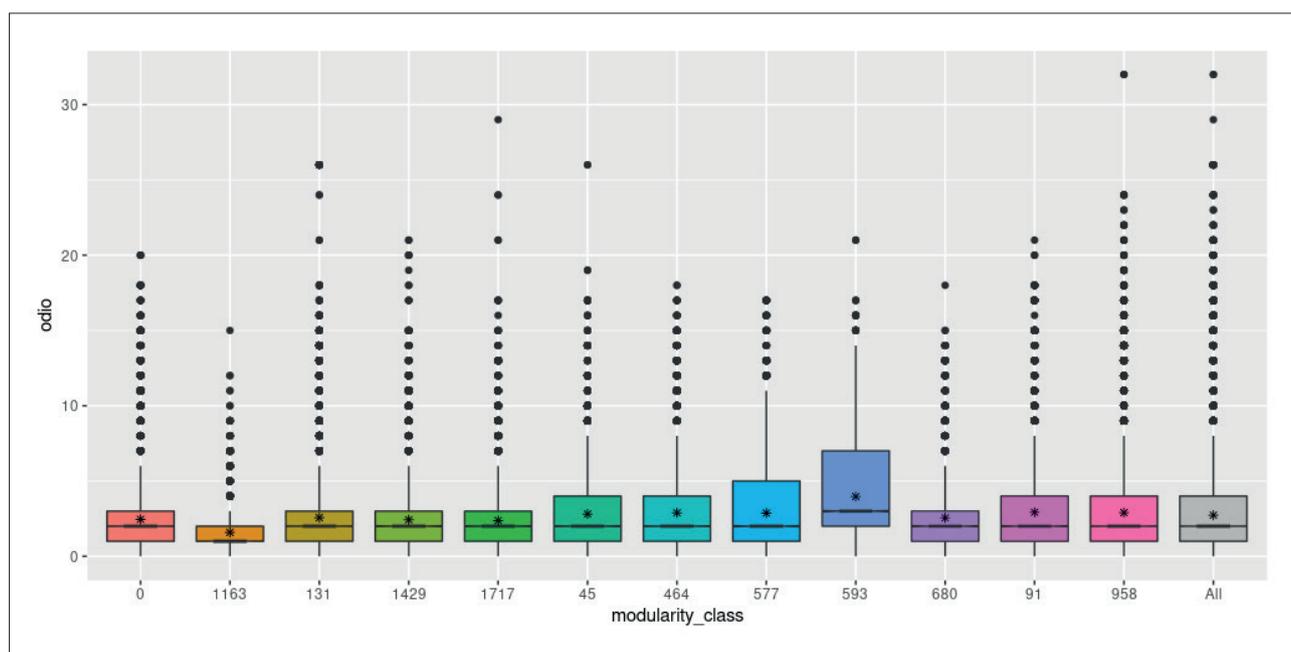


Figura 2. Presencia de odio por cada clúster principal

ceptualicen un tipo de asociaciones. La presencia tan común de algunas de ellas daría lugar al segundo razonamiento, con formación de palabras como “feminismoradical”, “supuestoantifa”, “pedoliteral” o “dominecabra”. De esta manera se estaría estableciendo un vocabulario creado para la asociación de determinadas ideas.

Los tweets con mayor carga de odio detectados son, a modo de ejemplo, los descritos en la tabla 2.

Tabla 2. Mensajes con mayor carga de odio detectada

Texto	Nº favoritos	Nº RTs	Clúster	Probabilidad de ser bot
Dice algo una persona trans con datos, con vivencias propias, etc... Respuesta de TERF: MÁTATE. ERES TONTO. NO TIENES RAZÓN CUIR DE MIERDA. MÁTATE. ERES IDIOTA. TONTO DE MIERDA. Se nota que tenéis muchos argumentos de peso.	1	0	958	3,96%
@dyJulieh Esta mierda es tan ridícula que ya no sabe uno ni que decir. Ahora todo tiene que ver con ser negro, trans, pobre, feo, mujer feminazi, gay. Que LADILLA hasta que nivel de estupidez nos esta llevando esta agenda progre basura (la cuenta @dyJulieh ya no existe)	0	0	1.717	44,10%
Hay un método muy simple pa saber si el humor es aceptable o no Pregunta(te) donde está la gracia, si la respuesta es “que es negro, mujer, gay, bi, trans, tonto etc...” el chiste no es aceptable. Tmbn puedes pensar en si ataca a un colectivo oprimido, pero no creo q os de... https://t.co/w7A2YWqqzq	0	0	958	81,27%
Yo pensaba así también, que perra vergüenza me da ahora. Le debemos muchos de nuestros derechos a una mujer trans vih+ y afrodescendiente. Así que hacerte tonto y discriminar solo porque tu crees que eres “menos gay” está de la verga. Los gays heteronormados no saben coger FACT https://t.co/o0pQxPEkqg	505	62	131	6,73%
@uwuk1ra @satogarca Sabías que las mujeres trans tienen el cerebro formado de una manera parecido al de una mujer biológica? Pero pues eso te lo pasas por el culo, solo eres un idiota de mierda (@uwuk1ra está suspendida por incumplimiento de normas, @satogarca ya no existe)	0	0	131	72,36%
@guirimadri @abixquertg @DieBatsuDie Somos del reino Animalia efectivamente, pero tratar a un ser humano cual animal salvaje y dar esa clasificación, más si vas a tocar fibra sensible a una persona trans por mierda que eres, me parece una falta. Calladita te ves más bonita, terf Soy estudiante de Biología por cierto (@abixquertg ya no existe)	0	0	958	13,52%
También es día contra la transfobia y qué puto coraje que le diga “este pendejo” a una mujer trans que únicamente está mostrando su apoyo al movimiento. Me tienen hasta la verga con sus exclusiones transfóbicas y retrógradas de mierda. https://t.co/otYORArzPY	5	2	131	73,96%
Hombre: Señorita, es usted muy hermosa. Mujer: MALDITO PERRO, DESGRACIADO, ACOSADOR MISÓGINO, VIOLADOR MAL PARIDO!! Gay/Trans: Ay perra k culote te cargas y esas pinches chichotas, te pasas maldita, te odio! mmm!! *la sabrosea toda* Mujer: Ay amiga jijí gracias, eres bien linda!	3	1	1.717	0,20%
@cxsmic_girl @_AlexClockwork @_erosgarcia @Dalma2206 LUEGO PEDIREIS RESPETO, VERDAD, PUTAS TERFS DE MIERDA, SOIS ESKORIA. ¿TÍOS TRANS?. TU PUTA MADRE, LOKA DEL COÑO. YA SABEIS A LO QUE VAIS A ENFRENTAROS, COMO SIGAIS DANDO X CULO CON LAS TRANS. LOS NAZIS DAIS ASKO, PERO NO HAY ALGO QUE DE MAS ASKO K UNA TERF RETRASADA DE MIERDA. 🗣️🗣️🗣️	0	0	91	36,92%
Que triste ver a gays, que disfrutaban de derechos y cierta libertad gracias a los trans que han dado la pelea por ellos en años anteriores, tirandolos bajo el bus y aun cuestionando quien es hombre o mujer por sus gónadas o algo tan tonto como la menstruación, basuras	3	0	593	4,36%

Los mensajes, en un número muy importante con odio asociado, que se recogen tanto entre los análisis de minería de texto, como entre los de mayor intensidad de odio, vienen a exponer un mensaje claro contra el feminismo, al que se acusa de intransigente y de limitar la definición de mujer. Así, recibe clasificaciones de secta y diferentes insultos, especialmente ligado al término despectivo “terf” (*trans exclusionary radical feminists*). Por contra, el término “trans” está asociado a los términos “cis”, “mujeres” y “hombres” y de necesidad de antidepresivos. Se expresa así un discurso muy visceral, lleno de odio, pero donde no se aprecian debates locales, sino entre colectivos independientemente del lugar donde se encuentren. Es de destacar que numerosas cuentas que participan o son interpeladas en los mensajes, un año después han sido eliminadas por los usuarios o suspendidas por infringir las normas de la red social *Twitter*.

Aunque el odio es intenso en casi todos los grupos, se detecta un porcentaje de probabilidad bot muy alto, especialmente en determinados clusters, como se expone en la figura 3. No se aprecia una correspondencia estadística entre la posición que ocupa la cuenta en la red, aunque algo con el discurso emocional que vierten dichas cuentas. De esta manera, los valores de correlación de Pearson entre el nivel de odio y otras variables de naturaleza y posición en la red son totalmente despreciables: -0,016 Probabilidad bot, 0,0002 Favorito, -0,081 RT, 0,152 de centralidad de intermediación, -0,008 con vector propio o *eigenvector*, -0,003 con grado y -0,005 con centralidad de cercanía.

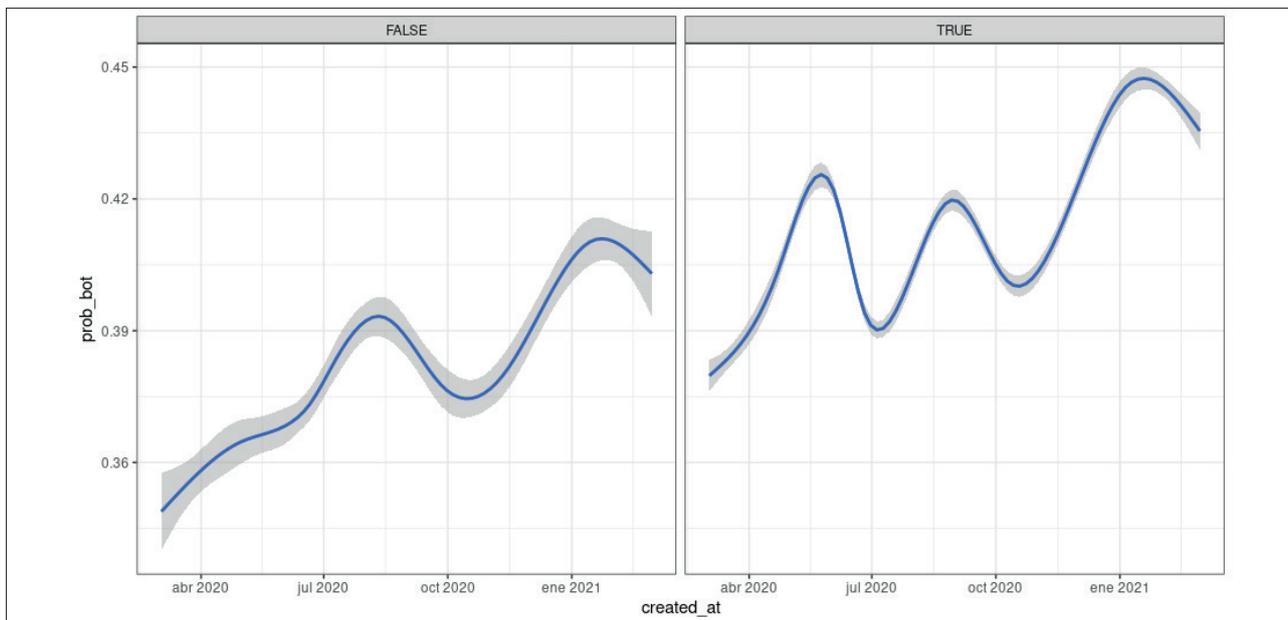


Figura 4. Probabilidad media de ser bot a lo largo del tiempo en cuentas que hacen mensajes directos (False) y las que hacen RTs (True)

Por tanto, el odio no tiene relación directa con ser un influencer y tener mayor seguimiento, ser más retweeteado, ser más marcado como favorito, o ser un bot. Tal como se puede ver igualmente en la tabla 2, los tweets caracterizados por el algoritmo muestran mensajes con mucho odio independientemente de su porcentaje de probabilidad bot, número de favoritos o de RTs. Por este motivo no puede caracterizarse que sean los bots los que generan odio, sino que existe un perfil de usuario que puede ir muy ligado o no a sus influencers (centralidad de cercanía), vertebran una red (intermediación) o son influencers de su red (vector propio).

En la figura 4 puede comprobarse como, a lo largo del año analizado, el número de cuentas que hacen RTs distribuyendo odio, independientemente de su intensidad, es superior a las cuentas que lanzan mensajes propios. Esto viene a corroborar lo dicho por Williams (2021), según lo cual los bots no crean odio, pero lo “hinchán”. El que no exista a su vez, como se dijo anteriormente, correlación entre probabilidad bot e intensidad viene también a confirmar lo expuesto por Vidgen (2021): las cuentas no exhiben grandes intensidades en su discurso para no ser detectadas fácilmente y así poder permanecer más tiempo en la red. Este aspecto en que los bots se encargarían de redifundir los mensajes de otros, más que en generarlos, enlazaría con lo expuesto por Keller (2019) en redes sociales durante las elecciones de Corea del Sur o la pandemia Covid-19 en España (Arce-García; Said-Hung; Mottareale-Calvanese, 2022).

La presencia de bots en función del clúster o grupo identificado inicialmente (figura 5), exponen en general un comportamiento medio, en muchos casos casi cercano al 50%, llegando varios grupos en su tercer cuartil a un 75% de posibilidad

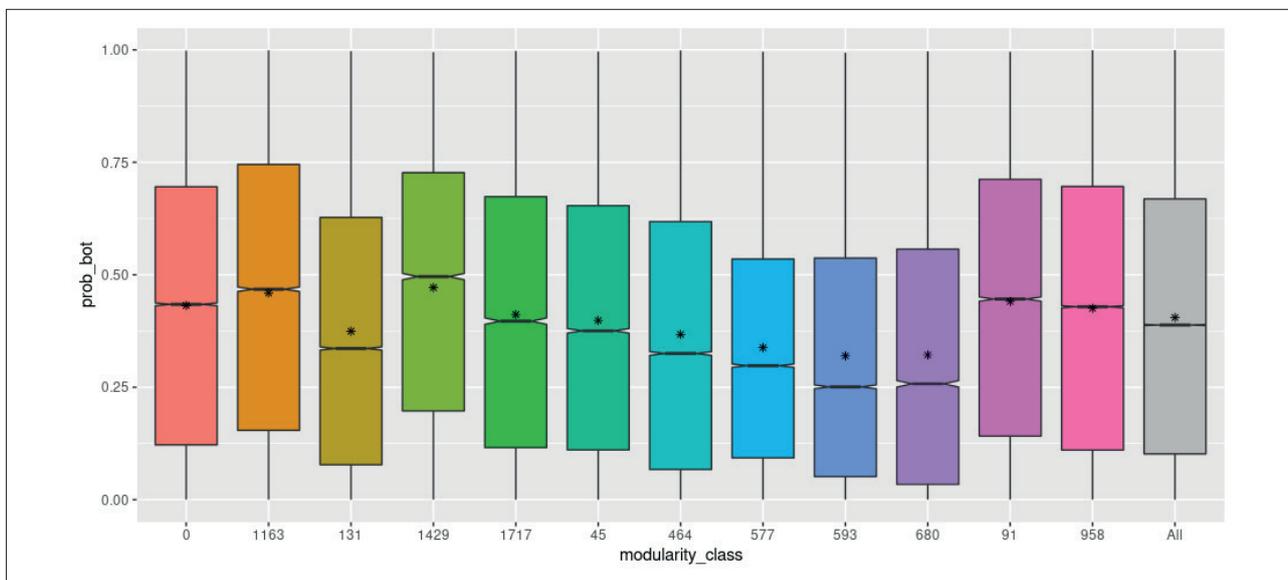


Figura 5. Distribución de probabilidad de presencia bots en cada clúster

de serlo. Esto demostraría la artificialidad en muchas de las cuentas que participan, especialmente en algunos de los grupos trans y de ultraderecha, para lo cual hay dos posibles interpretaciones:

- presencia de usuarios que van por libre y tienen un discurso intenso muy polarizado, y/o
- el uso de cuentas tipo troll diseñadas para crear polémicas, establecer diversos temas en la *agenda setting* o hacer campañas de desinformación mediante enfrentamientos entre grupos de “falsa bandera” (Bot Ruso, 2022), es decir, cuentas que se hacen pasar por grupos contrarios.

En cambio, el odio presenta, con respecto a las emociones básicas y sentimientos o polaridad vertidas, una ligera correspondencia estadística apreciable con la tristeza y, sobre todo, con el asco o aversión. De esta manera las correlaciones de Pearson son -0,273 con la polaridad o vector; 0,068 con la confianza; 0,080 con la alegría; 0,086 con la anticipación o pensamiento racional; 0,097 con la sorpresa; 0,192 con el miedo; 0,244 con enfado; 0,284 con tristeza; y 0,405 con el asco o aversión. Por tanto, puede comprobarse que las emociones, en este caso, que más llevan al odio, provienen del asco, principalmente, y de la tristeza.

El establecimiento de un árbol de regresión predictivo sobre el caso con respecto a todos los mensajes que tengan odio detectado, expuesto en la figura 6 (en las cajas aparece arriba el nivel medio de intensidad de odio y debajo el número de mensajes y lo que supone sobre el total), viene a corroborar la fuente principal del odio a través del asco o aversión y el miedo, que alcanza sus mayores valores cuando el sentimiento de asco es superior a 2 y el miedo es igual o superior a 2. El peso principal del odio proviene del asco en un 50%, un 13% a la polaridad negativa, y un 12% al miedo, cayendo el resto de emociones por debajo del 10% en importancia. Por lo tanto, una mezcla de niveles medio-elevado de asco y miedo son los que generan un mayor contenido de odio. Los mensajes de odio que entremezclan tristeza tienen valores inferiores en su intensidad. La correlación entre el valor predicho y el real es de 0,498 y el error absoluto medio (MAE) de 1,24, con lo que se puede estimar una aceptablemente buena predicción del algoritmo.

Según la rueda de las emociones básicas de **Plutchik (1980)**, el odio es el incremento de intensidad de la emoción básica de la aversión o asco, pero también puede lograrse a partir de “diadas” (combinaciones) de otras emociones básicas, como pueden ser tristeza e ira o mezclas de miedo con ira o tristeza. Las combinaciones de emociones, cuanto más alejadas además en su rueda (el asco y el miedo están alejadas en tres pasos, o diada terciaria, según su autor), son más difíciles de persistir en el tiempo. El odio, además, activa un comportamiento secundario, el rechazo. Por eso el odio generado a partir del asco o aversión (*disgust*), sería el más intenso y perdurable en el tiempo que el generado desde combinación de otras emociones. En el caso analizado, el odio se basa primariamente de asco o aversión, acompañado de niveles altos de miedo.

Sobre el último periodo de recogida, del 25/01/2021 al 01/03/2021, se obtuvieron 83.986 mensajes, pero también se realizó una recogida paralela sobre determinados países especificados en el apartado de metodología a través de geolocalización. Se pudo obtener así la ubicación de 33.795 mensajes (40,24% del total), pudiéndose determinar que, aunque se trataba de una conversación en español, la distribución fue mundial. Así se observó que en ese periodo, sobre los mensajes geolocalizados, el 40,21% procedía de España, el 15,86% de Estados Unidos, 8,26% de México, 6,76% de Colombia, 6,11% Brasil, 5,66% Venezuela, 4,27% Filipinas, 3,83% Argentina, 2,99% Rusia (zona Moscú), 2,93% Turquía, 0,65% República Dominicana, 0,43% Inglaterra (zona Londres), 0,35% Bolivia, 0,31% Italia (zona Roma), 0,24% Tailandia y 0,23% Albania. El resto de países contienen porcentajes inferiores. Se detectan por tanto varios países de habla no hispana con porcentajes no despreciables en la conversación, en los que estudios recientes caracterizan industrias de propaganda y desinformación dirigidas, entre otros, hacia España desde otros países (Arce-García; Said-Hung; Motta-reale-Calvanese, 2022).

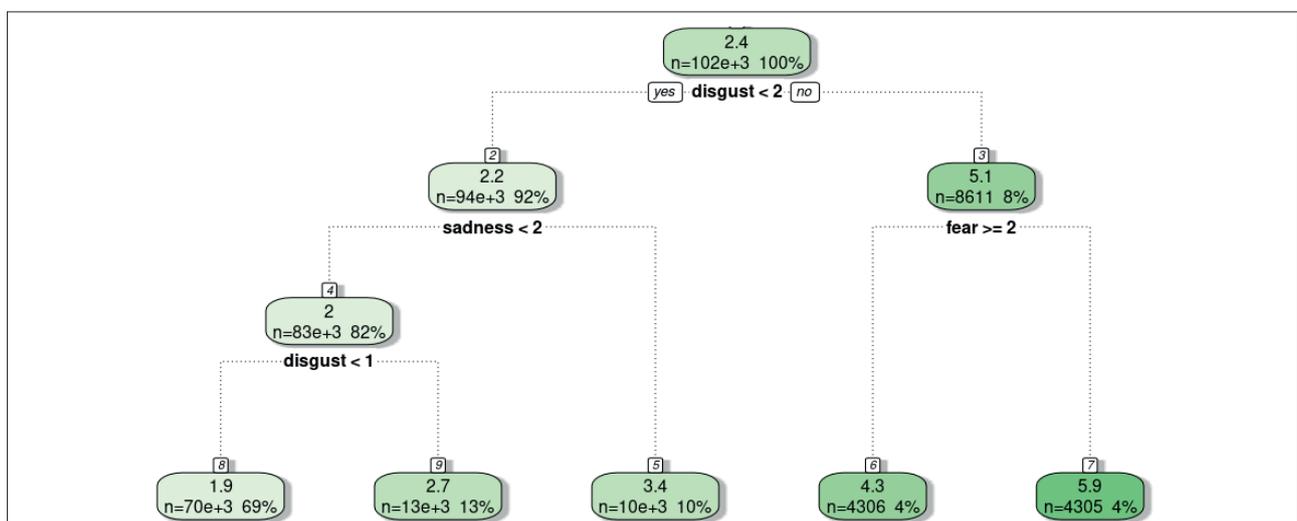


Figura 6. Árbol de regresión sobre las emociones con respecto a los mensajes de odio

5. Conclusiones

El análisis durante un año de la red social *Twitter* alrededor de la conversación entre grupos, principalmente feministas y trans, mostró un gran debate con un alto componente de odio. Aunque se siga la formación clara de grupos mediante el análisis de clusters, ante un tema tan polarizado debería verse claramente una separación de filtros burbuja (**Pariser**, 2011) bien marcados, pero en muchas ocasiones existe una gran indefinición entre diversos grupos y no se aprecian marcadas diferencias y separaciones, como se expone en la figura 1, al contrario que en otros casos muy polarizados. De todas formas, los altos niveles de polaridad y de odio demuestran la primera hipótesis planteada en este trabajo entre los grupos trans y feministas.

La red detectada, además de estar muy interrelacionada, no tiene claros referentes ni grandes influencers. Únicamente se exponen unos pocos grupos o clusters alejados del resto, uno de temática trans y dos de tendencia de ultraderecha, aunque aun así presentan extensas relaciones e interconexiones con los demás clusters. Existe una gran cantidad de discurso de odio en todos los grupos detectados, aunque los de nivel extremo son mucho más escasos, con un amplio ataque hacia el feminismo al que se acusa de radical y excluyente, que se asocia al término *terf* (acrónimo de *trans exclusionary radical feminists*), frente a un grupo trans que sufre (necesita antidepresivos), coincidiendo con el relato de autovictimización marcado por **Carratalá** (2021, p. 90). Se encuentra un perfil muy claro de asociación de palabras a lo largo de un año completo, por lo que se trata de ideas y discursos mantenidos en el tiempo, así como palabras creadas tales como “feminismoradical” o “pedoliteral”. Igualmente se intenta equiparar feminismo al término excluyente *terf*, palabra despectiva que, aunque minoritaria, tiene amplia difusión en redes (**Ayuso**, 2020, p. 219).

Alrededor de esta estructura de red tan poco marcada existen otros datos que podrían dar pie a una artificialidad en el debate: muchas cuentas meses después han sido eliminadas o suspendidas por infringir las normas de la red social. Igualmente se detecta un comportamiento tipo bot en muchos de los grupos, aunque no se encuentra relación estadística significativa entre mayor probabilidad de serlo y mayor difusión de odio o sentimientos negativos, pero sí se aprecia mayor presencia bot entre las cuentas que hacen reenvíos (RTs) de mensajes de menor probabilidad, pudiendo ser un típico comportamiento de tipo troll acompañado de bots de apoyo en todos los sentidos con posibles campañas de falsa bandera (simulación de pertenecer a una ideología contraria). El que los bots se dediquen especialmente a redifundir odio más que a establecer mensajes propios, encajaría con los estudios de **Keller** (2019), que marcan estrategias organizadas de difusión de campañas de desinformación o ataques a determinados grupos típicos de la técnica denominada *astroturfing*.

La presencia de grupos de ultraderecha y la cantidad de cuentas supuestamente trans suspendidas o borradas alrededor de los debates incidirían en esta idea de contaminar el debate del resto de clusters. Asimismo, la existencia de porcentajes no desdeñables desde diversos países, algunos conocidos por sus industrias de desinformación, de no habla hispana ahonda la sospecha. Por lo tanto, todo incide en la idea de un debate que está siendo alimentado y fomentado por grupos que presentan estrategias ya estudiadas y citadas.

La determinación de cuáles son las emociones que marcan al odio en este caso provienen principalmente del asco o aversión, unido principalmente al miedo. Tal como expone **Wylie** (2020) provocar odio, sobre todo basado en el asco o aversión, es el componente principal para introducir un pensamiento no racional que se mantenga en el tiempo frente al generado únicamente por el miedo.

Todo lo expuesto revela un discurso internacional, mantenido en el tiempo, que no sería aislado ni espontáneo (en este estudio se abarca todo el ámbito hispanohablante en español durante un año) tal como igualmente exponen **Villar-Aguilés** y **Pecourt-García** (2021, p. 43), que intenta introducir una serie de asociaciones de ideas a través de cuentas de escasa relevancia en la red aprovechando la teoría sociológica de **Granovetter** (1973). Con ella se estaría viendo que no hace falta incidir sobre todos los usuarios, sino solo sobre unos pocos que luego trasladarán sus nuevas ideas al resto del colectivo. Que un 20% de los mensajes de los 12 principales grupos emitan mensajes de odio intenso es un porcentaje importante como para acabar influenciando a toda la red. Este fenómeno ha sido contemplado en otras investigaciones académicas (**Alabao**, 2021; **Gutiérrez-Almazor**; **Pando-Canteli**; **Congosto**, 2020; **Etura-Hernández**; **Gutiérrez-Sanz**; **Martín-Jiménez**, 2017), pero igualmente podría encajar dentro de la “guerra cultural” hacia, entre otros, el colectivo feminista tal como expuso **Davis** (2019), confirmando la segunda hipótesis planteada.

Se podría hablar, por tanto, de la presencia de indicios bastante intensos de una estructura o de movimientos no naturales en las redes sociales en ataque al feminismo, encaminadas a provocar enfrentamientos entre grupos, justo cuando autores trans reconocen que cuando el feminismo está más abierto, el ambiente es más hostil (**Missé**, 2019, p. 155). Se trata de un discurso continuo donde se introduce odio (asociado principalmente al asco, que sería la emoción básica que hace que perviva más en el tiempo y en intensidad), polarización y asociaciones de ideas que se inserten en la sociedad, aunque sean falsas o minoritarias. Que el odio se genere básicamente a partir de la emoción básica propia, el asco, y generar por tanto rechazo, busca que dicho odio perviva y no se elimine claramente, ya que desplaza al pensamiento racional. Sirve también esta investigación para mostrar el esquema de posibles influencias en la sociedad a través de las redes sociales para provocar enfrentamientos y difusión de odio entre grupos.

Con todos estos datos y confirmación de las dos hipótesis planteadas, cabe entonces destacar que existe una fuerte polarización que provoca ataques entre grupos trans y feministas, en los que se aprecian algunos de naturaleza muy sospechosa (uso de cuentas bot, cuentas suspendidas o eliminadas, relaciones con grupos ajenos y procedencias no desdeñables desde países de no habla hispana), con metodologías apreciadas en otras campañas organizadas. El uso de

estas técnicas vendría a definir un ambiente en el que se persigue que se generen sentimientos de odio que perduren en el tiempo ya que el discurso repetitivo y la provocación de emociones básicas fuertes generan la base principal del odio social y que actualmente se vería con las nuevas corrientes de estudio en redes sociales del *gaslighting* (Shane; Willaert; Tuters, 2022). Según esta teoría la generación de emociones básicas fuertes en la comunicación origina en los receptores manipulación y les plantea dudas sobre la percepción de su realidad.

Se intenta equiparar feminismo al término excluyente *terf* (acrónimo de *trans exclusionary radical feminists*), palabra despectiva, que aunque minoritaria, tiene amplia difusión en redes

La posible utilización de la industria de la propaganda y de la desinformación, sustentada en perfiles y procedencias conocidas como sospechosas, estaría haciendo su labor de provocar desafección y desconfianza hacia los grupos trans y feminista, exponiéndolos como enfrentados y difusores de odio, en cuanto serían realmente las víctimas del ataque. El uso de bots principalmente para la redifusión y otros comportamientos ahondan en la sospecha de la presencia este tipo de industria que es muy difícil de detectar.

La generación de odio, planteamiento de dudas y de alta polarización hace que la sociedad se plantee a ambos grupos como conflictivos, que generan odio y desprecio, situando especialmente al grupo feminista como intransigente y exclusivo. Todo ello vendría entonces a plantear una posible campaña de largo recorrido, donde se sitúa al feminismo como un elemento poco atractivo ante un desplazamiento de la ventana de Overton, o sea, rango de ideas que la opinión pública asume como aceptables (Williams, 2021).

Las limitaciones de este trabajo vienen principalmente de la forma de detección de odio, ya que se trata de una metodología muy nueva en español que se basa en uno de los primeros léxicos existentes en este idioma validado como es *Hurtlex*. Tiene como mayor inconveniente la falta de calificación de intensidad entre el léxico por separado, pero se ha utilizado un algoritmo que sí tiene en cuenta la conjunción de varias palabras juntas que potencien o disminuyan su nivel, además de su buen éxito en castellano según los autores del léxico empleado. Existen varios proyectos en España para hacer muestras de entrenamiento por métodos de redes neuronales, pero ninguno en este momento finalizado. Igualmente, los léxicos emocionales empleados son casi inexistentes en lenguas latinas que, aunque presentan fiabilidades menores que en inglés (superiores al 80%), rondan al menos el 70% (Mohammad, 2016). La muestra se considera fiable al abarcar un año completo de recogida.

La presencia de textos que puedan tener doble sentido o irónico es igualmente un problema para estas metodologías, pero dado el volumen de mensajes analizados se considera que el global de resultados no se ve alterado o lo hace mínimamente. Como líneas futuras se propone estudiar más a fondo el esquema de ataque y confrontación provocado en las redes sociales, llevado no solo a este tema, sino también a otros de la sociedad donde se podría estar intentando cambiar a medio plazo el parecer de la opinión pública.

6. Referencias

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/mti510063>

- 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 filologia: 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://rdr.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 monographs*.
<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; Scharrow, 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/10.1177/205395172111033566>

- 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 "genderismo 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>
- Neubauer, 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.es/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, Alicia; 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