



Discurso de odio en redes: Intersección género, raza y clase. Estado del arte

Hate Speech on Social Media: The Intersection of Gender, Race, and Class. State-of-the-art

Assoc. Prof. Karen Hasleidy Machado-Mena, Universidad Tecnológica de Pereira (Colombia)
(khmm@utp.edu.co) (<https://orcid.org/0000-0003-0409-6647>)

RESUMEN

Esta revisión sistemática PRISMA de 49 estudios analiza la producción científica reciente (2017-2025), para dar cuenta del abordaje del carácter interseccional del discurso de odio en redes sociales, donde género, raza y clase socioeconómica se entrelazan produciendo experiencias de discriminación digital. El objetivo fue identificar los marcos teórico-metodológicos predominantes y evaluar su capacidad para capturar esta complejidad. Empleamos una metodología mixta que integró análisis bibliométrico (con Bibliometrix y VOSviewer) y análisis de contenido crítico de la literatura indexada en Scopus siguiendo los criterios PRISMA. El campo evidencia un crecimiento exponencial desde 2021, con una producción concentrada en España (44,9% del corpus) y un núcleo central de revistas especializadas. Predominan los métodos mixtos, combinando técnicas computacionales (PLN, aprendizaje automático) con análisis crítico del discurso. Si bien identificamos marcos psicosociales (Teoría de la amenaza integrada) y críticos (Interseccionalidad, Teoría crítica de la raza), el análisis revela ausencia de perspectivas como la gubernamentalidad de plataformas y las economías afectivas. La dimensión de clase social permanece subestudiada en comparación con el género y la raza. Concluimos que la investigación debe adoptar marcos más robustos y metodologías integradas que capturen simultáneamente la escala computacional, la profundidad estructural de las opresiones interseccionales, la dimensión afectiva y la gobernanza algorítmica, para desarrollar intervenciones que aborden las raíces del odio interseccional en los ecosistemas digitales contemporáneos.

ABSTRACT

This PRISMA systematic review of 49 studies analyzes recent scientific production (2017-2025) to account for approaches to the intersectional nature of hate speech on social media, where gender, race, and socioeconomic class intertwine, producing experiences of digital discrimination. The objective was to identify the predominant theoretical and methodological frameworks and evaluate their capacity to capture this complexity. We employed a mixed-methods approach that integrated bibliometric analysis (using Bibliometrix and VOSviewer) and critical content analysis of the literature indexed in Scopus, following the PRISMA criteria. The field shows exponential growth since 2021, with production concentrated in Spain (44.9% of the corpus) and a core of specialized journals. Mixed-methods approaches predominate, combining computational techniques (NLP, machine learning) with critical discourse analysis. While we identified psychosocial (Integrated Threat Theory) and critical (Intersectionality, Critical Race Theory) frameworks, the analysis reveals a lack of perspectives such as platform governmentality and affective economies. The social class dimension remains understudied compared to gender and race. We conclude that research must adopt more robust frameworks and integrated methodologies that simultaneously capture computational scale, the structural depth of intersectional oppressions, the affective dimension, and algorithmic governance to develop interventions that address the roots of intersectional hate in contemporary digital ecosystems.

PALABRAS CLAVE | KEYWORDS

Discurso de odio, redes sociales, interseccionalidad, estado del arte, interdisciplinariedad, producción científica.
Hate Speech, Social Media, Intersectionality, State of the Art, Interdisciplinarity, Scientific Production.

1. Introducción

Los discursos de odio en redes sociales constituyen una manifestación estructural de violencia simbólica que perpetúa desigualdades en las sociedades contemporáneas (Butler, 2021). En un contexto donde el 63.9% de la población mundial son usuarios activos de redes sociales (We Are Social, 2025), estas plataformas han dejado de ser espacios de socialización para convertirse en escenarios en los que se libran batallas por el reconocimiento y la representación. No obstante, la acción moderadora de las plataformas tecnológicas resulta insuficiente, como lo demuestran las 35.1 millones de piezas de contenido de odio eliminadas por Facebook en 2022, que no logran contener la proliferación de estos discursos (Ibrahimova, 2023).

Esta investigación surge de identificar una paradoja fundamental en el estudio del discurso de odio digital: si bien existe un crecimiento exponencial en publicaciones científicas desde 2017 (Tontodimamma et al., 2021), persiste una fragmentación disciplinar que restringe la comprensión de su complejidad interseccional. Estudios como el de Kwarteng et al. (2022) demuestran empíricamente que los clasificadores automáticos fallan sistemáticamente en detectar misoginia racializada porque no capturan intersecciones entre categorías de opresión; y Breazu (2023) evidencia que el racismo ciego al color en YouTube enmascara formas encubiertas de odio que escapan a los métodos de detección automatizada.

Revisiones bibliométricas previas (Izquierdo Montero et al., 2022; Ramírez-García et al., 2022) han señalado la escasez de investigaciones que integren perspectivas críticas desde comunicación y educación, pero ninguna ha proporcionado un análisis sistemático de los modos en que los marcos teóricos y metodológicos existentes abordan (u omiten) las interconexiones entre género, raza y clase social en el discurso de odio digital. Tal brecha es particularmente relevante en contextos como el colombiano, en el cual la ausencia de un marco legal específico, evidenciado por el archivado en 2015 del proyecto de ley que prohibía la apología al odio y la declaración de la Ley 1482 de 2011 por la Corte Constitucional como EXEQUIBLE (porque nació con vicios democráticos y procedimentales – además de ser concebida en una era pre-redes sociales masivas y pre-inteligencia artificial generativa), se combina con la identificación por parte del Ministerio TIC (2025) de que las plataformas digitales reproducen y amplifican desigualdades estructurales.

Ahora bien, esta investigación se fundamenta en la integración de cuatro dimensiones teóricas: 1) la conceptualización del discurso de odio como tecnología productora de realidad (Butler, 2021); una tecnología que institucionaliza prejuicios a través de repetidas performatividades discursivas que construyen al otro como amenaza y, por tanto, justifican su supresión simbólica o física; 2) la teoría interseccional (Collins, 2000; Crenshaw, 1989) que obliga a trascender visiones unidimensionales del odio, para entender que los discursos de odio no impactan de igual manera a todas las personas dentro de un grupo social: el género, la raza, la clase y la sexualidad se entrelazan para producir experiencias específicas de vulnerabilidad; 3) la economía afectiva de plataformas que analiza la monetización de la atención generada por contenido odioso (Martino, 2025); y 4) la gubernamentalidad algorítmica (Rouvroy y Berns, 2018) para explicar que los algoritmos moldean comportamientos, normas y valores a través del diseño de interfaces, términos de servicio y sistemas de moderación de contenido.

La singularidad de este estudio radica en aplicar un enfoque interseccional crítico que combina análisis bibliométrico cuantitativo con evaluación cualitativa de la capacidad explicativa de los marcos teóricos predominantes. De esta forma, cuando revisiones como la de Tontodimamma et al. (2021) identifican temas generales mediante modelado LDA, nosotros desagregamos la producción científica para evidenciar vacíos teóricos específicos y proponer perspectivas emergentes. En este marco, el artículo se propone realizar una revisión sistemática de la producción académica reciente (2017-2025) sobre discursos de odio en redes sociales con tres objetivos específicos: mapear el abordaje de la intersección entre género, raza y clase; identificar los marcos teóricos y metodológicos predominantes; y, señalar vacíos epistémicos y proponer perspectivas integradoras.

La relevancia de esta investigación trasciende el ámbito académico, ya que, como demostró el experimento de Hangartner et al. (2021), la efectividad del contradiscurso depende de una comprensión matizada de las intersecciones que aquí investigamos. En un escenario donde organismos internacionales como la UNESCO (2021) han desarrollado estrategias para contrarrestar el discurso de odio, explicar sus manifestaciones interseccionales resulta importante para diseñar intervenciones creativas y contextualmente relevantes.

2. Material y método

Aplicamos criterios PRISMA 2020 (Page et al., 2021) para seleccionar un corpus de alta calidad (49 estudios), combinado con la triangulación de enfoques cuantitativos y cualitativos para evaluar críticamente

la capacidad de los marcos existentes en dicho corpus para capturar la complejidad interseccional del discurso de odio en redes sociales. Esta aproximación dual nos permite describir tendencias de publicación e identificar vacíos teóricos y epistémicos fundamentales, particularmente la escasa representación de perspectivas como la gubernamentalidad de plataformas y las economías afectivas de corte posestructuralistas, y proponer una línea de investigación que nos ayude a renovar el tratamiento de la problemática que prevalece en la literatura actual.

La Figura 1 presenta el diagrama de flujo PRISMA que detalla el proceso completo de identificación, cribado, elegibilidad e inclusión de estudios.

Figura 1: Indicadores metodológicos del estudio.		
Indicador	Descripción	Justificación
Criterio de elegibilidad	1. Temporalidad: 9 años (2017-2025) 2. Tipología documental: toda la codificada en la base de datos	1. Primer año con publicaciones que abordan el análisis del tema, consistente con los resultados de otros estudios. 2a. Responder a los posibles países e instituciones involucradas. 2b. Distintos patrones de publicación y citación de investigadores según país e institución.
Fuentes de información	1. Scopus	Utilizado por otros estudios de naturaleza similar Mayor cobertura de las revistas Mayor volumen total de citación
Estrategia de búsqueda	TITLE-ABS-KEY ("discurso de odio" OR "hate speech" AND "redes sociales" OR "social media" AND "género" OR "gender" OR "etnia" OR "ethnicity" OR "raza" OR "race" OR "clase social" OR "social class") AND PUBYEAR > 2017 AND PUBYEAR < 2025.	Búsqueda genérica por términos focalizados de interés
Proceso de selección de los estudios	1. Fecha de consulta: Del 6/10/2024 a 12/08/2025 2. Campo de búsqueda: "Title" "Abstract" y "KeyWords" 3. Selección: existencia o no del trabajo indexado en la base de datos	1. Campos de contenido textual de artículos comúnmente utilizados en la recuperación de información bibliográfica.
Proceso de extracción de datos	1. Exportación: formato csv para su tratamiento en base de datos 2. Tratamiento posterior en Microsoft Excel 3. Uso del Software Bibliometrix – biblioshiny 5.0 4. Uso de VOSViewer Software 1.6.20	1. Detección y eliminación de registros duplicados. 2. Volcado de datos de producción y categoría para generar gráficos 3. Generación de mapas de palabras clave o red
Lista de los datos	Estructuración de los datos: N trabajos, N citas, % trabajos citados al menos una vez, ratio de citas por trabajo, impacto normalizado por categoría respecto al mundo, y % trabajos realizados en colaboración internacional. N=49	Ramírez-García et al. (2022) La búsqueda inicial identificó 317 registros. Tras la eliminación de duplicados y la aplicación de criterios de inclusión mediante cribado de títulos, resúmenes y palabras clave por dos investigadores independientes ($\kappa = 0.92$), obtuvimos un corpus final de 49 artículos.

En su componente cuantitativo, realizamos un análisis bibliométrico de la producción científica, extrayendo y analizando variables clave para trazar un panorama estructural del campo. Esto incluyó: 1) el mapeo de la producción científica, evaluando su evolución temporal, la productividad por países y regiones, y la identificación de los autores más productivos; 2) el estudio de las redes de colaboración a través del análisis de coautoría entre autores, instituciones y países; 3) la medición del impacto científico, considerando las citas por documento y el cálculo del índice h; y 4) un análisis conceptual basado en la co-ocurrencia de palabras clave (estableciendo un umbral de ≥ 2 repeticiones) y la posterior agrupación en clusters temáticos. En todos estos análisis utilizamos el software especializado Bibliometrix/Biblioshiny 5.0 (Aria y Cuccurullo, 2017) y VOSviewer 1.6.20 (van Eck y Waltman, 2010).

Este enfoque cuantitativo lo complementamos y enriquecimos con un análisis cualitativo temático y de contenido, centrado en una revisión crítica de la literatura, desglosada en dos ejes: un análisis temático para la codificación inductiva de los marcos teóricos y metodologías predominantes, y un análisis de

contenido para evaluar perspectivas teóricas, identificar vacíos temáticos y discernir tendencias discursivas. Para ello, desarrollamos un sistema de codificación *ad hoc* que permitió un escrutinio pormenorizado de cada estudio, focalizándonos en: la evaluación de los marcos teóricos y su aplicación interseccional, para valorar críticamente su capacidad explicativa; la solidez metodológica de cada investigación, para juzgar la coherencia entre sus objetivos, métodos y la evidencia empírica presentada; la extracción de hallazgos interseccionales específicos, para recoger la evidencia sobre las interacciones entre las categorías de análisis; y la identificación de brechas, para analizar tanto las limitaciones reportadas como las omisiones temáticas.

El proceso cualitativo fue riguroso e iterativo, e implicó un análisis en profundidad de cada uno de los 49 estudios, una codificación iterativa con sucesivas rondas de refinamiento temático, una triangulación aumentada entre los hallazgos bibliométricos y cualitativos, y una evaluación crítica de la contribución específica de cada trabajo al campo de estudio interseccional.

La decisión de trabajar con un corpus refinado de 49 estudios nos permitió implementar garantías de rigor metodológico de alto nivel, traducida en una profundidad analítica significativa en el examen de cada estudio con un mayor detalle. Aseguramos así una mayor validez de contenido, pues contamos con un corpus coherente y temáticamente focalizado, con lo cual la fiabilidad del proceso se vio aumentada, logrando un coeficiente de confiabilidad intercodificador excelente ($\kappa = 0.92$). Asimismo, mantuvimos una transparencia metodológica, con una trazabilidad completa del proceso de selección.

3. Análisis de resultados

3.1. Análisis cuantitativo: producción científica, redes de colaboración, impacto y temáticas relevantes

El análisis diacrónico de la producción científica revela una tendencia de crecimiento exponencial a partir del año 2021, que culmina en un pico máximo en 2023 con 17 publicaciones (Figura 2); el período inicial (2017-2020) muestra una producción embrionaria, con solo 3 artículos publicados en cuatro años. La proyección para 2025, con 9 artículos ya identificados en la mitad del año, sugiere una consolidación del campo como área de investigación emergente y de relevancia creciente.



El número de afiliaciones institucionales evidencia predominio de la producción científica española, que representa el 44,9% del corpus analizado (22 de 49 artículos). (Figura 3 - Tabla 1)

Tabla 1: Documento por afiliación.

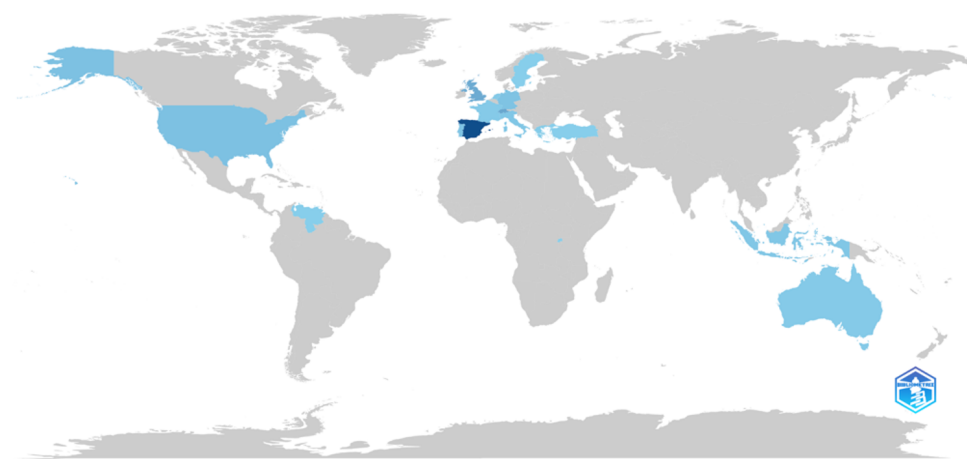
País	No.
España	22
Reino Unido	5
Portugal	3
Australia	2
Alemania	2
EE. UU	2
Australia; Suecia	1
España; Grecia; Italia	1
España; Venezuela	1

Tabla 1: Documento por afiliación.

País	No.
España; Reino Unido	1
Francia	1
Indonesia	1
Italia	1
Italia; España	1
Países Bajos	1
Países Bajos; Ruanda	1
Suiza	1
Suiza; Alemania	1
Turquía	1

Figura 3: Contribución de países según afiliación de autores.

Producción científica de los países



La producción científica muestra un predominio de colaboraciones nacionales (Figura 3 - Tabla 1) con el 83.7% de los artículos (41 de 49) resultantes de cooperaciones dentro de un mismo país. De estos, 24 estudios proceden de una única institución y 17 involucran entre dos y tres instituciones del mismo territorio. En contraste, la colaboración transfronteriza es menor, representando el 14.3% (7 artículos) entre dos países y un caso excepcional de cooperación trinacional. Este patrón evidencia dos modelos predominantes: consorcios nacionales multiinstitucionales, como los articulados entre universidades españolas, y colaboraciones internacionales estratégicas aún incipientes, ejemplificadas por alianzas como España-Grecia-Italia y España-Rusia.

En cuanto a la productividad individual, identificamos un grupo pequeño de investigadores productivos, con tres artículos (Arcila-Calderón, C. – Amores, J.) y dos artículos (Iranzo-Cabrera, M. – Arce-García, S. – Piñeiro-Otero, T. – Matamoras-Fernández, A. – Sánchez-Holgado, P.), tendencia que señala la vigencia de la Ley de Lotka, toda vez que una gran mayoría de autores (130) contribuyen con un solo trabajo, lo que evidencia una productividad altamente concentrada.

Respecto a las publicaciones más influyentes, consideramos las 10 que han recibido el mayor número de citas (Tabla 2).

El análisis de impacto revela una distribución desigual, con el trabajo de Matamoras-Fernández (2017) y Matamoras-Fernández y Farkas (2021) como seminal con 547 citaciones acumuladas. Este caso, junto con investigaciones como Mozafari et al. (2020) y Williams et al. (2020), confirma que la antigüedad temática es un factor importante para la acumulación de citas; no obstante, estudios recientes como Dickel y Evolvi (2023) y Arcila-Calderón et al. (2022b) muestran tasas de citación anual excepcionales, que sugieren un impacto inmediato que los posiciona como futuras referencias a pesar de su corta trayectoria.

Tabla 2: Publicaciones más citadas.

Título del documento	Autores	Citado por	Citas por año
Platformed racism: the mediation and circulation of an Australian race-based controversy on Twitter, Facebook and YouTube	Matamoros-Fernández (2017)	291	32,3333333
Racism, hate speech, and social media: a systematic review and critique	Matamoros-Fernández y Farkas (2021)	256	51,2
Hate speech detection and racial bias mitigation in social media based on BERT model	Mozafari et al. (2020)	189	31,5
Hate in the machine: anti-black and anti-muslim social media posts as predictors of offline racially and religiously aggravated crime	Williams et al. (2020)	159	26,5
Empathy-based counterspeech can reduce racist hate speech in a social media field experiment	Hangartner et al. (2021)	88	17,6
"Victims of feminism": exploring networked misogyny and #MeToo in the manosphere	Dickel y Evolvi (2023)	63	21
Hate speech and social acceptance of migrants in Europe: analysis of tweets with geolocation	Arcila-Calderón et al. (2022b)	46	11,5
Detecting ethnicity-targeted hate speech in Russian social media texts	Pronoza et al. (2021)	40	8
Step outside and say that: analysis of hate speech against women on twitter	Piñeiro-Otero y Martínez-Rolán (2021)	33	6,6
Xenophobia and hate speech towards refugees on social media: reinforcing causes, negative effects, defense and response mechanisms against that speech	Aldamen (2023)	25	8,33333333333333

Tabla 3: Impacto de la producción científica de los autores.

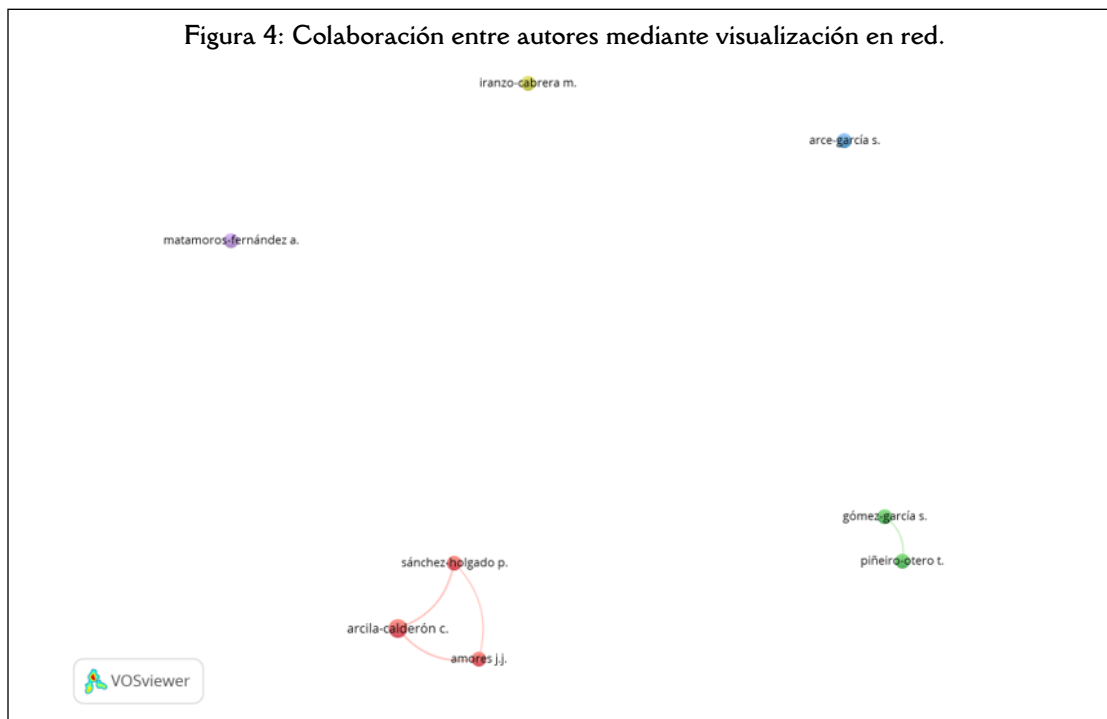
Autor	h_index	g_index	m_index	Total Citaciones	No. Publicaciones	Año inicial de publicación
Arce-García S	3	3	0,75	51	3	2022
Arcila-Calderón C	2	3	0,5	67	3	2022
Gómez-García S	2	2	0,66666667	11	2	2023
Iranzo-Cabrera M	2	2	1	6	2	2024
Matamoros-Fernández A	2	2	0,22222222	547	2	2017
Piñeiro-Otero T	2	2	0,4	42	2	2021
Sánchez-Holgado P	2	2	0,5	66	2	2022
Olmos-Alcaraz	1	1	0,33333333	6	1	2023
Aldamen Y	1	1	0,33333333	21	1	2023
Amores J	1	1	0,25	46	1	2022

El análisis del impacto científico (Tabla 3) revela un panorama diversificado: Matamoros-Fernández lidera, pero investigadores emergentes como Arcila-Calderón y Sánchez-Holgado muestran una influencia temprana, acumulando 67 y 66 citas respectivamente desde 2022; destaca particularmente Arce-García, quien alcanza el índice h más alto del grupo (3), lo que muestra una productividad de impacto sostenida. Estos datos señalan una comunidad científica en consolidación, en la que coexisten referentes consolidados con investigadores noveles que muestran un prometedor impacto inmediato.

Realizamos un análisis de coautoría para elaborar el mapa de colaboración (Figura 4), con los autores que tuvieran mínimo dos publicaciones en el corpus, sin considerar un número mínimo de citas. De los 136 autores, 8 cumplieron el parámetro.

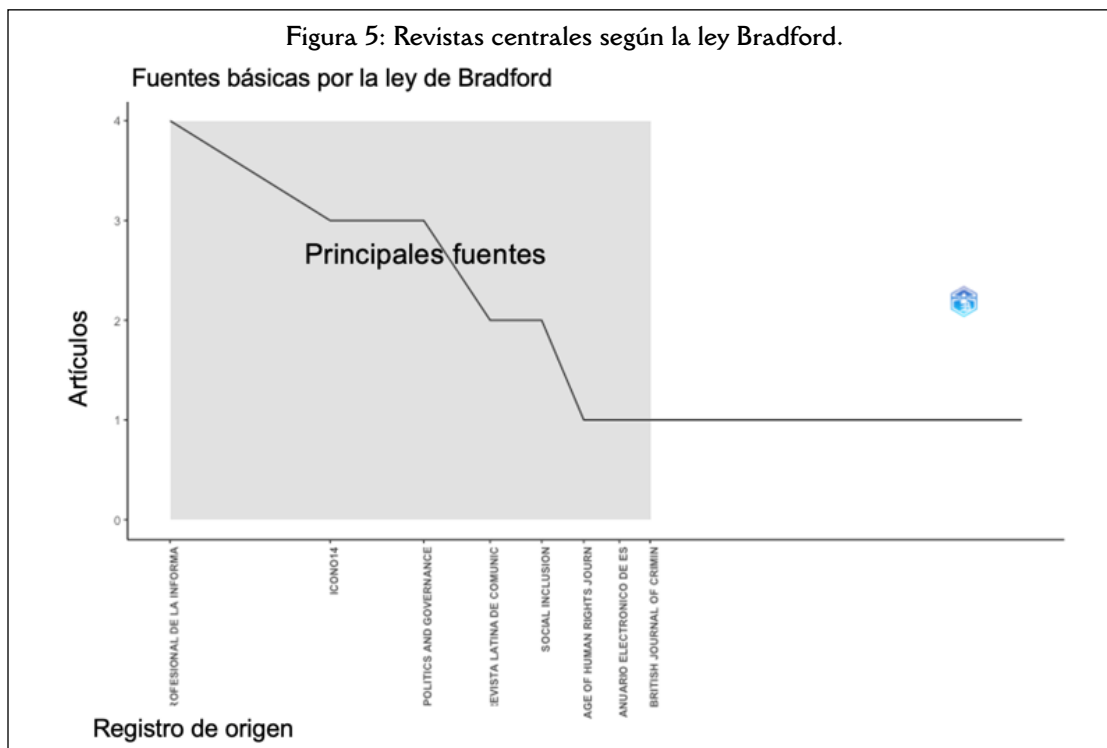
No identificamos la existencia de una red extensa de colaboración, de hecho, el mapa muestra dos grupos con mínimo dos autores colaboradores: el primer grupo, formado por Arcila-Calderón, Amores y Sánchez-Holgado, publica en áreas de comunicación, política y gobernanza, y estudia el discurso de odio hacia migrantes, refugiados y colectivo LGBTQ+, al igual que la polarización política; el segundo, formado por Gómez-García y Piñeiro-Otero, publica en áreas de información y comunicación e investiga el discurso de odio hacia las mujeres políticas en España.

Figura 4: Colaboración entre autores mediante visualización en red.



En cuanto a la distribución de la producción científica en las revistas académicas, evidenciamos el cumplimiento de la Ley de Bradford, manifestada en la concentración temática en un núcleo reducido de publicaciones (Figura 5 – Tabla 4).

Figura 5: Revistas centrales según la ley Bradford.



Revista	Frecuencia	Cúmulo de frecuencia	Zona
Profesional de la Informacion	4	4	Zone 1
Icono14	3	7	Zone 1
Politics and Governance	3	10	Zone 1
Revista Latina de Comunicacion Social	2	12	Zone 1
Social Inclusion	2	14	Zone 1
Age of Human Rights Journal	1	15	Zone 2
Anuario Electronico de Estudios en Comunicacion Social Disertaciones	1	16	Zone 2
British Journal of Criminology	1	17	Zone 2
Communication and Society	1	18	Zone 2
Comunicacao e Sociedade	1	19	Zone 2

Ahora bien, para identificar los temas más relevantes y sus interrelaciones en el campo de estudio, empleamos un mapa de red de co-ocurrencia de palabras clave. Del análisis de los 49 artículos extrajimos 177 palabras clave de autor. Para garantizar la claridad visual, establecimos una frecuencia mínima de dos ocurrencias por término. Previamente, desarrollamos un tesauro para unificar conceptos equivalentes (como “mujer” y “mujeres”) mediante lematización, asegurando consistencia terminológica. El procesamiento con VOSviewer permitió identificar ocho *clusters* temáticos, en los que los nodos representan palabras clave, su tamaño muestra la frecuencia, los colores indican la agrupación temática y los enlaces visualizan la fuerza de co-ocurrencia. Esta configuración gráfica señala los focos de investigación prioritarios y sus conexiones dentro del dominio analizado (Figura 6).

Figura 6: Palabras clave de coocurrencia en visualización de red.

Racimo 1. Discurso de odio, desinformación y COVID-19

Este grupo de investigaciones se centra en la intersección entre la desinformación, la pandemia de COVID-19 y el discurso de odio, particularmente en plataformas como Instagram y en espacios digitales como la manosfera, con un enfoque en el sexismo y la misoginia. Particularmente Hong et al. (2025) analizaron los efectos de la moderación de contenido sobre la desinformación y el odio anti-asiático en Instagram durante la pandemia; y Tarullo et al. (2024) estudiaron la narrativa #Feminazis en la misma plataforma, interesándose por la desinformación y el discurso de odio misógino como amplificadores de la retórica de la manosfera.

Racimo 2. Discurso de odio contra colectivos LGTBIQ+

Las investigaciones de este racimo analizan el discurso de odio dirigido hacia la comunidad LGTBIQ+ en España, en el contexto de debates políticos polarizados y las tensiones entre la libertad de expresión y el discurso de odio. Sánchez-Holgado et al. (2023) investigaron el odio y la polarización en Twitter en torno a la Ley Trans española, un tema que también es abordado por Arce-García y Menéndez-Menéndez (2023), quienes desarrollaron una metodología para rastrear el origen y las características del discurso de odio sobre la diversidad sexual y de género en la misma plataforma.

Racimo 3. Racismo y plataformas

Este racimo de estudios analiza las manifestaciones de racismo en plataformas de redes sociales específicas, por ejemplo, Breazu (2023) realizó un análisis crítico del racismo antigitano y la humillación hacia migrantes romaníes en YouTube.

Racimo 4. Feminismo, refugiados y Twitter

Las investigaciones agrupadas aquí utilizan predominantemente el análisis de contenido para estudiar en Twitter el discurso de odio vinculado a temas feministas y hacia colectivos de refugiados, como Sánchez-Meza et al. (2023) que investigaron la conversación digital en Twitter sobre la ministra española Irene Montero, preguntándose si las redes sociales se transforman en redes de odio.

Racimo 5. Misoginia y género en redes sociales

Este racimo se caracteriza por una focalización en la misoginia y la violencia de género como formas predominantes de discurso de odio en línea: en Asia, Li (2024) realizó un estudio de caso sobre el odio hacia las mujeres, vinculándolo con la desigualdad de género en entornos digitales; y en el contexto europeo, Iranzo-Cabrera et al. (2024) abordaron el discurso de odio contra mujeres políticas en redes sociales.

Racimo 6. Xenofobia y migración

Los estudios de este grupo aplican técnicas avanzadas de aprendizaje automático, como el aprendizaje profundo, para detectar y analizar el discurso de odio xenófobo dirigido a migrantes, por ejemplo, Arcila-Calderón et al. (2022a) desarrollaron y evaluaron un clasificador de discurso de odio racista y xenófobo utilizando aprendizaje superficial y profundo; y Pronoza et al. (2021) utilizaron el aprendizaje profundo para detectar el discurso de odio étnico en redes sociales rusas.

Racimo 7. Inmigración y polarización política

La investigación en este racimo utiliza el análisis de redes sociales para explicar el uso de la inmigración como un catalizador de la polarización política y el discurso de odio, es el caso de Kuric Kardelis et al. (2024), desde la polarización y el ciberodio en los medios sociales digitales en el contexto de los Menores Extranjeros No Acompañados (MENA); y Olmos Alcaraz (2023) desde el discurso político en Twitter durante la crisis migratoria de Ceuta en 2021.

Racimo 8. Delitos de odio

Este racimo, aunque más pequeño, se centra en la dimensión delictiva del discurso de odio conectando el odio online con problemáticas del mundo offline. Para el caso, Magano y D'Oliveira (2023) investigaron las expresiones de odio y racismo en las redes sociales como manifestaciones de antigitanismo y delitos de odio en Portugal; y Williams et al. (2020) fueron un paso más allá, utilizando publicaciones en redes sociales anti-negras y anti-musulmanas como predictores de delitos raciales y religiosos agravados en el mundo offline.

3.2. Análisis cualitativo: marcos teóricos y diseños investigativos en la producción científica

La investigación contemporánea sobre el discurso de odio en redes sociales desglosa un panorama en el cual la violencia digital se dirige de manera predominante hacia grupos demográficos específicos, ya que los estudios que analizamos evidencian una focalización en mujeres (especialmente aquellas en cargos políticos o de relevancia pública), inmigrantes, refugiados y comunidades racializadas como los asiáticos durante la pandemia, y el colectivo LGTBIQ+. Esta focalización se verifica en estudios como el de Boukemia et al. (2025), quienes cuantificaron que las diputadas reciben 47,03 respuestas insultantes adicionales por cada 10.000 seguidores en comparación con los diputados.

También encontramos que contextos políticos y geopolíticos decisivos operan como catalizadores del odio, y el estudio de Amores y Arcila-Calderón (2025) mostró que los debates legislativos sobre leyes migratorias en el sur de Europa actuaron como detonantes para picos de narrativas xenófobas en YouTube y X. Los autores señalan que las crisis migratorias, los debates legislativos polarizantes, los eventos deportivos o los ciclos de noticias sensacionalistas proveen el caldo de cultivo perfecto para la proliferación de narrativas de odio que, desde formatos como memes y bulos, amplifican su alcance, enmascaran su virulencia y se propagan estratégicamente a través de plataformas como Twitter, YouTube, Facebook e Instagram. Para explicar y combatir este problema, el campo académico está respondiendo con un cada vez más sofisticado equipamiento teórico y metodológico que sintetizamos a continuación.

Teorías explicativas sobre discurso de odio en redes sociales

Desde una perspectiva teórica, el análisis del corpus revela marcos explicativos que operan desde lo psicosocial hasta lo estructural (Tabla 5).

Tabla 5: Equipamientos teóricos.		
Teoría/Modelo	Conceptos	Autores
Teoría de la amenaza integrada del prejuicio	- Amenaza realista (económica/seguridad)	- Williams et al. (2020)
	- Amenaza simbólica (valores/identidad)	- Amores y Arcila-Calderón (2025)
Modelo motivacional de proceso dual	- Autoritarismo de Derecha (RWA) - Orientación a la Dominancia Social (SDO)	- Dellagioma et al. (2024)
Teoría del conflicto de grupo	- Competencia por recursos escasos	- Wu et al. (2025)
Interseccionalidad	- Sistemas de opresión superpuestos (raza, género, clase, etc.)	- Kwarteng et al. (2022)
Teoría crítica de la raza / racismo ciego al color	- Negación de la raza como mecanismo de enmascaramiento de desigualdades	- Breazu (2023) - Matamoros-Fernández y Farkas (2021)
Manosfera y reorganización digital de la misoginia	- Victimización masculina - Narrativas antifeministas	- Dickel y Evolvi (2023) - Tarullo et al. (2024) - Piñeiro-Otero y Martínez-Rolán (2021)
Trastorno de la información	- Desinformación: contenido falso creado deliberadamente para causar daño. - Misinformación: contenido falso, pero no creado deliberadamente para causar daño. - Malinformación: contenido basado en la realidad, pero usado para infligir daño.	- Lilleker y Pérez-Escobar (2023)
Teoría de la espiral del silencio / Teoría del chivo expiatorio / Efecto de fatiga por compasión	- Autocensura en climas de opinión hostiles - Culpar a los refugiados de robar oportunidades - Otros como amenazas potenciales al orden social existente - Publicación repetida de representaciones negativas que reduce la compasión - Sentimientos iniciales de simpatía en apatía, xenofobia y odio.	- Aldamen (2023)
Pragmática lingüística / actos de habla	- Odio encubierto - Illocución - Humor agresivo y estereotipos	- Parvareh (2023) - Silva y Carvalho (2023)
Modelo nacional-racista y la figura del monstruo	- Inevitable limitación e imposibilidad de la integración cultural - MENA - inocencia de la infancia con elementos de violencia y criminalidad	- Kuric-Kardelis et al. (2024)
Aporofobia / Pobre digno – pobre indigno	- Pobre que se esfuerza y merece - Victimización nativista - Sin ingresos, el pícaro, dependiente de subsidios	- Vélez-Coto et al. (2025) - Magano y D'Oliveira (2023)

Evidenciamos la coexistencia de marcos explicativos psicosociales y críticos:

Las teorías psicosociales demuestran empíricamente sus postulados: la Teoría de la Amenaza Integrada

se valida con el hallazgo de Williams et al. (2020) de una correlación significativa en el Reino Unido entre tuits que estigmatizan a inmigrantes musulmanes como amenaza y crímenes de odio offline. El Modelo de Proceso Dual es respaldado por Dellagiacoma et al. (2024), quienes con una muestra de 7.300 usuarios confirmaron que el Autoritarismo de Derecha predice apoyo a la censura, mientras que la Orientación a la Dominancia Social predice la expresión directa de odio. Asimismo, la Teoría del Conflicto de Grupo encuentra sustento en el análisis espacial de Wu et al. (2025), que evidenció en EE.UU. correlaciones entre variables socioeconómicas (pobreza, educación, diversidad) y el riesgo de odio anti-asiático.

Desde los marcos críticos, la Interseccionalidad es importante para comprender las limitaciones de las herramientas automatizadas, como demostró Kwarteng et al. (2022), cuando evidenció que los clasificadores fallan en detectar la misoginia racializada contra mujeres negras. El racismo velado se ejemplifica en el análisis de Breazu (2023) sobre creadores de contenido en YouTube, quienes mediante un lenguaje aparentemente neutral elogiaban para enmarcar a los romaníes como una amenaza. La reorganización digital de la misoginia en la Manosfera fue documentada por Dickel y Evolvi (2023), quienes identificaron en foros incels una narrativa de victimización masculina que redefine el #MeToo como una caza de brujas.

La propagación del odio se ve facilitada por la desinformación, como evidenciaron Lilleker y Pérez-Escolar (2023) al mostrar que actores de extrema derecha en Reino Unido y España utilizan datos falsos sobre criminalidad para legitimar narrativas de odio. A nivel discursivo, la pragmática lingüística revela mecanismos encubiertos: Parvaresh (2023) identificó que preguntas retóricas aparentemente inocuas invalidan quejas legítimas, y Silva y Carvalho (2023) constataron mediante grupos focales que el humor basado en estereotipos es vivido como violencia simbólica. Finalmente, la construcción del otro monstruoso e indigno, analizada por Kuric Kardelis et al. (2024) con la figura del MENA y por Magano y D'Oliveira (2023) con la dicotomía pobre digno/indigno, muestra que la aporofobia y la deshumanización operan para justificar moralmente el odio.

Abordaje metodológico del discurso de odio interseccional

Evidenciamos una evolución metodológica: los enfoques tradicionales se hibridan con técnicas computacionales avanzadas para abordar la escala y complejidad del problema en entornos digitales (Tabla 6)

Tabla 6: Equipamientos metodológicos.

Dimensión metodológica	Enfoques y técnicas principales	Autores
Diseño de investigación	Métodos mixtos (predominante): triangulación para capturar la multidimensionalidad del problema.	Iranzo-Cabrera et al. (2024); Aldamen (2023); Sánchez-Meza et al. (2023).
	Estudios cuantitativos a gran escala: aprendizaje automático (BERT, RoBERTa) para clasificar millones de datos y testear hipótesis.	Boukemia et al. (2025): diferencias de género. Williams et al. (2020): correlación odio online-crímenes offline. Arcila-Calderón et al. (2022a): validación de modelos.
	Estudios cualitativos en profundidad: análisis de mecanismos discursivos en corpus acotados.	Breazu (2023): análisis crítico del discurso. Dickel y Evolvi (2023): análisis temático en la manosfera.
	Enfoques innovadores: difracción (mezcla de métodos para abrazar la complejidad) y experimentos de campo (evidencia causal).	Giraud et al. (2025): enfoque difractivo. Hangartner et al. (2021): efectividad del contradiscurso.
Técnicas computacionales y PLN	APIs de plataformas: recolección de datos masivos de Twitter/X, Instagram, YouTube.	Boukemia et al. (2025); Breazu (2023).
	Clasificación automatizada: modelos transformer (BERT, RoBERTa) para detectar odio, toxicidad, emociones.	Hong et al. (2025); Pronoza et al. (2021).
	Análisis de sentimiento/emoción: medición de la carga afectiva (ira, miedo, asco) en el discurso.	Hong et al. (2025); Vélez-Coto et al. (2025).
	Modelado temático (LDA): identificación no supervisada de temas subyacentes en grandes volúmenes de texto.	Amores y Arcila-Calderón (2025).
	Análisis de redes sociales: mapeo de comunidades, influencers y estructura de la conversación.	Kuric Kardelis et al. (2024); Arce-García y Menéndez-Menéndez (2023).
Técnicas cualitativas	Análisis de contenido cualitativo: categorización de temas y estrategias discursivas.	Magano y D'Oliveira (2023); Lilleker y Pérez-Escolar (2023).
	Análisis crítico del discurso (ACD/MCDA): deconstrucción de ideologías, poder y construcción de la otredad en texto e imagen.	Ndahinda y Mugabe (2024).
	Análisis temático: identificación de patrones y temas recurrentes.	Dickel y Evolvi (2023).
	Grupos focales: indagación en percepciones y experiencias de colectivos objetivo.	Silva y Carvalho (2023).
	Netnografía: inmersión prolongada en comunidades online para entender cultura y normas internas.	Lopes Miranda (2023).

Tabla 6: Equipamientos metodológicos.

Dimensión metodológica	Enfoques y técnicas principales	Autores
Técnicas cuantitativas y experimentales	Encuestas: medición de percepciones y actitudes de la población general.	Sánchez-Holgado et al. (2023).
	Experimentos: medición de respuestas emocionales y evaluación causal de intervenciones.	Vélez-Coto et al. (2025); Hangartner et al. (2021).
	Análisis espacial (SIG/GWR): mapeo y predicción de la distribución geográfica del odio.	Wu et al. (2025).

En el ámbito de los diseños de investigación, la característica sobresaliente es el predominio de los métodos mixtos, con la triangulación como norma más que excepción. Investigaciones como las de Iranzo-Cabrera et al. (2024) y Aldamen (2023) ejemplifican esta tendencia, combinando aproximaciones cuantitativas y cualitativas para superar las limitaciones inherentes a cada enfoque por separado y capturar así la multidimensionalidad del problema. Paralelamente, los estudios cuantitativos computacionales a gran escala emplean aprendizaje automático supervisado mediante modelos avanzados como BERT, RoBERTa o XLM-RoBERTa que, entrenados con codificación humana previa, permiten analizar millones de tuits o comentarios para testear hipótesis complejas, como las diferencias de género en las reacciones (Boukemia et al., 2025) o la correlación espacial entre el odio online y los crímenes offline (Williams et al., 2020). La fiabilidad de estos abordajes se ve reforzada por prácticas como la validación externa de los modelos con nuevas muestras (Arcila-Calderón et al., 2022a).

Los estudios cualitativos puros mantienen su vigencia y se han especializado en el análisis en profundidad de los mecanismos discursivos en corpus acotados, desde técnicas como el análisis crítico del discurso (Breazu, 2023) y el análisis temático (Dickel y Evolvi, 2023). Estas técnicas son relevantes para deconstruir las narrativas, ideologías y las formas encubiertas de odio, como el humor o los elogios estereotipados, que los métodos automatizados suelen pasar por alto. Esta evolución metodológica también incluye abordajes innovadores, como el enfoque difractivo propuesto por Giraud et al. (2025), que en lugar de buscar simples correlaciones, centraliza el análisis en las tensiones entre los datos computacionales, cuantitativos y cualitativos, reconociendo así la complejidad irreductible del racismo digital. Asimismo, los experimentos de campo aleatorios (Hangartner et al., 2021) introducen evidencia causal sobre la efectividad de las estrategias de contradiscurso, lo que eleva el listón en la evaluación de intervenciones concretas.

En cuanto a la construcción y extracción de datos, el abanico técnico se ha expandido, por ejemplo, en el ámbito computacional y del Procesamiento del Lenguaje Natural (PLN), es habitual el uso de APIs de plataformas (siendo la API de Twitter/X la fuente principal), junto con herramientas específicas para YouTube para recolectar datos a escala. Junto a la recolección, el uso extensivo de modelos Transformer para la clasificación automatizada de odio, toxicidad o emociones, el análisis de sentimiento, el modelado temático no supervisado y el análisis de redes sociales para mapear comunidades y estructuras conversacionales, se han consolidado como pilares del análisis cuantitativo avanzado.

Por el lado cualitativo, las técnicas se han diversificado para captar la riqueza del problema, a partir del análisis de contenido cualitativo, el análisis crítico del discurso (ACD) y multimodal (MCDA), el análisis temático, los grupos focales con colectivos objetivo, la netnografía para la inmersión en comunidades online y el análisis retórico-argumentativo proveen las herramientas necesarias para comprender las dimensiones contextuales, ideológicas y experienciales del odio.

Finalmente, las técnicas cuantitativas tradicionales y experimentales, como las encuestas para medir percepciones generales, los experimentos controlados para medir respuestas emocionales o evaluar intervenciones, y el análisis estadístico espacial (SIG/GWR) para mapear la distribución geográfica del odio, completan un panorama metodológico robusto y multifacético, capaz de abordar el discurso de odio interseccional desde sus manifestaciones más vastas y superficiales hasta sus raíces más profundas y contextuales.

4. Conclusión y discusión

Nuestro estudio muestra un campo de investigación en plena consolidación, caracterizado por un crecimiento exponencial desde 2021 que corrobora y actualiza lo observado por Tontodimamma et al. (2021) en su análisis de treinta años. Sin embargo, su revisión identificaba temas generales mediante modelado LDA, pero nuestro enfoque interseccional crítico permite desagregar esa producción para evidenciar vacíos teóricos y epistémicos fundamentales que persisten a pesar del aumento volumétrico.

Ahora bien, la predominancia de la producción científica española (45%) y la concentración en redes de colaboración nacionales confirman y matizan los patrones identificados por Antona Jimeno et al. (2024) para España y LATAM. Coincidimos con su observación sobre la madurez del campo, pero nuestro análisis cualitativo revela que esta productividad no se traduce necesariamente en una sofisticación teórica. De otro lado, la evolución metodológica hacia métodos mixtos y técnicas computacionales avanzadas que documentamos es coherente con la explosión de investigación en aprendizaje automático documentada por Tontodimamma et al. (2021). Pero identificamos una paradoja importante: aunque el campo avanza técnicamente, reproduce el mismo desbalance disciplinar que Antona Jimeno et al. (2024) identificaban entre enfoques computacionales y sociales.

Estudios como los de Kwarteng et al. (2022) y Breazu (2023) proporcionan evidencia empírica contundente sobre las limitaciones que Izquierdo Montero et al. (2022) señalaban teóricamente: los clasificadores automáticos fallan sistemáticamente en detectar misoginia racializada porque no capturan intersecciones. Tal desconexión entre capacidades técnicas y comprensión teórica del problema explica por qué, como observaba Tontodimamma et al. (2021), el crecimiento en publicaciones sobre detección automática no se traduce necesariamente en efectividad para combatir el odio en su complejidad.

Nuestra investigación aporta entonces una triangulación metodológica que combina análisis bibliométrico cuantitativo con evaluación cualitativa crítica de marcos teóricos, lo cual nos permite plantear que los marcos teóricos existentes resultan insuficientes para capturar la complejidad interseccional del discurso de odio en redes sociales. Así, y basándonos en Martino (2025), Rouvroy y Berns (2018) y Ahmed (2015) proponemos desarrollar marcos teóricos que integren la economía afectiva de plataformas y la gubernamentalidad algorítmica. Lo anterior porque los estudios revisados muestran que el odio se propaga, pero no explican suficientemente por qué las plataformas tienen incentivos estructurales para una moderación laxa. La economía afectiva aporta este eslabón perdido, en tanto invita a considerar el análisis de la atención como mercancía del contenido odioso, pues generar alta interacción (enfado, indignación), es monetizable (Martino, 2025).

Los marcos psicosociales que identificamos (Teoría de la amenaza, RWA/SDO) explican disposiciones individuales, pero no los modos en que los algoritmos las activan, amplifican y moldean; y allí la gubernamentalidad algorítmica propuesta por Rouvroy y Berns (2018) permite estudiar esta ingeniería de la subjetividad pues desplaza la mirada de las cámaras de eco como agregados de individuos con ideas afines, para analizarlas como sistemas de gubernamentalidad que perfilan a usuarios con ciertas disposiciones y los agrupan por ideología, por perfiles demográficos e interseccionales complejos (hombres jóvenes, blancos, de zonas rurales con precariedad económica). Sistemas que además les suministran a los usuarios un flujo constante de contenido que construye una amenaza específica y personalizada (la feminazi, el inmigrante violento, el vándalo, etc.), con lo que optimizan la activación de sus sesgos; y proveen soluciones a través de la recomendación de contranarrativas, gurús digitales y comunidades (manosfera, grupos anti-inmigración) que ofrecen un sentido de pertenencia y un enemigo claro para gobernar así su malestar de manera productiva para la plataforma.

Por esta misma línea, planteamos que metodológicamente el análisis de sentimiento tradicional (positivo/negativo/neutral) es insuficiente, pero las economías afectivas (Ahmed, 2015) permiten mapear el valor social y político de las emociones en circulación. Desde esta perspectiva, consideramos que un análisis de emociones interseccionales, como marco integrado, desarrollaría taxonomías afectivas granulares (ira indignada, desprecio, resentimiento, placer sádico) y entrenaría modelos de PLN para detectarlas. Luego, analizaría los modos en que se asignan diferencialmente a grupos interseccionales (el desprecio hacia las mujeres políticas vs. el miedo hacia el migrante racializado). Adicionalmente, se investigaría la forma en que ciertos afectos (como la indignación virtuosa o el resentimiento identitario) funcionan como moneda social dentro de comunidades online, acumulando capital afectivo para los usuarios que los expresan de manera afectiva y efectiva, lo cual a su vez es amplificado por los algoritmos.

Para operacionalizar estos marcos, proponemos una metodología mixta en cuatro fases. Una primera fase computacional para mapear a gran escala de volumen, compromiso y emociones complejas. Una segunda fase de gubernamentalidad para realizar una auditoría algorítmica de patrones de recomendación y segmentación. Una tercera fase de economía política para analizar flujos de monetización y modelos de negocio. Y una cuarta fase crítico-interseccional para realizar etnografía digital y análisis de impacto experiencial.

Nuestro enfoque en un corpus refinado, aunque metodológicamente riguroso, comparte la limitación geográfica que Antona Jimeno et al. (2024) identificaban en el campo: la concentración en producción

española/europea puede excluir perspectivas del Sur Global relevantes para comprender la interseccionalidad en contextos postcoloniales.

Consideramos algunas implicaciones y soluciones factibles que trascienden lo académico, por ejemplo, para plataformas digitales la necesidad de sistemas que capturen dimensiones interseccionales más allá del odio explícito, para políticas públicas la urgencia de regulaciones que consideren la gubernamentalidad algorítmica como factor estructural, y para investigación futura la creación de observatorios interdisciplinarios para monitoreo continuo.

Recomendamos finalmente un cambio hacia marcos integrados que capturen simultáneamente la escala computacional, la profundidad estructural de las opresiones interseccionales, la dimensión afectiva y la gobernanza algorítmica, pues solo mediante esta aproximación integral podremos desarrollar intervenciones que aborden las raíces del odio interseccional en los ecosistemas digitales contemporáneos.

Apoyos

Este artículo es financiado mediante Beca Doctoral por el Ministerio de Ciencia, Tecnología e Innovación de Colombia y apoyado por la Vicerrectoría de Investigaciones, Innovación y Extensión, de la Universidad Tecnológica de Pereira.

Agradecimientos

La autora desea agradecer a la Dra. Ruth Amanda Cortes-Salcedo por sus comentarios durante las etapas finales de esta revisión. Este trabajo está dedicado a la memoria de la Dra. María Camila Ospina-Alvarado, cuyas ideas y legado intelectual fortalecieron este proyecto.

Referencias

- Ahmed, S. (2015). *La Política Cultural De Las Emociones*. Universidad Nacional Autónoma de México.
- Aldamen, Y. (2023). Xenophobia and Hate Speech towards Refugees on Social Media: Reinforcing Causes, Negative Effects, Defense and Response Mechanisms against That Speech. *Societies*, 13(4), 83. <https://doi.org/10.3390/soc13040083>
- Amores, J.-J. y Arcila-Calderón, C. (2025). What is the nature of the racist and xenophobic hate speech disseminated on social media in Southern Europe? Topic modelling of anti-immigration messages posted on X and YouTube in Spain, Italy and Greece. *Communication & Society*, 38(1), 163–180. <https://doi.org/10.15581/003.38.1.013>
- Antona Jimeno, T., Mayagoitia-Soria, A. y Dordević, J. (2024). Investigación sobre el Discurso de Odio. Una propuesta de análisis bibliométrico en España y LATAM entre 2021 y 2022. *Revista ICONO 14. Revista científica de Comunicación y Tecnologías emergentes*, 22(1), e2128. <https://doi.org/10.7195/ri14.v22i1.2128>
- Arce-García, S. y Menéndez-Menéndez, M.-I. (2023). Inflaming public debate: a methodology to determine origin and characteristics of hate speech about sexual and gender diversity on Twitter. *Profesional de la información*, 32(1), e320106. <https://doi.org/10.3145/epi.2023.ene.06>
- Arcila-Calderón, C., Amores, J.-J., Sánchez-Holgado, P., Vrysis, L., Vryzas, N. y Oller Alonso, M. (2022a). How to Detect Online Hate towards Migrants and Refugees? Developing and Evaluating a Classifier of Racist and Xenophobic Hate Speech Using Shallow and Deep Learning. *Sustainability*, 14(20), 13094. <https://doi.org/10.3390/su142013094>
- Arcila-Calderón, C., Sánchez-Holgado, P., Quintana-Moreno, C., Amores, J.-J. y Blanco-Herrero, D. (2022b). Hate speech and social acceptance of migrants in Europe: Analysis of tweets with geolocation. *Comunicar*, 30, 21–34. <https://doi.org/10.3916/C71-2022-02>
- Aria, M. y Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975. <https://doi.org/10.1016/j.joi.2017.08.007>
- Boukemia, J., Sältzer, M. y Boyer, S. (2025). Misogyny, politics, and social media determinants of hostile engagement against women parliamentarians on Twitter. *Legislative Studies Quarterly*, 50(3), e12486. <https://doi.org/10.1111/lsq.12486>
- Breazu, P. (2023). Entitlement Racism on YouTube: White injury—the licence to Humiliate Roma migrants in the UK. *Discourse, Context & Media*, 55, 100718. <https://doi.org/10.1016/j.dcm.2023.100718>
- Butler, J. (2021). *Palabras que hieren. Sobre el discurso de odio y la política de lo performativo*. Paidós.
- Collins, P. H. (2000). *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. Routledge. <https://doi.org/10.4324/9780203900055>
- Crenshaw, K. W. (1989). Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics. *The University of Chicago Legal Forum*, 139–167. https://scholarship.law.columbia.edu/faculty_scholarship/3007
- Dellagiacoma, L., Geschke, D. y Rothmund, T. (2024). Ideological attitudes predicting online hate speech: the differential effects of right-wing authoritarianism and social dominance orientation. *Frontiers in Social Psychology*, 2, 1389437. <https://doi.org/10.3389/frsps.2024.1389437>
- Dickel, V. y Evolvi, G. (2023). “Victims of feminism”: exploring networked misogyny and #MeToo in the manosphere. *Feminist Media Studies*, 23(4), 1392–1408. <https://doi.org/10.1080/14680777.2022.2029925>
- Giraud, E. H., Poole, E., de Quincey, E. y Richardson, J. E. (2025). Learning from online hate speech and digital racism: From automated to diffractive methods in social media analysis. *The Sociological Review*, 73(6), 1388–1407. <https://doi.org/10.1177/00380261241305260>

- Hangartner, D., Gennaro, G., Alasiri, S., Bahrich, N., Bornhoft, A., Boucher, J., Demirci, B. B., et al. (2021). Empathy-based counterspeech can reduce racist hate speech in a social media field experiment. *Proceedings of the National Academy of Sciences*, 118(50), e2116310118. <https://doi.org/10.1073/pnas.2116310118>
- Hong, T., Tang, Z., Lu, M., Wang, Y., Wu, J. y Wijaya, D. (2025). Effects of #coronavirus content moderation on misinformation and anti-Asian hate on Instagram. *New Media & Society*, 27(2), 931–954. <https://doi.org/10.1177/14614448231187529>
- Ibrahimova, M. (2023). *Desenmascarando los discursos de odio en el mundo digital*. UNESCO. <https://courier.unesco.org/es/articles/desenmascarando-los-discursos-de-odio-en-el-mundo-digital>
- Iranzo-Cabrera, M., Castro-Bleda, M. J., Simón-Astudillo, I. y Hurtado, L.-F. (2024). Journalists' Ethical Responsibility: Tackling Hate Speech Against Women Politicians in Social Media Through Natural Language Processing Techniques. *Social Science Computer Review*, 43(3), 475–502. <https://doi.org/10.1177/08944393241269417>
- Izquierdo Montero, A., Laforgue-Bullido, N. y Abril-Hervás, D. (2022). Hate speech: a systematic review of scientific production and educational considerations. *Revista Fuentes*, 24(2), 222–233. <https://doi.org/10.12795/revistafuentes.2022.20240>
- Kuric Kardelis, S., Sanmartín Ortí, A., Morañón Ferrer, X. y Guiteras Vila, X. (2024). MENAs y polarización en medios sociales digitales: la retroalimentación del odio. *Revista ICONO 14. Revista científica de Comunicación y Tecnologías emergentes*, 22(1), e2074. <https://doi.org/10.7195/ri14.v22i1.2074>
- Kwarteng, J., Perfumi, S. C., Farrell, T., Third, A. y Fernandez, M. (2022). Misogynoir: challenges in detecting intersectional hate. *Social Network Analysis and Mining*, 12(1), 166. <https://doi.org/10.1007/s13278-022-00993-7>
- Li, X. (2024). Hate speech against women on social media: Case study analysis in Asia. *Environment and Social Psychology*, 9(12), 1–8. <https://doi.org/10.59429/esp.v9i12.3220>
- Lilleker, D. y Pérez-Escobar, M. (2023). Demonising Migrants in Contexts of Extremism: Analysis of Hate Speech in UK and Spain. *Politics and Governance*, 11(2), 127–137. <https://doi.org/10.17645/pag.v11i2.6302>
- Lopes Miranda, S. (2023). Analyzing Hate Speech Against Women on Instagram. *Open Information Science*, 7(1), 20220161. <https://doi.org/10.1515/opis-2022-0161>
- Magano, O. y D'Oliveira, T. (2023). Antigypsyism in Portugal: Expressions of Hate and Racism in Social Networks. *Social Sciences*, 12(9), 511. <https://doi.org/10.3390/socsci12090511>
- Martino, B. (2025). Justicia Social, odio digital y construcción monstruosa del Otro. *Millcayac - Revista Digital de Ciencias Sociales*, 11(21), 1–12. <https://revistas.uncu.edu.ar/ojs3/index.php/millca-digital/article/view/8685>
- Matamoros-Fernández, A. (2017). Platformed racism: the mediation and circulation of an Australian race-based controversy on Twitter, Facebook and YouTube. *Information, Communication & Society*, 20(6), 930–946. <https://doi.org/10.1080/1369118X.2017.1293130>
- Matamoros-Fernández, A. y Farkas, J. (2021). Racism, Hate Speech, and Social Media: A Systematic Review and Critique. *Television & New Media*, 22(2), 205–224. <https://doi.org/10.1177/1527476420982230>
- Ministerio de Tecnologías de la Información y las Comunicaciones. (2025, Mar 21). *Acciones para frenar la discriminación en Internet y construir un entorno digital más inclusivo*. <https://www.mintic.gov.co/portal/inicio/Sala-de-prensa/Noticias/400125:Acciones-para-frenar-la-discriminacion-en-Internet-y-construir-un-entorno-digital-mas-inclusivo>
- Mozafari, M., Farahbakhsh, R. y Crespi, N. (2020). Hate speech detection and racial bias mitigation in social media based on BERT model. *PLOS ONE*, 15(8), e0237861. <https://doi.org/10.1371/journal.pone.0237861>
- Ndahinda, F. M. y Mugabe, A. S. (2024). Streaming Hate: Exploring the Harm of Anti-Banyamulenge and Anti-Tutsi Hate Speech on Congolese Social Media. *Journal of Genocide Research*, 26(1), 48–72. <https://doi.org/10.1080/14623528.2022.2078578>
- Olmos Alcaraz, A. (2023). El discurso político sobre las migraciones en Twitter durante la «crisis migratoria» de Ceuta (2021): De la corrección política al discurso del odio. *Cultura, Lenguaje y Representación*, 31, 13–30. <https://doi.org/10.6035/clr.6793>
- Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., et al. (2021). PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ*, 372, n160. <https://doi.org/10.1136/bmj.n160>
- Parvareh, V. (2023). Covertly communicated hate speech: A corpus-assisted pragmatic study. *Journal of Pragmatics*, 205, 63–77. <https://doi.org/10.1016/j.pragma.2022.12.009>
- Piñero-Otero, T. y Martínez-Rolán, X. (2021). Say it to my face: Analysing hate speech against women on Twitter. *Profesional de la información*, 30(5), e300502. <https://doi.org/10.3145/epi.2021.sep.02>
- Pronoza, E., Panicheva, P., Koltsova, O. y Rosso, P. (2021). Detecting ethnicity-targeted hate speech in Russian social media texts. *Information Processing & Management*, 58(6), 102674. <https://doi.org/10.1016/j.ipm.2021.102674>
- Ramírez-García, A., González-Molina, A. y Moyano-Pacheco, M. (2022). Interdisciplinarity of scientific production on hate speech and social media: A bibliometric analysis. *Comunicar*, 30(72), 129–140. <https://doi.org/10.3916/C72-2022-10>
- Rouvroy, A. y Berns, T. (2018). Gobernabilidad algorítmica y perspectivas de emancipación: ¿lo dispar como condición de individuación mediante la relación? *Ecuador Debate*, 104, 124–147. <http://hdl.handle.net/10469/15424>
- Sánchez-Holgado, P., Arcila-Calderón, C. y Gomes-Barbosa, M. (2023). Hate Speech and Polarization Around the “Trans Law” in Spain. *Politics and Governance*, 11(2), 187–197. <https://doi.org/10.17645/pag.v11i2.6374>
- Sánchez-Meza, M., Schlesier Corrales, L., Visa Barbosa, M. y Carnicé-Mur, M. (2023). ¿De redes sociales a redes del odio? Análisis de la conversación digital en Twitter sobre la ministra de Igualdad española Irene Montero. *Estudios sobre el Mensaje Periodístico*, 29(3), 717–736. <https://doi.org/10.5209/esmp.87271>
- Silva, C. y Carvalho, P. (2023). Quando É que os Elogios e o Humor Podem Ser Considerados Discurso de Ódio? Uma Perspetiva dos Grupos-Alvo em Portugal. *Comunicação e Sociedade*, 43, e023006. [https://doi.org/10.17231/comsoc.43\(2023\).4135](https://doi.org/10.17231/comsoc.43(2023).4135)
- Tarullo, R., Calvo, D. y Iranzo-Cabrera, M. (2024). #Feminazis on Instagram: Hate Speech and Misinformation, the Ongoing Discourse for Enlarging the Manosphere. *KOME*, 12(2), 91–110. <https://doi.org/10.17646/KOME.of.18>
- Tontodimamma, A., Nissi, E., Sarra, A. y Fontanella, L. (2021). Thirty years of research into hate speech: topics of interest and their evolution. *Scientometrics*, 126(1), 157–179. <https://doi.org/10.1007/s11192-020-03737-6>

- UNESCO. (2021). *Addressing hate speech on social media: contemporary challenges* (Discussion paper, 1). <https://unesdoc.unesco.org/ark:/48223/pf0000379177>
- van Eck, N. J. y Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538. <https://doi.org/10.1007/s11192-009-0146-3>
- Vélez-Coto, M., Morejón-Llamas, N. y Cárdenas-Rica, M. L. (2025). Sentimientos de odio ante fotografías racializadas: respuesta emocional y comportamental de adultos en redes sociales. *Palabra Clave*, 28(1), e2816. <https://doi.org/10.5294/pacla.2025.28.1.6>
- We Are Social. (2025). *Digital Report 2025*. We Are Social Ltd. <https://wearesocial.com/uk/blog/2025/02/digital-2025>
- Williams, M. L., Burnap, P., Javed, A., Liu, H. y Ozalp, S. (2020). Hate in the Machine: Anti-Black and Anti-Muslim Social Media Posts as Predictors of Offline Racially and Religiously Aggravated Crime. *The British Journal of Criminology*, 60(1), 93–117. <https://doi.org/10.1093/bjc/azz049>
- Wu, C.-Y., Chang, S.-Y., Liu, L.-Y. y Hohl, A. (2025). GIS-based analysis of anti-Asian hate speech and its socioeconomic and ideological drivers in the United States during the early COVID-19 pandemic. *GeoJournal*, 90(1), 10. <https://doi.org/10.1007/s10708-024-11264-7>