

Retroalimentación afectiva en Telegram: el papel de las reacciones emoji en la recepción de mensajes políticos hostiles

Nombre y apellidos:	Marco Antonio Ruiz-Pérez
Institución de afiliación:	Universidad Complutense de Madrid
Email:	marcoantoniorp@ucm.es
Personal investigador en formación y doctorando en Periodismo en la UCM. Máster en Comunicación, Cultura, Sociedad y Política por la UNED y graduado en Derecho y en Ciencias Políticas y de la Administración por la UGR. Fue becario de formación para posgraduados del Centro de Investigaciones Sociológicas en 2024. Sus líneas de investigación se centran en comunicación política, campañas y debates electorales, medios de comunicación, discurso de odio y extrema derecha.	

Nombre y apellidos:	Javier García-Marín
Institución de afiliación:	Universidad de Granada
Email:	jgmarin@ugr.es
Profesor Titular en el Departamento de Ciencia Política y de la Administración de la UGR. Doctor en Ciencia Política, licenciado en Historia Contemporánea y máster en Economía y Relaciones Internacionales. Ha sido profesor visitante en la London School of Economics, la Universidad de Glasgow, la Universidad de Ámsterdam, Sciences Po Bordeaux y la Universidad de Buenos Aires. Su investigación se centra en comunicación política y en la aplicación de técnicas de aprendizaje automático al análisis de textos y contenidos audiovisuales.	

Nombre y apellidos:	Francisco José Segado Boj
Institución de afiliación:	Universidad Complutense de Madrid
Email:	fsegado@ucm.es
Doctor en Periodismo por la UCM (2008) y Profesor Titular en el Departamento de Periodismo y Comunicación Global de la Facultad de Ciencias de la Información de la UCM. Sus líneas de investigación se centran en comunicación digital, redes sociales, meta-investigación y discursos de odio. Ha publicado más de 70 artículos científicos y cuenta con experiencia en encuestas, análisis de contenido y métodos bibliométricos.	

Resumen

Las redes sociales no solo difunden mensajes políticos, sino que también actúan como infraestructuras de interacción y retroalimentación afectiva. En Telegram, esta dimensión adquiere especial relevancia por la coexistencia de distintas *affordances* de participación, como los reenvíos, las respuestas y las reacciones mediante emojis. En este contexto, el presente trabajo analiza cómo dos mecanismos de hostilidad discursiva —el *othering* y el discurso de odio— se relacionan con las formas de interacción de las audiencias en el ecosistema español de Telegram vinculado a la extrema derecha. El estudio parte de un corpus de 4.113.119 mensajes textuales procedentes de 66 canales y combina clasificación automatizada de ambos mecanismos de hostilidad con el análisis de metadatos de difusión, respuesta y reacción. Los resultados muestran patrones diferenciados: el *othering* se asocia con una mayor difusión mediante reenvíos y con una mayor probabilidad de recibir reacciones, mientras que el discurso de odio presenta una relación más intensa con la respuesta conversacional. Asimismo, el análisis de los emojis revela repertorios afectivo-evaluativos distintos, con una mayor presencia de expresiones de elevada intensidad afectiva entre las reacciones asociadas al odio. En conjunto, los resultados muestran que la hostilidad en Telegram no solo se manifiesta en los contenidos emitidos, sino también en las formas diferenciadas mediante las que las audiencias los redistribuyen, responden y codifican afectivamente.

Palabras clave

emojis, extrema derecha, polarización, Telegram, discurso hostil

Introducción

La consolidación de las plataformas digitales como espacios de comunicación política ha transformado no solo las posibilidades de producción y circulación de contenidos, sino también las formas mediante las que las audiencias interactúan con ellos. El ciberespacio constituye hoy una extensión de la esfera pública en la que se articulan procesos de comunicación, movilización y formación de opinión (Mazzoleni, 2010; Cotarelo, 2013), mientras que las arquitecturas específicas de cada plataforma condicionan las posibilidades de visibilidad, circulación e interacción de los actores políticos. Estas transformaciones han resultado particularmente relevantes para la extrema derecha, que ha aprovechado los entornos digitales como espacios privilegiados para la organización, la movilización y la difusión de sus repertorios discursivos (Caiani y Parenti, 2013; O'Callaghan et al., 2013; Bennett y Kneuer, 2024).

Dentro de este ecosistema, Telegram ocupa una posición singular. Su combinación de canales de difusión, grupos de conversación, reenvíos y formas ligeras de participación permite observar simultáneamente la producción de mensajes políticos, su circulación y las respuestas que generan entre las audiencias. Esta arquitectura resulta especialmente pertinente para estudiar la hostilidad política, dado que permite distinguir entre mecanismos discursivos diferentes —como el discurso de odio y el *othering*— y analizar si estos producen patrones diferenciados de difusión, interacción y retroalimentación afectiva. El presente estudio aborda esta relación atendiendo específicamente a tres *affordances* de Telegram —reenvíos, respuestas y reacciones mediante emojis— y a las diferencias entre dos formas de hostilidad discursiva: el odio explícito y la construcción del «otro».

Far right o extrema derecha

La delimitación conceptual de la extrema derecha continúa siendo objeto de debate debido a la heterogeneidad de actores, organizaciones y subculturas situadas más allá de la derecha convencional (Pirro, 2023). Siguiendo la distinción propuesta por Mudde (2019), cabe diferenciar entre una *radical right* que acepta formalmente el sistema democrático, aunque cuestiona componentes esenciales de la democracia liberal —como el pluralismo, los derechos de las minorías o el Estado de derecho—, y una *extreme right* abiertamente contraria a la democracia. En este trabajo se adopta *far right* o extrema derecha como concepto paraguas para ambas familias, una opción especialmente pertinente cuando se analizan ecosistemas digitales caracterizados por la coexistencia e interconexión de actores ideológicamente heterogéneos (Mudde, 2019; Pirro, 2023). En su denominada cuarta ola, además, estos actores han experimentado una creciente normalización política, mediática y social, incrementando su capacidad de penetración en la esfera pública contemporánea (Mudde, 2019).

A pesar de su diversidad, la literatura identifica un conjunto relativamente estable de asuntos y repertorios discursivos compartidos. Entre ellos aparecen el antielitismo, las narrativas conspirativas, el racismo y la victimización blanca, el rechazo de la inmigración, los discursos *anti-woke* y las posiciones contrarias al feminismo, el género y las comunidades LGBTIQ+ (Willaert et al., 2022; Curley et al., 2022). La crítica a las élites políticas nacionales e internacionales, los partidos tradicionales y la izquierda constituye uno de los elementos más recurrentes (Winberg, 2017; Vaughan y Heft, 2023; Pérez-Díaz y

Arroyas Langa, 2025), mientras que las posiciones antiinmigración se observan de forma transversal en diferentes contextos nacionales (Bastos, 2020; Indelicato y Magalhães Lopes, 2024; Edilashvili, 2025). Asimismo, las teorías conspirativas suelen integrarse con estos asuntos, como ocurre con la vinculación entre antielitismo y conspiracionismo (DiMaggio, 2022) o entre inmigración y la teoría del «gran reemplazo» (Svatoňová y Doerr, 2024).

Esta orientación discursiva tiene un marcado componente antagonista. Las estrategias comunicativas de la extrema derecha se articulan recurrentemente mediante la identificación de responsables, amenazas y enemigos —de la nación, de determinados grupos sociales o de un orden moral considerado legítimo— (Cervi et al., 2023). La literatura ha relacionado esta construcción de enemigos con el proceso de *enemization* (Salvatore et al., 2018), mediante el cual determinados actores dejan de ser presentados simplemente como adversarios políticos para convertirse en grupos moralmente deslegitimados, amenazantes o incompatibles con el *ingroup*. Este antagonismo se combina además con un uso estratégico de componentes emocionales, especialmente apelaciones al miedo, el resentimiento, la amenaza o el agravio (Wodak, 2021). En consecuencia, estudiar la comunicación de la extrema derecha exige considerar no solo sobre qué asuntos habla, sino también los mecanismos de hostilidad mediante los que construye las fronteras entre grupos y las respuestas que esos mensajes generan entre sus comunidades.

Telegram y extrema derecha

El desarrollo de la extrema derecha digital se encuentra estrechamente relacionado con la diversificación del ecosistema de plataformas. El endurecimiento de las políticas de moderación de servicios convencionales como Facebook, Twitter o Instagram y la suspensión de cuentas por incumplimiento de sus términos de uso han favorecido procesos de *deplatforming* (Rogers, 2020). Paralelamente, han adquirido protagonismo plataformas alternativas como Gab, Parler, BitChute o Telegram, integradas en lo que Donovan et al. (2018) conceptualizan como ecosistema *alt-tech*. Estos espacios proporcionan infraestructuras digitales en las que comunidades ideológicamente afines pueden reconstruir redes de información, sociabilidad y movilización tras su expulsión o desplazamiento desde plataformas convencionales (Rogers, 2020; Urman y Katz, 2020).

Telegram ocupa, no obstante, una posición híbrida dentro de este ecosistema. No se trata de una plataforma concebida específicamente para comunidades políticas radicales, sino de un servicio generalista de mensajería que combina funciones propias de aplicaciones privadas, redes sociales y sistemas de difusión masiva. Buehling y Heft (2023) la caracterizan precisamente como una «plataforma híbrida»: orientada a un público general, pero utilizada simultáneamente como espacio de refugio por formas de activismo consideradas desviadas o excluidas de otros entornos digitales. En este sentido, diferentes investigaciones sitúan Telegram dentro de las infraestructuras utilizadas por la extrema derecha para reconstruir comunidades, mantener repertorios políticos y articular redes de difusión después del *deplatforming* (Rogers, 2020; Heslep y Berge, 2021; Walther y McCoy, 2021).

Esta migración no responde únicamente a menores niveles de moderación, sino también a las características sociotécnicas de la plataforma. Telegram integra chats privados, grupos y

supergrupos muchos-a-muchos y canales uno-a-muchos capaces de alcanzar audiencias amplias. Van Raemdonck y Pierson (2021) señalan que esta arquitectura combina funciones de «evangelización» mediante canales caracterizados por una comunicación asimétrica con espacios de coordinación y conversación grupal, articulados entre sí mediante los reenvíos. A ello se suman *affordances* infraestructurales asociadas a la resiliencia frente a bloqueos y censura, que han contribuido a reforzar su percepción como espacio orientado a la privacidad y la libertad de expresión (Ermoshina y Musiani, 2021; Santos, Saldaña y Tsyganova, 2021).

Estas características han permitido a la extrema derecha desarrollar en Telegram una auténtica infraestructura comunicativa (Urman y Katz, 2020), empleada tanto para la circulación de información como para la producción y difusión de contenidos hostiles (Al-Rawi, 2021; Stewart et al., 2023). En España, Ruiz-Pérez (2026) identifica precisamente un ecosistema de extrema derecha en Telegram constituido por actores diversos —*influencers* y líderes de opinión, pseudomedios, partidos y colectivos y espacios de discusión— que participan en la producción, circulación y recepción de contenidos políticos. Esta heterogeneidad hace especialmente relevante analizar Telegram no únicamente como repositorio de mensajes, sino como un espacio en el que la arquitectura de la plataforma permite observar cómo determinados contenidos son difundidos, respondidos y afectivamente validados por sus audiencias.

Hostilidad discursiva: odio y *othering*

La producción de hostilidad constituye una dimensión central de la comunicación de la extrema derecha. La literatura ha señalado una asociación persistente entre estos actores y el discurso de odio, hasta el punto de identificar el odio tanto como resultado de procesos de radicalización como componente característico de determinados entornos comunicativos de extrema derecha (Marwick et al., 2022). Sin embargo, la hostilidad política no se manifiesta únicamente mediante insultos, amenazas o expresiones abiertamente degradantes. Una parte relevante de ella opera a través de mecanismos discursivos más implícitos que delimitan quién pertenece al grupo y quién debe ser situado fuera de él.

En este punto resulta útil distinguir analíticamente entre discurso de odio y *othering*. El primero remite a formas comparativamente explícitas de agresión, degradación, estigmatización o amenaza dirigidas hacia individuos o colectivos. Aunque conceptos como discurso de odio, lenguaje tóxico, abuso verbal, agresividad o incivilidad se han utilizado de manera parcialmente solapada en la literatura (Anjum y Katarya, 2024), el componente distintivo de estas formas de hostilidad se encuentra en su orientación hacia la degradación o agresión simbólica del objetivo (Pradel et al., 2024). La investigación computacional se ha centrado especialmente en estas expresiones explícitas, tanto mediante estudios sobre toxicidad como mediante sistemas automatizados de detección de odio (Vidgen et al., 2021).

El *othering*, en cambio, describe un mecanismo discursivo de diferenciación mediante el cual se construye y refuerza la frontera entre «nosotros» y «ellos» (Van Dijk, 1993). Estas estrategias permiten atribuir inferioridad, ilegitimidad o amenaza a determinados grupos y fijar simbólicamente quién pertenece a la comunidad y quién queda situado fuera de ella. En la comunicación de la extrema derecha, este proceso se dirige habitualmente hacia mujeres

(Southern y Harmer, 2019), migrantes y minorías étnicas o religiosas (Molla, 2024), colectivos LGBTIQ+ (Nartey, 2021) y también hacia élites políticas e institucionales (Resende y Freire, 2025; Machuy, 2025). Novais y Christofolletti (2025) distinguen, en este sentido, entre un *othering downward*, dirigido hacia grupos subordinados o vulnerables, y un *othering upward*, orientado hacia grupos percibidos como poderosos, como élites políticas, económicas o institucionales.

El *othering* se encuentra estrechamente relacionado con los procesos de construcción de enemigos (*enemization*) característicos de la comunicación polarizada (Salvatore et al., 2018). García-Marín et al. (2025) vinculan estas estrategias con la oposición binaria entre *ingroup* y *outgroup*, la evaluación moral y la estereotipación de grupos subordinados, mientras que Alorainy et al. (2018) muestran que el *othering* puede constituir un marcador especialmente relevante del odio implícito. De este modo, permite capturar una «zona gris» de hostilidad que los sistemas orientados exclusivamente hacia expresiones explícitas pueden tener dificultades para identificar (Assimakopoulos et al., 2017; Young, 2025).

Ambos mecanismos son, por tanto, conceptualmente próximos pero no equivalentes. Mientras que el *othering* puede operar mediante deslegitimación, sospecha moral, exclusión simbólica o representación de un grupo como amenaza, el odio implica una expresión más directa de rechazo, degradación o agresión. Esta diferencia permite plantear que no todas las fronteras simbólicas construidas mediante *othering* alcanzan necesariamente la misma intensidad afectiva que las expresiones explícitas de odio. Al mismo tiempo, las estrategias de *othering* pueden contribuir a intensificar la hostilidad al representar al *outgroup* no solo como diferente, sino como peligroso o invasivo (Stephan y Stephan, 2000; Zamri et al., 2023). La relación entre ambos mecanismos debe entenderse, en consecuencia, como parcialmente solapada: el *othering* puede contribuir a estructurar las fronteras simbólicas sobre las que se desarrolla la hostilidad, mientras que el odio representa una forma comparativamente más explícita e intensa de expresión de esa hostilidad.

Esta distinción adquiere especial importancia cuando la atención se desplaza desde la producción del mensaje hacia su recepción. Si el odio contiene una manifestación más explícita de agresión y una mayor activación de componentes evaluativos y afectivos, cabe preguntarse si también genera una respuesta más intensa de las audiencias que formas más implícitas de delimitación del «otro».

Affordances de Telegram: reenvíos, respuestas y reacciones

Las plataformas no constituyen espacios neutrales en los que circula el contenido político. Sus arquitecturas habilitan y restringen determinados cursos de acción, condicionando cómo los usuarios producen, distribuyen y responden a los mensajes. El concepto de *affordance* permite analizar precisamente esta relación entre las características de una tecnología, sus usuarios y las posibilidades de acción que emergen de su interacción (Evans et al., 2017). Desde esta perspectiva, Telegram ofrece una configuración especialmente relevante porque integra en un mismo entorno mecanismos de difusión, conversación y retroalimentación de baja intensidad.

Lou et al. (2021) señalan que Telegram facilita el consumo y la circulación de información mediante *affordances* relacionadas con la accesibilidad, la personalización y la

simultaneidad: los usuarios pueden recibir información, reenviarla, comentarla y discutirla sin abandonar la aplicación. En este trabajo se atiende particularmente a tres de estas posibilidades de acción: reenvíos, respuestas y reacciones. Aunque las tres pueden considerarse indicadores de interacción con el contenido, representan comportamientos diferentes. Los reenvíos expresan una acción de redistribución mediante la cual el usuario contribuye a ampliar la circulación del mensaje más allá de su contexto inmediato; las respuestas implican una intervención conversacional más costosa, que puede adoptar la forma de apoyo, discusión o contestación; y las reacciones mediante emojis ofrecen un mecanismo de retroalimentación inmediato y de bajo coste que permite comunicar una evaluación sin necesidad de producir una respuesta textual.

La investigación previa muestra que la circulación de contenidos en Telegram no depende únicamente del tamaño de la audiencia, sino también de las propiedades discursivas del propio mensaje. Jost y Dogruel (2023) encuentran que determinadas características —como las apelaciones a la acción *online* y *offline* o el antielitismo— incrementan la retransmisión en canales radicales, mientras que otros elementos discursivos presentan efectos diferentes. Esto sugiere que las *affordances* de interacción y difusión pueden funcionar como indicadores de la capacidad de distintos repertorios discursivos para activar a las audiencias. Aplicado a la hostilidad, cabe esperar que mecanismos discursivos diferentes no generen necesariamente la misma intensidad de interacción.

Esta expectativa resulta especialmente relevante en la comparación entre odio y *othering*. Si el discurso de odio representa una manifestación más explícita e intensamente afectiva de la hostilidad que las formas más amplias de delimitación simbólica asociadas al *othering*, resulta razonable esperar una mayor activación de conductas de interacción y retroalimentación. En términos observables, esta activación puede manifestarse tanto mediante respuestas y reacciones como mediante una mayor propensión a redistribuir el contenido. De este razonamiento se deriva la primera hipótesis:

H1. Los mensajes que contienen discurso de odio presentarán una mayor probabilidad de generar interacción y retroalimentación de los usuarios, así como una mayor difusión, que los mensajes que contienen *othering*.

Las reacciones mediante emojis permiten profundizar específicamente en la dimensión afectiva de esa respuesta. Los emojis constituyen señales no verbales capaces de expresar y modular emociones, evaluaciones y orientaciones relacionales, reducir ambigüedades y desempeñar funciones interpersonales y de trabajo emocional (Fischer y Herbert, 2021; Riordan, 2017). En este sentido, las reacciones pueden conceptualizarse como repertorios afectivos de participación: mecanismos mediante los cuales los miembros de una comunidad hacen visibles orientaciones emocionales o evaluativas sin elaborar una intervención discursiva completa.

No obstante, la interpretación de una reacción no puede reducirse mecánicamente a la emoción convencionalmente asociada a su emoji. Tardelli et al. (2025) muestran que en Telegram las reacciones consideradas positivas pueden predominar incluso ante mensajes negativos o neutrales. Un corazón o un pulgar hacia arriba pueden representar aprobación del contenido, solidaridad con el emisor, afiliación grupal o reconocimiento, sin implicar que el usuario comparta literalmente la emoción expresada en el mensaje. Por ello, más que

inferir directamente estados emocionales individuales, resulta metodológicamente más prudente analizar los emojis como un repertorio observable de respuesta afectiva y evaluativa.

Desde esta perspectiva, el interés no reside únicamente en determinar si los mensajes hostiles reciben más reacciones, sino también en establecer si distintos mecanismos de hostilidad producen composiciones diferentes de emojis. Si el odio constituye una forma más explícita e intensamente afectiva de hostilidad que el *othering*, cabe esperar que esta diferencia se proyecte también sobre el repertorio de señales empleado por las audiencias. De esta expectativa se deriva la segunda hipótesis:

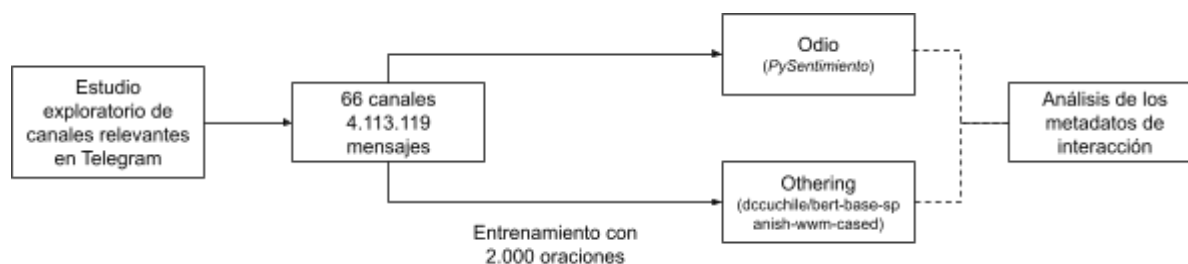
H2. Los mensajes que contienen discurso de odio estarán asociados con un repertorio de reacciones mediante emojis caracterizado por una mayor presencia de expresiones de alta intensidad afectiva que los mensajes que contienen *othering*.

En conjunto, ambas hipótesis desplazan el análisis desde una perspectiva centrada exclusivamente en el contenido emitido hacia la relación entre hostilidad discursiva, arquitectura de plataforma y comportamiento observable de las audiencias. Los reenvíos permiten aproximarse a la difusión; las respuestas, a la interacción conversacional; y las reacciones, a la retroalimentación afectiva y a su composición. Esta articulación permite examinar no sólo qué formas adopta la hostilidad en el ecosistema español de extrema derecha en Telegram, sino también qué capacidad tienen sus diferentes modalidades para activar y estructurar la participación de las comunidades digitales.

Metodología

La Figura 1 resume las distintas fases del diseño de investigación, desde la construcción del corpus hasta la clasificación automatizada de la hostilidad discursiva y el análisis de las distintas formas de interacción disponibles en Telegram. El corpus se obtuvo mediante un procedimiento de identificación de canales desarrollado en dos etapas. En primer lugar, se realizó una exploración inicial destinada a localizar los canales más relevantes y visibles vinculados con la extrema derecha en el ecosistema hispanohablante de Telegram. A partir de esta selección inicial se aplicó un procedimiento de bola de nieve, incorporando progresivamente nuevos canales cuando aparecían mencionados, enlazados o como origen de mensajes reenviados por los canales previamente identificados. La inclusión atendió a cuatro criterios: la existencia de una vinculación sustantiva con actores, discursos o comunidades asociadas con la extrema derecha; un volumen aproximado mínimo de 1.000 suscriptores; el uso predominante del español y la presencia de referencias sustantivas al contexto político, social o cultural español; y la exclusión de espacios centrados principalmente en contextos latinoamericanos o cuyas audiencias y contenidos remitieran de forma predominante a realidades distintas de la española. Esta valoración se realizó a partir del contenido publicado y de la información disponible en el nombre, descripción e imagen identificativa de cada canal.

Figura 1. Diseño de la investigación



Fuente: elaboración propia.

La muestra final estuvo compuesta por 66 canales públicos. Mediante acceso programático a la API de Telegram a través de la librería Telethon (Lonami, 2024) se recuperaron 4.960.330 mensajes publicados desde la creación de cada canal hasta el 15 de agosto de 2025. El procedimiento conservó información correspondiente al canal y a las características de los mensajes, pero no identificadores personales de los usuarios. Los posibles metadatos relativos a emisores individuales se emplearon únicamente durante el procesamiento y no fueron incorporados al conjunto de datos analítico. Antes del análisis se excluyeron los registros constituidos exclusivamente por enlaces, menciones, *hashtags*, emojis o elementos visuales sin contenido textual. Asimismo, se eliminaron del texto enlaces, menciones, *hashtags* y emojis susceptibles de introducir ruido en los procedimientos de clasificación. Tras este procesamiento, el corpus analítico quedó constituido por 4.113.119 mensajes textuales.

El *othering* se operacionalizó como la presencia de mecanismos discursivos mediante los cuales se establece una diferenciación entre un grupo propio y un grupo externo, coherente con la construcción de una frontera simbólica entre «nosotros» y «ellos» (Van Dijk, 1993). Para su detección se desarrolló un clasificador supervisado binario mediante ajuste fino de BETO (dccuchile/bert-base-spanish-wwm-cased) (Cañete et al., 2020), modelo basado en BERT y preentrenado específicamente para español. El entrenamiento utilizó 2.000 oraciones extraídas del propio corpus y codificadas manualmente, distribuidas de forma equilibrada entre 1.000 ejemplos positivos y 1.000 negativos. La oración se adoptó como unidad de entrenamiento para aislar la estructura semántica mediante la que se construye la oposición intergrupala y reducir el ruido procedente de mensajes extensos que pueden combinar contenidos heterogéneos.

Para aplicar el clasificador al corpus completo, los mensajes se segmentaron en oraciones o unidades lingüísticamente coherentes mediante spaCy (Montani et al., 2023). Los fragmentos que excedían la longitud máxima admitida por el modelo se subdividieron adicionalmente mediante tokenización. Cada segmento se clasificó de forma independiente y las puntuaciones se agregaron posteriormente al nivel del mensaje mediante una regla disyuntiva, conservando como probabilidad final la puntuación máxima obtenida entre todos los segmentos. Por tanto, un mensaje se consideró positivo cuando al menos uno de sus fragmentos superaba el punto de corte establecido.

El modelo se validó sobre una muestra independiente de 800 mensajes codificados por dos evaluadores humanos, utilizando el consenso entre ambos como referencia. A partir de esta validación se estableció un umbral de 0,734. En dicho punto de corte, el clasificador alcanzó un acuerdo observado de 0,939, κ de Cohen = 0,754, α de Krippendorff = 0,708, F1 = 0,789

y *balanced accuracy* = 0,845. La precisión fue de 0,894 y el *recall* de 0,706, lo que refleja un comportamiento relativamente conservador en la identificación de casos positivos.

La presencia de discurso de odio se estimó de manera independiente mediante el modelo en español robertuito-hate-speech, incluido en la librería *pysentimiento* (Pérez et al., 2021). Este clasificador está basado en RoBERTuito (Pérez et al., 2022) y fue entrenado con textos procedentes de Twitter del corpus español de HatEval (Basile et al., 2019). Su utilización se consideró apropiada por haber sido desarrollado específicamente para la identificación de discurso de odio en textos breves en español procedentes de redes sociales, un formato que comparte características comunicativas con una parte sustantiva de los mensajes analizados en Telegram. El modelo devuelve probabilidades correspondientes a las dimensiones *hateful*, *targeted* y *aggressive*; en este estudio se utilizó exclusivamente la puntuación *hateful*, al ser la dimensión directamente vinculada con la presencia de discurso de odio.

La puntuación obtenida mediante *pysentimiento* se contrastó con la misma muestra independiente de 800 mensajes empleada para validar el clasificador de *othering*. A partir del consenso humano se estableció un umbral de 0,746, con un acuerdo observado de 0,957, κ de Cohen = 0,712, α de Krippendorff = 0,672, $F1 = 0,736$ y *balanced accuracy* = 0,859. La precisión (0,730) y el *recall* (0,742) presentaron valores similares, indicando un comportamiento equilibrado entre precisión y capacidad de recuperación de casos positivos.

Ambas variables se mantuvieron conceptualmente diferenciadas. Mientras que el discurso de odio identifica expresiones explícitamente hostiles dirigidas hacia personas o colectivos, el *othering* captura un mecanismo más amplio de construcción, diferenciación y deslegitimación de un exogrupo. Por ello, ambos fenómenos pueden aparecer de manera independiente o concurrir en un mismo mensaje. Para los análisis comparativos se distinguieron los mensajes que presentaban exclusivamente *othering*, aquellos que presentaban exclusivamente odio y una categoría de referencia formada por mensajes en los que no se detectaba ninguno de los dos mecanismos. Los mensajes en los que coexistían *othering* y odio fueron excluidos de estos modelos con el objetivo de aislar el comportamiento específico de cada forma de hostilidad.

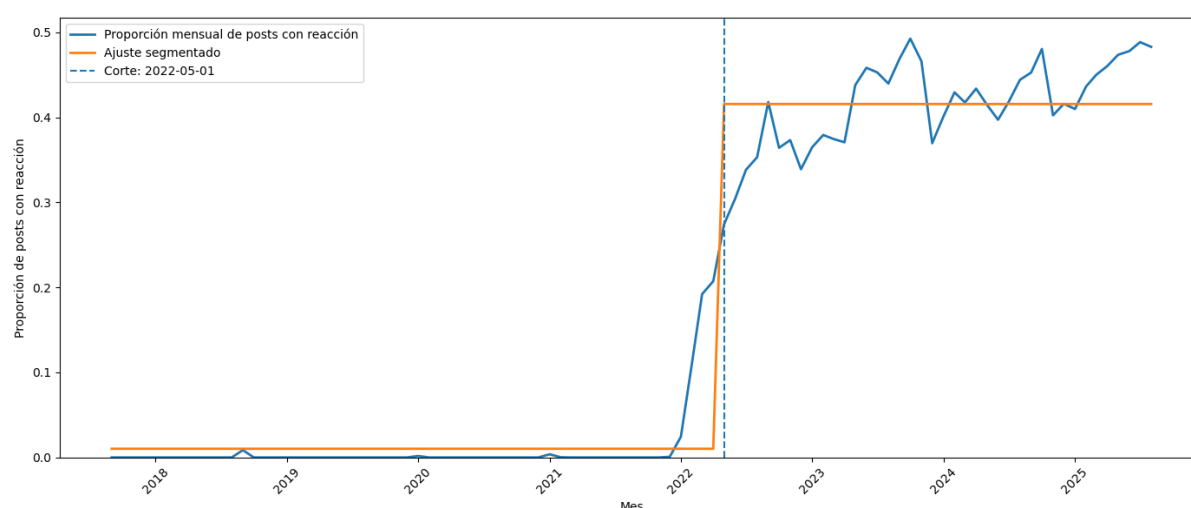
El análisis de las reacciones requiere una delimitación temporal específica, ya que esta *affordance* no estuvo disponible ni fue utilizada de manera homogénea durante todo el periodo cubierto por el corpus. Telegram introdujo oficialmente las reacciones a mensajes el 30 de diciembre de 2021. Estas quedaron habilitadas por defecto en los chats privados, mientras que en grupos y canales su disponibilidad dependía inicialmente de que los administradores las activaran (Telegram, 2021). Por esta razón, la ausencia de reacciones durante los primeros años del corpus no puede interpretarse como ausencia de respuesta de los usuarios, sino que responde fundamentalmente a una diferencia estructural en las posibilidades de interacción disponibles en la plataforma.

Para determinar empíricamente el momento a partir del cual las reacciones podían considerarse una práctica consolidada de participación, se calculó mensualmente la proporción de mensajes con actividad sustantiva de reacción entre aquellos que habían alcanzado al menos 1.000 visualizaciones. Se consideró que un mensaje presentaba actividad sustantiva cuando acumulaba un mínimo de 50 reacciones y una ratio de al menos

cinco reacciones por cada 1.000 visualizaciones. Estos requisitos permitieron excluir interacciones residuales y evitar que mensajes con una exposición extraordinariamente elevada fueran clasificados como activos únicamente por su volumen absoluto de reacciones.

Sobre esta serie mensual se aplicó un procedimiento de detección de un único punto de cambio mediante programación dinámica, utilizando una función de coste basada en la variación cuadrática y estableciendo una duración mínima de doce meses para cada segmento. Como muestra la Figura 2, el procedimiento situó el punto de cambio en mayo de 2022. Antes de esta fecha, la proporción media de mensajes elegibles con actividad sustantiva de reacción se situaba en torno al 1 %, mientras que posteriormente ascendía aproximadamente al 41,5 %. En consecuencia, todos los análisis relacionados con reacciones se restringieron al periodo comprendido entre el 1 de mayo de 2022 y el final del corpus. Esta segmentación evita comparar periodos estructuralmente distintos y permite interpretar la ausencia o presencia de reacciones dentro de una ventana temporal en la que esta funcionalidad ya constituía un mecanismo consolidado de participación.

Figura 2. Segmentación temporal de la proporción mensual de mensajes con reacción



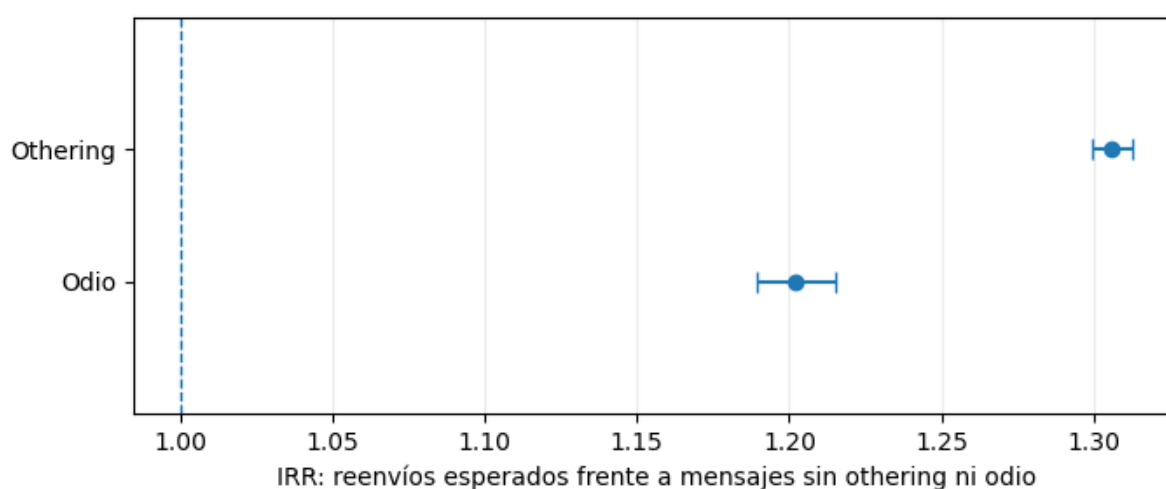
Fuente: elaboración propia.

Para evaluar la relación entre los mecanismos de hostilidad y las distintas formas de interacción se estimaron modelos de regresión adaptados a la naturaleza de cada variable dependiente. El volumen de reenvíos se analizó mediante una regresión binomial negativa, mientras que la probabilidad de recibir una respuesta o al menos una reacción se estimó mediante modelos logísticos binomiales. Para el análisis específico de los emojis se ajustaron, por separado para cada reacción, modelos binomiales que estimaron tanto su probabilidad de aparición como su sobrerrepresentación relativa dentro del conjunto de reacciones recibidas. En todos los modelos se compararon los mensajes con *othering* y con discurso de odio frente a aquellos que no contenían ninguno de los dos mecanismos, excluyendo los casos de coexistencia y controlando por el volumen de exposición mediante el logaritmo de las visualizaciones.

Resultados

Los resultados muestran que ambas formas de hostilidad se asocian positivamente con la difusión de los mensajes una vez controlado su nivel de exposición. Como se observa en la Figura 3, los mensajes que contienen exclusivamente *othering* presentan una tasa esperada de reenvíos un 30,6 % superior a la de los mensajes en los que no se detecta ni *othering* ni odio (IRR = 1,306; IC 95 % [1,300, 1,313]; $p < .001$). El discurso de odio también se relaciona con una mayor circulación, aunque con una magnitud algo menor: los mensajes que contienen exclusivamente odio presentan un incremento esperado del 20,2 % en los reenvíos (IRR = 1,202; IC 95 % [1,190, 1,215]; $p < .001$).

Figura 3. Efecto del *othering* y el odio sobre el volumen esperado de reenvíos en mensajes hostiles



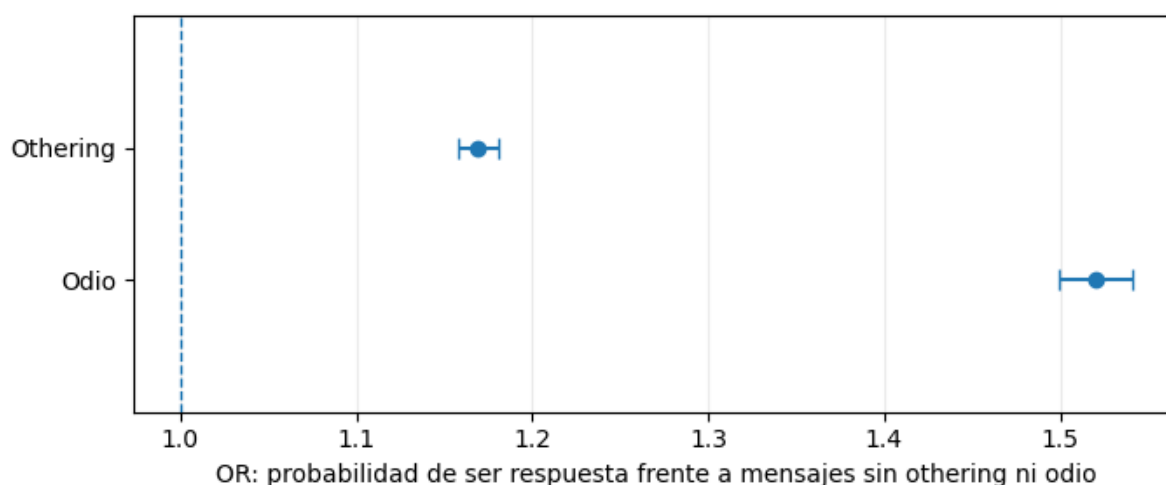
Nota. Estimaciones obtenidas mediante un modelo de regresión logística controlado por visualizaciones con errores estándar agrupados por canal.

Fuente: elaboración propia.

La Figura 3 refleja, por tanto, un patrón convergente para ambos mecanismos: una vez mantenido constante el volumen de visualizaciones, la presencia de hostilidad se relaciona con una mayor retransmisión del contenido. No obstante, la magnitud de la asociación es superior para el *othering*, cuyo IRR se sitúa aproximadamente diez puntos porcentuales por encima del correspondiente al discurso de odio. Esta diferencia indica que las formas de construcción y deslegitimación del exogrupo presentan una asociación particularmente intensa con la circulación posterior de los mensajes.

Por otro lado, la Figura 4 muestra igualmente asociaciones positivas y estadísticamente significativas entre ambos mecanismos de hostilidad y la probabilidad de interacción mediante respuesta. Los mensajes que contienen exclusivamente *othering* presentan unas *odds* de respuesta un 16,9 % superiores a las de los mensajes sin *othering* ni odio (OR = 1,169; IC 95 % [1,158, 1,180]; $p < .001$). La asociación es considerablemente mayor en el caso del discurso de odio: las *odds* de recibir una respuesta son un 52,0 % superiores a las de la categoría de referencia (OR = 1,520; IC 95 % [1,499, 1,541]; $p < .001$).

Figura 4. Efecto del othering y el odio sobre la probabilidad de respuesta en mensajes hostiles



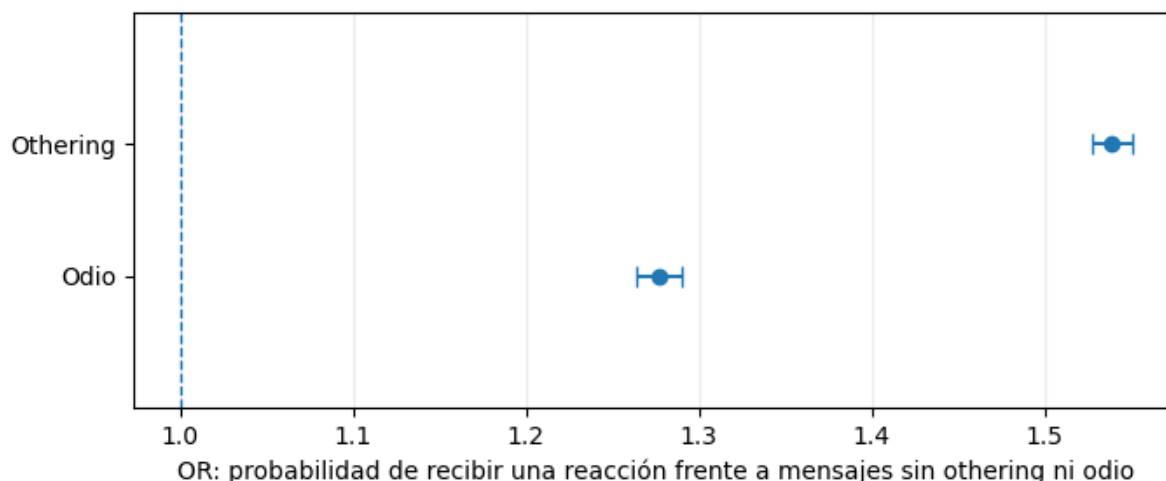
Nota. Estimaciones obtenidas mediante un modelo de regresión logística controlado por visualizaciones con errores estándar agrupados por canal.

Fuente: elaboración propia.

A diferencia de lo observado para los reenvíos, la comparación representada en la Figura 4 sitúa al discurso de odio como el mecanismo asociado con una mayor respuesta conversacional. La distancia entre ambas estimaciones es apreciable: mientras que el efecto del *othering* permanece relativamente próximo a la unidad, el correspondiente al odio alcanza un OR de 1,52. Los resultados apuntan así a que ambas formas de hostilidad favorecen la intervención de la audiencia, aunque la asociación con la respuesta explícita es sustancialmente más intensa cuando el mensaje contiene discurso de odio.

El patrón cambia nuevamente al considerar las reacciones como mecanismo de retroalimentación de bajo coste. Como muestra la Figura 5, los mensajes con exclusivamente *othering* presentan unas *odds* de recibir al menos una reacción un 53,9 % superiores a las de los mensajes sin ninguno de los dos mecanismos de hostilidad (OR = 1,539; IC 95 % [1,527, 1,551]; $p < .001$). El discurso de odio también está positivamente asociado con esta forma de participación, aunque con una intensidad menor: las *odds* aumentan un 27,7 % respecto a la categoría de referencia (OR = 1,277; IC 95 % [1,263, 1,290]; $p < .001$).

Figura 5. Efecto del othering y el odio sobre la probabilidad de reacción en mensajes hostiles



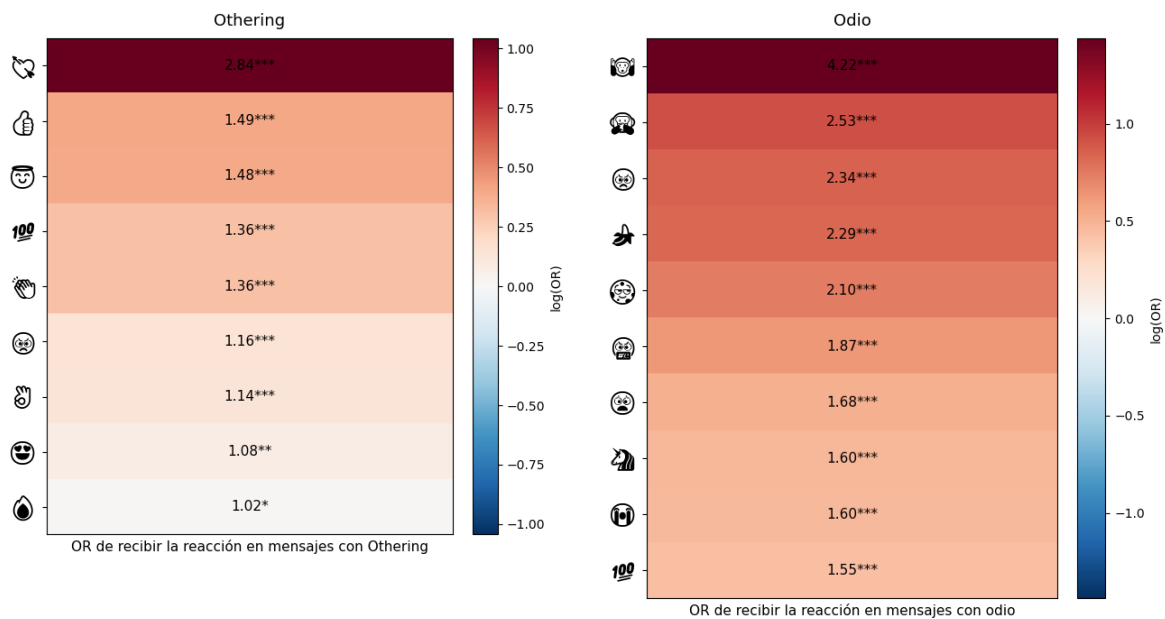
Nota. Estimaciones obtenidas mediante un modelo de regresión logística controlado por visualizaciones con errores estándar agrupados por canal.

Fuente: elaboración propia.

En consecuencia, la Figura 5 evidencia que la presencia de hostilidad se relaciona con una mayor propensión de las audiencias a reaccionar mediante emojis, pero la magnitud de la asociación difiere entre mecanismos. En este caso, el *othering* presenta el efecto comparativamente más elevado, con una estimación aproximadamente 26 puntos porcentuales superior a la correspondiente al odio. Esta diferencia resulta especialmente relevante porque indica que la capacidad de activar una reacción no sigue necesariamente el mismo patrón que la participación conversacional observada mediante las respuestas.

El análisis desagregado de las reacciones revela repertorios claramente diferenciados. En los mensajes con *othering*, la Figura 6 muestra que la asociación de mayor magnitud corresponde a 🍷 (OR = 2,840; $p < .001$), seguida a cierta distancia por 👍 (OR = 1,491), 😊 (OR = 1,481), 🙌 (OR = 1,360) y 🙌 (OR = 1,360), todas ellas estadísticamente significativas. También aparecen asociaciones positivas con 😡 (OR = 1,155), 🔥 (OR = 1,140), 😍 (OR = 1,080) y 🔥 (OR = 1,015). En conjunto, los emojis más probables ante *othering* combinan señales convencionalmente asociadas con aprobación, apoyo o reconocimiento con algunas respuestas de contenido afectivo negativo, particularmente 😡.

Figura 6. Top 10 de reacciones emoji más probables con el othering y el odio



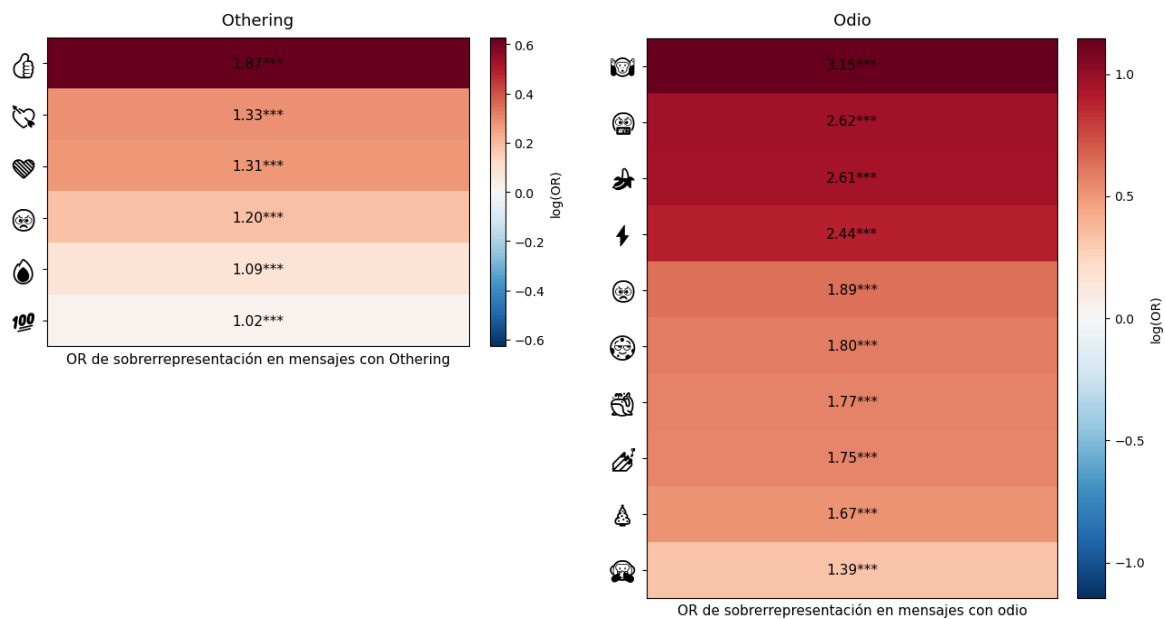
Nota. Se presentan las diez asociaciones estadísticamente significativas con mayor OR para cada mecanismo de hostilidad, ordenadas de mayor a menor. Estimaciones obtenidas mediante modelos de regresión GLM-PPML por emoji controlados por visualizaciones con errores estándar agrupados por canal.

Fuente: elaboración propia.

El repertorio asociado al odio presenta un perfil diferente y asociaciones generalmente de mayor magnitud. Como recoge la Figura 6, 🤔 es la reacción con mayor incremento relativo (OR = 4,220; $p < .001$), seguida por 🙄 (OR = 2,529), 😡 (OR = 2,339), 🙏 (OR = 2,287), 😊 (OR = 2,095) y 🙇 (OR = 1,868). También se incrementa significativamente la probabilidad de 🙈 (OR = 1,678), 🙌 (OR = 1,604), 🙌 (OR = 1,596) y 🙏 (OR = 1,548). Frente al patrón observado para el *othering*, las asociaciones más intensas en los mensajes de odio incluyen una mayor presencia de expresiones convencionalmente vinculadas con enfado, censura, miedo o aflicción, aunque también aparecen emojis cuyo significado depende fuertemente de usos irónicos o contextuales.

Cuando el análisis se centra no simplemente en la probabilidad de aparición, sino en la sobrerrepresentación de cada emoji dentro del repertorio de reacciones, las diferencias entre ambos mecanismos se hacen todavía más nítidas. Como muestra la Figura 7, el *othering* se caracteriza principalmente por 👍 (OR = 1,874; $p < .001$), seguido de 🙏 (OR = 1,331), ❤️ (OR = 1,312), 😡 (OR = 1,202), 🙌 (OR = 1,087) y 🙏 (OR = 1,021). El número de emojis sobrerrepresentados es, además, relativamente reducido, lo que indica una mayor concentración de su perfil específico en unas pocas formas de respuesta.

Figura 7. Top 10 de reacciones emoji más características con el othering y el odio



Nota. Se presentan las diez asociaciones estadísticamente significativas con mayor OR para cada mecanismo de hostilidad, ordenadas de mayor a menor. Estimaciones obtenidas mediante modelos de regresión GLM-PPML por emoji controlados por visualizaciones con errores estándar agrupados por canal.

Fuente: elaboración propia.

El discurso de odio presenta en la Figura 7 un repertorio más amplio y con asociaciones de mayor magnitud. Destacan 🤬 (OR = 3,151; $p < .001$), 🤬 (OR = 2,621), 🍌 (OR = 2,609) y ⚡ (OR = 2,439), seguidos por 😡 (OR = 1,887), 😞 (OR = 1,801), 🦋 (OR = 1,769), 🍌 (OR = 1,750), 🌲 (OR = 1,674) y 🐒 (OR = 1,388). La comparación con el *othering* muestra así dos perfiles composicionales distintos: mientras que este último presenta una sobrerrepresentación especialmente marcada de señales como 👍 y ❤️, el odio se caracteriza por un conjunto más heterogéneo de emojis y por efectos relativos considerablemente superiores en varias de sus reacciones distintivas.

Discusión

Los resultados muestran, en primer lugar, que la hostilidad discursiva no produce un efecto homogéneo sobre la participación de las audiencias, sino que su relación con el *engagement* depende tanto del mecanismo discursivo como de la *affordance* considerada. Esta constatación amplía la evidencia de Jost y Dogruel (2023), quienes muestran que las características discursivas de los mensajes condicionan su capacidad de circulación en canales radicales de Telegram. Los resultados obtenidos aquí sugieren que esta heterogeneidad no se limita a diferentes temas o apelaciones movilizadoras, sino que aparece también entre distintas formas de hostilidad. Además, el hecho de que *othering* y odio intercambien su posición relativa según se analicen reenvíos, respuestas o reacciones refuerza la conveniencia de no tratar el *engagement* como una categoría unidimensional. En línea con la perspectiva de las *affordances* (Evans et al., 2017; Lou et al., 2021), cada

posibilidad técnica implica una conducta diferente: redistribuir un contenido, intervenir conversacionalmente o proporcionar una señal de retroalimentación de bajo coste.

Esta diferenciación resulta especialmente visible en los reenvíos y las reacciones. Aunque ambos mecanismos de hostilidad incrementan la circulación respecto a los mensajes que no contienen ninguno de ellos, el *othering* presenta una asociación superior con los reenvíos y también con la probabilidad de recibir al menos una reacción. Este resultado resulta relevante a la luz de la conceptualización del *othering* como una forma más amplia y potencialmente menos explícita de hostilidad que delimita simbólicamente la frontera entre «nosotros» y «ellos» (Van Dijk, 1993; García-Marín et al., 2025). La literatura ha destacado precisamente que estas construcciones pueden operar en una zona discursiva menos manifiesta que el odio explícito, mediante procesos de deslegitimación, amenaza o exclusión del exogrupo (Alorainy et al., 2018; Young, 2025). Los resultados sugieren que este carácter menos explícito no implica una capacidad menor para activar a las audiencias. Por el contrario, las formas de diferenciación intergrupales asociadas al *othering* parecen especialmente compatibles con acciones de baja fricción como reaccionar y, sobre todo, con la redistribución del contenido.

Una interpretación posible es que la eficacia participativa de la hostilidad no aumente linealmente con su explicitud. El *othering* puede proporcionar marcos de pertenencia y exclusión suficientemente claros para activar procesos de identificación intergrupales sin requerir necesariamente expresiones abiertamente agresivas. Esto resulta coherente con los trabajos que sitúan la construcción del *outgroup* en la base de mecanismos más amplios de hostilidad y amenaza (Stephan y Stephan, 2000; Zamri et al., 2023). En el ecosistema de extrema derecha, donde la literatura ha documentado precisamente la centralidad de la construcción de enemigos y grupos amenazantes (Cervi et al., 2023; Novais y Christofolletti, 2025), este tipo de formulación puede facilitar la circulación de marcos antagonistas más allá de las expresiones inequívocamente identificables como odio. En ese sentido, los resultados complementan los estudios que han documentado la presencia de contenidos hostiles en Telegram (Al-Rawi, 2021; Stewart et al., 2023), al mostrar que su relevancia no depende únicamente de su presencia, sino también de la manera específica en que las audiencias interactúan con diferentes grados y formas de hostilidad.

El comportamiento de las respuestas introduce, sin embargo, una diferencia importante. En esta modalidad de interacción, el discurso de odio presenta una asociación considerablemente superior a la del *othering*. A diferencia de una reacción mediante emoji, responder implica una intervención explícita y comparativamente más costosa en la conversación. Por tanto, la mayor capacidad del odio para generar respuestas puede interpretarse como indicio de una mayor capacidad para suscitar intervención conversacional. Este resultado es compatible con la caracterización teórica del odio como una manifestación más explícita de agresión, degradación o estigmatización que el *othering*, aunque los datos no permiten determinar si esas respuestas representan adhesión, refuerzo, crítica o confrontación. Precisamente por ello, no sería adecuado interpretar automáticamente una mayor probabilidad de respuesta como mayor aprobación del mensaje. Lo que sí muestran los resultados es una mayor activación conversacional ante las expresiones de odio.

En conjunto, este patrón obliga a rechazar la H1 en su formulación general, que anticipaba que el discurso de odio presentaría una mayor interacción, retroalimentación y difusión que el *othering*. La expectativa solo encuentra respaldo en una de las tres dimensiones estudiadas: las respuestas, donde el odio presenta una asociación sustancialmente mayor. En cambio, el *othering* supera al odio tanto en reenvíos como en probabilidad de reacción. Más que una relación ordinal en la que una mayor intensidad hostil produzca sistemáticamente más *engagement*, los resultados apuntan a una relación condicionada por la naturaleza de la acción disponible. El odio parece estar particularmente asociado con la intervención conversacional, mientras que el *othering* lo está con formas de circulación y retroalimentación menos costosas. Esta diferenciación aporta además evidencia empírica al planteamiento de Van Raemdonck y Pierson sobre la importancia de analizar Telegram a partir de los distintos cursos de acción que habilita su arquitectura, en lugar de considerar la participación como un comportamiento único.

El análisis de los emojis permite profundizar en esta diferencia atendiendo no solo a la cantidad de retroalimentación, sino también a su composición. El repertorio asociado al *othering* se caracteriza fundamentalmente por señales como 👍, ❤️, 💖, 🔥 o 100, mientras que el odio presenta asociaciones particularmente elevadas con emojis como 🤬, 😡, 🙄, 🙊, 😬 o 😭. La distinción es especialmente clara en el análisis de las reacciones características: frente a un repertorio de *othering* concentrado en un número relativamente reducido de señales de aprobación o reconocimiento, el odio se asocia con un repertorio más diverso en el que aparecen expresiones convencionalmente relacionadas con ira, enfado, impacto o aflicción. Estos resultados son compatibles con la concepción de los emojis como recursos capaces de transmitir y modular contenido afectivo (Fischer y Herbert, 2021; Riordan, 2017) y sugieren que la diferencia entre odio y *othering* no reside únicamente en cuánto reaccionan las audiencias, sino también en cómo codifican esa reacción.

Esta evidencia ofrece apoyo a la H2, aunque con una cautela interpretativa importante. El repertorio asociado al odio contiene más señales convencionalmente vinculadas con estados de elevada intensidad afectiva, particularmente 🤬 y 😡, y muestra asociaciones de mayor magnitud para varias de ellas. En contraste, el *othering* aparece más estrechamente relacionado con señales convencionalmente vinculadas con aprobación o respaldo, como 👍 y ❤️. Por tanto, el patrón observado es consistente con la expectativa de que el odio genere repertorios de reacción de mayor intensidad afectiva. No obstante, este respaldo debe entenderse en términos composicionales y no como demostración directa de los estados emocionales de los usuarios. La investigación de Tardelli et al. (2025) advierte precisamente que en Telegram el significado pragmático de las reacciones no coincide necesariamente con la emoción prototípicamente representada por el emoji y que determinadas señales positivas pueden funcionar fundamentalmente como indicadores de aprobación, solidaridad o afiliación social.

Esta precaución es particularmente necesaria para emojis como 🙊, 🙄, 🍌, 🤬 o 🗿, cuyo significado puede estar fuertemente condicionado por códigos comunitarios, ironía o usos contextuales. Por ello, los resultados permiten hablar con mayor seguridad de repertorios afectivo-evaluativos diferenciados que de emociones individuales directamente observadas. Esta distinción resulta teóricamente relevante: las reacciones no constituyen simplemente una medición de cuánto gusta o disgusta un contenido, sino un vocabulario paralingüístico

mediante el que las audiencias hacen visibles posiciones relacionales respecto al mensaje y respecto a la comunidad. En este sentido, el predominio de 👍 y ❤️ en el repertorio característico del *othering* puede ser compatible con mecanismos de aprobación y afiliación grupal, mientras que la sobrerrepresentación de señales de ira e impacto en el odio apunta hacia una expresión afectiva comparativamente más intensa. Esta lectura conecta los resultados con la consideración de Telegram como una infraestructura que no solo facilita la difusión de contenidos, sino también procesos de validación, intervención y negociación normativa dentro de las comunidades (Van Raemdonck y Pierson, 2024).

En términos más generales, los hallazgos permiten matizar la relación entre hostilidad y participación dentro del ecosistema español de extrema derecha en Telegram. La literatura previa ha mostrado que la plataforma constituye una infraestructura relevante para la circulación y reproducción de comunidades de extrema derecha (Urman y Katz, 2020; Al-Rawi, 2021; Stewart et al., 2023), así como que sus características sociotécnicas facilitan formas diferenciadas de difusión y conversación. Los resultados de este estudio añaden que esas *affordances* no actúan de manera indiferenciada ante cualquier contenido hostil. El *othering* parece especialmente eficaz para activar mecanismos ligeros de reconocimiento y circulación, mientras que el odio genera proporcionalmente una mayor intervención conversacional y un repertorio de respuestas más intensamente afectivo. Telegram aparece así no solo como una infraestructura para transmitir hostilidad, sino también como un entorno en el que las comunidades pueden redistribuirla, responderla y codificar afectivamente su relación con ella.

Esta distinción constituye una de las principales aportaciones del estudio porque cuestiona una equiparación frecuente entre intensidad de la hostilidad e intensidad del *engagement*. Los resultados indican que una forma discursiva comparativamente menos explícita como el *othering* puede alcanzar una capacidad de difusión y de activación de reacciones superior a la del odio manifiesto, mientras que este último destaca por la intensidad y composición de la respuesta que suscita. La hostilidad digital debe entenderse, por tanto, no únicamente como una propiedad del contenido, sino como un fenómeno relacional que emerge de la interacción entre forma discursiva, arquitectura de plataforma y repertorios participativos de las audiencias.

Conclusiones

Este estudio muestra que el *othering* y el discurso de odio operan como formas diferenciadas de hostilidad dentro del ecosistema español de extrema derecha en Telegram y que sus efectos no son equivalentes en términos de participación. El *othering* se asocia con una mayor difusión mediante reenvíos y con una mayor probabilidad de recibir reacciones, mientras que el odio presenta una relación más intensa con la respuesta conversacional. Estos resultados sugieren que la explicitud de la hostilidad no determina de manera lineal su capacidad de generar *engagement*: formas menos directas de exclusión simbólica pueden resultar especialmente eficaces para circular y recibir reconocimiento, mientras que el odio parece activar con mayor fuerza formas de intervención más explícitas.

Las reacciones mediante emojis permiten además observar que ambas formas de hostilidad generan repertorios afectivo-evaluativos distintos. El *othering* aparece más asociado con señales de aprobación, reconocimiento o afiliación, mientras que el odio presenta un

repertorio más heterogéneo y con una mayor presencia de expresiones convencionalmente vinculadas con ira, impacto o aflicción. Esta diferenciación refuerza la utilidad de analizar los emojis no solo como una métrica adicional de *engagement*, sino como una forma específica de retroalimentación mediante la que las audiencias hacen visibles orientaciones afectivas y evaluativas. Al mismo tiempo, estos resultados deben interpretarse con cautela, ya que el significado de un emoji depende de convenciones comunitarias, usos irónicos y contextos concretos de interacción.

Entre las principales limitaciones del estudio debe señalarse, en primer lugar, que la clasificación automatizada de *othering* y discurso de odio está sujeta a errores de medición inherentes a cualquier sistema supervisado, especialmente en el caso del modelo de odio, entrenado originalmente con textos de Twitter y orientado a categorías específicas de odio. En segundo lugar, el análisis de las reacciones se restringe al periodo posterior a mayo de 2022, por lo que no permite comparaciones homogéneas con etapas anteriores y depende además de que esta funcionalidad estuviera habilitada en los canales. Asimismo, el significado de los emojis no es unívoco y puede variar según el contexto, la ironía o las convenciones de cada comunidad, de modo que sus asociaciones deben interpretarse como patrones agregados de respuesta y no como indicadores directos de emociones individuales. Finalmente, los modelos estiman asociaciones entre hostilidad e interacción, pero no permiten establecer relaciones causales ni generalizar automáticamente los resultados más allá del ecosistema público español de extrema derecha en Telegram.

No obstante, a modo de conclusión, este estudio pretende subrayar la importancia de analizar la hostilidad digital como un fenómeno relacional entre contenido, arquitectura de plataforma y comportamiento de las audiencias. Reenviar, responder y reaccionar constituyen prácticas distintas y, por tanto, ofrecen información complementaria sobre cómo los mensajes hostiles son amplificados, discutidos y validados dentro de comunidades digitales. Futuras investigaciones deberían profundizar en la interpretación contextual de los repertorios emoji, examinar su evolución temporal y combinar análisis automatizados con aproximaciones cualitativas que permitan comprender con mayor precisión cómo se negocian y normalizan estas formas de hostilidad en Telegram.

Referencias

Alorainy, W., Burnap, P., Liu, H., et al. (2018). The enemy among us: Detecting hate speech with threats based 'othering' language embeddings. *arXiv preprint arXiv:1801.07495*.

Al-Rawi, A. (2021). Telegramming hate: Far-right themes on dark social media. *Canadian Journal of Communication*, 46(4), 821–851.

Anjum, & Katarya, R. (2024). Hate speech, toxicity detection in online social media: A recent survey of state of the art and opportunities. *International Journal of Information Security*, 23(1), 577–608. <https://doi.org/10.1007/s10207-023-00755-2>

Assimakopoulos, S., Baider, F. H., & Millar, S. (2017). *Online hate speech in the European Union: A discourse-analytic perspective*. Springer Nature.

Basile, V., Bosco, C., Fersini, E., Nozza, D., Patti, V., Rangel-Pardo, F. M., Rosso, P., & Sanguinetti, M. (2019). SemEval-2019 Task 5: Multilingual detection of hate speech against

immigrants and women in Twitter. In *Proceedings of the 13th International Workshop on Semantic Evaluation* (pp. 54–63). Association for Computational

Bastos, E. (2020). Estratégias discursivas na construção do espírito nacionalista: a Hungria 'de' Viktor Orbán. *REDIS: Revista de Estudos do Discurso*, (9), 12–42. <https://doi.org/10.21747/21833958/red9a1>

Bennett, W. L., & Kneuer, M. (2024). Communication and democratic erosion: The rise of illiberal public spheres. *European Journal of Communication*, 39(2), 177–196.

Buehling, K., & Heft, A. (2023). Pandemic protesters on Telegram: How platform affordances and information ecosystems shape digital counterpublics. *Social Media + Society*, 9(3). <https://doi.org/10.1177/20563051231199430>

Caiani, M., & Parenti, L. (2013). *European and American extreme right groups and the Internet*. Routledge. <https://doi.org/10.4324/9781315580845>

Cervi, L., Tejedor, S., & Villar, M. (2023). Twitting against the enemy: Populist radical right parties discourse against the (political) “other”. *Politics and Governance*, 11(2), 235–248. <https://doi.org/10.17645/pag.v11i2.6391>

Cotarelo, R. (2013). *Ciberpolítica: Las nuevas formas de acción y comunicación políticas*. Tirant lo Blanch.

Curley, C., Siapera, E., & Carthy, J. (2022). Covid-19 protesters and the far right on Telegram: Co-conspirators or accidental bedfellows? *Social Media + Society*, 8(4). <https://doi.org/10.1177/20563051221129187>

DiMaggio, A. R. (2022). Conspiracy theories and the manufacture of dissent: QAnon, the ‘Big Lie’, Covid-19, and the rise of rightwing propaganda. *Critical Sociology*, 48(6), 1025–1048. <https://doi.org/10.1177/08969205211073669>

Donovan, J., Lewis, B., & Friedberg, B. (2018). Parallel ports: Sociotechnical change from the alt-right to alt-tech. In M. Fielitz & N. Thurston (Eds.), *Post-digital cultures of the far right: Online actions and offline consequences in Europe and the US* (pp. 49–66). Transcript Verlag. <https://doi.org/10.1515/9783839446706-004>

Edilashvili, N. (2025). Anti-immigration rhetoric in Donald Trump’s presidential speeches: A corpus-assisted comparative analysis of US presidential discourse (1946–2021) populist and emotionally charged language versus constructive, policy-oriented framing. *Przegląd Politologiczny*, (2), 141–156.

Ermoshina, K., & Musiani, F. (2021). The Telegram ban: How censorship «made in Russia» faces a global Internet. *First Monday*, 26(5). <https://doi.org/10.5210/fm.v26i5.11704>

Evans, S. K., Pearce, K. E., Vitak, J., & Treem, J. W. (2017). Explicating affordances: A conceptual framework for understanding affordances in communication research. *Journal of Computer-Mediated Communication*, 22(1), 35–52. <https://doi.org/10.1111/jcc4.12180>

Fischer, B., & Herbert, C. (2021). Emoji as affective symbols: Affective judgments of emoji, emoticons, and human faces varying in emotional content. *Frontiers in Psychology*, 12, 645173. <https://doi.org/10.3389/fpsyg.2021.645173>

García-Marín, J., Fernández-García, B., & Luengo, O. (2025). Otherness and othering. In A. Nai, M. Grömping, & D. Wirz (Eds.), *Elgar encyclopedia of political communication* (Vol. 3, pp. 131–134). Edward Elgar Publishing. <https://doi.org/10.4337/9781035301447.vol3.00036>

Heslep, D. G., & Berge, P. (2021). Mapping Discord's darkside: Distributed hate networks on Disboard. *New Media & Society*, 26(1), 534–555. <https://doi.org/10.1177/14614448211062548>

Indelicato, M. E., & Magalhães Lopes, M. (2024). Understanding populist far-right anti-immigration and anti-gender stances beyond the paradigm of gender as 'a symbolic glue': Giorgia Meloni's modern motherhood, neo-Catholicism, and reproductive racism. *European Journal of Women's Studies*, 31(1), 6–20. <https://doi.org/10.1177/13505068241230819>

Jost, P., & Dogruel, L. (2023). Radical mobilization in times of crisis: Use and effects of appeals and populist communication features in Telegram channels. *Social Media + Society*, 9(3). <https://doi.org/10.1177/20563051231186372>

Lou, C., Tandoc, E. C., Jr., Hong, L., Pong, X.-Y., Lye, W.-X., & Sng, N.-G. (2021). When motivations meet affordances: News consumption on Telegram. *Journalism Studies*, 22(7), 934–952. <https://doi.org/10.1080/1461670X.2021.1906299>

Machuy, C. (2025). The enemy on the left: The construction of the "evil communist" in Brazilian far-right discourse. In R. A. Novais & R. Christofolletti (Eds.), *The Palgrave handbook on right-wing populism and otherness in global perspective*. Palgrave Macmillan.

Marwick, A., Clancy, B., & Furl, K. (2022). Far-right online radicalization: A review of the literature. *The Bulletin of Technology & Public Life*. <https://doi.org/10.21428/bfcb0bff.e9492a11>

Mazzoleni, G. (2010). *La comunicación política*.

Molla, T. (2024). Racial othering: Structural roots and anti-racist actions from below. *SN Social Sciences*, 4(7). <https://doi.org/10.1007/s43545-023-00806-4>

Mudde, C. (2019). *The far right today*. Polity.

Nartey, M. (2021). Marginality and otherness: The discursive construction of LGBT issues/people in the Ghanaian news media. *Media, Culture & Society*, 44(4), 785–801. <https://doi.org/10.1177/01634437211045552>

Novais, R. A., & Christofolletti, R. (Eds.). (2025). *The Palgrave handbook on right-wing populism and otherness in global perspective*. Palgrave Macmillan. <https://doi.org/10.1007/978-3-031-77868-1>

O'Callaghan, D., Greene, D., Conway, M., Carthy, J., & Cunningham, P. (2013). An analysis of interactions within and between extreme right communities in social media. In M. Atzmüller, A. Chin, D. Helic, & A. Hotho (Eds.), *Ubiquitous social media analysis* (Vol. 8329, pp. 88–107). Springer. https://doi.org/10.1007/978-3-642-45392-2_5

Pérez, J. M., Furman, D. A., Alonso-Alemany, L., & Luque, F. M. (2022). RoBERTuito: A pre-trained language model for social media text in Spanish. In *Proceedings of the*

Thirteenth Language Resources and Evaluation Conference (pp. 7235–7243). European Language Resources Association. <https://aclanthology.org/2022.lrec-1.785/>

Pérez-Díaz, P. L., & Arroyas Langa, E. (2025). El populismo disruptivo de Javier Milei: Análisis de su estrategia discursiva en redes sociales. *European Public & Social Innovation Review*, 10, 1–23. <https://doi.org/10.31637/epsir-2025-2064>

Pirro, A. L. P. (2023). Far right: The significance of an umbrella concept. *Nations and Nationalism*, 29(1), 101–112. <https://doi.org/10.1111/nana.12860>

Pradel, F., Zilinsky, J., Kosmidis, S., & Theocharis, Y. (2024). Toxic speech and limited demand for content moderation on social media. *American Political Science Review*, 118(4), 1895–1912.

Riordan, M. A. (2017). Emojis as tools for emotion work: Communicating affect in text messages. *Journal of Language and Social Psychology*, 36(5), 549–567. <https://doi.org/10.1177/0261927X17704238>

Resende, E., & Freire, C. M. (2025). Populism, enemies and threat construction, and otherness: How populists use othering rhetoric to get reelected to remain in power. The case of Viktor Orbán 2028 election. In R. A. Novais & R. Christofolletti (Eds.), *The Palgrave handbook on right-wing populism and otherness in global perspective*. Palgrave Macmillan. https://doi.org/10.1007/978-3-031-77868-1_12

Rogers, R. (2020). Deplatforming: Following extreme Internet celebrities to Telegram and alternative social media. *European Journal of Communication*, 35(3), 213–229. <https://doi.org/10.1177/0267323120922066>

Ruiz-Pérez, M. A. (2026). Alt-tech in Spain: Mapping the Spanish far-right ecosystem on Telegram. *Observatorio (OBS*)*, 20(SI), 127–148. <https://doi.org/10.15847/OBS20262875>

Salvatore, S., Mannarini, T., Avdi, E., et al. (2019). Globalization, demand of sense and enemization of the other: A psychocultural analysis of European societies' sociopolitical crisis. *Culture & Psychology*, 25(3), 345–374. <https://doi.org/10.1177/1354067X18779056>

Santos, M., Saldaña, M., & Tsyganova, K. (2021). Subversive affordances as a form of digital transnational activism: The case of Telegram's native proxy. *New Media & Society*, 26(1), 131–153. <https://doi.org/10.1177/14614448211054830>

Southern, R., & Harmer, E. (2019). Othering political women: Online misogyny, racism and ableism towards women in public life. In K. Lumsden & E. Harmer (Eds.), *Online othering*. Palgrave Macmillan. https://doi.org/10.1007/978-3-030-12633-9_8

Stephan, W. G., & Stephan, C. W. (2000). An integrated threat theory of prejudice. In S. Oskamp (Ed.), *Reducing prejudice and discrimination* (pp. 23–45). Psychology Press.

Stewart, N. K., Al-Rawi, A., Celestini, C., et al. (2023). Hate influencers' mediation of hate on Telegram: "We declare war against the anti-white system". *Social Media + Society*, 9(2). <https://doi.org/10.1177/20563051231177915>

Svatoňová, E., & Doerr, N. (2024). How anti-gender and gendered imagery translate the Great Replacement conspiracy theory in online far-right platforms. *European Journal of Politics and Gender*, 7(1), 83–101. <https://doi.org/10.1332/25151088Y2023D000000006>

Tardelli, S., Alvisi, L., Cima, L., Cresci, S., & Tesconi, M. (2025). *Emoji reactions on Telegram often reflect social approval over emotional resonance* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2508.06349>

Telegram. (2021, December 30). *Reacciones, spoilers, traducción y códigos QR*. <https://telegram.org/blog/reactions-spoilers-translations/es?ln=a>

Urman, A., & Katz, S. (2020). What they do in the shadows: Examining the far-right networks on Telegram. *Information, Communication & Society*, 25(7), 904–923. <https://doi.org/10.1080/1369118X.2020.1803946>

van Dijk, T. A. (1993). *Elite discourse and racism*. SAGE Publications. <https://doi.org/10.4135/9781483326184>

Van Raemdonck, N., & Pierson, J. (2024). Contentious content on messaging apps: Actualising social affordances for normative processes on Telegram. In H. S. Dunn, M. Ragnedda, M. L. Ruiu, & L. Robinson (Eds.), *The Palgrave handbook of everyday digital life* (pp. 143–167). Palgrave Macmillan. https://doi.org/10.1007/978-3-031-30438-5_9

Van Raemdonck, N., & Pierson, J. (2021). Taxonomy of social network platform affordances for group interactions. In *2021 14th CMI International Conference—Critical ICT Infrastructures and Platforms (CMI)* (pp. 1–8). IEEE.

Vaughan, M., & Heft, A. (2023). Anti-elitism in the European radical right in comparative perspective. *JCMS: Journal of Common Market Studies*, 61, 76–94. <https://doi.org/10.1111/jcms.13347>

Vidgen, B., Thrush, T., Talat, Z., & Kiela, D. (2021). Learning from the worst: Dynamically generated datasets to improve online hate detection. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing* (pp. 1667–1682).

Walther, S., & McCoy, A. (2021). US extremism on Telegram. *Perspectives on Terrorism*, 15(2), 100–124.

Willaert, T., Peeters, S., Seijbel, J., & Van Raemdonck, N. (2022). Redes de desinformación: una investigación cuali-cuantitativa de canales antagónicos de Telegram en neerlandés. *First Monday*, 27(5). <https://doi.org/10.5210/fm.v27i5.12533>

Winberg, O. (2017). Insult politics: Donald Trump, right-wing populism, and incendiary language. *European Journal of American Studies*, 12(12-2).

Wodak, R. (2021). *The politics of fear*. SAGE Publications.

Young, J. (2025). Dogwhistles, discrimination, humour and the law: Regulating implicit messaging. *Open Library of Humanities*, 11(2). <https://doi.org/10.16995/olh.19789>

Zamri, N. A. K., Mohamad Nasir, N. A., Hassim, M. N., et al. (2023). Digital hate speech and othering: The construction of hate speech from Malaysian perspectives. *Cogent Arts & Humanities*, 10(1). <https://doi.org/10.1080/23311983.2023.2229089>