

Hatedom multiplataforma: estrategias de los influencers para gestionar la aversión afectiva

Cross-platform hatedom: influencers' strategies for managing affective aversion

Hatedom multi-plataforma: estratéгias dos influencers para gerir a aversão afectiva

Arantxa Vizcaíno-Verdú, Universidad Internacional de la Rioja, Logroño, España
(arantxa.vizcaino@unir.net)

Paloma Contreras-Pulido, Universidad Internacional de la Rioja, Logroño, España
(paloma.contreras@unir.net)

RESUMEN | Ante el crecimiento de comportamientos tóxicos en las comunidades de fans en línea, este estudio examina la construcción del *hatedom* en múltiples plataformas de redes sociales y las estrategias que los influencers emplean para navegar en entornos digitales hostiles. Mediante un enfoque de métodos mixtos, la investigación analiza 16.215 comentarios dirigidos a seis influencers, utilizando técnicas avanzadas de procesamiento de lenguaje natural y modelos de inteligencia artificial preentrenados. Paralelamente, por medio de entrevistas cualitativas estructuradas, se obtuvo información sobre las percepciones de estos influencers respecto de sus interacciones con seguidores y detractores, presentando las estrategias empleadas para gestionar el odio recibido. Los hallazgos revelan que el *hatedom*, caracterizado por una fuerte aversión afectiva, presenta variaciones significativas en sus niveles de toxicidad y polarización emocional en diferentes plataformas, como YouTube, X, Facebook, Instagram, TikTok y Twitch. Los influencers adoptan diversas estrategias de resistencia, entre las que destacan ignorar interacciones negativas, bloquear a usuarios malintencionados y utilizar el sarcasmo como mecanismo defensivo. Asimismo, subrayan la importancia de proteger su salud mental, limitando su exposición digital para salvaguardar su bienestar. Este estudio resalta la necesidad de un enfoque integral y multifacético para comprender el fenómeno del *hatedom*, combinando el análisis de la polarización de las comunidades digitales con la implementación de herramientas tecnológicas avanzadas para la moderación. Al priorizar la salud mental y la resiliencia de los influencers, es posible promover un entorno digital más saludable en una evolución constante del comportamiento de los fans en redes sociales.

PALABRAS CLAVE: *fandom*, discurso de odio, plataforma, redes sociales, toxicidad, seguidores, comunidad en línea, *influencers*

FORMA DE CITAR

Vizcaíno-Verdú, A. & Contreras-Pulido, P. (2025). Cross-platform *hatedom*: influencers' strategies for managing affective aversion. *Cuadernos.info*, (60), 68-92. <https://doi.org/10.7764/cdi.60.85872>

ABSTRACT | *Amidst the rise of toxic behavior in online fan communities, this study examines the construction of cross-platform hatedom through social media and the strategies influencers use to navigate hostile digital environments. Using a mixed methods approach, this study analyzes 16,215 comments directed at six influencers using advanced natural language processing techniques and pre-trained AI models. In addition, structured qualitative interviews provided insights into the influencers' perceptions of their interactions with fans and haters and shed light on their strategies for dealing with hatedom. The results show that hatedoms, characterized by strong affective dislike, vary significantly in their levels of toxicity and emotional polarization on different platforms, including YouTube, X, Facebook, Instagram, TikTok and Twitch. Influencers employ various coping mechanisms, such as ignoring negative interactions, blocking malicious users, and using sarcasm as a defense mechanism. They also emphasize the importance of mental health and often limit their personal engagement to protect their wellbeing. This study highlights the need for a comprehensive, multi-faceted approach to understanding hatedom that combines insights into the polarization of the digital community with the implementation of advanced technological moderation. By prioritizing the mental health and resilience of influencers, it is possible to foster a healthier digital environment amidst the ever-evolving dynamics of fan behavior on social media.*

KEYWORDS: *fandom, hate speech, platform, social media, toxicity, followers, online community, influencers*

RESUMO | Perante o aumento de comportamentos tóxicos nas comunidades de fãs online, este estudo examina a construção do hatedom em múltiplas plataformas de redes sociais e as estratégias que os influencers utilizam para navegar em ambientes digitais hostis. Utilizando uma abordagem de métodos mistos, a pesquisa analisa 16.215 comentários dirigidos a seis influencers, utilizando técnicas avançadas de processamento de linguagem natural e modelos de inteligência artificial pré-treinados. Paralelamente, foram utilizadas entrevistas qualitativas estruturadas para obter informações sobre as percepções destes influencers em relação às suas interações com seguidores e detractores, apresentando as estratégias utilizadas para gerenciar o ódio recebido. Os resultados revelam que o hatedom, caracterizado por uma forte aversão afectiva, apresenta variações significativas nos seus níveis de toxicidade e polarização emocional em diferentes plataformas, incluindo YouTube, X, Facebook, Instagram, TikTok e Twitch. Os influencers adoptam uma variedade de estratégias de resistência, incluindo ignorar interações negativas, bloquear utilizadores maliciosos e usar o sarcasmo como mecanismo de defesa. Salientam também a importância de proteger a sua saúde mental, limitando a sua exposição digital para salvaguardar o seu bem-estar. Este estudo realça a necessidade de uma abordagem holística e multifacetada para compreender o fenômeno do hatedom, combinando a análise da polarização das comunidades digitais com a implementação de ferramentas tecnológicas avançadas de moderação. Ao dar prioridade à saúde mental e à resiliência dos influencers, é possível promover um ambiente digital mais saudável em uma constante evolução do comportamento dos fãs nas redes sociais.

PALAVRAS CHAVE: *fandom, discurso de ódio, plataforma, redes sociais, toxicidade, seguidores, comunidade online, influencers*

INTRODUCCIÓN Y MARCO TEÓRICO

El Oxford English Dictionary seleccionó el término tóxico como la palabra del año 2018, reflejando la creciente preocupación en torno a las actitudes y comportamientos nocivos en contextos sociales, especialmente en las relaciones en línea (Schuessler, 2018). Este fenómeno ha sido analizado por diversos autores, entre los que destaca Arouh (2020), quien exploraba el uso y las implicancias del concepto en la cultura contemporánea, específicamente en el *fandom*. En estas comunidades, las conductas tóxicas suelen manifestarse mediante críticas o prejuicios que pueden intensificarse hasta convertirse en formas más graves de agresión, llegando a propagar discursos de odio (Díaz et al., 2019).

Los discursos de odio se consideran uno de los problemas más apremiantes de la sociedad actual. En un reciente informe de las Naciones Unidas se destacaba cómo los algoritmos de las redes sociales pueden amplificar este tipo de situaciones digitales (OHCHR, 2023) priorizando contenidos que generan respuestas emocionales intensificadas como, por ejemplo, la indignación, la sorpresa o la ira (Maarouf et al., 2024). Si bien los discursos de odio conllevan actos hostiles para dañar o violentar individuos con base en características como la etnia, el género o la religión, se ha observado que los mecanismos algorítmicos extienden su influencia a las comunidades de fans, manifestándose en formas más diversas impulsadas por una interacción dinámica entre hostilidad, validación social y polarización grupal (Riemer & Peter, 2021).

Las comunidades de fans son particularmente susceptibles a estas dinámicas debido al intenso compromiso emocional que define sus interacciones sociales. Se ha evidenciado que los seguidores muestran un alto grado de inversión afectiva (Théberge, 2006) y se adhieren a normas autoritarias (Luo & Li, 2024), proyectando una profunda devoción emocional hacia el *fandom*. Sin embargo, este entusiasmo no siempre se manifiesta de manera positiva. Como indica Jenkins (1992), la misma intensidad que impulsa el afecto hacia un *fandom* puede derivar en comportamientos intensos o tóxicos. Barnes (2022) ilustra cómo esta intensidad afectiva puede provocar un cambio repentino desde la admiración al odio, especialmente cuando figuras públicas o creadores de contenido son percibidos como desertores.

Esta dinámica se ha observado con la aparición de los anti-fans (Duffy et al., 2022), quienes adoptan una postura antagonista hacia algunos creadores, motivados con frecuencia por agravios percibidos de índole moral o social (Aburime, 2022). Los anti-fans participan en interacciones nocivas caracterizadas por la obsesión hacia un *influencer* o celebridad en línea. Vizcaíno-Verdú y sus colegas (2020) definían esta relación como *fanbullying*, definiéndola como un vínculo inherentemente abusivo entre fans y figuras públicas. Esta relación implica que los roles de los personajes ficticios dejan de ser percibidos como tales y el odio de la comunidad se

dirige en su lugar hacia actores, actrices e *influencers*. En este sentido, Valenzuela-García y sus colegas (2023) señalaban que los *influencers* no suelen considerarse víctimas de este tipo de odio ya que, a diferencia de otras personas –como individuos anónimos afectados por el ciberacoso–, no son percibidos como tales debido a su estatus público y popular, lo que dificulta que generen empatía y aprecio.

Según Ouvrein y sus colegas (2023), los *influencers* suscitan sentimientos de envidia y desprecio, emociones que generan actitudes de menosprecio y ciberacoso hacia ellos. Valenzuela-García y sus colegas (2023) concluían que los agresores expresan una aversión hacia el estilo de vida de los *influencers*, al percibirlo como injusto y carente de esfuerzo para alcanzar tales privilegios. Esta percepción de incompetencia o injusticia se ha asociado con la creencia de que su riqueza se obtiene mediante la comercialización de su vida.

Asimismo, se ha determinado que las plataformas digitales facilitan este odio debido a la desinhibición (Suler, 2006) que difumina la línea entre la ficción y la realidad (Vizcaíno-Verdú et al., 2020). A ello se suman los efectos de los algoritmos sobre la libertad de expresión en las redes sociales (Riemer & Peter, 2021), donde amplifican o suprimen el discurso para optimizar la difusión de contenido, lo que distorsiona el intercambio justo y libre de ideas en conversaciones públicas (Martínez-Valerio & Mayagoitia-Soria, 2021; Riemer & Peter, 2021).

Se ha documentado que la opacidad algorítmica desplaza la cuestión de la libertad de expresión desde el qué puede decirse hacia el qué será escuchado y por quién (Riemer & Peter, 2021). Como consecuencia, los *influencers* recurren frecuentemente a la autocensura. En los últimos años, las figuras públicas expresaban sus opiniones políticas o ideológicas a través de redes sociales. Sin embargo, estudios recientes, entre ellos el de Martínez-Valerio y Mayagoitia-Soria (2021), sugieren una disminución en esta práctica debido a la prevalencia del ciberacoso y a la necesidad de proteger la reputación de la marca personal (Von-Mettenheim & Wiedmann, 2022).

En este contexto, los *influencers* se ven forzados a comunicarse con sus seguidores mediante estrategias efectivas y adaptadas a diferentes plataformas desarrollando, en la mayoría de los casos, tácticas para gestionar el discurso y hacer frente al ciberacoso (Abidin, 2019). Entre estas se incluyen los sentimientos de culpa o la vergüenza ante mensajes recibidos –los que suelen permanecer ocultos debido a bloqueos, filtros o eliminaciones previas–, la falta de denuncia o la aceptación silenciosa de actos potencialmente delictivos, la autocensura en determinados temas, o el abandono de plataformas donde experimentan mayores ataques (Martínez-Valerio & Mayagoitia-Soria, 2021).

Según el estudio de Walther (2022), las plataformas que permiten retroalimentación de los usuarios (por ejemplo, Me gusta en Facebook, favoritos en X (Twitter) o upvotes en Reddit) refuerzan el comportamiento de los receptores. Estas interacciones pueden desencadenar tres procesos fundamentales: 1) incentivar la hostilidad de individuos para publicar más mensajes de odio, 2) intensificar el odio hacia sus objetivos, puesto que las personas tienden a defender públicamente sus posturas cuando reciben retroalimentación positiva, y 3) desarrollar relaciones más profundas que refuercen los mensajes tóxicos mediante interacciones continuas con individuos de pensamiento afín, inclusive si estas relaciones se dan en entornos puramente virtuales (Walther, 2022).

Ante este panorama, se ha observado un creciente interés académico en el estudio de las relaciones entre fans, seguidores e *influencers*, especialmente con la proliferación de plataformas y las oportunidades de interacción y participación comunitaria. Un problema urgente que requiere de atención es el auge del *fandom* tóxico. Mientras que estudios recientes han analizado los efectos personales, físicos y psicológicos de este tipo de odio en personalidades del *fandom* (Walther, 2022; Martínez-Valerio & Mayagoitia-Soria, 2021; Vizcaíno-Verdú et al., 2020), este estudio se centra en cómo los anti-fans construyen el *hatedom* y cómo los *influencers* lo gestionan en las plataformas digitales. Los objetivos específicos que guiaron este enfoque son: O1) identificar la polaridad y toxicidad de los comentarios dirigidos a *influencers* en diferentes plataformas, O2) explorar el rol y la relevancia de las comunidades de fans en el desempeño laboral de los *influencers*, O3) analizar la interacción entre *influencers*, fans y *haters*, e O4) investigar las estrategias empleadas por los *influencers* para gestionar el odio en redes sociales.

MÉTODO

Se empleó un enfoque metodológico mixto que integró técnicas cuantitativas y cualitativas. En primer lugar, se empleó procesamiento de lenguaje natural (PLN) y modelos de inteligencia artificial preentrenados para explorar la probabilidad de negatividad y toxicidad en los comentarios presentes en los perfiles de redes sociales de seis *influencers*. Posteriormente, se llevaron a cabo entrevistas con estos perfiles para profundizar en sus percepciones y estrategias para gestionar interacciones hostiles en sus comunidades.

Modelos de PLN para analizar la polaridad y toxicidad en los comentarios

En la primera fase del análisis se examinó la polaridad y la toxicidad de los comentarios recibidos por los *influencers*. Los comentarios fueron recopilados mediante técnicas de *web scraping* (Mysiuk & Shuvar, 2023) y los datos se extrajeron de manera sistemática en cada perfil (véase la sección sobre la muestra de *influencers*). Se seleccionó un conjunto amplio de plataformas a fin de garantizar una muestra

de comentarios aleatoria y representativa para el estudio. Específicamente, para cada *influencer* se seleccionaron las tres publicaciones más recientes de mayo de 2024 en todos sus perfiles activos (tabla 2), recopilando un total de 7236 comentarios de Instagram, 5131 de YouTube, 4013 de TikTok, 1664 de X, 1380 de Facebook y 1072 de Twitch.

Antes de realizar el análisis, los comentarios fueron preprocesados con la biblioteca Pandas para corregir posibles errores de codificación y filtración de datos. La función de filtrado se diseñó para eliminar emojis, caracteres no alfanuméricos, texto posterior al símbolo @ (para garantizar el anonimato de los usuarios que comentaban o etiquetaban a otros usuarios), enlaces (spam) y líneas en blanco. No obstante, se conservaron caracteres especiales y relevantes del español, como acentos (por ejemplo, á), la letra ñ y los signos de puntuación. La muestra final incluyó 5341 comentarios de Instagram, 4782 de YouTube, 3031 de TikTok, 1472 de X, 1033 de Twitch y 556 de Facebook, con un cómputo total de 16.215 unidades de análisis.

Para detectar y clasificar la polaridad y la toxicidad de los comentarios se utilizaron dos modelos de PLN: 1) BETO para el análisis de sentimiento y 2) Multilingual BERT para la detección de la toxicidad. Ambos se basan en la técnica de preentrenamiento *Bidirectional Encoder Representations from Transformers* (BERT), conocida por su capacidad para comprender el contexto de una palabra en todas sus posiciones dentro de una oración (Devlin & Chang, 2018). La selección de los modelos se fundamentó en los siguientes factores:

1. BETO. Especializado en español, este modelo ofrece alta precisión en la clasificación de emociones (negativa, neutra y positiva), fundamental para interpretar emociones en comentarios dirigidos a *influencers* hispanohablantes.
2. Multilingual BERT. Su capacidad multilingüe y entrenamiento especializado en la detección de la toxicidad lo convierten en una herramienta sólida para identificar y clasificar lenguaje ofensivo y agresivo.
3. Arquitectura *Transformer*. Ambos modelos se sostienen sobre arquitectura *Transformer*, lo que les permite captar relaciones contextuales complejas y mejorar la detección de matices tanto en el análisis de sentimiento como en la identificación de la toxicidad.

La implementación técnica del análisis de sentimiento y toxicidad se llevó a cabo en diversas etapas. En primer lugar, se cargaron los modelos preentrenados junto con sus respectivos tokenizadores en Google Colab empleando la biblioteca *Transformers*. El tokenizador de BETO convirtió los comentarios en secuencias de tokens procesables, lo que permitió ejecutar el análisis de sentimiento. De manera similar, el tokenizador de Multilingual BERT segmentó los comentarios para la exploración de toxicidad.

Objetivos	Marco teórico	Preguntas abiertas
O1. Explorar el rol y relevancia de la comunidad de fans para la profesión de los influencers.	(Valenzuela et al., 2023; Abidin, 2019; Vizcaíno-Verdú et al., 2020)	<p>(1) ¿Qué opinas sobre tus seguidores de redes sociales? ¿Qué representan para ti?</p> <p>(2) ¿Cuál es tu perspectiva sobre el rol que desempeñan los seguidores en tu profesión de influencer?</p>
O2. Examinar cómo interactúan los influencers con la comunidad de fans en múltiples redes sociales.	(Aburrime, 2022; Abidin, 2019; Riemer & Peter, 2021; Vizcaíno-Verdú et al., 2020; Walther, 2022)	<p>(3) ¿Has notado alguna diferencia significativa entre las comunidades de seguidores que tienes en diferentes plataformas? ¿En qué aspectos varían?</p> <p>(4) ¿Consideras que tu comunidad de seguidores se caracteriza por el respeto y la positividad, o has experimentado algún tipo de controversia o conflicto en tus perfiles?</p> <p>(5) ¿Cómo gestionas los comentarios negativos o las críticas constructivas que recibes en tus publicaciones? ¿Has tenido que lidiar con situaciones de odio en alguna ocasión?</p> <p>(6) ¿Qué medidas tomas para fomentar un ambiente respetuoso y positivo entre tu comunidad de seguidores?</p>
O3. Profundizar en las estrategias que utilizan los influencers para gestionar el hatedom en redes sociales.	(Arouh, 2020; Abidin, 2019; Martínez-Valerio & Mayagoitia-Soria, 2021; Valenzuela et al., 2023)	<p>(7) ¿Cuál es tu enfoque al interactuar con tus seguidores? ¿Intentas responder a todos los comentarios y mensajes que recibes o priorizas ciertos tipos de interacciones?</p> <p>(8) ¿Crees que es importante establecer límites claros en tu relación con los seguidores en línea? ¿Cómo mantienes un equilibrio saludable entre tu vida personal y tu presencia en redes sociales?</p> <p>(9) ¿Has presenciado o sido testigo de casos en los que tus seguidores se enfrenten entre sí debido a diferencias de opinión o lealtades a otros creadores? ¿Cómo gestionas este tipo de situaciones?</p> <p>(10) ¿Has experimentado cambios en la dinámica de tu comunidad a lo largo del tiempo? ¿Ha habido momentos en los que sentiste que perdiste o ganaste seguidores debido a ciertas acciones o declaraciones?</p>

Tabla 1. Entrevistas estructuradas

Fuente: Elaboración propia.

Se generaron matrices *logits* para cada comentario, representando la probabilidad de que perteneciera a diferentes categorías de sentimiento y toxicidad. Para el análisis de sentimiento, las categorías incluyeron positivo, neutro y negativo, mientras que, en la clasificación de la toxicidad, se consideraron tóxico, insulto, amenaza y obscenidad. Por ejemplo, si un comentario obtenía *logits* de 2,0, 1,0 y 0,1 para las categorías de sentimiento (positivo, neutro y negativo, respectivamente),

la función *softmax* transformaba los valores en probabilidades de 0,65, 0,24 y 0,11. Esto expresaba, en un 65% de probabilidad, que el comentario fuera positivo, un 24% que fuera neutro y un 11% que fuera negativo. Igualmente, en el análisis de toxicidad, los *logits* se convertían en probabilidades que aludían a la naturaleza del comentario. Por ejemplo, un comentario podía presentar un 45% de probabilidad de ser considerado tóxico, un 30% de contener un insulto, un 20% de expresar una amenaza y un 5% de incluir obscenidades.

Los resultados fueron almacenados en un *DataFrame* y exportados a una hoja de cálculo para su interpretación, a fin de obtener una visión integral de las interacciones hostiles en los perfiles de los *influencers*.

Entrevistas estructuradas a *influencers*

A continuación, se enfocó la atención en las perspectivas de los *influencers*, llevando a cabo entrevistas estructuradas entre junio de 2023 y febrero de 2024 para profundizar en sus relatos en torno a sus experiencias con el *hatedom* (Sjoerd-Jeroen et al., 2024).

Las entrevistas siguieron un modelo estructurado con diez preguntas abiertas, diseñadas en concordancia con los objetivos de la investigación e inspiradas en los estudios de Valenzuela-García y sus colegas (2023), Martínez-Valerio y Mayagoitia-Soria (2021), Walther (2022) y Vizcaíno-Verdú et al. (2020). Este diseño estandarizado (tabla 1) se empleó para minimizar la duración de las entrevistas, ya que estos profesionales mediáticos, debido a sus agendas, cuentan con escaso tiempo para participar en interacciones más extensas con las investigadoras.

Muestra de *influencers*

Debido a la alta demanda de los *influencers* y su reducida tasa de respuesta (Senft, 2013), se invitó a 22 perfiles españoles a participar en el estudio para alcanzar una tasa de respuesta suficiente para obtener datos cualitativos significativos (Creswell, 2013). La selección se llevó a cabo mediante un muestreo aleatorio no probabilístico, considerando el contexto cultural y operativo de las investigadoras y ajustándose a la actividad y al reconocimiento de los *influencers* en diferentes plataformas, caracterizadas fundamentalmente por dinámicas comunitarias (Abidin, 2018).

Seis *influencers* aceptaron la invitación. Dados sus compromisos profesionales y personales, así como la sensibilidad del tema –relativo al odio y los conflictos con sus comunidades–, se garantizó el anonimato para proteger su profesión y reputación digital. La tabla 2 entrega información detallada sobre el contenido temático de cada *influencer*, las plataformas que utiliza y su número de seguidores (hasta junio de 2024).

Pseudónimo del influencer	Tema principal	Plataforma	Seguidores
INFLU1	Arte y belleza	TikTok	398.000
		Instagram	162.000
		YouTube	545.000
INFLU2	Humor	TikTok	7.949.000
		Instagram	207.000
		TikTok	1,4 millones
INFLU3	Estilo de vida, canto y baile	Facebook	697.000
		Instagram	260.000
		YouTube	14.600
INFLU4	Humor	X	915
		TikTok	397.600
		Instagram	276.000
INFLU5	Humor	X	27.500
		YouTube	675
		TikTok	2,4 millones
INFLU6	Divulgación científica	YouTube	922.000
		Instagram	378.000
		Twitch	5.786
INFLU6	Divulgación científica	YouTube	786.000
		Instagram	331.000
		X	256.756
INFLU6	Divulgación científica	TikTok	162.400
		Twitch	61.741

Tabla 2. Muestra de influencers

Fuente: Elaboración propia.

Análisis de las transcripciones mediante teoría fundamentada

Tras la transcripción de las entrevistas con el apoyo automatizado de Microsoft Office y la posterior filtración manual de los argumentos para garantizar su claridad, se tradujeron del español al inglés para el análisis cualitativo mediante el software Atlas.ti 24. La teoría fundamentada (*grounded theory*) de Strauss y Corbin (1997) guio el análisis inductivo-deductivo a lo largo de las siguientes fases: 1) codificación abierta, 2) codificación axial, y 3) construcción de memos.

En la fase de codificación abierta, las transcripciones se segmentaron en unidades de texto con significado, asignando a cada una un código descriptivo. Mediante este proceso se identificaron patrones y conceptos iniciales, tales como a) apoyo laboral, b) comentarios negativos, c) interacción personal, d) estrategias de gestión, y e) entorno positivo.

La codificación axial consistió en agrupar los códigos en categorías más amplias para establecer relaciones y estructurar los datos hacia la exploración de conexiones conceptuales. Los códigos axiales incluyeron a) la relación con la comunidad de seguidores, b) la gestión de comentarios negativos, c) las diferencias entre plataformas, y d) el fomento de un entorno respetuoso.

En la fase de memos se identificaron categorías clave de acuerdo con el marco y los objetivos de la investigación para llevar a cabo la interpretación narrativa y comprender la dinámica comunitaria entre *influencers*, fans y *haters*. Se extrajeron tres memos fundamentales: 1) relaciones con comunidades de seguidores, 2) diferencias en la interacción con comunidades entre plataformas, y 3) estrategias para gestionar comentarios hostiles y conflictos de odio dentro de una comunidad.

En cumplimiento con los principios éticos de la investigación y con el objetivo de garantizar la protección de los derechos humanos y civiles conforme al Código Ético de la Association of Internet Researchers (2012), tanto los *influencers* como los usuarios que comentaban en su contenido fueron anonimizados por la naturaleza sensible de los hallazgos.

RESULTADOS Y DISCUSIÓN

Objetivo 1. Polaridad y toxicidad en los perfiles de los *influencers*

Comenzando por el análisis computacional para comprender el contexto comunitario de los *influencers*, especialmente en lo que concierne a la polaridad y a la toxicidad, se identificaron diferencias significativas en la recepción de los comentarios en función del perfil (figura 3). INFLU1, cuyo contenido se centra en el arte y la belleza, recibía principalmente comentarios neutros (59%), con una proporción considerablemente baja de comentarios negativos (8%) y tóxicos (6%). INFLU2, dedicado al humor, registró un 24% de comentarios negativos y una toxicidad global de 11%. En este caso, se observó que los comentarios contenían con frecuencia humor negro, sarcasmo o ironía, lo que contribuía a una mayor percepción de toxicidad. Algunos ejemplos con alto porcentaje de negatividad y toxicidad fueron: “Menudo grupo de personas sin vida angustiadas por todo”, “Todo el mundo le guarda rencor al vecino, el mío es medio pervertido y drogadicto” y “¿Por qué tenéis vidas tan aburridas y perezosas como para hablar con vuestro vecino sobre vuestras aburridas vidas?”.

INFLU3, por otro lado, recibía un 5% de comentarios negativos y 11% de toxicidad. En este caso, las críticas estaban relacionadas con su orientación sexual –“¿Eres gay? Qué desperdicio, eres tan guapo”– o su estilo de vida –“¿Qué haces viviendo en el asqueroso Londres? Tienes una casa espectacular”–. INFLU4, cuyo contenido se enfoca en el humor y en los monólogos, registró un 24% de comentarios negativos con un nivel de toxicidad de 9%. Aquí, los comentarios introducían expresiones como “Increíble cómo algunas personas son tan pelotas” o “Que vuelva el Caudillo, qué asco”, en referencia a Francisco Franco, quien gobernó España como dictador entre 1939 y 1975, expresando nostalgia por el régimen autoritario.

Se observó que INFLU5 también presentaba un nivel moderado de negatividad (21%) y toxicidad (9%). Las críticas estaban dirigidas hacia la influencer tanto en aspectos personales como en relación con su contenido – “Hija desagradecida” y “Nadie quiere verte”–, así como en términos de género –“Qué difícil es entender a las mujeres” y “Muchos hombres están preocupados por el ahorro y el futuro, y les parece una estupidez gastar dinero en cosas innecesarias”–.

En relación con INFLU6, cuyo enfoque se orientaba hacia la divulgación científica, se registró el mayor porcentaje de comentarios negativos (50%) y toxicidad (16%). Las expresiones tóxicas presentaron niveles mínimos de lenguaje soez (3%) e insultos (2%), poniendo de manifiesto la controversia y polarización inherentes al contenido (figura 1).

En cuanto a la polarización y toxicidad en diferentes plataformas, se observó que en X y YouTube el porcentaje de comentarios negativos era superior (55,5% y 46%, respectivamente) en comparación con Instagram (24,6%), Twitch (19,9%), TikTok (11,2%) y Facebook (1,9%). En Facebook (81,4%) y TikTok (62,6%) predominaban los comentarios positivos, reflejando una comunidad más respetuosa. Los comentarios neutros eran más frecuentes en Twitch (65,2%) e Instagram (47,8%) (figura 2).

Asimismo, los valores de toxicidad variaban significativamente entre plataformas. X y YouTube registraron la mayor proporción de comentarios tóxicos, con 15,8% y 15,7% respectivamente. En contraste, Facebook, Instagram, TikTok y Twitch presentaban puntuaciones inferiores, con 12,1%, 11,3%, 10,5% y 8,7%. El uso de lenguaje obsceno se observó con mayor frecuencia en YouTube (3,8%) y mínimamente en Facebook (1%). Tanto las amenazas como los insultos también se mantenían en niveles consistentemente bajos en todas las plataformas, con insultos en un rango de 0,3% a 1,9%, y amenazas por debajo de 0,1%.

El alto porcentaje de comentarios negativos en X y YouTube sugiere la existencia de una base de usuarios más crítica o conflictiva, probablemente influenciada por la comunicación abierta y en muchos casos anónima, que caracteriza a estas plataformas.

Ataque personal

 *Eres fan de ese tipo o qué. O simplemente te cae mal ella y críticas todo lo que hace.*
Te hartas a follar, ¿eeh?
¿Qué es peor? ¿Estafar o copiar el contenido que a otros les ha llevado meses de investigación?

Racismo

 *Pro colonialistas no son capaces de admitir el genocidio que cometió el gobierno español en América.*
Veo a colombianos y argentinos metidos en esta mierda.
La mayoría de los ignorantes que siguen a este son de Latinoamérica. Da qué pensar.

Sexismo/ misoginia

 *Si fuese mujer todos dirían que es un ejemplo de empoderamiento. He visto los vídeos del tipo y no son para tanto [...] no le vi nada diabólico.*
Las mujeres no entienden la mentalidad masculina.
Deberían investigar al feminismo. A eso decirle secta es quedarse muy corto.

Ideología

 *Preocúpate por tus mierdas. Aunque viendo esto ya veo cuáles son tú intenciones, llegar a gente progre que piensa que hacer eso es de sectario y de no tener razón.*
La ideología de género no tiene respaldo científico y menos sentido común. Qué dices a eso.

Figura 1. Comentarios negativos y tóxicos hacia INFLU6.

Fuente: Elaboración propia.

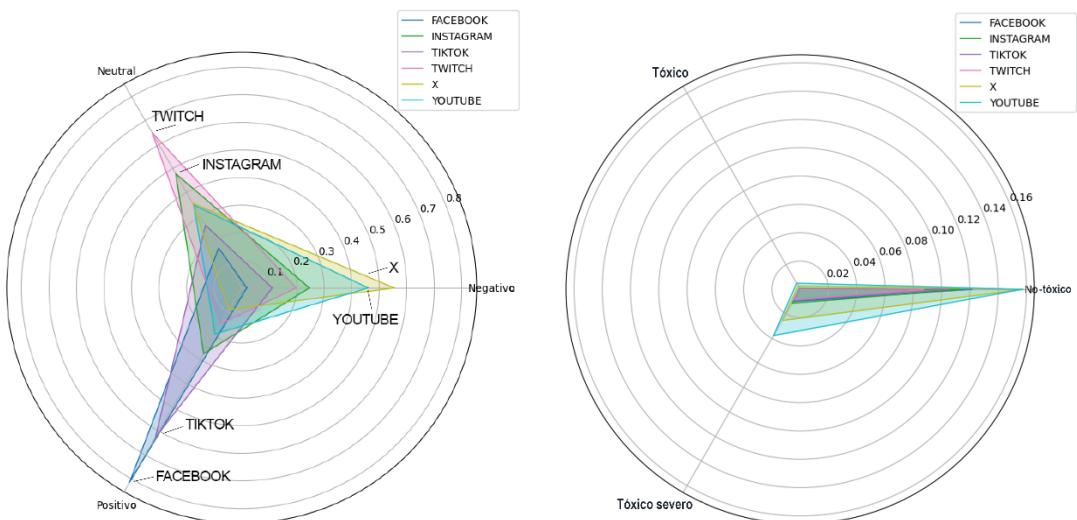


Figura 2. Polaridad y toxicidad multiplataforma entre influencers

Fuente: Elaboración propia.

En contraste, la predominancia de comentarios positivos en TikTok, Facebook e Instagram sugiere que estos espacios propician un entorno comunitario más

positivo y cooperativo en los casos analizados. En los casos de Twitch y TikTok, los niveles de toxicidad más bajos (en términos de lenguaje abusivo, palabras soeces y amenazas) podrían atribuirse a políticas de moderación de contenido y directrices comunitarias más eficaces para reducir interacciones y comportamientos nocivos.

Objetivo 2. Relevancia de las comunidades de fans para los *influencers*

A continuación, se exploraron las percepciones de los *influencers* en su relación con sus seguidores, considerando las P1 y P2. En una constante, las celebridades digitales percibían a sus seguidores como una fuente de apoyo, principalmente en términos de validación emocional y retroalimentación constructiva.

Para mí, ellos hacen que mi contenido sea valioso y se aprecie porque me brindan una retroalimentación constante y una motivación para seguir mejorando (P1-INFLU2).

Son un apoyo fundamental para mí (P1-INFLU4).

Según estas figuras, la comunidad es un pilar esencial para la sostenibilidad de su profesión, desempeñando un papel cocreativo en la generación de contenido. En este sentido, su proyecto colaborativo resalta el rol central de los seguidores, específicamente de los fans. Los entrevistados reconocieron que, en su ausencia, el contenido perdería impacto y relevancia, sugiriendo que su éxito depende de una relación positiva. Esta dependencia se evidenció en declaraciones como las que siguen:

Lo significan todo para mí. Sin ellos, no sería nada. La humildad es crucial. Un creador de contenidos sin seguidores no es nada (P2-INFLU1).

Los seguidores son esenciales. Si no hubiera una audiencia interesada en mi trabajo, no lo habría conseguido profesionalmente (P2-INFLU6).

La relación entre el *influencer* y sus seguidores implicaba, asimismo, un componente económico básico. Estos no solo respaldan el contenido o participan en su creación, sino que se convierten en el segmento clave de las marcas que buscan colaboraciones sostenibles y rentables. Esta dinámica convierte a los seguidores en una fuente indirecta de ingresos, esencial para la viabilidad financiera del *influencer*. Dicho vínculo, tanto a nivel personal como lucrativo, se destaca en los ejemplos siguientes:

Los seguidores representan dos cosas claras: una conexión real y un motor económico. [...] De alguna manera, no dejas de “trabajar para ellos”. Si no les gustan tus videos, ese mes no comes. Así que es una relación bastante intensa que hay que aprender a gestionar (P2-INFLU3).

Si no tuviera una comunidad de seguidores comprometidos que interactuaran con mis publicaciones sobre colaboraciones remuneradas con marcas, las marcas no me contratarían (P2-INFLU4).

Los fans también brindan un apoyo significativo a los *influencers* en lo que a bienestar emocional se refiere. La expresión de afecto puede ejercer un impacto positivo en su salud mental. Esto es crucial en un entorno digital en el que la presión pública y el escrutinio pueden volverse obsesivos (Théberge, 2006), y contribuye con frecuencia al fortalecimiento de la autoestima de estos creadores digitales. Algunos señalaban que:

Siempre me dan ánimos y me ayudan cuando comparto mis problemas (P1-INFLU5).

Las historias que me cuentan y mi propio carácter emocional hacen que mi relación con ellos sea no solo cercana, sino también sincera; me he emocionado al leer comentarios y disfruto hablando con ellos, por lo que cuando creo contenido, realmente quiero hacerles sonreír y que obtengan algo al consumir mi contenido (P2-INFLU3).

Como apuntaba Jenkins (1992), la relación entre *influencers*, seguidores y fans es multidimensional y altamente interdependiente. Esta comunidad proporciona una validación significativa, una retroalimentación constructiva y apoyo emocional, aspectos fundamentales para la motivación de los *influencers* y la continuidad de su labor. No obstante, dicha relación genera discrepancias al analizar esta dinámica si se atiende a las diferencias entre redes sociales.

Objetivo 3. Diferencias comunitarias entre plataformas de redes sociales

Las diferentes experiencias de los *influencers* con respecto a la demografía y el comportamiento de los usuarios en las plataformas resultaron notables entre la P2 y la P4. Por ejemplo, en YouTube y X suelen aparecer más comentarios negativos, generalmente derivados de usuarios no seguidores. En contraste, TikTok promueve una relación más respetuosa y juvenil e Instagram genera opiniones mixtas. Algunos *influencers* asociaban esta última plataforma con una comunidad madura e íntima, mientras que otros la percibían como enjuiciadora y ofensiva (figura 3).

En TikTok, los comentarios solían parecer inmaduros, aunque no malintencionados. Una base de usuarios predominantemente joven, compuesta por *millennials* y la generación Z, tiende a manifestar una comunicación menos elaborada. Como resultado, sus comentarios, aunque a veces negativos, no presentan la intencionalidad perjudicial que suele observarse en YouTube. Esta distinción subraya la influencia de los factores demográficos y las características de cada plataforma, como los filtros de discursos de odio y las herramientas para denunciar dinámicas interactivas.



Figura 3. Percepciones sobre el fandom y el hatedom de los influencers en redes sociales

Fuente: Elaboración propia.

Esto subraya también el efecto de desinhibición, intensificado por el entorno digital (Suler, 2006).

Por el contrario, Instagram se destacaba como una plataforma que fomenta conexiones más cercanas y personales entre influencers y seguidores, aunque dicha interacción varía según la experiencia del usuario. Algunos perfiles describían la aplicación como un entorno privado, íntimo y respetuoso, cimentado en una dinámica de mensajes directos bidireccionales y chats grupales unidireccionales. Otros la consideraban una plataforma perjudicial, en la que los usuarios adultos atacan indiscriminadamente tanto el contenido como a las personas.

Esta relación pone de manifiesto la importancia de diferenciar el perfil del no seguidor (usuario externo), caracterizado predominantemente por la negatividad, del del seguidor o fan, que se percibe como una figura que contribuye positivamente a la comunidad. Expresado de otro modo, el primero es un grupo social periférico marcado por la negatividad, el conflicto y el discurso de odio. Por ejemplo, en relación con temas feministas, progresistas o controvertidos, los no seguidores, especialmente aquellos que interactuaban de manera anónima, proyectaban su hostilidad a través de comentarios que “incluso superan los límites de la legalidad” (P4-INFLU6).

Los algoritmos de las plataformas ejercen una influencia significativa en esta relación, lo que respalda los hallazgos de Riemer y Peter (2021) y Whalter (2022). Uno de los entrevistados señaló diferencias notables en la manera en la que los sistemas algorítmicos de diferentes plataformas detectan automáticamente contenido malintencionado para evitar o restringir su difusión, reconociendo que “el algoritmo filtra el odio de manera diferente en cada plataforma” (P3-INFLU3).

En este contexto, las diferencias entre plataformas en cuanto a la tipología de comentarios y la demografía de seguidores resultaron esenciales para comprender la dinámica de exo y endogrupos. Para afrontarlo, se comienzan a adoptar estrategias de comunicación orgánica y gestión de comunidades (Ouvrein et al., 2023), lo que invita a enfocar integralmente el análisis de las dinámicas de confrontación del odio en redes sociales más allá de la mera victimización (Abidin, 2019; Vizcaíno-Verdú et al., 2020).

Objetivo 4. Gestión estratégica y mediadora del *hatedom*

La mayoría de los *influencers* señaló contar con una base de seguidores caracterizada por una marcada atmósfera de respeto y positividad (P4-P10). Destacaron que un entorno saludable es fundamental para su bienestar emocional, así como para la sostenibilidad de su labor y su permanencia en el ámbito digital.

En general, mi perfil recibe muy poco odio porque se caracteriza principalmente por humor blanco (P4-INFLU2).

Mi comunidad de seguidores siempre es igual en todas las plataformas; es decir, son fans leales que disfrutan de lo que hago (P4-INFLU4).

Mis seguidores siempre son respetuosos, incluso cuando no están de acuerdo con algo que hago (P4-INFLU5).

Creo que las personas que me siguen son constructivas (P4-INFLU6).

Este tipo de interacciones positivas no solo refuerza la motivación del creador, sino que también promueve un sentido de pertenencia y cohesión dentro de la comunidad (Barnes, 2022). En lo que respecta a la gestión de comentarios hostiles, los *influencers* declararon emplear estrategias orgánicas que complementan los sistemas de control de los discursos de odio facilitados por las plataformas (figura 4). Esta gestión proactiva resulta crucial para mantener la calidad de la relación y proteger el bienestar tanto de los creadores de contenido como de sus seguidores.

De forma generalizada, se adoptaba la estrategia (1) de ignorar comentarios negativos, en muchos casos altamente eficaz. Al evitar la visibilidad de estos mensajes hostiles, los *influencers* previenen la propagación del odio y mantienen un enfoque positivo tanto en interacciones en vivo como en conversaciones asincrónicas.

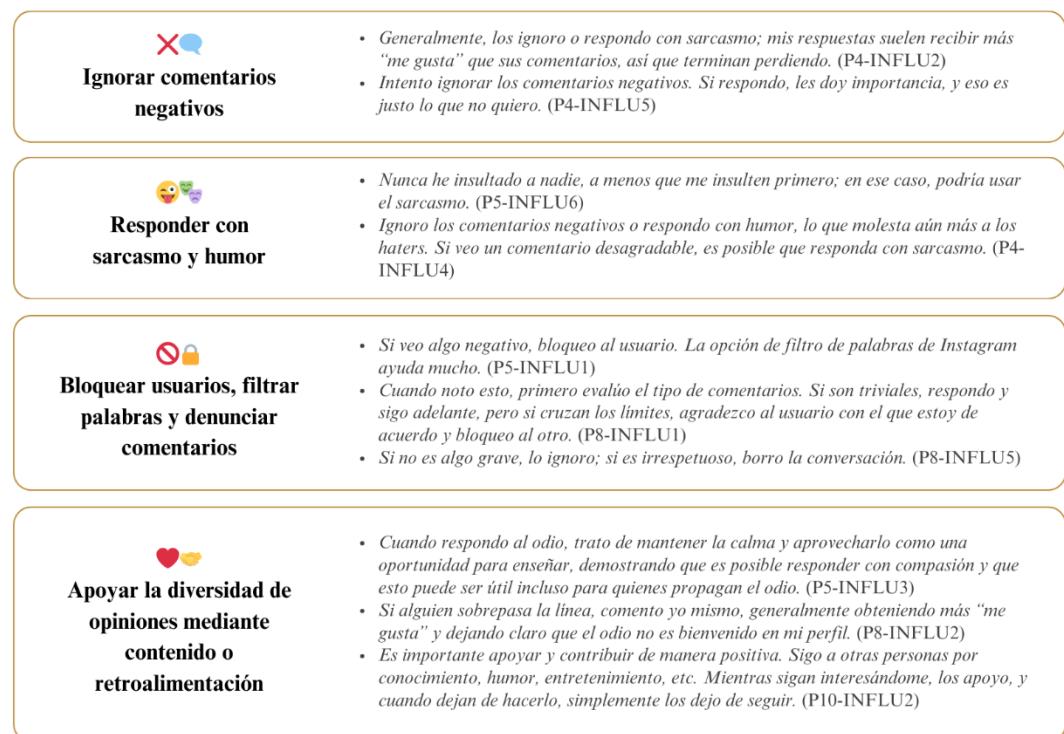


Figura 4. Estrategias adoptadas por los influencers para gestionar el hatedom

Fuente: Elaboración propia.

Otra táctica que empleaban era (2) la respuesta con sarcasmo o humor. Afirmaban que este enfoque puede desarmar a los usuarios más críticos y perjudiciales, transformando una situación potencialmente hostil en una oportunidad de entretenimiento o posicionamiento. En este sentido, uno de los *influencers* destacaba: "No puedes permitir que los comentarios de tus seguidores dominen tu vida" (INFLU6-P10).

Como consecuencia, en estos casos, recurrían a (3) bloquear al usuario, filtrar palabras o denunciar comentarios a través del sistema de la plataforma. Estos métodos de defensa frente al *hatedom* se alinean con las estrategias cognitivas y conductuales introducidas por Ouvrein y sus colegas (2023).

A pesar de estas medidas, los comentarios negativos eran asociados generalmente con el exogrupo, al que se etiquetaba como ignorante y carente de empatía debido a su incapacidad para reconocer el impacto potencial de sus mensajes de odio tanto para la comunidad como el *influencer*. Para ello, adoptan rutinas de autocuidado comunitario (4) que complementan aquellas identificadas en estudios previos (Ouvrein et al., 2023; Valenzuela-García et al., 2023). A través de estas prácticas, los *influencers* afirmaba que:

Doy espacio a las personas que están sufriendo para sanar. Es obvio que son ellos los que están llenos de odio, no yo. Simplemente, vivo en paz (P6-INFLU4).

Fomento el debate y el pensamiento crítico y doy ejemplo con ello. Nunca he insultado a nadie (P4-INFLU5).

Los conflictos entre seguidores y comunidades debido a diferencias de opinión constituían otra realidad entre los *influencers*, denominándolos “batallas campales” (P9-INFLU3). Estas confrontaciones pueden surgir por diversas razones, como discrepancias en relación con el contenido, la apariencia física, la orientación sexual, la lealtad a otros perfiles o malentendidos. En estos casos, recurrían a la mediación y la diplomacia por medio de (a) la validación de diversas sensibilidades y argumentos con respeto, (b) la confrontación directa de reacciones hostiles con el propósito de expresar una opinión y restar importancia al tema, o (c) la eliminación del hilo de comentarios que exacerbaba el discurso como adelantaban Ouvrin y sus colegas (2023) y Martínez-Valerio y Mayagoitia-Soria (2021). Un *influencer* explicaba que:

El tiempo me enseñó que no vale la pena intentar convencer a nadie en línea.

Nadie se conoce y muy pocas personas están dispuestas a cambiar de opinión y escuchar de verdad (P9-INFLU3).

Esta gestión comunitaria se integraba, por ende, en las actividades diarias de los *influencers* (Abidin, 2019; Ouvrein et al., 2023; Valenzuela-García et al., 2023), sostenida por un odio ciego (Aburime, 2022). No obstante, la abrumadora cantidad de comentarios los condujo al establecimiento de límites naturales entre las comunidades de seguidores y sus detractores.

Límites en torno a las expectativas: equilibrar lo íntimo y lo laboral

Ya sean públicos o privados, los mensajes que reciben los *influencers*, especialmente dentro de comunidades extensas, dificultan la comunicación con los usuarios. Se señaló que la comunidad presenta expectativas excepcionalmente altas y exige interacciones rápidas y constantes. Con frecuencia, expresaba uno de los entrevistados, esto conduce al fomento de “sentimientos de culpa” (P7-INFLU5). Se destacó asimismo que esta dinámica puede provocar con facilidad que los seguidores se conviertan en *haters*, lo que provoca un incremento de la agresividad y de la obsesión. Por ejemplo, algunos *influencers* describían que:

A veces, intentas ser amable con tus seguidores y responderles, pero terminan tomándose demasiadas libertades e incluso se enfadan (P8-INFLU1).

Algunas personas tienen dificultades para entender que no somos amigos. Tengo aproximadamente 20 seguidores que me hablan todos los días y son

habituales en mis espectáculos. Una chica de 18 años asistía a mis shows cada mes desde noviembre. Era el mismo espectáculo, lo había visto unas 20 veces y seguía volviendo. Incluso me ha traído regalos. Intenté hablar con ella y su madre porque me hacía sentir muy culpable (P8-INFLU2).

Tuve una pareja y solía compartir mucho contenido sobre nosotros en redes sociales. Cuando terminamos, todos comenzaron a preguntar qué había pasado y por qué nos separamos. Algunos incluso enviaron mensajes a mi nueva pareja para conseguir información. Esto me afectó, pero en realidad fue culpa mía por no establecer límites. Si no hubiera compartido tanto mi vida privada esto no habría ocurrido (P8-INFLU5).

Estas experiencias demostraron la relevancia de establecer límites sobre la vida privada de los *influencers* para evitar un agotamiento emocional o la invasión de su intimidad. Algunos explicaban emplear estrategias complementarias de creación de contenido con el fin de mitigar la sensible transición de un seguidor desde el afecto al rechazo. Entre estas se incluían 1) la publicación de contenido relacionado con su vida personal, como sesiones de preguntas y respuestas en las que se atienden consultas de manera directa y 2) la desconexión temporal de redes sociales mediante el establecimiento de rutinas de acceso que no excedan el tiempo dedicado a su vida cotidiana. El objetivo era “no sentirse obligado a mostrar cada aspecto de la vida, sino solo aquello con lo que me sienta cómodo compartiendo” (P8-INFLU6).

La responsabilidad del *influencer*: batalla de influencias

Los *influencers* desempeñan un rol significativo en la vida de sus seguidores y, en consecuencia, asumen una gran responsabilidad en cuanto al discurso y los valores promovidos. Son conscientes de su influencia y de que esta puede impactar de manera sustancial las actitudes, el comportamiento y las decisiones de su comunidad (Jenkins, 1992, 2006). Uno de ellos señaló explícitamente que

Creo que debemos ser cuidadosos con lo que mostramos, promovemos y decimos. Tenemos mucha visibilidad, y considero que debemos ser responsables con lo que publicamos (P2-INFLU5).

La responsabilidad de los *influencers* trasciende la mera creación de contenido atractivo o promocional (von-Mettenheim & Wiedmann, 2022), pues también implica la promoción de valores que los definen como seres humanos. Esta manifestación de intereses, opiniones y estilos de vida genera una disonancia cognitiva, lo que con frecuencia conduce a la pérdida de seguidores. Por ejemplo, algunos entrevistados explicaban que

Perdí seguidores de la comunidad vegana porque hice una campaña con El Pozo, pero fue una salida bastante silenciosa. Simplemente me dejaron de seguir y algunos expresaron su descontento, aunque de buena manera (P10-INFLU2)

Me mantengo fiel a lo que siento en cada momento y no siempre creo el mismo contenido porque prefiero hacer lo que disfruto [...]. Esto no gusta a todos mis seguidores, ya que a veces quieren que haga lo que ellos desean ver. Como resultado, muchos se van, pero no me preocupa. Si alguien se va, es porque así tiene que ser (P10-INFLU3).

Así, el éxito de un *influencer* depende necesariamente de la generación de entornos positivos, lo que requiere un compromiso comunicacional constante, para el cual, indicaban, “no existe manual de instrucciones” (P2-INFLU3). Una comunidad saludable exige “la promoción de contenido valioso” (P2-INFLU2), a pesar de la influencia de los sistemas algorítmicos de recomendación y la polarización fluctuante en las conexiones, de manera que los *influencers* “debemos lidiar con ello” (P10-INFLU5), tal y como evidencian los estudios de Martínez-Valerio y Mayagoitia-Soria (2021), y de Riemer y Peter (2021). A causa de esto, los *influencers* se encuentran sometidos a una presión social que desdibuja su rol de figuras influyentes (Barnes, 2022), en un contexto en el que una base de seguidores altamente demandante y obsesiva puede convertirse en su mayor desafío, derivando en un discurso de odio multiplataforma.

CONCLUSIONES

El *hatedom*, que engloba comunidades de (anti)fans, se caracteriza por una crítica persistente e intensa hacia figuras públicas, especialmente *influencers*. Este fenómeno trasciende el odio irracional (Duffy et al., 2022; Aburime, 2022) y representa una relación comunitaria de compromiso obsesivo que reacciona de manera aguda al contenido en línea generado por las celebridades digitales, determinado por una polarización emocional extrema (Barnes, 2022). Este estudio contribuye sobre las dinámicas de estas interacciones polarizadas entre *influencers* y seguidores en múltiples plataformas sociales, resaltando la complejidad y los mecanismos subyacentes involucrados en la relación.

Si bien el *fandom* tóxico y el *hatedom* comparten algunas similitudes, como la inversión afectiva (Théberge, 2006) y las reacciones intensas que difuminan la frontera entre la realidad y la ficción (Vizcaíno-Verdú et al., 2020), sus dinámicas y manifestaciones difieren considerablemente. El *fandom* tóxico emerge sobre una base de fans específica, en la que los miembros pueden participar mediante

conductas perjudiciales que derivan en exclusión, crítica y luchas de poder, pese a compartir intereses comunes (Walther, 2022; Arouh, 2020).

En contraste, el *hatedom* canalizan su proyección negativa y colectiva hacia el exterior, enfocándose principalmente en figuras públicas o influencers con quienes establecen relaciones parasociales obsesivas, gestando una suerte de pseudo amistad unilateral (Hoffner & Bond, 2022). Estas comunidades se agrupan para compartir su desdén u hostilidad hacia un *influencer* que consideran que ha obtenido fama de manera inmerecida (Ouvrein et al., 2023). En consecuencia, el acoso multiplataforma a estos perfiles para obtener detalles íntimos de sus vidas transforma la admiración en toxicidad compulsiva y aversión. Este comportamiento no solo busca la crítica o el perjuicio, sino la ruptura de barreras que los influencers intentan equilibrar para su bienestar y privacidad. Esto implica que el desarrollo de contenido en múltiples plataformas –para cultivar una comunidad más amplia y heterogénea en sentido demográfico– puede dar lugar a la conformación de una macro-comunidad tóxica. Esta interacción prolongada intensifica las relaciones y dificulta que los *influencers* controlen los efectos negativos.

En relación con el primer objetivo (O1), enfocado en la identificación de la polaridad y toxicidad multiplataforma en torno a los comentarios recibidos por *influencers*, los resultados mostraron diferencias en función del perfil del *influencer* y el entorno. X y YouTube presentaban un mayor porcentaje de comentarios negativos y tóxicos, mientras que TikTok, Facebook y Twitch introducen espacios más positivos y colaborativos. En contraste, en Instagram se observaban discrepancias dependiendo de las herramientas de comunicación empleadas para interactuar con los seguidores (salas de chat, stories bidireccionales, publicaciones estáticas, etc.). Esta variabilidad destaca cómo la naturaleza de cada plataforma, en lo que a sistemas de retroalimentación se refiere, puede posibilitar el *hatedom* y enfatizar una necesaria revisión e implementación de estrategias de moderación. Estas disparidades sugieren que la arquitectura de la plataforma y sus normas comunitarias (Luo & Li, 2024) influyen en la expresión del odio y la toxicidad, lo que requiere una comprensión matizada de la demografía única de cada una de ellas para desarrollar intervenciones más efectivas.

Respecto del segundo objetivo (O2), sobre el rol y la relevancia de las comunidades de fans para la profesión de los *influencers*, los resultados sugerían que estas comunidades pueden impactar profundamente en sus decisiones personales y profesionales (Abidin, 2019; Valenzuela-García et al., 2023). Se evidenció cómo los fans proporcionan validación y apoyo emocional, constituyéndose como la base financiera para las marcas con las que los *influencers* colaboran. Sin embargo, la inversión afectiva del *fandom* (Théberge, 2006) parece evolucionar hacia una

aversión afectiva hacia las actividades diarias de los *influencers*. Esta aversión los incita a aplicar estrategias de afrontamiento para mitigar comportamientos negativos y conflictivos (Ouvrein et al., 2023). Esta dualidad afectiva explica el complejo entramado emocional que realizan los *influencers* al navegar en un panorama donde existe un delicado equilibrio entre admiración y hostilidad.

De acuerdo con el tercer objetivo (O3), relativo a la manera en la que los *influencers* interactúan con sus seguidores, el *hatedom* ejerce presión para priorizar ciertas plataformas o actividades para satisfacer a su audiencia. Las decisiones de los *influencers* se moldean dependiendo del patrón *hatedom* desplegado en su perfil, impulsado por la edad y madurez de la comunidad, el grado de intimidad, las funciones de la plataforma y el nivel de anonimato o transparencia identitaria de los (anti)fans involucrados. Así, los *influencers* utilizan diferentes estrategias para lidiar con el odio y la toxicidad en línea (Valenzuela-García et al., 2023). Estas estrategias abarcan desde la omisión de comentarios negativos hasta respuestas humorísticas o sarcásticas, así como el bloqueo y denuncia de usuarios ofensivos en algunos casos. La mediación y la diplomacia desempeñan un rol fundamental en la resolución de los conflictos entre seguidores.

Se observó que las estrategias adoptadas por los *influencers* están significativamente influenciadas por la polaridad y toxicidad de los comentarios en sus perfiles. Aquellos *influencers* que recibían un mayor número de comentarios negativos y tóxicos, como INFLU2 e INFLU6, implementaban medidas más proactivas para gestionar la hostilidad, como el bloqueo de usuarios y el uso del sarcasmo como mecanismo de respuesta. Por el contrario, los *influencers* cuya retroalimentación predominante era neutral o positiva, como INFLU1, centraban sus esfuerzos en fomentar una interacción respetuosa y en fortalecer una comunidad de apoyo. Estos hallazgos destacan que la naturaleza de los comentarios influye de sobremodo en las estrategias de gestión de su presencia e interacción en línea. Sin embargo, resulta fundamental que los *influencers* establezcan límites evidentes para proteger su bienestar emocional y mantener un equilibrio saludable entre su vida personal y su actividad en redes sociales.

En lo que respecta al cuarto objetivo (O4), orientado a las estrategias que los *influencers* emplean para afrontar el *hatedom* en redes sociales, se observó que el impacto psicológico de este fenómeno se alinea con análisis previos en torno al acoso digital, enmarcado por la ansiedad (Lehto, 2022), el estrés (Smith et al., 2017) y la depresión (Lind & Wickström, 2024). Los *influencers* no solo son percibidos como víctimas merecedoras de este odio (Valenzuela-García et al., 2023), sino que experimentan un proceso de profunda deshumanización. Se espera que respondan a la intensa e incontrolable presión del *hatedom* multiplataforma. Esta

dinámica de amor-odio persiste incluso en su esfera privada y se mantiene de forma constante, viral y pública.

LIMITACIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN

Este estudio tiene diversas limitaciones. En primera instancia, la muestra se restringió a seis *influencers* españoles, lo que puede no considerarse representativo en lo que a experiencias en otras regiones o contextos culturales se refiere. Adicionalmente, la recopilación de comentarios se realizó en mayo de 2024, lo que podría no haber considerado variaciones temporales en la polaridad y en la toxicidad de los comentarios, ni posibles avances en la moderación de cada plataforma.

Desde una perspectiva metodológica, si bien los modelos avanzados de procesamiento del lenguaje natural empleados para el análisis cuantitativo de los comentarios son herramientas de potencial evidente, dificultan la captación de sutilezas contextuales y culturales propias del discurso digital. Del mismo modo, aunque las entrevistas estructuradas proporcionaron información cualitativa fundamental, su profundidad pudo verse limitada por la disponibilidad de tiempo de los *influencers*.

Pese a ello, los resultados sugieren nuevas líneas de investigación y recomendaciones para un análisis más exhaustivo sobre las dinámicas del *hatedom* y su impacto en la experiencia comunitaria en redes sociales. Por ejemplo, el desarrollo de algoritmos mejorados para la detección de toxicidad y la moderación de contenido podría contribuir a una gestión más eficaz de los discursos negativos. Asimismo, estudios longitudinales permitirían examinar la evolución de la relación entre *influencers* y *hatedom* en un período extendido, introduciendo variaciones contextuales y temporales. Para una mayor representatividad, futuras investigaciones deberían incluir una muestra diversa de *influencers* que abarque regiones, culturas, temáticas de contenido e identidades de género más heterogéneas.

Abordar el fenómeno del *hatedom* requiere, en suma, de un enfoque multidimensional que permita comprender con mayor profundidad la polarización de las comunidades digitales en torno a los *influencers*. Esta línea no solo optimizaría los sistemas tecnológicos de las plataformas para moderar con mayor eficacia los discursos de odio digitales, sino que contribuiría a consolidar la importancia de la salud mental frente a la toxicidad en línea. Es fundamental que los *influencers* que expresan resiliencia, asertividad, humanidad y humildad en la gestión de conflictos comunitarios no sean eclipsados por la rápida y automatizada dinámica de las redes sociales y, en menor medida, reducidos a figuras frívolas sin identidad. Un enfoque holístico de esta problemática avanzaría en la protección del bienestar de los creadores de contenido y fomentaría la construcción de comunidades digitales más respetuosas y solidarias.

APOYOS

Este estudio cuenta con el apoyo del grupo de investigación COYSODI de la Universidad Internacional de la Rioja y del proyecto *La incidencia del marketing de influencers en la construcción de la identidad digital de los adolescentes (INFLUIDENTITY)* (The impact of influencer marketing on the construction of teenagers' digital identity (INFLUIDENTITY)). El tiempo dedicado a esta investigación también se enmarca en los estudios de la TikTok Cultures Research Network y el IERLab (Influencer Ethnography Research Lab) de la Curtin University (Australia).

REFERENCIAS

- Abidin, C. (2018). *Internet celebrity: Understanding fame online*. Emerald Publishing Limited. <https://doi.org/10.1108/9781787560765>
- Abidin, C. (2019). Victim, rival, bully: Influencers' narrative cultures around cyberbullying. In H. Vandebosch & L. Green (Eds.), *Narratives in research and interventions on cyberbullying among young people* (pp. 199-212). Springer. https://doi.org/10.1007/978-3-030-04960-7_13
- Aburime, S. (2022). Hate narratives, conditioned language and networked harassment: A new breed of anti-shipper and anti-fan-antis. *Journal of Fandom Studies*, 10(2-3), 135-155. https://doi.org/10.1386/jfs_00060_1
- Association of Internet Researchers. (2012). *Ethical decision-making and internet research. Recommendations from the AoIR Ethics Working Committee (version 2.0)*. <https://aoir.org/reports/ethics2.pdf>
- Arouh, M. (2020). Toxic fans: Distinctions and ambivalence. *Ex-centric Narratives: Journal of Anglophone Literature, Culture and Media*, (4), 67-82. <https://ejournals.lib.auth.gr/ExCentric/article/view/7917>
- Barnes, R. (2022). Loving to hate: Fandom fueling polarized behavior. In R. Barnes (Ed.), *Fandom and polarization in online political discussion* (pp. 61-86). Palgrave Macmillan. https://doi.org/10.1007/978-3-031-14039-6_4
- Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches*. SAGE.
- Devlin, J. & Chang, M. W. (2018, November 2). Open sourcing BERT: State-of-the-art pre-training for natural language processing. *Google Research*. <https://research.google/blog/open-sourcing-bert-state-of-the-art-pre-training-for-natural-language-processing/>
- Díaz, J. E., Cortés, M., & Hinojosa, C. (2019). *Detecting toxic interactions on the Twitter Star Wars fandom using machine learning*. Tecnológico de Monterrey.
- Duffy, B. R., Miltner, K. M., & Wahlstedt, A. (2022). Policing "fake" femininity: Authenticity, accountability, and influencer antifandom. *New Media & Society*, 247, 1657-1676. <https://doi.org/10.1177/14614448221099234>

- Hoffner, C. A. & Bond, B. J. (2022). Parasocial relationships, social media, & well-being. *Current Opinion in Psychology*, 45, 101306. <https://doi.org/10.1016/j.copsyc.2022.101306>
- Jenkins, H. (1992). *Textual Poachers: Television fans and participatory culture*. Routledge
- Jenkins, H. (2006). *Fans, bloggers, and gamers. Exploring participatory culture*. New York University Press.
- Lehto, M. (2022). Ambivalent influencers: Feeling rules and the affective practice of anxiety in social media influencer work. *European Journal of Cultural Studies*, 25(1), 201-216. <https://doi.org/10.1177/1367549421988958>
- Lind, J. & Wickström, A. (2024). Representations of mental health and mental health problems in content published by female social media influencers. *International Journal of Cultural Studies*, 27(2), 217-233. <https://doi.org/10.1177/13678779231210583>
- Luo, Z. & Li, M. (2024). Participatory censorship: How online fandom community facilitates authoritarian rule. *New Media & Society*, 26(7), 4236-4254. <https://doi.org/10.1177/14614448221113923>
- Maarouf, A., Pröllochs, N., & Feuerriegel, S. (2024). The virality of hate speech on social media. In J. Nichols (Ed.), *Proceedings of the ACM on Human-Computer Interaction* (Vol. 8) (pp. 1-22). <https://doi.org/10.1145/3641025>
- Martínez-Valerio, L. & Mayagoitia-Soria, A. M. (2021). *Influencers y mensajes de odio: jóvenes y consumo de contenidos autocensurados*. Revista Prisma Social (Influencers and hate messages: Young people and consumption of self-censored content). *Prisma Social*, (34), 4-39. <https://hdl.handle.net/10115/30379>
- Mysiuk, I. & Shuvar, R. (2023). Automating web scraping of user comments for sentiment analysis in social networks. In *2023 IEEE 13th International Conference on Electronics and Information Technologies (ELIT)* (pp. 77-81). IEEE. <https://doi.org/10.1109/ELIT61488.2023.10310867>
- OHCHR. (2023). *Report of the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia, and related intolerance on combating glorification of Nazism, neo-Nazism, and other practices that contribute to fueling contemporary forms of racism, racial discrimination, xenophobia, and related intolerance*. United Nations.
- Ouvrein, G., Jorge, A., Cabral, J., & Vandebosch, H. (2023). Coping comes with the job: An exploratory study into the selection and use of coping strategies for online aggression among social media influencers. *Telematics and Informatics Reports*, 10, 100052. <https://doi.org/10.1016/j.teler.2023.100052>
- Riemer, K. & Peter, S. (2021). Algorithmic audiencing: Why we need to rethink free speech on social media. *Journal of Information Technology*, 36(4), 409-426. <https://doi.org/10.1177/02683962211013358>
- Schuessler, J. (2018, November 14). 'Toxic' is Oxford's word of the year. No, we're not gaslighting you. *The New York Times*. <https://bit.ly/3WdwxGh>
- Senft, T. M. (2013). Microcelebrity and the branded self. In J. Hartley, J. Burgess, & A. Bruns (Eds.), *A companion to new media dynamics* (pp. 346-354). Blackwell Publishing Ltd. <https://doi.org/10.1002/9781118321607.ch22>

- Sjoerd-Jeroen, M., Basten, F., Taran, G., Panagoulia, A., Coughlan, G., & Duarte, J. (2024). The structured narrative interview. *Narrative Inquiry*, 34(2), 307-334. <https://doi.org/10.1075/ni.23079.moe>
- Smith, M. M., Speth, T. A., Sherry, S. B., Saklofske, D. H., Stewart, S. H., & Glowacka, M. (2017). Is socially prescribed perfectionism veridical? A new take of the stressfulness of perfectionism. *Personality and Individual Differences*, 110, 115-118. <https://doi.org/10.1016/j.paid.2017.01.031>
- Strauss, A. & Corbin, J. M. (1997). *Grounded theory in practice*. SAGE.
- Suler, J. (2006). The online disinhibition effect. *International Journal of Applied Psychoanalytic Studies*, 2(2), 184-188. <https://doi.org/10.1002/aps.42>
- Théberge, P. (2006). Everyday fandom: Fan clubs, blogging, and the quotidian rhythms of the internet. *Canadian Journal of Communication*, 30(4), 485-502. <https://doi.org/10.22230/cjc.2005v30n4a1673>
- Valenzuela-García, N., Maldonado-Guzmán, D. J., García-Pérez, A., & Del-Real, C. (2023). Too lucky to be a victim? An exploratory study of online harassment and hate messages faced by social media influencers. *European Journal on Criminal Policy and Research*, 29, 397-421. <https://doi.org/10.1007/s10610-023-09542-0>
- Vizcaíno-Verdú, A., Contreras-Pulido, P., & Guzmán-Franco, M. D. (2020). Construcción del concepto fanbullying: Revisión crítica del acoso en redes sociales (Building the fanbullying concept: Critical review of social media harassment). *Pixel-Bit. Revista de Medios y Educación*, (57), 211-230. <https://doi.org/10.12795/pixelbit.2020.i57.09>
- Von-Mettenheim, W. & Wiedmann, K. P. (2022). Hate speech and bad language: The ugly face of social influencers and its impact on brands. An abstract. In B. Jochims & J. Allen (Eds.), *Optimistic marketing in challenging times: Serving ever-shifting customer needs* (pp. 291-292). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-24687-6_116
- Walther, J. B. (2022). Social media and online hate. *Current Opinion in Psychology*, 45, 101298. <https://doi.org/10.1016/j.copsyc.2021.12.010>

SOBRE LAS AUTORAS

ARANTXA VIZCAÍNO-VERDÚ, profesora contratada doctora en la Universidad Internacional de La Rioja, España. Obtuvo su doctorado en Comunicación en la Universidad de Huelva. Actualmente es parte de diversos grupos y redes de investigación en España y Australia, en los que analiza la cultura popular en redes sociales. Especializada en alfabetización mediática e informacional, su investigación se centra en el contenido generado por usuarios en Internet, el fandom y las narrativas transmedia desde las perspectivas de la comunicación, publicidad, educación y humanidades digitales.



<https://orcid.org/0000-0001-9399-2077>

PALOMA CONTRERAS-PULIDO, periodista, educadora social y doctora en Educación por la Universidad de Huelva. Es profesora titular en la Universidad Internacional de La Rioja, especializada en tecnologías aplicadas a la educación y didáctica, y profesora invitada en másteres sobre comunicación audiovisual y producción televisiva. Actualmente supervisa tesis doctorales sobre medios, tecnología, educación e inclusión social. Es miembro de la Red Alfamed e investigadora del Grupo COYSODI (UNIR), contribuyendo al estudio de la alfabetización mediática y la inclusión social.



<https://orcid.org/0000-0002-6206-7820>