



UNIVERSIDADE ESTADUAL PAULISTA JÚLIO DE MESQUITA FILHO

FACULDADE DE FILOSOFIA E CIÊNCIAS

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

MIRELYS PUERTA-DÍAZ

TESIS DE DOCTORADO

**REPRESENTACIÓN DEL DOMINIO DE LA CAMPAÑA #STOPHATEFORPROFIT
A PARTIR DEL ANÁLISIS DE REDES SOCIALES**

Marília, SP

2022

MIRELYS PUERTA-DÍAZ

**REPRESENTACIÓN DEL DOMINIO DE LA CAMPAÑA #STOPHATEFORPROFIT
A PARTIR DEL ANÁLISIS DE REDES SOCIALES**

Tesis de doctorado presentada al Programa de Pos-graduación en Ciencia da Informação (PPGCI) de la Universidad Estadual Paulista “Júlio de Mesquita Filho”, como requisito para la obtención del título de doctor en Ciencia de la Información.

Área de Concentración: Información, Tecnología y Conocimiento.

Línea de Investigación: Producción y Organización de la Información.

Orientador: Prof. Dr. Daniel Martínez-Ávila.

Coorientadora: Prof. Dra. María-Antonia Ovalle-Perandones.

Financiamiento: PROEX/CAPES (Becaria AUIP 2018)

Marília, SP

2022

Dados Internacionais de Catalogação na Publicação

P977r Puerta-Díaz, Mirelys
Representación del dominio de la campaña
#StopHateforProfit a partir del Análisis de Redes Sociales /
Mirelys Puerta-Díaz. -- Marília, 2022
144 p.: il., tabs.

Tese (doutorado) - Universidade Estadual Paulista (Unesp),
Faculdade de Filosofia e Ciências, Marília
Orientador: Daniel Martínez-Ávila
Coorientadora: María-Antonia Ovalle-Perandones

1. Discurso do ódio. 2. Ciência da Informação. 3. Stop Hate
for Profit. 4. Análise de Redes Sociais. 5. Análise de
sentimentos. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da
Faculdade de Filosofia e Ciências, Marília. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

MIRELYS PUERTA-DÍAZ

**REPRESENTACIÓN DEL DOMINIO DE LA CAMPAÑA #STOPHATEFORPROFIT
A PARTIR DEL ANÁLISIS DE REDES SOCIALES**

Tesis de doctorado presentada al Programa de Pos-graduación en Ciencia da Informação (PPGCI) de la Universidad Estadual Paulista “Júlio de Mesquita Filho”, como requisito para la obtención del título de doctor en Ciencia de la Información.

Área de Concentración: Información, Tecnología y Conocimiento.

Línea de Investigación: Producción y Organización de la Información.

BANCA EXAMINADORA

Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação da Faculdade de Filosofia e Ciências da Universidade Estadual Paulista Júlio de Mesquita Filho, como requisito parcial para obtenção do título de Doutora em Ciência da Informação, aprovado em 11 de 03 de 2022, pela Banca Examinadora constituída pelos seguintes professores:

Prof. Dr. Daniel Martínez Ávila

Orientador y Presidente de la Banca Examinadora
Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP/Marília - SP)

Prof. Dra. Maria Cláudia Cabrini Grácio
Miembro Titular Interno

Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP/Marília - SP)

Prof. Dr. João Batista Ernesto de Moraes – PPGCI/Unesp
Membro Titular Interno
Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP/Marília - SP)

Prof. Dr. Maria Aurora Cuevas
Miembro Titular Externo
Universidad Complutense de Madrid (Comunidad de Madrid - España)

Prof. Dr. José Alberto Benítez Andrades
Miembro Titular Externo
Universidad de León (Espanha)

Prof. Dr. Pablo Parra Valero
Miembro Suplente Externo
Universidad Complutense de Madrid (Comunidad de Madrid - España)

Prof. Dr. Yohannis Marti-Lahera
Miembro Suplente Externo
Universidad de La Habana (Cuba)

Marília, 11 de 03 de 2022.

**“El que tiene un derecho no obtiene el de violar el
ajeno para mantener el suyo”.**
José Martí

Dedicatoria:

A mis padres, hermana, abuelos y padrinos,
deseo que sepan que este es también vuestro logro.

AGRADECIMIENTOS

Todo proyecto recibe colaboraciones a lo largo de su desarrollo y la presente tesis de doctorado no fue una excepción, hoy se las entrego para lectura, gracias a la ayuda y esfuerzo de todos los que menciono a continuación:

El presente trabajo fue realizado con el apoyo de la Coordinación de Perfeccionamiento de Personal de Nivel Superior -Brasil (CAPES) - código de Financiamiento 001 y la Fundación Carolina (España). A la AUIP por fomentar este intercambio académico en Brasil. A la UNESP por brindarme todos los recursos y herramientas que fueron necesarios para llevar a cabo el proceso investigativo.

Quiero expresar un sincero agradecimiento a mis tutores Daniel Martínez Ávila y María Antonia Ovalle Perandones (Toñi) los principales pilares de este estudio, quienes, a pesar de la distancia física que nos separaba, con su dirección, conocimiento, enseñanza y colaboración me permitieron hilvanar el entretejido dialógico de esta tesis y la enriquecieron enormemente durante mi estancia en Madrid. Gracias por el compromiso, la amabilidad, la paciencia, el tiempo dedicado y el conocimiento transmitido durante todos estos años.

Existen personas especiales que siempre están ahí para ti y aunque estén ‘detrás del telón’ son esenciales para su ejecución. Ese lindo papel lo cumple mi prima Cary, sin ella no hubiese podido realizar tantos sueños profesionales. Gracias por siempre querer verme feliz y exitosa.

A Edgar Bisset y Yoa Martí, que fueron eslabones claves en mi crecimiento como investigadora y se mantuvieron a mi lado en este camino profesional ayudándome a hacerlo realidad.

Expreso mi más profunda gratitud a Ilídio y Paloma, por ser amigos, excelentes colegas y sobre todo por ser mis referentes de actuar moral y compromiso profesional. Sus aportes están cimentados en las páginas de este estudio.

A Larissa Pavarini, Larissa de Mello y Natalia, las hermanitas brasileñas que gané con el doctorado, gracias por ayudarme incondicionalmente en estos años, gracias por todos los

ánimos, sin ellos no hubiese sido posible sacar adelante esta investigación. Chicas, nos debemos lindas tardes de risas, cafés y panes de queso.

A Alex y Edisnel por ser de extrema ayuda en la dimensión tecnológica de la presente investigación.

Este resultado también se lo dedico a los profesores de FCOM Radamés Linares, Gloria Ponjuán, Rojas, Débora, Yunier, Maria del Carmen Villardefrancos, Estrella, Magda, Zoia, ellos fueron claves en mi formación universitaria en Cuba y a los cuales tanto admiro. En especial agradecer a mis tutores de licenciatura y maestría Ania R. Hernández Quintana, Dionnys Maldonado y Mayra Mena por conducirme en el camino correcto de la investigación científica.

Al equipo interdisciplinar de OpLaDyn, en especial a sus líderes Eunice y Mariana, gracias por la confianza y dejarme participar de este increíble proyecto que me permitió salir de la zona de confort y me desafió en tantos aspectos profesionales. A Ana, Edna, Valdirene y Mari Vitti solo tengo palabras de agradecimiento por hacerme sentir parte esencial de este gran equipo.

A la profesora María Claudia Cabrini por las valiosas contribuciones, así como a los integrantes del grupo EMI, en especial a Helen, Fernanda y Bianca gracias por el cariño y la bienvenida al grupo.

Asimismo, agradezco a mis compañeros del Grupo de Pesquisa FTI, especialmente a su líder Carlos Almeida por los conocimientos que aportó en cada criterio que emitió sobre la investigación.

A los profesores Aurora y Pablo de la UCM y José Alberto de la ULe por sus contribuciones en la banca de cualificación y defensa final, gracias por aceptar ser parte de este resultado y por compartir ideas y conocimientos en pos de mejorar la calidad del estudio.

A los colegas del PPGCI Caio, Sandra, Ana María, Jacquelin, y a la profesora Silvana por todas las gestiones de apoyo y bienvenida a Marília.

A la profesora Marta Valentín y María José Jorente mi más sincero agradecimiento, gracias por tanta amabilidad y apoyo profesional.

A todos los colegas y profesores del PPGCI, por los intercambios de ideas e debates fructíferos durante las disciplinas del doctorado y a quienes echaron un ojito a mis textos para corregir mi portugués.

Gracias a Jose, Viena y China que siempre me apoyaron a pesar de tener sus propios desafíos personales.

A JARR por sus sabios consejos, sin su impulso constante no estaría defendiendo esta investigación.

A Eliane Pellegrini tu apoyo significó mucho en la última y más difícil etapa de realización de esta tesis. Gracias por todo el cariño y apoyo incondicional. Gracias por los buenos *rolês* en Madrid, Barcelona, Portugal. Que bien contar contigo en este recomienzo.

A todos, muchas gracias, GRATIDÃO!

RESUMEN

El presente estudio aborda el fenómeno de la desinformación y propagación del discurso del odio en los sitios de redes sociales como antesala de gestación de la campaña boicot a Facebook #StopHateforProfit. Expone los principios teóricos conceptuales del discurso del odio, así como los enfoques disciplinares presentes en la literatura internacional. Describe al análisis de dominio y Análisis de Redes Sociales como métodos sobre los cuales la investigación se respalda para contribuir a la representación de un dominio marcado por el activismo y una actividad social complementaria a la académica que se ve representada en sitios de redes sociales de Internet. Se describe las estrategias, técnicas y herramientas de procesamiento de lenguaje natural empleadas para el procesamiento y análisis de los datos extraídos de Twitter. Los resultados ilustran cómo la red social internacional construida mediante las relaciones de interacción en este sitio de red social contribuyó a que Stop Hate for Profit alzara el éxito como campaña boicot a Facebook y creciera rápidamente a amplia escala internacional a solo días de ser lanzada. La narrativa general en Twitter en torno a la campaña mostró mensajes positivos para generar conciencia a las principales compañías que publicitan en Facebook sobre el discurso del odio y la desinformación como fenómenos dañinos a la democracia, y también cómo la opinión pública hizo uso de vocabulario cargados de sentimientos negativos para con prevalencia de emociones de miedo para alertar sobre las consecuencias que este fenómeno causa a la sociedad.

Palabras Clave: Stop Hate for Profit; Facebook; Discurso del odio, Análisis de Redes Sociales; Análisis de dominio; Análisis de sentimientos.

RESUMO

A pluralidade de aspectos relacionados ao discurso de ódio na web tem levado ao aumento do seu estudo científico em áreas como sociologia, psicologia, linguística, filosofia, e recentemente estendeu-se para áreas menos tradicionais em sua análise como a ciência da informação e a ciência da computação. Isso levou a um rápido crescimento da produção intelectual sobre o tema nos últimos anos e abrindo cada vez mais espaço nos debates (públicos) da opinião pública em geral, o que mostra uma forte ligação entre eles. A análise bibliométrica dos trinta anos de pesquisa sobre o tema desenvolvida por Tontodimamma et al. (2021) mapeou os tópicos de interesse e sua evolução no discurso científico e aplicou métodos de análise de co-palavras. No entanto, este estudo, assim como outros publicados até o momento, não considera o método de análise de domínio para abordar o discurso de ódio, focando não apenas no sistema de conhecimento científico, mas também nas complexas interações sociais não científicas que contribuem para a construção do conhecimento do domínio do discurso de ódio.

Os perigos que a democracia enfrentou com o aumento do índice de alertas de discurso de ódio amplamente propagados na plataforma do Facebook por meio de anúncios, alarmaram organizações e instituições de defesa dos direitos civis, como Common Sense Media, Free Press, Anti-Defamation League (ADL) e a Associação Nacional para o Avanço das Pessoas de Cor (NAACP). Essa situação levou essas organizações a se unirem e lançarem a campanha Stop Hate for Profit em 17 de junho de 2020. Seu objetivo era muito claro, fazer uma ligação internacional para que as empresas suspendessem sua publicidade paga na rede social Facebook.

Este boicote é uma reivindicação à empresa Facebook, para que assuma seu papel de responsabilidade social corporativa e se oriente para a proteção dos direitos civis da sociedade. Esse fenômeno não apenas levou a uma participação massiva de empresas nesse movimento ativista contra o discurso de ódio, mas estima-se que gerou por trás dele um fenômeno social e informacional que exigia ser revelado. Estima-se que tenha levado à formação de comunidades a partir das interações de menção ou troca de informações entre os participantes que utilizaram a *hashtag* da campanha. Também não estão claros quais

narrativas tiveram maior impacto no sucesso ou não da campanha e se houve liderança de opinião nela além dos membros da coalizão e os assuntos de interesse público que foram formados e disseminados a partir da participação na campanha. Diante desse cenário, surge a necessidade de responder às seguintes questões: Quais são os principais contornos teóricos do domínio de conhecimento do 'discurso de ódio'? Como os métodos de análise de domínio e Análise de Redes Sociais (ARS) podem contribuir, como formas de pesquisa na Organização do Conhecimento, para o estudo do domínio do 'discurso de ódio'?

Para responder a essas questões, o presente estudo traça como objetivo geral representar o domínio do #StopHateforProfit em seus aspectos sociais, informativos e de interesse da opinião pública. Para atingir esse objetivo, são definidos objetivos específicos:

1. Estudar a aplicação dos métodos de análise de domínio e Análise de Redes Sociais, como perspectivas de pesquisa na Organização do Conhecimento, ao estudo do domínio da campanha #StopHateforProfit.
2. Estudar o comportamento temático do discurso de ódio no interesse público a partir de sua representação no discurso jornalístico das pesquisas do New York Times e do Google.
3. Representar e discutir a estrutura social da campanha #StopHateforProfit.
4. Analisar a estrutura temática da campanha #StopHateforProfit.

Em primeiro lugar, este estudo fornece uma estrutura oportuna para a inserção de uma perspectiva de pesquisa crítica que primeiro determina a estrutura social e conceitual gerada na campanha #StopHateforProfit, contando com ferramentas de análise visual. Nesse sentido, esta pesquisa contribui para a identificação, descrição e interpretação das principais diretrizes teóricas que definem o discurso de ódio como domínio de conhecimento e o delineamento metodológico que permite estudar os comportamentos sociais e informacionais em campanhas de protesto contra o discurso de ódio como #StopHateforProfit.

Em segundo lugar, as comunidades que compõem o domínio contra o discurso de ódio são analisadas e representadas sob a perspectiva sociocognitiva e a Análise de Redes Sociais, visando uma compreensão sociocultural para compreender as complexas interações sociais que contribuem para a construção do conhecimento de combate a esse fenômeno. O interesse

e a opinião pública pela campanha são estudados ao longo do tempo, com o objetivo de identificar padrões e tendências da abordagem dessa temática no cenário internacional. Sem dúvida, a Internet oferece inúmeras vantagens para pessoas, empresas, organizações, o estado e a sociedade como um todo, porém, uma de suas características é a falta de controle centralizado, o que implica que aqueles usuários que decidem abusar dessa rede para seus próprios propósitos podem facilmente ter sucesso, oferecendo-lhes muitos caminhos que garantem seu anonimato e privacidade sem nenhum custo. Waldron (2012) não trata o discurso de ódio como a raiz causal da desigualdade social baseada na identidade, mas apenas observa que em sociedades onde o discurso de ódio circula livremente (na Internet, no rádio, em livros e folhetos), torna-se uma característica fixa do meio social.

De acordo com a classificação proposta por Esquivel Alonso (2016), o termo discurso de ódio é utilizado com o claro propósito de se afastar de questões criminais e constitucionais, motivadas por questões étnicas, raciais, religiosas, violência, hostilidade, entre outras. Por outro lado, dada a sua relação com os limites da liberdade de expressão, Rodríguez-Izquierdo (2015) argumenta que a maior parte da literatura se dedica a analisar até que ponto certas mensagens depreciativas e ofensivas, dirigidas aos grupos diferenciados em razão de sua raça, religião, sexo ou orientação sexual, são constitucionalmente permitidos.

Waldron (2012) afirma que a publicação de discursos de ódio tem como principal objetivo “comprometer a dignidade daqueles a quem se dirige, tanto aos seus próprios olhos como aos olhos de outros membros da sociedade”. Brow (2017) nesse sentido argumenta que tal discurso leva a problemas de dano, dignidade, segurança, diálogo cultural saudável, democracia e legitimidade, para citar os mais relevantes. O discurso de ódio é então usado como linguagem pejorativa ou discriminatória com referência a uma pessoa ou grupo com base em quem eles são, ou seja, sua religião, etnia, nacionalidade, raça, cor, ascendência, gênero ou outro fator de identidade. Isso geralmente está enraizado e gera intolerância e ódio e, em certos contextos, pode ser humilhante e excludente.

Com base na proposta de Waldron (2012), o pesquisador Marques (2020) defende que se trata de um “ataque sistemático à segurança da dignidade de alguém”. Este autor não fala apenas sobre a finalidade do discurso de ódio, mas também sobre a natureza sistemática de

seus efeitos (MAITRA, 2009). Nessa perspectiva, Marques também afirma que declarações que expressam emoções odiosas episódicas não são discursos de ódio, pois não atendem à condição de sistematicidade. O ódio como uma emoção episódica "se desenvolve quando outros maltratam ou humilham alguém, ou cujas ações deliberadas se tornam um obstáculo aos objetivos de alguém". Especificamente, Marques defende que nem todas as expressões deste tipo de emoção são reconhecidas como discurso de ódio e que o ódio pode ser expresso mesmo quando o emissor não possui esse sentimento. O ódio, por um lado, pode ter um caráter episódico, dirigido a alguém que é percebido como malicioso, e por outro, também pode se tornar um sentimento de longo prazo, principalmente em contextos intergrupais. Em sua análise, a autora conclui que o discurso de ódio é ilocucionalmente expressivo e pressupõe o ódio contínuo como um sentimento que "organiza o mundo social das pessoas".

O estudo tem como foco a análise de domínio (HJØRLAND, 2002, 2018) e a análise de redes sociais como métodos nos quais a pesquisa se baseou para contribuir para a representação de um domínio marcado pelo ativismo e uma atividade social que complementa o acadêmico que está representado no sites de redes sociais na Internet. O caráter da pesquisa é definido como descritivo considerando os métodos de pesquisa utilizados no mapeamento da estrutura social e o uso de técnicas de processamento de linguagem natural para atingir seu objetivo central.

Seguindo essa perspectiva, possíveis vieses são minimizados pela integração de diferentes métodos de pesquisa e perspectivas teórico-disciplinares para abordar o domínio da campanha de boicote #StopHateforProfit. Considera-se pelo autor citado que para compreender um domínio de forma abrangente, recomenda-se que os investigadores utilizem mais do que uma abordagem de análise de domínio, uma vez que um estudo enquadrado neste método só é considerado se forem consideradas as diferentes teorias ou “paradigmas” do domínio em profundidade. Nessa linha, o presente estudo contribui para a compreensão do domínio da campanha #StopHateforProfit por meio da análise de seu ativismo no Twitter durante o ano de 2020 na perspectiva da Análise de Redes Sociais. Além disso, são descritas as estratégias, técnicas e ferramentas de processamento de linguagem natural utilizadas para o processamento e análise dos dados extraídos do Twitter.

Para analisar o interesse do público em relação à campanha, os dados fornecidos pela ferramenta Google Trends são escolhidos como primeira fonte, que, por meio de algoritmos, capta o interesse global em determinado tema (MOURA, 2009). No caso da estratégia de busca da campanha, foi selecionado o termo “Stop Hate for Profit” e o período de 01/06/2020 a 31/11/2020. Em um segundo momento, utilizou-se como fonte a base de dados do jornal NYTimes utilizando a mesma estratégia de busca cujos sete artigos recuperados foram processados com o software de mineração de dados Orange. Os 25.480 tweets e 121.291 retuítes para um total de 146.771 tweets foram coletados por meio de uma chamada à API do Twitter com a linguagem R, os quais foram classificados com o hashtag da campanha #StopHateforProfit e serviram como fonte de dados para a Análise de Redes Sociais. Esse método de pesquisa nos permitiu determinar a estrutura social das comunidades #StopHateforProfit. Seu uso justifica-se por medir, descrever e analisar as variáveis relacionais sociais de um conjunto de atores para representar a estrutura desse grupo (WASSERMAN; FAUST, 1994).

Nesta perspectiva, a ARS ofereceu a esta investigação um conjunto de conceitos e métodos analíticos para identificar e analisar os tipos de relações existentes entre as diferentes entidades que constituem o domínio participante nesta campanha. Segundo Wasserman e Faust (1994), os padrões nas relações estabelecidas entre os atores que compõem esse domínio podem ser estudados a partir da análise de suas estruturas, portanto, neste ponto buscamos identificar semelhanças e papéis nas relações entre os atores envolvidos na campanha.

Os resultados ilustram como a rede social internacional, construída por meio de relacionamentos de interação neste site de rede social, ajudou o Stop Hate for Profit a alcançar o sucesso como uma campanha de boicote do Facebook e rapidamente crescer para uma ampla escala internacional poucos dias após o lançamento. Confirmou-se a estreita relação entre os temas de interesse público e os temas centrais da cobertura jornalística internacional.

A campanha não só despertou o interesse dos internautas sobre o significado da campanha, suas demandas e o fenômeno do discurso de ódio, mas também sobre o

significado do fenômeno financeiro que está na base de sua demanda, dizem as implicações do boicote publicitário do Facebook. A mineração de dados de notícias produziu essa relação causal com as principais preocupações de interesse público sobre empresas relacionadas a campanhas, como Starbuck e Unilever.

A narrativa geral no Twitter em torno da campanha mostrou mensagens positivas para conscientizar as principais empresas que anunciam no Facebook sobre o discurso de ódio e a desinformação como fenômenos nocivos à democracia, e também como a opinião pública fez uso de vocabulário carregado de sentimentos negativos com prevalência de emoções de medo para alertar sobre as consequências que esse fenômeno causa à sociedade.

Palavras-chave: Stop Hate for Profit, Facebook, Discurso de ódio, Twitter, Análise de Redes Sociais, Análise de Domínio, Análise de Sentimentos.

ABSTRACT

This study addresses the phenomenon of misinformation and the spread of hate speech on social networking sites as a prelude to the #StopHateforProfit Facebook boycott campaign. It exposes the conceptual theoretical principles of hate speech, as well as the disciplinary approaches present in international literature. It also describes domain and social network analysis as the research and how these contribute to the representation of a domain marked by activism and a social activity. These are complementary to the academic ones represented on Internet social networking sites. The strategies, techniques and natural language processing tools used for the processing and analysis of the data extracted from Twitter are described. The results illustrate how the international social network built through interactional relationships on this social networking site gave “Stop Hate for Profit” a successful spread to the Facebook boycott campaign and rapidly grew to a wide international scale within days of its launch. The general narrative on Twitter around this campaign showcased positive messages to raise awareness among the main companies which advertise on Facebook presenting hate speech and disinformation as harmful phenomena to democracy. Likewise, it demonstrates how public opinion made use of loaded vocabulary with negative feelings, emotions where fear prevailed to alert the public about the consequences that this phenomenon causes to society.

Keywords: Stop Hate for Profit; Facebook; Hate speech; Twitter; Social Network Analysis; Domain Analysis; Sentiment Analysis.

LISTA DE FIGURAS

Figura 1. Cobertura temática sobre ARS en Ciencias de la Información de Brasil.....	33
Figura 2. Flujo de la minería de datos de los artículos del NYTimes en Orange.	60
Figura 3. Script de recuperación de tuits #StopHateforProfit.....	66
Figura 4. Flujo de Minería de datos de los retuits en #StopHateforProfit	78
Figura 5. Tuit de mayor difusión en la campaña #StopHateforProfit.....	84
Figura 6. Tuit de Sleeping Giants del 27/06/2022.	85
Figura 7. Regiones geográficas de interés y alcance público en Stop Hate for Profit	89
Figura 8. Cobertura temática de los términos sobre #StopHateforProfit en NYTimes.	92
Figura 9. Personalidades, empresas y organizaciones sin fines de lucro firmantes.	96
Figura 10. Grafo de las cinco principales comunidades en #StopHateforProfit.....	100
Figura 11. Subgrafo de los nodos con un valor de grado entrante > 1.000	102
Figura 12. Subgrafo de los nodos con un valor de grado saliente > 50	105
Figura 13. Nodos en #StopHateforProfit con grado saliente = 0.....	107
Figura 14. Red de los 100 actores más influyentes en #StopHateforProfit.	108
Figura 15. Actores de #StopHateforProfit con valor de grado saliente = 0.....	109
Figura 16. Red de centralidad de intermediación.	111
Figura 17. Términos más frecuentes en tuits de la campaña. Fuente: elaboración propia en Wordart.com con datos de pesquisa.....	114
Figura 18. Términos más frecuentes en los retuits de la campaña #StopHateforProfit. Fuente: elaboración propia en Orange con datos de pesquisa.	114
Figura 19 Regiones que fueron temas de debate en #StopHateforProfit (retuits)	116
Figura 20. Mapa MDS mostrando la probabilidad de tópico marginal de los tópicos descubiertos en #StopHateforProfit utilizando LDA.....	119
Figura 21. Ejemplo de retuit con mayor adherencia al tópico 1	120
Figura 22. Distribución de sentimientos y emociones para #StopHateforProfit en 2020.....	122
Figura 23. Mapa de calor de la clasificación de sentimientos en los retuits.....	126
Figura 24. Ejemplo de visualización de retuits con mayor puntuación negativa del mapa de calor.....	127

LISTA DE TABLAS

Tabla 1. Tipología de discursos del odio.	46
Tabla 2. Diez grupos de leyes/reglamentos/códigos que restringen el discurso del odio.	49
Tabla 3 Métricas y estadísticas de Gephi para análisis de la red como un todo.	68
Tabla 4. Temáticas populares relacionadas al interés público por Stop Hate for Profit.	90
Tabla 5 Términos prominentes sobre #SobreHateforProfit en NYTimes.	93
Tabla 6. Medidas de red de #StopHateforProfit.	97
Tabla 7. Tamaño de las cinco mayores subcomunidades en #StopHateforProfit.	101
Tabla 8. Relación de los 30 nodos con mayor valor de grado saliente.	106
Tabla 9. Medidas de la red formada por los nodos más influyente #StopHateforProfit.	108
Tabla 10. Relación de nodos con mayor valor de intermediación.	111
Tabla 11 Actores con mayor valor de grado de vector propio.	112
Tabla 12. Relación de los cinco tópicos latentes principales que describen el debate de #StopHateforProfit.	118
Tabla 13. Script de R con la función 'reemplazar' para sustituir términos en el corpus.	121
Tabla 14. Clasificación de emociones y sentimientos de 30 principales tuits con mayor peso positivo.	123
Tabla 15. Cinco principales tuits con mayor ocurrencia de vocabulario positivo.	123
Tabla 16. Clasificación de emociones y sentimientos de vocabularios extraídos de los cinco tuit con mayor con mayor valor positivo.	124
Tabla 17. Principales tuits con mayor ocurrencia de vocabulario negativo.	124
Tabla 18. Clasificación de emociones y sentimientos de vocabularios extraídos de los cinco tuit con mayor con mayor peso negativo. Fuente: elaboración propia a partir de datos extraídos con RStudio.	125

LISTA DE GRÁFICOS

Gráfico 1. Evolución temporal de los tuits orgánicos emitidos #StopHateforProfit: 25/06-06/08/2020	82
Gráfico 2. Evolución temporal de los tuits y retuits emitidos #StopHateforProfit: 25/06-06/08/2020	83
Gráfico 3. Comportamiento temporal del interés y alcance público Web de Stop Hate for Profit.	88
Gráfico 4. Distribución de las comunidades de #StopHateforProfit según su tamaño.	99

LISTA DE SIGLAS

AD – Análisis del Discurso

ADL – Anti-Defamation League

ARS – Análisis de Redes Sociales

BDTD – Biblioteca Digital Brasileña de Tesis y Disertaciones

BRAPCI – Base de Dados em Ciência da Informação

CI – Ciencia de la Información

NAACP – National Association for the Advancement of Colored People

OC – Organización del Conocimiento

ONU – Organización de Naciones Unidas

OSP – Proveedores de servicios en línea

REBIUN – Red de Bibliotecas Universitarias

TESEO – Base de datos de Tesis Doctorales de España

OSP – Proveedores De Servicios En Línea (del inglés)

LDA – Latent Dirichlet Allocation

TABLA DE CONTENIDO

1. INTRODUCCIÓN	25
1.1 Objeto y problema de la investigación	27
1.2 Objetivos y justificación de la investigación	31
1.3 Estructura capitular	35
2. DISCURSOS DE ODIOS: CONTEXTO HISTÓRICO Y TEÓRICO	37
2.1 Análisis conceptual de ‘discursos de odio’ y términos relacionados	41
2.2 Dimensiones del discurso del odio	42
2.3 Regulación legislativa	43
3. PROCEDIMIENTOS METODOLÓGICOS	50
3.1 Análisis de dominio	50
3.2 Análisis del interés público	53
3.2.1 Fuente de datos de opinión pública	58
3.3 Análisis de Redes Sociales	63
3.3.1 Fuente de datos para el Análisis de Redes Sociales	65
3.3.2 Tipo de matriz	66
3.3.3 Medidas de Análisis de Redes Sociales	66
3.3.3.1 Análisis para toda la red	66
3.3.3.2 Análisis de la red a partir de sus actores	68
3.3.4 Pre-procesamiento de los datos obtenidos de Twitter	70
3.4 Análisis de sentimientos	71
3.4.1 Niveles y enfoques del análisis de sentimientos	73
3.4.2 Fase de limpieza de datos	75
3.4.3 Los paquetes y funciones para el análisis de sentimientos	76
3.5 Análisis de sentimientos para los retuits	77
3.5.1 Flujo de Trabajo	77
4. RESULTADOS DEL ANÁLISIS DE LA CAMPAÑA STOP HATE FOR PROFIT	81
4.1 Stop Hate for Profit e interés público	81
4.1.1 Alcance público de la campaña #StophateforProfit en Google Trends	87
4.1.2 Alcance público de la campaña #StophateforProfit en The New York Times	91
4.2 Análisis de Redes Sociales de la campaña #StopHateforProfit	95
4.2.1 Actores participantes	95

4.2.2 Características de la estructura de red.....	96
4.2.3 Centralidad de grado	101
4.2.3.1 Actores centrales: grado entrante.....	102
4.2.3.2 Actores centrales: grado saliente.....	104
4.2.3.3 Subred de los actores más influyentes: centralidad de grado.....	107
4.2.4 Centralidad de intermediación	110
4.3 Análisis temático de la campaña #StopHateForProfit.....	113
4.4 Análisis de sentimientos en #StopHateforProfit.....	120
CONCLUSIONES.....	128
REFERENCIAS BIBLIOGRÁFICAS	131

1. INTRODUCCIÓN

Los avances en Internet y los sitios de redes sociales han propiciado el desarrollo de un nuevo ambiente de actuación de la sociedad en el que su potencial democratizador constituye un catalizador en la dinámica comunicativa e informacional de los actores sociales. El uso extendido de los sitios de redes sociales (ej. Facebook, Twitter, Instagram, etc.) en la cotidianidad humana tiene eco en el aumento de la participación democrática de sus usuarios en aspectos sociales, políticos y organizativos, así como en la creación de narrativas y movimientos de acción colectiva.

Los actores políticos se comunican directamente con los ciudadanos. Casi en tiempo real periodistas e investigadores monitorean la difusión de noticias, estudian la formación y dinámica de opiniones. El desarrollo de Internet y ha propiciado que los ciudadanos tengan la oportunidad de participar de formas activas en el debate y de que puedan presentar sus propias ideas directamente a los actores involucrados en los procesos de toma de decisiones. Por tanto, en este escenario es necesario que se difunda una información confiable, ya que la ausencia de confiabilidad en la información (y la desinformación) traería implicaciones negativas para el funcionamiento de la democracia (KALSNES, 2018; CAPOZZI et al, 2018).

En estos escenarios existe entonces el riesgo de conflicto entre la libertad de expresión y la injuria al prójimo, la cual puede adoptar diversas formas, como por ejemplo, la difamación o la discriminación, siendo usadas para ejercer la libertad de manifestar odio hacia el otro.

A pesar de que la desinformación y los discursos de odio proceden históricamente del ambiente social físico, ambos son fenómenos que se refuerzan en este nuevo ecosistema socio-tecnológico. Por una parte, cada nueva tecnología de comunicación permite nuevas formas de manipular y amplificar la desinformación para las personas y las sociedades (KALSNES, 2018), lo cual se extiende igualmente a las crecientes manifestaciones de expresiones de odio e intolerancia.

Gu, Kropotov y Yarochkin (2017) afirman que estos fenómenos son motivados por un problema social mayor: la manipulación de la opinión pública para afectar a la sociedad. En

los últimos años, la desinformación y el discurso de odio difundidos en línea han provocado crecientes daños físicos, estos sin limitaciones geográficas o tipo de ambiente tecnológico. Dadas las características de sus herramientas de difusión informacional, estas tecnologías resultan herramientas eficaces para la expresión, socialización de opiniones ofensivas y espacios de acción organizada de grupos extremistas.

El odio es descrito por Royzman, McCauley y Rozin (2005) como el fenómeno afectivo más destructivo en la historia de la naturaleza humana. Es precisamente la cierta impunidad en la comunicación que brinda el anonimato en estas plataformas sociales lo que ha llevado a su más amplia proliferación.

Revenge Sánchez (2015) afirma que la expresión en español ‘discursos del odio’ es una traducción literal de “*hate speech*” y se encuentra firmemente asentada en la literatura académica. El autor advierte la falta de congruencia que supone utilizar el término ‘discurso’ para contextos que hacen alusión exclusivamente a contenidos irracionales, cargados de una clara intención injuriosa que busca provocar una afectación en la dignidad de un grupo de personas a través de expresiones hirientes. Señala que el Consejo de Europa ha definido el discurso del odio como “aquél que cubre todas las formas de expresión que extiendan, incitan, promuevan o justifican el odio racial, la xenofobia, el anti-semitismo o cualquier otra forma de odio basada en la intolerancia” (RAVENGA SÁNCHEZ, 2015, p.13).

Al ser un fenómeno extremadamente dañino para la sociedad, por su relación directa con la ejecución de actos de violencia, suicidios o la vulneración de determinados derechos humanos, el discurso del odio ha dado lugar a la necesidad de su estudio crítico y detección automatizada en los ambientes virtuales. La identificación de este tipo de comunicación y manifestación prejuiciosa en el ambiente digital resulta efectiva para identificar grupos de personas en riesgos, pero esta tarea se complejiza por el fenómeno de Big Data, fenómeno que ha conducido a un desarrollo acelerado de técnicas algorítmicas para su detección y posterior control o eliminación.

Todo lo anterior mencionado conlleva a que no exista una definición universalmente aceptada del término “discurso de odio” y las discusiones teóricas sobre el tema se encuentren lejos de concluir.

1.1 Objeto y problema de la investigación

Los debates sobre el discurso del odio en los últimos años han ganado mayor espacio en el discurso científico, configurándose como áreas muy productivas en su estudio la ciencia de la comunicación, las ciencias políticas, las ciencias del comportamiento como la sociología, la psicología, la lingüística, la filosofía, y recientemente extendiéndose hacia áreas menos tradicionales en su estudio, como la ciencia de la información y ciencia de la computación/informática.

La pluralidad de aspectos relacionados con los discursos de odio en la web obtiene mayor atención en los estudios científicos, llevando a un rápido crecimiento de la producción intelectual sobre el tema en los últimos años y abriendo cada vez más espacio en los debates de la opinión pública en general, lo que muestra una fuerte conexión entre ellos.

Varias investigaciones han tendido a examinar las fronteras entre las manifestaciones de odio y su relación directa con el derecho democrático de libertad de expresión (CAMMAERTS, 2009; WALDRON, 2012; BLEICH, 2014; BUYSE, 2014; GREENE, SIMPSON, 2017). En algunas se discuten las prácticas y los discursos odio efectuados por grupos o individuos mediante el uso de Internet (CAMMAERTS, 2009; CITRON & NORTON, 2011; BURNAP; WILLIAMS, 2015; COHEN-ALMAGOR, 2018). Estudios emergentes analizan el uso de discursos de odio para promover crímenes y violencia de género y la etnoviolencia (KAPRALSKI, 2016; COLLINS, 2017; COHEN-ALMAGOR, 2018; VELDHUIS et al., 2018). Recientemente Tontodimamma et al. (2021) han considerado los treinta años de investigación sobre la temática, desde una perspectiva bibliométrica que mapeó los temas de interés y su evolución en el discurso científico y aplicando métodos de análisis de co-palabras. Sin embargo no existen trabajos que consideren el método del análisis de dominio para estudiar el discurso de odio no solo enfocando al sistema de conocimiento científico, sino también a las complejas interacciones sociales no científicas que tributan a la construcción del conocimiento del discurso del odio como dominio.

Puerta-Díaz, Ovalle-Perandones y Martínez-Ávila (2020) demuestran el crecimiento ininterrumpido y elevado de publicaciones sobre discurso de odio desde el año 2013 (siendo 2018 el año más productivo) y advierten que los autores de las diferentes áreas argumentan

que este crecimiento y tratamiento multidisciplinar en la literatura académica es una evidencia de las preocupaciones internacionales por combatir estas manifestaciones que generan delitos de odio, al mismo tiempo que reflejan una respuesta científica y social a los tumultuosos acontecimientos políticos en diversos ámbitos.

Por otro lado, en el discurso periodístico, Trindade (2020) identifica un aumento significativo de la expresión “discurso de odio” entre los años 2012-2018 e identifica que este pasó a destacarse específicamente a partir del 2012 – año en que Facebook también alcanzó la marca de un billón de usuarios activos mensuales.

Ante el descrito escenario en el entorno digital global, tanto el Consejo de Derechos Humanos de Naciones Unidas como la Comisión Europea contra el Racismo y la Intolerancia, han expresado su preocupación por el aumento del discurso de odio en el escenario digital y el alto riesgo de los efectos de su trivialización. En 2019 en el informe “La Estrategia y Plan de Acción de las Naciones Unidas para la lucha contra el discurso de odio” se afirma que se están explotando los medios sociales para promover la intolerancia, a la vez se alerta del avance de los movimientos neonazis a favor de la supremacía blanca y se confirman las dimensiones de este fenómeno a escala social planteadas por Gu, Kropotov y Yarochkin en su estudio del 2017. Para la ONU el discurso público llega a transformarse en un arma para cosechar ganancias políticas con una retórica incendiaria que estigmatiza y deshumaniza a las minorías, los migrantes, los refugiados, las mujeres y todos aquellos etiquetados como “los otros”.

Los peligros que enfrentaba la democracia con el aumento del índice de alertas de discursos de odio, en el contexto mundial de pandemia por el nuevo coronavirus a inicios del año 2020, alarmó a organizaciones e instituciones de defensa de los derechos civiles, como la Common Sense Media, Free Press, Anti-Defamation League (ADL) y la National Association for the Advancement of Colored People (NAACP), sobre su propagación en la plataforma de Facebook mediante anuncios publicitarios. Esta situación conlleva a que las organizaciones se unieran en una coalición y lanzaran la campaña Stop Hate for Profit el 17 de junio de 2020.

La ADL es una organización líder en lucha contra el odio, fundada en 1913 en respuesta al creciente clima antisemita e intolerante, y siendo su misión troncal proteger al

pueblo judío y garantizar la justicia y el trato justo para todos. Actualmente su alcance es amplio en la lucha contra todas las formas de odio, siendo incluso considerada líder mundial en exponer el extremismo, brindar educación contra los prejuicios y capacitación de las fuerzas del orden (ADL, 2021).

La campaña #StopHateforProfit tuvo un objetivo claro que fue el de hacer un llamamiento internacional a las grandes y pequeñas empresas para que pausaran su publicidad pagada en Facebook. De esta forma se enviaba un mensaje a la compañía para que implementase nuevas iniciativas que favorecieran dinámicas orientadas a eliminar la difusión del discurso del odio y la desinformación de su plataforma. Este boicot es un reclamo a Facebook, a asumir su rol en la responsabilidad social corporativa y a orientarse hacia la protección de los derechos civiles de la sociedad. Esta campaña surge para exigir a las empresas anunciantes en Facebook retirar su publicidad como vía para presionar a la compañía a tomar medidas contra los grupos centrados en la supremacía blanca, la milicia, el antisemitismo, la islamofobia y las conspiraciones violentas que constituye una amenaza para los valores democráticos, la estabilidad social y la paz (ONU, 2019).

Los miembros de la coalición organizadora de #StopHateforProfit se unieron no sólo a grandes corporaciones sino también a pequeñas empresas, organizaciones sin fines de lucro, así como a reconocidas personalidades públicas y grupos de justicia social, derechos humanos y trabajo, organizaciones religiosas e innumerables personas.

El principal objetivo de #StopHateforProfit fue promover la paralización de toda la publicidad en Facebook con el fin de bloquear sus ingresos y el aumento de las acciones de la empresa en la bolsa. Cabe mencionar que Facebook es la segunda plataforma publicitaria más grande del mundo, solo por detrás de Google, con unos ingresos publicitarios anuales estimados en 69,7 mil millones de dólares. El éxito de esta campaña protegió la democracia de la sociedad americana en el contexto de las elecciones presidenciales de los Estados Unidos del mismo año. En este escenario, la coalición demanda a la compañía de Facebook lo siguiente (ADL, 2021):

- Eliminar grupos centrados en la supremacía blanca, la milicia, el antisemitismo, la islamofobia y las conspiraciones violentas.

- Aumentar los recursos enfocados en el monitoreo de grupos que incitan el odio y la violencia.
- Cambiar la política de la plataforma para prohibir cualquier página de evento con un llamado a las armas, como lo destacan *Muslim Advocates* y la coalición *Change the Terms*.
- Dedicar el 5% de sus ingresos anuales a un fondo administrado de forma independiente para apoyar iniciativas, académicos y organizaciones que luchan contra el racismo, el odio y la división causada por la inacción de Facebook.

Si bien se conoce que a solo un mes de iniciada la campaña más de 1.100 empresas y algunos de los anunciantes líderes del mundo se unieron y suspendieron sus anuncios de Facebook, se desconoce la representatividad desempeñada por la mayor parte de los inversores publicitarios, los millones de pequeños anunciantes de la plataforma, así como de sus figuras políticas. Este fenómeno no solo llevó a una masiva participación de empresas en este movimiento activista contra el discurso del odio, sino que se estima que generó por detrás un fenómeno social e informacional que demanda ser develado, pues la participación activa o pasiva en la campaña se estima propicie la formación de comunidades a partir de las interacciones de mención o intercambio de información entre participantes que usaron el hashtag de la campaña. Tampoco queda claro cuales narrativas resultaron de mayor impacto al éxito o no de la campaña y si existió un liderazgo de opinión en la misma más allá de los miembros de la coalición y los temas de interés público que se formaron y difundieron a partir de la participación en la campaña.

Ante este escenario, surgen las siguientes preguntas de investigación:

¿Cuáles son los principales delineamientos teóricos del dominio de conocimiento del ‘discurso de odio’?

¿Cómo pueden contribuir los métodos de análisis de dominio y Análisis de Redes Sociales, como formas de investigación en la Organización del Conocimiento, al estudio del dominio del ‘discurso del odio’?

¿Cómo se comportan en el tiempo las necesidades informacionales sobre *Stop Hate For Profit* reflejadas a partir del interés y opinión pública internacional sobre el discurso del odio?

¿La estructura social participante en la campaña *Stop Hate For Profit* es lo suficientemente fuerte como para ser representativa del estudio de movimientos contra el discurso del odio en los sitios de redes sociales?

¿Cómo se describe *Stop Hate For Profit* a partir de la representación temática de la opinión pública participante?

1.2 Objetivos y justificación de la investigación

Los objetivos de la presente investigación presentados a continuación buscan dar respuesta a estas preguntas de investigación.

Objetivo general

Representar el dominio de #StopHateforProfit en su vertiente social, informacional y de interés de la opinión pública.

Para alcanzar ese objetivo principal, se plantean un grupo de objetivos específicos, que son:

5. Estudiar la aplicación de los métodos de análisis de dominio y Análisis de Redes Sociales, como perspectivas de investigación en la Organización del Conocimiento, al estudio del dominio de la campaña #StopHateforProfit.
6. Estudiar el comportamiento temático del discurso del odio en el interés público a partir de su representación en el discurso periodístico del New York Times y las búsquedas en Google.
7. Representar y discutir la estructura social de la campaña #StopHateforProfit.
8. Analizar la estructura temática de la campaña #StopHateforProfit.

Justificación de la investigación

Con el objetivo de identificar estudios que se aproximen al análisis de este fenómeno de movimiento de protesta contra el discurso del odio y la desinformación ocurrido en 2020, desde los preceptos de la Ciencia de la Información (en adelante CI), se realizó una búsqueda en estudios de posgrado en Brasil y España, contextos donde se desarrolla gran parte de la investigación. La consulta bibliográfica recuperó tesis de doctorado y disertaciones sobre el discurso del odio, sin embargo no se obtuvieron evidencias del enfoque de dicha campaña como objeto de estudio.

En la Biblioteca Digital Brasileña de Tesis y Disertaciones (BDTD) están disponibles 97 investigaciones sobre el discurso del odio como fenómeno, encuadradas en las áreas del Derecho, Comunicación y Semiótica, Psicología, Ciencias Sociales, Lingüística y Computación, sin embargo no se identificaron enfoques desde los presupuestos de la Ciencia de la Información. Por otro lado, en la BDTD existe un total de 502 estudios que aplican o teorizan sobre el Análisis de Redes Sociales, de los cuales 16 son procedentes de programas de posgraduación en Ciencia de la Información. La figura 1 representa los temas principales de dichas investigaciones, donde el tamaño del término representa el total de trabajos con adherencia a la misma. Como muestra la siguiente figura 1, estos se enfocan en el ARS como método para la CI integrado sobre todo al área de los estudios métricos de la información, específicamente en desvelar redes de colaboración en el discurso científico.



Figura 1. Cobertura temática sobre ARS en Ciencias de la Información de Brasil.
Fuente: Elaboración propia a partir de datos de BDTD. Elaborado en <http://www.word.art>

Para identificar aquellos trabajos con interconexión entre ambos temas fueron recuperados 6 disertaciones y una tesis de doctorado en la BDTD, de los cuales una tesis de maestría aplica como método el Análisis de Redes Sociales. Cappi (2017), en su trabajo debate sobre las dinámicas de interacción en Twitter en los nuevos entornos del debate político brasileiro, tomó como caso de estudio las publicaciones sobre la condena del expresidente de Brasil Luis Inácio Lula da Silva para demostrar la existencia de las burbujas ideológicas que sugieren cómo la construcción de identificaciones mutuas entre estos grupos de usuarios que participan en esta conversación digital termina por autorizar el discurso sistemático de la falta de respeto a la dignidad.

Por otro lado, en la *Base de Dados em Ciência da Informação* (BRAPCI), que constituye la mayor base de datos de la Ciencias de la Información en Brasil, no se recuperaron estudios sobre la temática con enfoques en el estudio de los flujos informacionales generados a partir de interacciones sociales en movimientos de activistas contra el discurso del odio en general. Los trabajos publicados en el área de la CI sobre el discurso del odio enmarcan su alcance en la disciplina de gestión, la mediación y el uso de la información. Si bien una de las tres propuestas busca la representación de la información de

expresiones de odio para la construcción de una ontología sobre discurso de odio, no considera la perspectiva en el enfoque de analizar las interacciones.

Un resultado similar a la BDTD fue identificado en el contexto español en una búsqueda de información realizada en el Catálogo Colectivo de REBIUN (Red de Bibliotecas Universitarias) y TESEO (Base de datos de Tesis Doctorales de España). En estas plataformas no se identificaron tesis de maestría o de doctorado sobre el tema vinculado al campo de la CI o con el enfoque de estudio de dominio de conocimiento.

Estando esta investigación enmarcada en la línea de investigación del Programa de Postgrado en Ciencia de la Información de la UNESP Producción y Organización de la Información, como subárea de la CI que se ocupa de la representación de la información y el conocimiento, el presente estudio se interesó por develar la estructura social de las comunidades discursivas del dominio de conocimiento de *Stop Hate for Profit*. Para ello, se pretende analizar el estado de opinión e interés público internacional sobre #StopHateforProfit, propia de líneas de investigación del campo de la Ciencias de la Información, que promueva el desarrollo de diversas iniciativas en el plano político, académico, entre otros, que ayude a contrarrestar sus implicaciones nocivas del discurso del odio en la sociedad. Los resultados del presente trabajo son altamente relevantes, ya que permitirán replicar la metodología delineada para determinar el estado y evolución de la opinión pública en relación a los discursos de odio en contextos globales.

Por lo anterior expuesto, la presente investigación contribuye al desarrollo del campo de estudio en dos vertientes:

Primeramente, supone un marco oportuno para la inserción de una perspectiva crítica de investigación que determina primeramente la estructura social y conceptual generada en la campaña #StopHateforProfit apoyándose en herramientas de analítica visual. En este sentido esta investigación tributa a la identificación, descripción e interpretación de los principales delineamientos teóricos que definen al discurso del odio como dominio de conocimiento y el delineamiento metodológico que permita estudiar los comportamientos sociales e informacionales en campañas de protesta contra el discurso del odio como lo es #StopHateforProfit.

En segundo lugar, se analiza y representa comunidades que componen el dominio contra el discurso del odio bajo la perspectiva socio-cognitiva y del Análisis de Redes Sociales, visando un entendimiento sociocultural para comprender las complejas interacciones sociales que tributan a la construcción del conocimiento de lucha contra este fenómeno. Se estudia el interés y opinión pública en la campaña lo largo del tiempo, objetivando identificar patrones y tendencias sobre esta cobertura temática en el escenario internacional.

1.3 Estructura capitular

En la sección de la **Introducción** se contextualizó el problema de investigación a partir del fenómeno de la desinformación y propagación del discurso del odio en los sitios de redes sociales y cómo surge la campaña objeto de estudio en respuesta a ello. Se justifica la investigación a partir del estado de su estudio en el contexto científico brasileiro y español a nivel de posgrado y se asocian elementos disciplinares y estrategias metodológicas empleadas por estudiosos del discurso del odio y movimientos de protesta. Se define el encuadramiento disciplinar de su abordaje teórico y se exponen los criterios que justifican su análisis. Los objetivos y preguntas de investigación que guían esta investigación son expuestas en este epígrafe.

En la **sección 2** se exponen los principios teóricos conceptuales del discurso del odio así como los enfoques disciplinares y metodológicos desde su principal aparición en bases de datos internacionales. Además se discuten los principales intereses temáticos de la comunidad científica en su estudio y las tipologías de discursos de odio conocidas.

La **sección 3** describe los métodos de análisis de dominio como teoría y un enfoque de la Bibliotecología y Ciencia de la Información y la Organización del Conocimiento y Análisis de Redes Sociales, técnicas y herramientas seleccionadas para desarrollar el análisis de dominio de la campaña boicot #StopHateforProfit. La investigación se respalda en una metodología mixta tanto cualitativa como cuantitativa propia de los estudios de análisis de dominio en la organización del conocimiento en escenarios de opinión pública en sitios de redes sociales. Se describe las estrategias, técnicas y herramientas tecnológicas para la

recuperación de los mensajes que definen la conversación de #StopHateforProfit y el flujo que define el procesamiento de minería de datos de Twitter.

En la **sección 4** se analizan los resultados de la estrategia metodológica empleada para el análisis de dominio de la campaña de #StopHateforProfit, los cuales son discutidos a partir de la integración metodológica empleada para su análisis.

La **sección de las conclusiones** sistematiza las discusiones finales de la investigación.

2. DISCURSOS DE ODIOS: CONTEXTO HISTÓRICO Y TEÓRICO

El principio rector de Internet es la libertad, los internautas son libres de ejercer su autonomía expresándose y publicando ideas, contenido y opiniones en la Web (COHEN-ALMAGOR, 2015). El resultado de su desarrollo, su naturaleza flexible y multipropósito hacen de Internet un sistema de comunicación único y sin precedentes en la historia de la humanidad. La red de redes se trata de comunicación y los inicios del siglo XXI marcó una nueva etapa en su evolución: el lanzamiento de los sitios de redes sociales (COHEN-ALMAGOR, 2015).

El primero conocido fue llamado SixDegrees.com y se lanzó en 1997, pero fue a partir de 2003 que ocurrió la revolución y aceptación de la mayoría de los sitios de redes sociales *online* (*online social networks* – OSN, término en inglés) más populares en la actualidad (PALLIS; ZEINALIPOUR-YAZTI; DIKAIKOS, 2011), redes de profesionales (por ejemplo, LinkedIn), redes de amigos (por ejemplo, MySpace, Facebook, Orkut), y redes para o compartir contenido específico como mensajes breves (p. ej., Twitter), diarios y blogs (p. ej., LiveJournal), fotos (p. ej., Flickr), y vídeos (p. ej., YouTube) (DI FELICE, 2012).

Facebook fue lanzado el 4 de febrero de 2004 por Mark Zuckerberg, Eduardo Saverin, Dustin Moskovitz y Chris Hughes. Este sitio, si bien comenzó como una red social para colegas universitarios de Harvard en septiembre de 2006 extendió su uso por miles de usuarios y de forma que en ella era permitido compartir información, fotos, diarios privados, pasatiempos e intereses con sus redes sociales *online*. Plataformas como Twitter permiten a otros usuarios no solo ver lo que se dijo, sino amplificar esas conversaciones con miles de otros usuarios, aun no siendo parte de lo que se discutió inicialmente (SELLARS, 2016).

Sin lugar a duda Internet brinda un sinnúmero de ventajas para las personas, empresas organizaciones, el estado y la sociedad como un todo, sin embargo, una de sus características es la falta de control centralizado, lo que implica que aquellos usuarios que deciden abusar de esta red para sus propios fines pueden tener éxito fácilmente al ofrecerles no pocas vías que garantizan su anonimato y privacidad sin costes.

Taddeo y Floridi (2017) afirman que mientras todos esos avances tecnológicos sigan creciendo cabe profundizar en la naturaleza, las implicaciones menos visibles y su impacto en el bienestar humano y ambiental en la actual era informacional. Una de las expectativas que se espera es el actuar ético de los proveedores de servicios en línea (del inglés *online service provider* – OSP), como Facebook, Google, Microsoft y Twitter, para que dichas compañías realicen sus tareas y alineen sus objetivos con las necesidades de las sociedades con base en principios éticos que deben guiar sus acciones; refiriéndose a los “principios de eficiencia, justicia, equidad y respeto de los valores sociales y culturales vigentes” de McQuail (1992, 47) en vista a preservar los derechos de sus usuarios (MADELIN, 2011; TADDEO; FLORIDI, 2017).

Waldron (2012) no trata el discurso de odio como la raíz causal de la desigualdad social basada en la identidad, más bien solo observa que en las sociedades en las que el discurso de odio circula libremente (en Internet, en la radio, en libros y folletos), este se convierte en una característica fija del entorno social.

Según la clasificación propuesta por Esquivel Alonso (2016), el término discursos de odio se emplea con el claro propósito de alejarse de cuestiones penales y constitucionales, motivado por cuestiones étnicas, raciales, religiosas, violencia, hostilidad, entre otras. Por otro lado, dada su relación con los límites de la libertad de expresión, Rodríguez-Izquierdo (2015) plantea que la mayor parte de la literatura se dedica a analizar hasta qué punto ciertos mensajes despreciativos e hirientes, dirigidos hacia colectivos diferenciados por razón de su raza, religión, sexo o tendencia sexual, son constitucionalmente permisibles.

Esta problemática, aún latente, se traduce en la necesidad de delimitar las fronteras de lo inaceptable e intolerable en una sociedad democrática (ESQUIVEL ALONSO, 2016) y lo que se identifica de un contenido expresado socialmente como de odio, de ahí la necesidad de su abordaje interdisciplinar y no limitado a su análisis jurista.

Para dar respuesta a la pregunta de qué se entiende por discurso del odio, y a pesar de no existir una definición consensuada internacionalmente (STURGES, 2015; UN, 2019), debido en parte a la común indefinición de fronteras entre lo que puede considerarse como expresiones de ‘odio’ y ‘discurso de odio’, así como la polémica en torno a lo que es

considerado como “odioso”, varias definiciones emergen desde la academia y organizaciones intergubernamentales:

Waldron (2012) afirma que la publicación de discursos de odio está dirigida a socavar la dignidad de una persona, siendo su principal objetivo “comprometer la dignidad de aquellos a quienes se dirige, tanto a sus propios ojos como a los ojos de otros miembros de la sociedad”. Brow (2017) en este sentido alega que tal discurso conlleva a problemas de daño, dignidad, seguridad, diálogo cultural saludable, democracia y legitimidad, por mencionar los más relevantes. El discurso del odio utiliza un lenguaje peyorativo o discriminatorio con referencia a una persona o un grupo en función de quiénes son, es decir, de su religión, etnia, nacionalidad, raza, color, ascendencia, género u otro factor de identidad. Esto a menudo está arraigado y genera intolerancia y odio y, en ciertos contextos, puede ser degradante y divisivo.

Marques (2020) apoya su definición del término, “ataque sistemático a la seguridad de la dignidad de alguien”, con base en la propuesta de Waldron (2012) sobre la finalidad del discurso de odio y la condición de sistematicidad de los efectos del discurso de Maitra (2009). Desde esta perspectiva, Marques también afirma que las declaraciones que expresan emociones de odio episódicas no son un discurso de odio, ya que no cumplen con la condición de sistematicidad. El odio como emoción episódica, “se desarrolla cuando otros maltratan o humillan a alguien, o cuyas acciones deliberadas se han convertido en un obstáculo para los objetivos de alguien”. En concreto, Marques sostiene que no todas las expresiones de este tipo de emoción se reconocen como discursos de odio y que se puede expresar odio incluso cuando el emisor carece de ese sentimiento. El odio, por un lado, puede tener un carácter episódico, dirigido hacia alguien que se percibe como malicioso y, por otro lado, también puede convertirse en un sentimiento a largo plazo, particularmente en contextos intergrupales. En su análisis, la autora concluye que el discurso de odio es ilocucionariamente expresivo y presupone el odio continuo como un sentimiento que “organiza el mundo social de las personas”.

Fischer et al. (2018) expresan que cuando las personas experimentan sentimientos de odio, más allá de perseguir intenciones maliciosas e inmorales, este viene acompañado de sentimientos de falta de control o impotencia.

En el plano de la identificación del discurso de odio, según Meyer-Pflug (2009) el concepto de incitación a la discriminación es el elemento fundamental. Sin embargo, la producción de odio también pasa por etapas preparatorias, como el fomento del prejuicio en el sentido de instigar en el grupo dominante “percepciones mentales negativas frente a individuos y grupos socialmente inferiores”.

El discurso de odio busca estigmatizar y “marcar al enemigo” basándose en la segregación. De esta forma, entona un mensaje articulado y atractivo para un determinado grupo que articula las formas de opresión. En general, los sujetos que no están incluidos en el modelo dominante de blanco, hombre, cristiano, heterosexual, de clase media a alta, se convierten en posibles víctimas del discurso de odio (SCHÄFER et al, 2015). Los locutores de este tipo de manifestaciones consideran algunos grupos sociales como inferiores e indignos de una misma ciudadanía y recurren a diferentes estrategias de persuasión principalmente relacionadas con la publicidad y la publicidad para ganar adeptos, a saber:

“la creación de estereotipos, la sustitución de nombres, la selección exclusiva de hechos favorables a su punto de vista, la creación de “enemigos”, la apelación a la autoridad y la afirmación y repetición (...) Aliado a esto, el discurso busca aumentar su probabilidad de aceptación debido al uso de argumentos emocionales y la ausencia de oposición directa e inmediata a tales mensajes. Combinados estos rostros, el que insulta y el que instiga, tenemos que este discurso, además de expresar, busca incrementar la discriminación” (SILVA et al., 2011, p. 448).

El discurso de odio impone obstáculos a las investigaciones y la actuación de los medios de control por cuestiones como el anonimato, múltiples direcciones en un mismo

sitio, la creación de perfiles falsos y páginas de foros privados, además de la falta de preparación de los agentes de investigación con respecto al uso de las nuevas tecnologías.

2.1 Análisis conceptual de ‘discursos de odio’ y términos relacionados

Si llegamos a analizar una de las principales acepciones del discurso, encontramos que esta llega a ser una categoría clave en los desarrollos teóricos de la arqueología epistémica de Foucault, el análisis del discurso anglosajón en el estudio de los procesos de construcción identitarios, la naturaleza lingüística del discurso de Barthes o la semiótica narrativo-discursiva de Greimas. En un sentido “foucaultiano”, podríamos entender un discurso como formaciones discursivas o sistema de dispersión para enunciados, no necesariamente textuales. Diversos autores han señalado que para Foucault un discurso no consiste en textos y el análisis del discurso no simplemente análisis textual ya que los textos, como tales, no están suficientemente delimitados como para funcionar como única base del análisis del discurso (ANDERSEN, 2003; MARTÍNEZ-ÁVILA, 2012).

Por otro lado, la noción de ‘discurso’ ha llegado a popularizarse como evento comunicativo específico. Van Dijk (2001) comenta que estamos en presencia de ello cuando se denota un tipo de discurso como una colección de discursos o de una clase de géneros discursivos (discurso médico, político, racista).

En este mismo sentido para Charaudeau (2009) el discurso no es apenas un lenguaje, su significado depende también de la identidad social de quien habla.

El sentido semiótico es más amplio que la noción de que el término se refiere ‘a cualquier clase de manifestación lingüística’. Van Dijk (2001) amplía este alcance de sus formas de expresión, afirmando que los discursos pueden incluir también expresiones no verbales tales como dibujos, cuadros, gestos, mímica, etc.

El discurso para los fines de la presente investigación adopta este sentido semiótico del estudio del discurso de Van Dijk (2019), que debe abordarse no solo como forma, significado y proceso mental, sino también como estructuras y jerarquías complejas de

interacción y prácticas sociales, incluyendo sus funciones en el contexto, la sociedad y la cultura.

Por otra parte, la noción del término ‘odio’ se define por Ravenga Sánchez (2015, p.37) atendiendo a su significado semántico “como una violenta repulsión o aversión a algo o a alguien que llega hasta el punto de causarle o desear que le ocurra daño”. Agrega el autor que para que pueda ser calificada de expresión de odio, en escenarios jurídicos, el lenguaje ha de ser extremo. Su identificación responde a criterios de forma o al contenido de la expresión. Ravenga Sánchez resalta que debe considerarse esa ‘injuria absoluta’, en cuanto a forma, ya sea esta expresada mediante el lenguaje o la actuación de modo que manifieste un desprecio intolerable al otro, proceso que resulta muy complejo dado su carácter circunstancial e indefinido, sujeto a una multitud de variaciones.

Los planteamientos de Ravenga llevan a retomar el debate de lo que se entiende entonces por ‘discursos de odio’. Llegado este punto, en la literatura podemos encontrar dos vertientes en las tendencias conceptuales que lo definen, una que atiende exclusivamente a su contenido y forma, como se ha abordado en análisis anteriores, y una segunda que considera el contexto histórico-social en el que se emite (CABO ISASI; GARCÍA JUANATEY, 2016).

Se considera aquí el posicionamiento de Brow (2015) cuando afirma que el estudio de discursos de odio debe atender a la triada forma-contenido-contexto en el que ocurre. Es en este sentido que cobra especial importancia el análisis de este fenómeno cuyo alcance debe incluir la perspectiva inferencial de las categorías contextuales en que surge. Este enfoque permitirá determinar el grado del daño que determinadas expresiones, símbolos y manifestaciones pueden llegar a ocasionar, las cuales pueden oscilar, según Cabo Isasi y García Juanatey (2016), desde el emocional hasta la provocación de escaladas de violencia en masa.

2.2 Dimensiones del discurso del odio

Para facilitar el análisis conceptual de la variedad de definiciones que existen sobre discursos de odio, Fortuna y Nunes (2018) sistematizan diferentes elementos: a) el discurso

de odio tiene objetivos específicos; b) el discurso de odio es incitar a la violencia o al odio; c) el discurso de odio es atacar o menospreciar; y d) el humor tiene un estatus específico.

- **El discurso de odio tiene objetivos específicos:** en este punto los autores señalan que todas las definiciones citadas en su trabajo explicitan que tiene objetivos específicos y se basa en características específicas de los grupos, como el origen étnico, religión u otra.

- **El discurso de odio es incitar a la violencia o al odio:** las diversas definiciones analizadas utilizan ligeramente diferentes términos para describir cuándo se produce el discurso de odio. La mayoría de las definiciones señalan que el discurso de odio es incitar a la violencia o al odio hacia una minoría (definiciones del Código de Conducta, ILGA, YouTube, Twitter)

- **El discurso de odio es atacar o menospreciar:** el análisis de otras definiciones establece que el discurso de odio es usar un lenguaje que ataca o menosprecia a los grupos anteriormente (definiciones de Facebook, YouTube, Twitter).

- **El humor tiene un estatus específico:** los autores indican sobre este aspecto que Facebook señala que algunos contenidos humorísticos y ofensivos son permitidos. Alegan que el estatus excepcional del humor hace que los límites sobre lo que está prohibido en la plataforma sean más difíciles de entender. A pesar de las similitudes entre las definiciones, se concluye que hay algunos matices para distinguirlos (por ejemplo, el uso del humor). Además, el análisis de contenido presentado permite sugerirlo para la definición de discurso de odio.

2.3 Regulación legislativa

A pesar que desde la década de 1960 democracias liberales han instituido leyes que penalizan la incitación al odio y los delitos motivados por el odio de forma que limitan la libertad de expresión (BLEICH, 2014), actualmente aún existen desacuerdos en la comunidad internacional sobre su regulación. Según Brugger (2007):

“En la jurisprudencia estadounidense dominante, la libertad de expresión, incluido el derecho a expresar

mensajes de odio, es un derecho prioritario que suele prevalecer sobre los intereses opuestos de dignidad, honor, civismo e igualdad. En los Estados Unidos, el discurso de odio se considera enteramente como una forma de discurso, no como una conducta, a pesar de que ese discurso puede ser realmente doloroso para otros. El derecho internacional y la mayoría de los sistemas legales no estadounidenses brindan mayor protección a la dignidad, el honor y la igualdad de los destinatarios del discurso de odio” (BRUGGER, 2007, p. 118).

Sobre el polémico modelo legislativo americano de Estados Unidos de América, en concreto, Rodríguez-Izquierdo (2015) afirma que el debate doctrinal gira en torno a la amplia protección de la libertad de expresión favorecida por la primera enmienda a la constitución americana sin interferencia o restricción del gobierno. De esta forma resulta casi imposible encontrar una justificación constitucional para la formulación de medidas restrictivas a manifestaciones de discurso del odio. Por su parte, en la comunidad europea y en otros países, como Canadá, la preocupación se centra, según la autora, en contener la emisión y los efectos nocivos del discurso del odio por razones que pueden variar según las culturas constitucionales y las experiencias históricas de cada estado.

La regulación legislativa sobre el discurso del odio ha propiciado ciertos desafíos dada la ausencia de una definición consensuada en determinados contextos (STURGES, 2015), lo que debilita sus fronteras en relación a las restricciones de la ‘libertad de expresión’ que es el principal término asociado a los debates teóricos sobre la temática. La libertad de expresión propicia la creación y pluralidad de la opinión pública y el debate que en ella se gesta, sin embargo, su abuso genera serios conflictos con otros derechos sociales (ESQUIVEL ALONSO, 2016). Por ello, en la literatura científica aparecen estrechamente relacionados ambos términos que además se encuentran asociados también a expresiones como ‘libertad del discurso’, ‘ética del discurso’ y ‘crímenes de odio’, entre otros.

Las leyes que protegen de este fenómeno tienen como foco las identidades diana de estos discursos. En la discusión conceptual se han analizado definiciones que delimitan el concepto del discurso de odio como fenómeno dirigido hacia una única identidad. Según Seglow (2016) este discurso puede estar dirigido al género, la religión, la raza y también la discapacidad, mientras que Chetty y Alathur (2018), al revisar los trabajos publicados hasta la fecha de su estudio, afirman la existencia de un discurso de odio híbrido, es decir, un discurso que no se dirige a una sola identidad en particular, sino puede tener más de una identidad como objetivos.

Las identidades referidas en Seglow (2016) y Chetty y Alathur (2018) son recogidas en la literatura por Esquivel Alonso (2016) como expresiones de odio por motivos de diversa índole: étnicos, raciales, religiosos, por hacer apología de la violencia, por enaltecer el terrorismo, por negar el Holocausto o por insultar, entre otros. Existen expresiones, símbolos y manifestaciones considerados ofensivos por determinadas comunidades o personas, como puede ser la destrucción pública de imágenes religiosas o símbolos patrios, así como las imágenes que estigmatizan la sumisión y subordinación de la mujer, comúnmente presentes en los discursos machistas, sexistas o la pornografía. En la tabla 1 se sistematizan algunas tipologías de discursos de odio propuestas por Esquivel Alonso (2016), establecidas a partir de la diversidad de motivos que lo conducen:

Motivos	Expresiones/manifestaciones	Ejemplos
Étnicos y raciales	El color de piel, la pertenencia a un grupo étnico e incluso la nacionalidad	Exterminios masivos de indígenas cometidos por los conquistadores del continente americano. Humor racista
	La intolerancia a la diferencia que ha provocado discriminación, segregación e incluso masacres	forma de nacionalismo agresivo contra ciertas minorías o inmigrantes
	Las atrocidades cometidas por las falsas ideas de la existencia de una “raza superior” que han llevado a la humanidad a cometer abominables actos en contra de sus semejantes	La sentencia <i>Kühnen contra Alemania</i> , de 12 de mayo de 1988
Religiosos	Históricos choques entre los diferentes credos (como por ejemplo entre cristianos y musulmanes fanáticos y ateos, etc.), algunos que han llevado a manifestaciones de reprobación y se han	<i>Versos satánicos</i> de Salman Rushdie de 1988 <i>La última tentación de Cristo</i> en 1988 Caricaturas de Mahoma en 2005

	materializado en muerte y desolación	Seminario satírico <i>Charlie Hebdo</i>
Apología del delito, violencia y la hostilidad	En el libre ejercicio de la libertad de expresión puede ocurrir que, por la comisión de algunos delitos, se emitan mensajes ofensivos, ultrajantes o que denoten un desprecio por una etnia, grupo o sector poblacional determinado (opinión pública después de un acto de terrorismo)	Sentencias <i>Özgür Gündem</i> contra <i>Turquía</i> , de 16 de marzo de 2000 <i>Medya FM Reha Rado ve İletişim Hizmetleri A. S. contra Turquía</i> , de 14 de noviembre de 2006
Discurso negacionista	Consiste en cuestionar o negar la realidad del genocidio cometido por los nazis durante la II Guerra Mundial, con el propósito declarado de borrar de la memoria colectiva la huella de esa infamia	Puesta en duda de la existencia de las cámaras de gas, el número de víctimas judías, las insinuaciones de una campaña de publicidad negativa fabricada por la comunidad judía, entre otras situaciones

Tabla 1. Tipología de discursos del odio.
Fuente: elaboración propia a partir de Esquivel Alonso (2016)

A partir de un análisis de esta propuesta de Esquivel Alonso (2016) se puede advertir que, en lo que respecta al alcance del discurso del odio, este es un fenómeno mucho más complejo que las tipologías recogidas en la tabla 1, como los discursos misóginos, homofóbicos entre otras categorías basadas en ciertas características identitarias.

En su obra más citada *The Hate Speech Law*, Brow (2015) categoriza las leyes, reglamentos o códigos que restringen el uso del discurso del odio en diez grupos principales según su adherencia a casos de difamación, estereotipos o estigmatización, expresiones e incitación al odio, amenazas al orden público, negacionismo etc. Estos grupos son recogidos en la tabla 2:

Grupo	Descripción	Países/regiones cuyos códigos penales /estatutos criminales recogen estas leyes
Difamación grupal	Leyes que proscriben el discurso público u otra conducta expresiva cuando equivale a la difamación grupal de miembros de grupos o clases de personas identificadas basada en ciertas características adscriptivas (por ej. hacer declaraciones falsas de hechos que dañan la reputación).	Holanda, Eslovaquia, España, algunos estados de los Estados Unidos de América, Francia, Alemania, Israel

Grupo	Descripción	Países/regiones cuyos códigos penales /estatutos criminales recogen estas leyes
Estereotipos negativos o estigmatización	Leyes que regulan los estereotipos negativos y la estigmatización basados en características protegidas (se encuentran regularmente en la esfera de la ley y la regulación de los medios). Por ejemplo, artículos de código de prácticas que regulan el uso injustificado de estereotipos o contenidos discriminatorios que podrían interpretarse razonablemente como condonación o fomento de los prejuicios.	Australia (ej. <i>Australian Broadcast Company - ABC</i>)
La expresión del odio	Leyes que prohíben el uso de insultos, calumnias, epítetos despectivos o ridiculizar de otra manera a determinadas personas; difundir públicamente ideas basadas en la inferioridad de tales personas; el uso público de cualquier palabra, signo o símbolo que sea profundamente insultante u ofensivo para dichas personas en función de las características protegidas.	Bélgica, Bolivia, Cuba, Croacia, Dinamarca, Ecuador, Grecia, Indonesia, Italia, Noruega, Ruanda, Suecia, Turquía, algunas regiones de Reino Unido y el estado de Connecticut en Estados Unidos.
Incitación al odio	Leyes que prohíben suscitar, incitar o promover sentimientos de odio u hostilidad hacia o entre miembros de grupos o clases de personas con base en características protegidas	Armenia, estados australianos, Azerbaiyán, Bangladesh, China, Timor Oriental, Alemania, Hungría, Irlanda, Kenia, Kirguistán, Lituania, Macedonia, Malasia, Países Bajos, Pakistán, Rusia, Singapur, Suiza, Reino Unido, Uruguay, Uzbekistán, Vietnam, Sudáfrica
Amenazas al orden público	Leyes que prohíben el discurso u otra conducta expresiva de miembros de grupos o clases de personas identificadas por características protegidas cuando sea una amenaza para el orden público.	Canadá, Egipto, Etiopía, Alemania, India, Malasia, Singapur, Turquía, Venezuela.
Negar, etc. actos de crueldad masiva, violencia, o genocidio	Leyes que penalizan la negación, trivialización flagrante, aprobación, justificación, condonación o glorificación de actos de crueldad masiva, violencia o genocidio perpetrados contra miembros de grupos	Canadá, República Checa, Israel, Rumanía, España, Suiza.

Grupo	Descripción	Países/regiones cuyos códigos penales /estatutos criminales recogen estas leyes
	o clases de personas en base a características protegidas (ej. negacionismo del holocausto)	
Delitos o agravios de dignitario	Leyes que restringen el discurso u otra conducta expresiva dirigida a miembros de grupos o clases de personas identificadas por características protegidas cuando constituye la promulgación de un delito digno o agravio	Provincia canadiense de Saskatchewan, Costa Rica, Alemania, Suiza.
Violaciones de los derechos civiles o humanos	Leyes que prohíben el discurso u otra conducta expresiva cuando se trata de una conducta que viola o interfiere con el ejercicio de los derechos civiles o humanos de las personas con base en características protegidas	Canadá, Nueva Zelanda, Reino Unido.
Delitos de odio orientados a la expresión	Leyes que prohíben el discurso u otra conducta expresiva que constituya un crimen de odio basado en características protegidas	En Croacia, Francia, Italia, Rusia, Reino Unido, Estados Unidos son recogidas en agravamientos de las penalizaciones a los declarados culpables de dichos delitos (ej. sentencias de prisión o multas más altas). También en estatutos y códigos penales nacionales (ej., Argentina, Bélgica, Brunéi, Camboya, Cuba, Timor Oriental, Ecuador, Finlandia, Alemania, Grecia, Italia, Kenia, Países Bajos, España).
Restricciones de tiempo, lugar y forma¹	Leyes que restringen el discurso u otra conducta expresiva dirigida a miembros de grupos o clases de personas identificadas por características protegidas a través de restricciones de tiempo, lugar y forma (ej. con respecto a las protestas en funerales y entierros,	Algunos estados de los Estados Unidos de América (Arizona, Kansas, Missouri, Ohio)

¹ Este grupo de leyes no constituyen leyes *per se*, pero Brow las incluye dada algunas circunstancias en que las autoridades han intentado utilizar restricciones de tiempo, lugar y forma a las prácticas de quien las emite (ej. cuando involucran medios o circunstancias de comunicación que no cumplen con las restricciones de tiempo, lugar o forma pertinentes).

Grupo	Descripción	Países/regiones cuyos códigos penales /estatutos criminales reconocen estas leyes
	específicamente, creando zonas de amortiguamiento o distancias mínimas).	

Tabla 2. Diez grupos de leyes/reglamentos/códigos que restringen el discurso del odio.

Fuente: Brow (2015, 2016), traducción propia

En cuanto a los tipos de identidad o características que deben ser protegidas, Brow (2016, p.281) sistematiza las siguientes:

- **estados afectivos o patrones de los mismos** (por ejemplo, orientación sexual, preferencia sexual);
- **afiliaciones relacionadas con comunidades, culturas, grupos sociales o familias** (por ejemplo, ciudadanía, etnia, idioma, estado civil, nacionalidad, estado parental, identidad regional, religión);
- **disposiciones de actitud, creencias o formas de pensar** (por ejemplo, política, religión);
- **atributos biológicos, genotípicos, fisiológicos o físicos** (-fenotípicos) (p. ej., estado médico, raza, sexo);
- **conducta, comportamiento** (fenotípico), **desempeño o formas de vida** (por ejemplo, desempeño de edad, educación, situación laboral, desempeño de género, estado civil, profesión, religión, comunidad de viajeros, historial de guerra).

3. PROCEDIMIENTOS METODOLÓGICOS

La investigación se respalda en una metodología mixta tanto cualitativa como cuantitativa propia de los estudios métricos y el análisis de dominio en la organización del conocimiento (HJØRLAND; ALBRECHTSEN, 1995; HJØRLAND, 2002, 2017; SMIRAGLIA, 2015). Su carácter es descriptivo considerando los métodos de investigación empleados en el mapeo de la estructura social y el empleo de técnicas de procesamiento del lenguaje natural para alcanzar el objetivo central de la presente investigación. Siguiendo esta perspectiva se minimizan posibles sesgos integrando distintos métodos de investigación y perspectivas teórico-disciplinares para abordar el dominio de la campaña boicot #StopHateforProfit.

3.1 Análisis de dominio

El análisis de dominio (AD) es una teoría y un enfoque de la Biblioteconomía y Ciencia de la Información (LIS, en sus siglas en inglés) y la Organización del Conocimiento introducida en esta área por Hjørland y Albrechtsen en 1995. Es promovida como perspectiva teórica dentro de la CI que surgió a partir del artículo científico en la *Journal of the American Society for Information Science* y cuyo título es *Toward a new horizon in information science: domain-analysis*. Supone satisfacer la carencia en nuestro campo disciplinar de una teoría global superadora de los enfoques fisicalistas y cognitivos que demarcaron sus primeras décadas evolutivas (DAMUS; ACUÑA, 2019). A partir de su adopción en el campo de la CI, por su importancia y relevancia, deviene un punto neurálgico para la investigación en CI (ROMERO QUESADA, 2013). La comunidad de organización del conocimiento ha adoptado el análisis de dominio como un paradigma metodológico académico para el descubrimiento de bases ontológicas y para el análisis continuo de la evolución de las comunidades académicas.

Anteriormente, fue Rubén Prieto-Díaz, del campo de la Ingeniería de software, quien proporcionó otra definición de “análisis de dominio” desde una perspectiva técnica dentro del área de desarrollo de software. Hjørland (2017, p.439), mediante un análisis crítico de dicha obra, afirma que posiblemente este sea el primer investigador en relacionarlo con la Biblioteconomía y Ciencia de la Información. De forma casual, el AD como término se

introdujo también en 1980 en otro ámbito por James M. Neighbors, investigador del *Department of Information and Computer Science* de la Universidad de California, como propuesta plasmada en su informe de investigación para optar el grado de Doctor en Filosofía de la Información y Ciencias de la Computación.

El análisis de dominio, como paradigma para la determinación de los requisitos de sistemas, se abre camino en la ciencia de la información desde la lectura de Smiraglia (2015, p.2) como “un medio de orientar los sistemas para la recuperación de información en función de los requisitos específicos o las necesidades de información de las comunidades de investigación”.

Desde una perspectiva más próxima a la comunidad científica de la información, específicamente al área de la organización del conocimiento en la cual encuadra la presente investigación, el AD se establece por Hjørland y Albrechtsen en su artículo de 1995 como un enfoque programático o un "nuevo horizonte" para la ciencia de la información, definido como “el estudio de los dominios del conocimiento o las comunidades de pensamiento o discursivas que son partes de la división del trabajo de la sociedad” (HJØRLAND; ALBRECHTSEN, 1995, p.400).

En un artículo publicado en la revista *Knowledge Organization* que da continuidad a su propuesta, Hjørland y Albrechtsen (1999) identifican el análisis de dominio como una tendencia en la investigación de clasificación enmarcada en esta disciplina. En el análisis hecho por los autores, se observa la influencia de la vertiente epistemológica pragmatista cuando proponen el enfoque de la organización del conocimiento hacia la clasificación del conocimiento en contextos históricos, sociales y culturales. De esta forma, la propuesta de dichos autores se aleja de los enfoques racionalistas del análisis de facetas, que busca encontrar dimensiones de conocimiento universalmente aplicables, y, en la visión de Smiraglia (2015), se aproximan a enfoques no documentales o basados en la intertextualidad (enfoques empíricos) para el descubrimiento de estructuras de conocimiento. En este sentido Smiraglia advierte una evolución hacia enfoques analíticos de dominio para el descubrimiento de ontologías centradas en el dominio y hacia órdenes de conocimiento epistémicos de dominio.

El análisis de dominio según Smiraglia (2015, p. 19) es “un paradigma metodológico dentro de la ciencia de la organización del conocimiento”. Su propósito es “revelar los contornos del conocimiento mantenido, ya sea este en forma de discurso en vivo o documentación registrada, mediante el análisis de los elementos de comunidades específicas que comparten una ontología común o una base de conocimiento”. Bajo este presupuesto tiene como objetivo mapear y visualizar los parámetros intelectuales del conocimiento compartido en una determinada comunidad, haciendo posible según el autor el uso de dichos resultados en los sistemas de organización del conocimiento para el avance de la comunidad.

Esta investigación asume los supuestos descritos en los enfoques de "Análisis de dominio" de Hjørland (1999, 2002, 2017), que intentan analizar los fenómenos como elementos estructurantes que permiten comprender las articulaciones dentro de una comunidad discursiva. Por lo tanto, los autores que investigan ‘discursos de odio’ son tratados en esta investigación como individuos que comparten ciertos rasgos de identidad, como el pensamiento teórico y metodológico sobre el tema.

Según Hjørland (2002, 2018), para comprender un dominio de manera integral, es recomendable que los investigadores utilicen más de un enfoque de análisis de dominio, puesto que solo se considera un estudio encuadrado dentro este método si considerar las diferentes teorías o "paradigmas" del dominio de forma profunda. Smiraglia (2014) también comparte esta línea de pensamiento, argumentando la necesidad de aplicación de métodos con enfoques mixtos.

El presente estudio contribuye al entendimiento del dominio de la campaña #StopHateforProfit a través del análisis de su activismo en Twitter durante el 2020 desde una perspectiva de Análisis de Redes Sociales. Esto se justifica considerando que Hjørland (2017, p.455) dice "Keilty y Smiraglia (2016) es un estudio de la comunicación homosexual masculina en un sitio de contacto de Internet, que proporciona un argumento para considerar esto como un dominio. Es claramente un ejemplo de un dominio que es una alternativa a una disciplina académica". Por lo anterior, en esta investigación se utilizarán los siguientes enfoques de análisis de dominio propuestos por Hjørland (2002, 2017): de estructuras e

instituciones en comunicación no científicas con apoyo de un análisis alométrico, los cuales son desarrollados en las siguientes subsecciones.

3.2 Análisis del interés público

El debate sobre la opinión pública ha sido enfocado en la literatura científica desde diferentes perspectivas epistemológicas, áreas de conocimiento y temas de investigación (SILVA et al., 2018). El uso extendido de plataformas de redes sociales (Facebook, Twitter, Instagram, etc.) ha contribuido al creciente interés por la temática en el siglo XXI. La posibilidad de una participación activa de los usuarios de estas plataformas en aspectos sociales, políticos y organizativos facilitó la creación de narrativas y redes de acción colectiva (BENNETT; SEGERBERG, 2012). Este fenómeno se constituye también como uno de los pilares fundamentales del escenario de big data a la vez que tiene una incidencia directa en nuevas dimensiones de estudios, ya que se dispone de datos de opinión a gran escala.

En el desarrollo histórico de los estudios sobre opinión pública han surgido diversos episodios controvertidos caracterizados por una falta de consenso teórico entre la comunidad científica internacional, principalmente en el pasado siglo XX. Estas incertezas que han rodeado su conceptualización y sedimentación como área y los diversos intentos por alcanzar un consenso internacional han incidido directamente en el cúmulo de publicaciones científicas sobre el tema, lo que muestra que las discusiones sobre el concepto están aún lejos de concluir. Lippmann (1966, p.81), en su seminal trabajo *Public Opinion*, sostiene que mientras para la teoría ortodoxa esta categoría constituye “un juicio moral sobre un grupo de hechos”, desde su criterio, una opinión pública “es principalmente una versión moralizada y codificada de los hechos [...] que el patrón de estereotipos en el centro de nuestros códigos determina en gran medida qué grupo de hechos veremos y con qué luz los veremos”. El autor hace especial énfasis en la incidencia de los estereotipos sociales (referido en su texto como ‘*pictures of our head*’) en la formación de la opinión pública, y en este sentido considera relevante para su análisis reconocer la tríada conformada por “la escena de la acción, la imagen humana de esa escena y la respuesta humana a esa imagen que se desarrolla en la escena de la acción” (LIPPMANN, 1966, p. 11). Otras definiciones de opinión pública plasmadas en la literatura incluyen variantes más simplistas como “las preferencias de la

población adulta en temas de relevancia para el gobierno” (ERIKSON; TEDIN, 2015, p. 41), o conceptos más amplios como el planteado por Shang et al. (2015) (“la opinión pública es la respuesta de las personas a fenómenos sociales, preguntas, temas importantes, actitudes, emociones que reflejan los problemas actuales del enfoque de la sociedad”). En este sentido, la teoría de “la espiral del silencio” de Noelle-Neumann (1991), aún después de 40 años, es considerada fundamental en los estudios de opinión pública para construir un marco de conceptos existentes. La autora agrupa los conceptos existentes hasta ese momento en dos categorías: (1) la opinión pública como racionalidad, lo que la hace 'instrumental'... en el proceso de formación de opinión y toma de decisiones en una democracia; y (2) la opinión pública como control social, en el que "su papel es promover la integración social y asegurar que exista un nivel suficiente de consenso en el que se puedan basar acciones y decisiones" (NOELLE-NEUMANN, 1995, p. 34, citado por SCHEUFLE; MOY, 2000).

En la línea de Glynn et al. (2018), se considera que la investigación sobre la opinión pública corresponde con un campo muy amplio de estudio en el que académicos de diversas disciplinas buscan comprender cómo se forman, comunican y miden las actitudes sobre determinados asuntos públicos. Al analizar los diferentes intereses disciplinarios de su estudio se revelan áreas prolíficas como la Comunicación, las Ciencias Sociales, la Historia de la Sociología, la Gestión Empresarial, las Relaciones Internacionales, la Psicología, la Ecología, las Ciencias Ambientales y, en la última década principalmente, la informática y la CI. El estudio de Silva et al. (2018) sobre el frente de investigación del dominio también confirma su carácter diversificado en términos teóricos y metodológicos. Sin embargo, y a pesar de esta diversificación de los métodos de estudio, Glynn et al. (2018) han identificado las técnicas más importantes que utilizan los investigadores para evaluar el estado de ánimo del público: los resultados de las elecciones, el comportamiento del consumidor, las fluctuaciones del mercado de valores, las reuniones públicas y las protestas, así como otros indicadores de comportamiento similares. Estas mediciones se tornan más complejas en función de las fuentes de datos utilizadas, según modelos orientados a tratamiento de datos masivos. Por otro lado, el interés científico en el análisis de la opinión pública se ha incrementado en los últimos cuatro años debido al eco de los acontecimientos políticos y el uso de técnicas de manipulación basadas en big data ligadas a la ultraderecha (por ej. en el caso del Brexit y las

campañas o elecciones presidenciales de Estados Unidos, Brasil y otros países asesoradas por Steve Bannon y empresas como Cambridge Analytica). Según Glynn et al. (2018) existen aún muchas preguntas abiertas sobre la opinión pública que necesitan ser resueltas: cómo se sienten los ciudadanos sobre la intervención injerencista norteamericana en conflictos extranjeros, cómo se evalúa al presidente y a otros líderes políticos, cómo difieren las opiniones entre varios grupos sociales, etc. En este sentido, cabe cuestionarse desde una dimensión informacional cuáles serían las grandes interrogantes que interesan a la ciencia para contrarrestar estas campañas de desinformación vinculadas a fines económicos.

Noelle-Neumann (1991) sostiene en su teoría de la espiral del silencio que la opinión pública no apareció por primera vez en el siglo XVIII, sino que ha existido durante miles de años en todas las sociedades humanas como un fenómeno que ejerce fuerza sobre gobiernos e individuos que crea y mantiene los consensos necesarios para el funcionamiento de la sociedad. Mientras que Rousseau fue uno de los pioneros del uso del término (*l'opinion publique*), con el significado de las costumbres y formas de actuar de todos los miembros de la sociedad, escritores franceses en la década de 1780 hicieron uso también de este término de forma indistinta junto a otras expresiones como 'voluntad común', 'espíritu público' y 'conciencia pública', todas ellas con connotaciones de los aspectos políticos de la opinión de las masas (ERIKSON; TEDIN, 2015).

Otras líneas de discusión intelectual se han centrado en los términos que componen la expresión: "opinión" y "público". La postura crítica de Noelle-Neumann (1991) sugiere que el término "público" en el concepto debe interpretarse en el sentido de "ojo público", es decir, lo que es "visible para todos", como control social. "Opinión", por su parte, se refiere a las expresiones de opinión que son públicamente visibles y audibles, así como al comportamiento del público en relación con temas cargados de valor.

Según indican Erikson y Tedin (2015), es notoriamente difícil encontrar una definición del concepto que sea satisfactoria para un número sustancial de investigadores. A pesar de ser un problema crítico definicional de la investigación sobre la opinión pública que alcanza la existencia de variaciones teóricas, incluidas las relacionadas con las construcciones conceptuales de los términos que lo componen, como bien plantean Glynn et al. (2018), este

sigue siendo un campo con fronteras y es válido afirmar que, independientemente de la diversidad en las definiciones, estas también presentan elementos que se superponen en cierta medida. Es posible construir en cierta medida una imagen global del concepto a través de las visiones particulares y complejas en los diferentes dominios científicos.

Si analizamos las intersecciones disciplinares de la opinión pública no puede dejar de mencionarse la influencia de la religión, la política de Estado y el discurso político de los candidatos presidenciales. En esta línea, Djupe y Calfano (2013) analizaron una metodología experimental para identificar cómo funciona esta influencia. Tal como señala Chapp (2015), estos autores expresan la comprensión tradicional de la religión en una visión más integral y cambiante, partiendo de la premisa de que la dinámica de la opinión pública influenciada por las religiones generalmente se forma, evoluciona o retrocede más rápido que las creencias, comportamientos y pertenencias de los individuos.

La opinión pública es también de interés en el ámbito de la Gestión. Para Shang et al. (2015) su estudio es útil para inferir lo que sucederá a corto plazo, siendo que sus resultados facilitan la toma de decisiones de gobiernos y empresas. Además, ante la llegada de Internet y el desarrollo de las tecnologías que trabajan con big data, su punto de vista proporciona una ventaja competitiva que posibilita el análisis profundo de la situación gracias al tratamiento de los datos masivos que están disponibles en esta esfera pública.

Estos casos son ejemplos de lo que parece ser otro gran desafío en los estudios de la opinión pública: su medición. Noelle-Neumann (1991), por su parte, incidió en las reservas sobre la práctica de la investigación de la opinión pública basada en cuestionarios. Para Linsay Rogers (1891-1971), politóloga reconocida de la Universidad de Columbia y autora del libro “The Pollsters: Public Opinion, Politics, and Democratic Leadership” que recoge un profundo debate crítico sobre la legitimación de los sondeos en los estudios de opinión pública, los principales cuestionamientos sobre la validez de este método tradicional de medición parten de las posibles distorsiones que podría traer al proceso político, ya que los cuestionarios sobre la opinión pública no la miden en su conjunto (FRIED, 2006; ERIKSON; TEDIN, 2015).

En la misma línea de pensamiento, Erikson y Tedin (2015) sostienen que la opinión pública y los resultados de los sondeos de opinión pública no son lo mismo, ya que estos últimos corresponden con una medida indirecta del primero. Pese a todo, Glynn et al. (2018) identifican cuatro métodos formales utilizados tradicionalmente por líderes políticos y periodistas para buscar información sobre la opinión pública:

1. Investigación con encuestas o sondeos,
2. Grupos focales (*focus group*),
3. Investigación experimental,
4. Análisis del contenido de los medios de comunicación.

Los estudios de medición de la opinión pública siguen siendo un gran desafío en la ciencia debido al carácter dinámico y complejo del escenario informativo actual. En la nueva coyuntura tecnológica, cognitiva y cultural de la era de la post-verdad en la que vivimos conviven flujos masivos de información falsa que tienen como principal objetivo afectar las dinámicas de la opinión pública. Teorías y métodos del campo de estudio de la opinión pública han sido enriquecidos no sólo por los debates intelectuales y los fenómenos políticos descritos en esta introducción, sino también por la dimensión infotecnológica que ha jugado un papel importante en las transformaciones de su marco teórico-metodológico.

Con la emergencia del fenómeno del big data en el presente siglo XXI, científicos, políticos y otros analistas encuentran nuevas fuentes de datos en Internet para medir la opinión pública, lo que ha llevado a que surjan nuevos métodos de medición y estudios científicos que consideran las nuevas fuentes de datos de opinión disponibles. Todo lo expuesto estimula el crecimiento de los métodos de medición, la emergencia de nuevos medios y tecnologías de la comunicación y la intensificación de la complejidad de las redes sociales (GLYNN et al., 2018). Un ejemplo de estas herramientas y métodos sería Google Trends. McCallum y Bury (2013), por ejemplo, analizaron con esta herramienta los patrones de búsqueda de los usuarios y hallaron una disminución en el interés por el medio ambiente; McLaren y Shanbhogue (2011) utilizaron esta fuente para estudiar indicadores económicos; Anderegg y Goldsmith (2014) la usaron para medir el interés público en el cambio climático durante la última década y los efectos del evento mediático 'Climategate'. En definitiva,

Google Trends se ha utilizado para analizar los intereses de los usuarios en varios campos. Por otra parte, los estudios que utilizan esta herramienta también enriquecen el análisis con métodos adicionales como el análisis de sentimiento, principalmente de los mensajes en redes sociales. La segunda fuente de datos por excelencia en las investigaciones sobre opinión pública son las redes sociales como Twitter y Facebook. En el plano de los estudios económicos, por ejemplo, se han rastreado indicadores macroeconómicos, como la inflación (MCLAREN; SHANBHOGUE, 2011), y se han analizado las proyecciones de consumo privado (VOSEN; SCHMIDT, 2011).

El prestigio, la influencia y la extensa referencia académica internacional del New York Times (NYTimes) (HICKS; WANG, 2013) explican su preferencia sobre otros periódicos como fuente de datos, dado los fines de la presente investigación. Una de las ventajas de su uso es que no se requiere pago para el acceso a los resúmenes, mientras que otro criterio interesante es que mantiene una cobertura desde el año 1851 hasta la actualidad, lo que viabiliza el estudio de la dinámica de opinión pública sobre discursos de odio desde siglos anteriores.

3.2.1 Fuente de datos de opinión pública

Para analizar el interés público en relación con la campaña se escoge como primera fuente los datos que facilita la herramienta de Google Trends, la cual, a través de algoritmos, capta el interés mundial sobre un tema determinado (MOURA, 2009). Siendo amplio su uso en diversas áreas de conocimiento, dicha aplicación de Big Data ha abierto nuevas oportunidades para el análisis científico sobre un determinado tema (JUN et al., 2018). Google Trends ofrece a la presente investigación datos de frecuencia de búsqueda en Google de un término de interés y permite establecer una comparación con otros términos de búsqueda o temáticas, en diferentes regiones e idiomas, en un periodo de tiempo definido.

Las temáticas de búsqueda empleada en Google Trends fueron:

- En el caso de la estrategia de búsqueda de la campaña objeto de estudio se utilizaron los términos “Stop Hate for Profit”, la región geográfica seleccionada fue ‘mundialmente’, y el período de tiempo del 01/06/2020 al 31/11/2020. Se amplía el

periodo de análisis en esta fuente de datos para identificar la fecha de inicio más aproximada al interés público por la campaña y la extensión hacia finales de noviembre para captar aquellas noticias sobre la campaña que pudieran estar relacionadas con las elecciones presidenciales en ese mes.

Se utilizó el paquete de Excel de Microsoft Office para el procesamiento de los datos extraídos en formato csv de la Plataforma Google Trends.

En un segundo momento se utilizó como fuente la base de datos del periódico NYTimes. La búsqueda y recuperación de las noticias fue realizada mediante una conexión a la API disponible en la plataforma de desarrolladores (<https://developer.nytimes.com/docs/articlesearch-product/1/routes/articlesearch.json/ge>). Se utilizó la herramienta *Orange Data Mining Software*² para este proceso y subsecuentes de la minería de datos del NYTimes sobre la campaña. La siguiente figura 2 muestra el flujo de procesos definidos en esta potente herramienta para el procesamiento del lenguaje natural.

² Este software fue desarrollado en la facultad de informática de la Universidad de Ljubljana y es un potente programa que permite realizar minería de datos y análisis predictivo de varias fuentes y tipologías de datos.

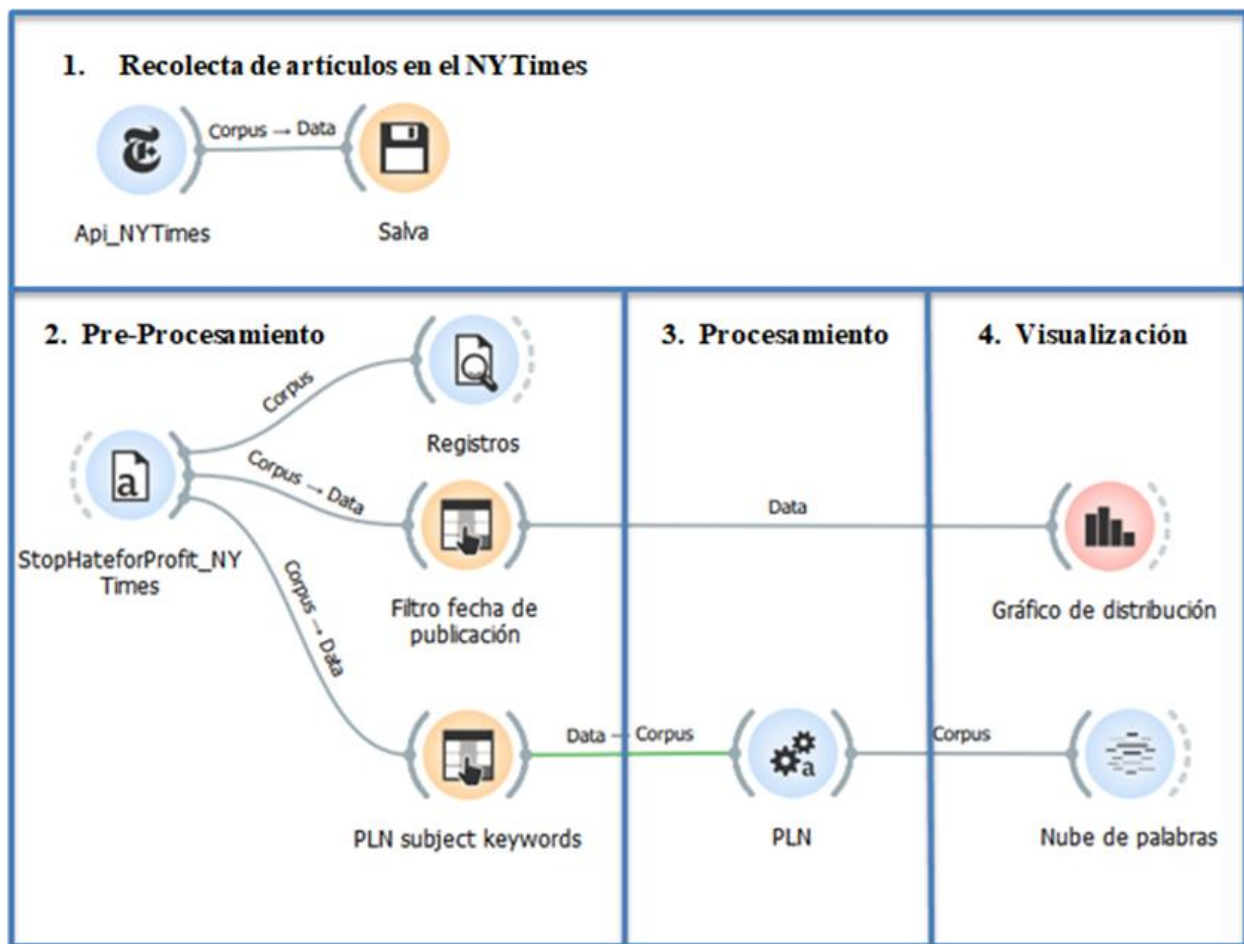


Figura 2. Flujo de la minería de datos de los artículos del NYTimes en Orange.
Fuente: Elaboración propia con el flujo de minería extraído del proyecto Orange para el NYTimes.

El flujo integra un grupo de acciones que son descritas en los próximos párrafos:

1. Recolección de artículos del NYTimes: se utilizó el widget *NYTimes article search* de Orange para recuperar datos de la API de búsqueda de artículos del New York Times. La recolección de los registros en formato json atendieron a dos filtros la búsqueda por palabras clave y periodo de tiempo. En el caso de esta fuente se mantiene el período de búsqueda que en Google Trends. Con la aplicación de filtros, se delimita el universo de las noticias de la gran base de datos de su plataforma. Se definieron los siguientes parámetros:

- fecha_comienzo: 20200601
- fecha_finalización: 20201131

- facet_fields: nulo
- consulta (query): “StopHateforProfit”

De los diez campos posibles para la extracción de datos de cada artículo solo se consideraron los siguientes: *headlines*, *abstract*, *date publication*, *lead paragraph* y *subject keywords*. Los registros recuperados se exportan con el widget *save*, en formato csv (Anexo 1, <https://doi.org/10.6084/m9.figshare.19307660>).

2. Pre-Procesamiento: se utilizó el widget *corpus* para importar los registros recuperados en formato csv exportados en la fase anterior. El *corpus viewer* permitió explorar los datos recuperados y escoger el de mayor completamiento informacional, para su tratamiento con algoritmos de procesamiento de lenguaje natural. En esta fase se realizan dos etapas de filtrado usando el widget *select columns* para la selección manual de atributos del dominio de datos (ORANGE, 2015):

- a) fecha de publicación: para el posterior análisis de la distribución temporal de la cobertura periodística a la campaña.
- b) temáticas claves: este campo contiene múltiples valores para cada artículo por lo que la riqueza de análisis será mayor para determinar los temas asociados a la campaña StopHateforProfit.

3. Procesamiento: en esta etapa se aplican técnicas de procesamiento de lenguaje natural al campo temáticas clave extraído de la fase anterior:

- a) Transformación: realiza modificaciones iniciales a los datos de entrada. Aplica un proceso de transformación de los términos a minúsculas y se eliminan los acentos del texto. La tokenización es el método empleado para dividir el texto en componentes más pequeños, como palabras y bigramas. La función aplicada es *Regex* que dividirá el texto por la expresión regular proporcionada, omitiendo las puntuaciones (UNIVERSITY OF LJUBLJANA, 2015).

- b) Filtrado 1: esta función fue definida para ser ejecutada en 2 ocasiones, en un primer momento se aplica el *POS Tagger*, que ejecuta el etiquetado de parte del discurso en *tokens*. Los sustantivos y los verbos fueron los parámetros configurados para el algoritmo (UNIVERSITY OF LJUBLJANA, 2018). Este proceso previo a la normalización mejora la precisión del proceso de tokenización.
- c) Normalización: aplica derivación y lematización a las palabras del campo extraído. Se selecciona el algoritmo *WordNet Lemmatizer* para aplicar una red de sinónimos cognitivos a *tokens* consultando en una gran base de datos léxica del idioma inglés. Su selección se basa en que toma en cuenta el contexto y convierte la palabra en su forma básica significativa, evitando así comunes errores ortográficos y resultados incorrectos en el análisis (BENGHUZZI; ELSHEH, 2020).
- d) Filtrado 2: se emplea el algoritmo de eliminación de palabras vacías del inglés (stop words), como “o”, “el”, “un”, “en”, “y”, etc., que aparecen con mucha frecuencia en textos, y que se hace necesario eliminar (BENGHUZZI; ELSHEH, 2020).
- e) Rango N-gram: crea n-grams a partir de los *tokens* obtenidos anteriormente. El rango de n-gramas determinado fue 1-2. Este valor devolverá un término o y dos que aparezcan juntos frecuentemente en el campo temática clave.
- f) Bag of Words: el algoritmo de este widget crea un corpus con recuentos de palabras para cada instancia de datos (registros recuperados). El método empleado para determinar la frecuencia del término es el conteo, esto constituye una ventaja para la visualización de los términos más frecuentes en una nube de palabras.

4. Visualizaciones: se procedió a la representación visual de los datos obtenidos en la fase de filtrado del pre-procesamiento de fecha de publicación y procesamiento del campo temáticas clave usando dos widgets:

- a) Distribuciones: que muestra gráficamente el comportamiento en el tiempo según su fecha de publicación, lo que ayuda a entender los hitos de la campaña y determinar posibles relaciones causales entre ambos discursos.
- b) Nube de palabras: modelo de visualización al conjunto de términos obtenidos del procesamiento de las temáticas claves.

3.3 Análisis de Redes Sociales

El Análisis de Redes Sociales (ARS), como método de investigación, permitió determinar la estructura social de comunidades de #StopHateforProfit. Su empleo se justifica ya que mide, describe y analiza las variables relacionales sociales de un conjunto de actores para representar la estructura de dicho grupo (WASSERMAN; FAUST, 1994).

En esta perspectiva, el ARS ofrece a la presente investigación un conjunto de conceptos y métodos analíticos para identificar y analizar los tipos de relaciones existentes entre las diferentes entidades que constituyen el dominio participante en esta campaña. Según Wasserman y Faust (1994), los patrones en las relaciones establecidas entre los actores que conforman este dominio pueden ser estudiados a partir del análisis de sus estructuras, por tanto, en este punto se busca identificar similitud y roles en las relaciones existentes entre los actores.

El ARS se preocupa por la comprensión de los vínculos entre entidades sociales y las implicaciones de esos vínculos. Uno de sus niveles de análisis son *los actores*, determinadas entidades sociales, no restrictas a individuos, estos pueden ser subgrupos, unidades sociales colectivas o corporativas, agencias de servicios públicos, estados-naciones, entre otros (WASSERMAN; FAUST, 1994, 17).

De este concepto se desprende que “los actores están vinculados entre sí por vínculos sociales” (WASSERMAN; FAUST, 1994, 18), es decir, es lo que establece una relación entre pares de actores como por ejemplo, transferencia de recursos materiales, emisión de evaluaciones, asociación o afiliación a eventos, interacciones comportamentales, movimientos migratorios, conexiones físicas como puentes que conectan dos puntos, relaciones formales o biológicas (descendencia), entre otros (WASSERMAN; FAUST, 1994).

El ARS, como técnica para el análisis de dominio, puede conducir a visualizaciones más complejas del mismo, siendo la teoría de redes una forma de mapear las relaciones entre los objetos en un conjunto de datos basándose en la simetría o asimetría de su proximidad relativa (SMIRAGLIA, 2014).

El escenario digital contemporáneo, propio de los sitios de redes sociales, entendidos por Boyd y Ellison (2007) como “una plataforma de comunicación en red”, genera a cada segundo una infinidad de conexiones y nodos de relación que requiere nuevas técnicas y métodos para el estudio de la dinámica y movimientos de las redes sociales como plataformas usadas por la sociedad (MACEDO; THURLEY; CAVALCANTI, 2020).

En este ámbito el sitio de red social Twitter se configura como fuente de datos de preferencia desde el ámbito académico y empresarial. Una posible explicación es que proporciona datos multimodales que contienen texto, imágenes y videos, junto con metadatos contextuales y sociales, como lo es la información temporal y espacial, e información sobre la conectividad y las interacciones del usuario. Estos datos enriquecidos y generados por los usuarios (KURSUNCU et al., 2019) son esenciales para el análisis de dichos dominios desde una perspectiva de Análisis de Redes Sociales.

La presente investigación pretende responder a interrogantes específicas como cuáles son los actores de la campaña que se pudieran considerar como centrales durante los dos primeros meses de su existencia o si existen personalidades públicas cuya fortaleza de vínculo en red, en la campaña, es mayor a la de las organizaciones que conforman la coalición inicial de la misma. Los epígrafes a continuación establecen las estrategias metodológicas para su resolución en este sentido.

3.3.1 Fuente de datos para el Análisis de Redes Sociales

En el presente estudio los datos de Twitter se utilizaron como fuente de información. El análisis de estos datos demanda el uso de técnicas de procesamiento de lenguaje natural y minería de datos, para ello se recopilaron mediante una solicitud a la API de Twitter con el uso de la herramienta RStudio. La estrategia de búsqueda fue configurada de la siguiente forma:

- los datos fueron filtrados por el *hashtag* #StopHateforProfit. Se descartaron los filtrados por ubicación, idiomas o las limitaciones a cuentas de usuarios específicas. El script de recuperación de tuits con R se incluye en la figura 3.
- se recuperaron los tuits como un objeto de anotación de JavaScript (JSON), los cuales contienen 16 campos de descripción que permiten aplicar un análisis enriquecido (fecha y horario de publicación, corpus textual del tuit, descripción del usuario, ubicación, datos sobre el *engagement*, etc.)
- Se define la cifra numérica de 5, del total de tuits a recuperar por cada consulta y
- Se capturan mediante el software R y exportan en formato valores separados por comas (csv en sus siglas en inglés) (ver Anexo 2, <https://doi.org/10.6084/m9.figshare.19307900>) para su posterior procesamiento y limpieza de datos en Excel (ver Anexo 3, <https://doi.org/10.6084/m9.figshare.19308029.v1>).

El script³ de recuperación de tuits queda por tanto definido de la siguiente forma, como figura 3:

³ Los datos de la fecha de inicio y fin fueron modificados en cada consulta para recuperar así el histórico de tuits en el rango de fecha delimitado en el estudio.

```
tw = twitteR::searchTwitter('#stophateforprofit', n = 1e5, since = '2020-06-25', until = '2020-06-26', retryOnRateLimit = 1e3)
```

Figura 3. Script de recuperación de tuits #StopHateforProfit.

Fuente: Elaboración propia en RStudio.

3.3.2 Tipo de matriz

Se trabajó el conjunto de datos recuperados para una red de tipo modo uno (un único conjunto de entidades), expresada a través de una matriz, la cual se clasifica, atendiendo a su dirección e intensidad, como una asimétrica ponderada (OVALLE-PERANDONES ET AL., 2020). El carácter asimétrico de la matriz de la red estriba en que, si bien los emisores de los tuits establecen una relación de mención con otros actores de la red, esta relación no ocurre necesariamente en la dirección inversa. En cuanto a la ponderación de la matriz se da a partir de que se registra el valor de cada reacción de mención, es decir, se explicita la intensidad de las menciones de manera cuantitativa.

3.3.3 Medidas de Análisis de Redes Sociales

En cuanto a los niveles de análisis, interesa caracterizar la red como un todo en un primer momento y las medidas de sus principales actores, determinar cuáles son sus vecinos, qué tipo de conexiones establece con sus vecinos, su centralidad, sus posiciones en la red, es decir, comprender a los nodos como conjuntos (instituciones, empresas, ONG, etc.) de agentes intencionales con acción e influencia en otros (FRESNO GARCÍA; DALY; SEGADO SÁNCHEZ-CABEZUDO, 2016).

A continuación, se presentan los indicadores que caracterizan las propiedades de una red.

3.3.3.1 Análisis para toda la red

En cuanto a las mediciones de propiedades de la red, interesa identificar las siguientes características de la estructura general de la red, estas estadísticas son proporcionadas por la herramienta Gephi y se relacionan en la siguiente tabla 3:

Medida de red	Descripción
Número total de nodos	Indica el número total de puntos (actores) partícipes de la red en un determinado período de tiempo.
Número total de aristas	Indica el número total de líneas (conexiones)
Grado promedio	Mide la cohesión estructural de una red. Esta es una mejor medida de la cohesión general que la densidad porque no depende del tamaño de la red (NOOY; MRVAR; BATAGELJ, 2018)
Grado	En un grafo el grado es el “número de nodos adyacentes a él” (WASSERMAN; FAUST, 1994, 125).
Densidad del grafo	Densidad de “un grafo dirigido es la proporción de aristas presentes en el dígrafo respecto de las posibles” (WASSERMAN; FAUST, 1994, 129).
Diámetro de la red	Se define como la mayor distancia entre cualquier par de sus puntos en la red (SCOTT, 2017). El número máximo de conexiones requeridas para atravesar la red hacia el nodo más alejado (CHERVEN, 2015).
Promedio <i>Path length</i>	Proporciona una medida de la eficiencia de la comunicación para toda una red, midiendo la ruta más corta posible entre todos los nodos de la red (CHERVEN, 2015).
Coefficiente de agrupación	Esta medida “se puede calcular a partir de aristas de la red” (NOOY; MRVAR; BATAGELJ, 2018, 361). El coeficiente de agrupamiento local de un vértice es la proporción de los pares entre sus vecinos que están directamente vinculados
Número de triadas	El censo de triadas que ocurren en la red
Modularidad	Es el número de aristas que surgen dentro de los grupos menos el número esperado en una red, equivalente con aristas colocadas al azar

Medida de red	Descripción
	(NEWMAN, 2006).
Número total de comunidades identificadas	Valor total de las comunidades en el grafo.

Tabla 3 Métricas y estadísticas de Gephi para análisis de la red como un todo
Fuente: elaboración propia.

Este primer análisis permite describir la red como un todo, completándose tras conocer el número de nodos que participan en una determinada red cuántos de ellos quedan aislados.

Si medimos el número de relaciones existentes y las comparamos con las relaciones posibles se podría -densidad-, en futuras campañas, establecer estrategias de interconexión entre grupos de actores que fortalecerán el posicionamiento de la narrativa central de la campaña.

La noción de **proximidad** es aquella calculada desde la suma de las distancias geodésicas hacia todos los puntos en el grafo (SABIDUSSI, 1966). Según Scott (2000) es el recíproco de la suma de las distancias.

3.3.3.2 Análisis de la red a partir de sus actores

En este nivel de análisis se busca caracterizar a los actores de la red, la naturaleza de las relaciones y su rol en la campaña. Para ello se considera en un primer momento la medida de centralidad de grado de los actores mediante la identificación del número de nodos adyacentes (vecinos) a él y a su vez el número de relaciones incidentes. Para Newman (2010, 169), aquellas personas “que tienen muchas conexiones con terceros pueden tener más influencia, más acceso a la información o más prestigio que los que tienen menos conexiones”. Por tanto, medir el número de conexiones a un nodo nos mostrará cómo de bien o no está conectado en términos de conexiones directas (FRESNO GARCÍA; DALY; SÁNCHEZ-CABEZUDO, 2016).

Los ‘líderes de opinión’ en este escenario se denominan como *social media* influyentes y el ARS proporciona la base analítica para su identificación al enfocar en las

relaciones, su fortaleza y dirección, al segmentar e identificar los grupos, subgrupos y comunidades que la conforman y la posición estructural que los actores tienen dentro de la red (FRESNO GARCÍA; DALY; SÁNCHEZ-CABEZUDO, 2016).

Con vistas a identificarlos, la presente investigación medirá el grado de centralidad de los nodos (actores) en la red. La centralidad se refiere a las posiciones de los nodos individuales dentro de la red (NOOY; MRVAR; BATAGELJ, 2005), permitiéndonos así visualizar aquellos actores que establecieron mayores vínculos de mención hacia otros nodos de la red y los que recibieron mayor número de menciones durante la etapa de análisis de la campaña. El grado de un nodo, es el número de líneas que son incidentes con él y equivalentemente es el número de nodos adyacentes a él (WASSERMAN; FAUST, 1994).

La medida de grado de un nodo en la red se calcula como la suma de grado entrante (*in-degree*) y grado saliente (*out-degree*). Estas medidas ayudarán a identificar los líderes de opinión dentro de la campaña.

La popularidad de los actores en ARS, se mide por sus grados de entrada (WASSERMAN, 1977, 1980); si un actor recibe muchas menciones, a menudo se dice que son prominentes o que tienen un gran prestigio, lo que podría significar que muchos otros actores buscan vínculos directos con ellos, e indicar su importancia en la red (SCOTT; CARRINGTON, 2011). Como consecuencia, un alto grado de *in-degree* de un nodo puede suponer un mayor prestigio en la red.

Mientras que, según Scott & Carrington (2011), los actores que tienen grados salientes (*out-degree*) inusualmente altos son actores que pueden intercambiar con muchos otros o hacer que muchos otros conozcan sus puntos de vista, siendo a menudo reconocidos como actores más influyentes.

Otra medida interesante para el análisis del rol de los actores en la red se mide calculando su grado de intermediación (*betweenness*) Freeman (1977), para caracterizar la relación entre los otros actores participantes de la campaña. En ARS se interpreta como la capacidad de alcance de llegar a otros actores de la red. Este indicador que se basa en cómo de importante es un nodo en términos de la vinculación con otros nodos, muestra con qué

frecuencia un actor se encuentra en el camino más corto entre pares de nodos. Freeman et al. (1991) amplió el enfoque básico de esta medida al considerar todos los caminos entre los actores (SCOTT; CARRINGTON, 2011) que interpreta la posición favorecida y poder de un actor en la red, en la medida en que más gente dependa de él para tener conexiones con otras personas. Por tanto, una posible interpretación sería que a mayor grado de intermediación más poder tiene el nodo en dicha red.

Para reforzar estos últimos análisis se mide la centralidad o poder del vector propio (*eigenvector centrality*), que tiene en cuenta cuán centrales son aquellos a quienes está conectado un *ego*, es la medida en que está vinculado a nodos con alta centralidad del vector propio (NOOY; MRVAR; BATAGELJ, 2005, 2018). Las redes *ego*, en esta investigación se apoyan en la perspectiva de Everett y Borgatti (2005) definida como aquellas “redes que consisten en un solo actor (*ego*) junto con los actores a los que están conectados (*alters*) y todos los vínculos entre estos *alters*”. La finalidad de analizar la red *ego* respondería además a analizar la información sobre los *alters* a los cuales está conectado y su influencia sobre ellos.

Se identifica el número de nodos débilmente fuertemente conectados:

- Nodos débilmente conectados: un componente (débil) es una subred conectada al máximo (débilmente). Una red está (débilmente) conectada si cada par de nodos está conectado por un semicamino (NOOY; MRVAR; BATAGELJ, 2005, 2018).
- Nodos fuertemente conectados: un componente fuerte es una subred fuertemente conexas. Una red está fuertemente conectada si cada par de nodos está conectado por un camino (NOOY; MRVAR; BATAGELJ, 2005, 2018).

3.3.4 Pre-procesamiento de los datos obtenidos de Twitter

Esta fase es realizada en dos subetapas:

1. Mediante el uso del paquete de software de Microsoft Excel las menciones de cada texto del tuit son separadas por columnas mediante el separador @. El uso de la fórmula de procesamiento de texto *left* de Excel permite la limpieza de datos dejando en cada celda únicamente las menciones realizadas por cada emisor del tuit. Esta fase es respaldada por el uso del editor de texto Sublime para una mayor calidad en la limpieza de los datos.
2. Creación de las relaciones en Microsoft Access: los datos obtenidos del proceso anterior son procesados en esta fase para definir los nodos emisores y receptores, así como las relaciones entre ellos, asignándoles un identificador único. Su empleo permitió generar una lista de relaciones entre cada nodo a partir de sus identificadores. La matriz obtenida del establecimiento de estas relaciones y definiciones de id de los nodos fue exportada coincidiendo en formato y estructura para su importación como fichero .net en Gephi (Anexo 4, <https://doi.org/10.6084/m9.figshare.19308059>).

3.3.5 Herramienta para el Análisis y visualización de Redes Sociales

Gephi en su versión 0.9.2 es el software escogido para la obtención de datos de propiedades de la red para posterior análisis y discusión de los resultados, además de su visualización. Su selección se basa en el criterio de que constituye una poderosa herramienta para la visualización de grafos y cuenta con diferentes opciones de filtrado, distintos cálculos estadísticos y de medidas de redes, además de variados algoritmos para la representación visual de los grafos (GOUVEIA, 2020) que serán empleados en el ARS del presente estudio.

3.4 Análisis de sentimientos

Si bien en la literatura especializada sobre el análisis de sentimientos es posible encontrar diversas referencias a la denominación de este campo de estudio, tales como: minería de opiniones, análisis de opiniones, extracción de opiniones, minería de sentimientos, análisis de subjetividad, análisis de afectos, análisis de emociones, minería de reseñas (LIU, 2015) y orientación semántica (TABOADA et al., 2011), existe un consenso al considerar cada una de ellas bajo el paraguas del término más aceptado que es el de análisis de sentimientos.

A este campo le interesa el análisis de las opiniones, sentimientos, valoraciones, actitudes y emociones de las personas hacia las entidades (ej. productos, servicios, organizaciones, individuos, acontecimientos, problemas o temáticas) y sus atributos expresados en texto escrito (MEDHAT; HASSAN; KORASHY, 2014; LIU, 2015), mediante el uso de varias técnicas de procesamiento del lenguaje natural (PLN), recuperación de información (RI), minería de datos estructurada y no estructurada (DM) (RAVI; RAVI, 2015).

Emociones básicas han seguido diferentes clasificaciones según diversos teóricos. Según Liu (2015), el término aparece por primera vez en la literatura profesional en el año 2013, mencionado en el trabajo Dave et al. (2003). Sin embargo, como resultado de la revisión bibliográfica realizada en este trabajo, puede constatarse que aparece también anteriormente en el estudio de Minkov (2002) titulado *Trading system applied to large mutual fund company* cuyo resultado de investigación fue presentado dos años antes como parte de la agenda científica del *evento International Workshop on Empirical Science of Financial Fluctuations* desarrollado en Tokyo, Japón.

La aplicación de esta técnica para resolver diversos problemas llevó a que fuera rápidamente acogida por otros campos más allá de los estudios enmarcados en el área de la computación/informática. En este sentido, se destacan entre los más prolíficos, siguiendo las categorías de la Web of Science, las Telecomunicaciones, la Biblioteconomía y Ciencia de la Información (LIS), la Lingüística, Ciencia de la Gestión de Investigación de Operaciones, Neurociencias, la Gestión y las Ciencias de la Salud, lo que hizo que a finales de la primera década del siglo XXI se obtuviera un acumulado de 105 publicaciones. Este valor fue superado como valor único de publicaciones del año 2011, y a partir del cual se evidencia un comportamiento de aumento en las publicaciones sobre el tema en la literatura profesional.

El análisis de sentimientos se considera comúnmente como una subárea del Procesamiento del Lenguaje Natural (PLN). Desde sus orígenes este ha expandido significativamente los estudios del PLN, introduciendo desafiantes problemas de investigación, hasta el momento no planteados; y es que el análisis de sentimientos pasa por cada una de sus áreas troncales: semántica léxica, la resolución de correferencia, la

desambiguación del sentido de las palabras, el análisis del discurso, la extracción de información y el análisis semántico (LIU, 2015).

3.4.1 Niveles y enfoques del análisis de sentimientos

En cuanto a la granularidad del análisis de sentimiento, este se ha desarrollado en tres niveles principales: (a) nivel basado en documentos, (b) nivel basado en sentencias y c) nivel basado en entidades o aspectos (PANG; LEE; VAITHYANATHAN, 2002; HU; LIU, 2004; PRABOWO; THELWALL, 2009; TAN ET AL., 2011; LIU, 2015; DO et al., 2019). La finalidad que se busca en cada uno de ellos se describe a continuación:

- a) Nivel basado en documentos: su objetivo es clasificar si un documento de opinión completo (comentarios en blogs, tuits, etc.) expresan sentimientos con valencia positiva o negativa (HU; LIU, 2004; LIU, 2015).
- b) Nivel basado en sentencias: o determinar si cada oración, expresada en una opinión (ej. en forma de tuits) es positiva, negativa o neutral (HU; LIU, 2004; LIU, 2015).
- c) Nivel basado en aspectos: a diferencia de los niveles anteriores, permite advertir exactamente en la expresión de opinión lo que a las personas les gusta y no les gusta. Su análisis es más completo y permite relacionar las entidades y aspectos identificados en el corpus analizado, así como la clasificación de las emociones del público asociados a ellas (ej. productos, servicios, temas, problemas, personas, organizaciones o eventos, los cuales a su vez generalmente presentan varios aspectos a ser analizados) (JIMÉNEZ-ZAFRA et al., 2016; LIU, 2015; DO et al., 2019). El análisis de opinión de aspectos o entidades comparativos es posible en este nivel a través de las siguientes tareas: dado un conjunto de textos evaluativos, que identifica oraciones comparativas a partir de ellos, se clasifican las oraciones comparativas identificadas en clases y posteriormente se extraen las relaciones comparativas en las oraciones identificadas (implica la extracción de entidades, sus características y las palabras clave comparativas) (JINDAL; LIU, 2006).

Las técnicas de clasificación de sentimientos se pueden dividir aproximadamente en (a) enfoque basado en léxicos, (b) enfoque basado en aprendizaje de máquina y (c) enfoque híbrido (LIU, 2015; MEDDEB et al., 2019).

La forma tradicional de realizar un análisis de sentimientos sin supervisión (*unsupervised sentiment analysis*, en su término en inglés) es el método basado en el léxico (HU et al, 2013). El proceso de análisis de sentimiento que sigue este enfoque depende de un léxico de sentimientos, es decir, de una colección de términos de sentimientos predefinidos (construidos manual o automáticamente). La lista de tales palabras y frases aparece en la literatura bajo el término 'léxico de sentimientos' o léxico de opinión (LIU, 2015). Este enfoque se divide en uno basado en diccionarios (lista de términos), en un corpus (frases y modismos) y también aparecen bajo listas de emoticonos (MEDDEB et al., 2019).

Bajo este enfoque se utilizan métodos semánticos o estadísticos para determinar la polaridad o valencia de sentimientos del texto (PAWAR; SHRISHRIMAL; DESHMUKH, 2015). Además de los diccionarios de léxico, los trabajos de investigación de sentimientos sobre mensajes de microblogging también han utilizado conjuntos de emoticonos positivos y negativos para detectar clases de sentimientos, a pesar de que las palabras subjetivas se pueden interpretar de manera diferente de un anotador a otro.

Según Taboada et al. (2011), una de las ventajas de aplicación de este modelo se da en análisis del contexto lingüístico para el tratamiento de marcadores de valencias (ej. intensificadores, downtoners, negación e irrealis), presente en la opinión pública.

La complejidad de este modelo parte primeramente de las propias características de los datos de redes sociales. Los textos en las redes sociales, principalmente en Twitter, son cortos (ampliados este año a 280, el doble de su valor original) y es común que la carencia de suficiente información agregada dificulte la calidad de la clasificación de la polaridad de los sentimientos expresados en ella (HU et al, 2013). En segundo lugar, la popularidad, en estas plataformas, de expresiones del lenguaje: "es geniallllllllll" y "good 9t:)", evolucionan rápidamente, a la par de los neologismos, todo lo cual resultan en desafíos para el procesamiento lingüístico.

Liu (2015) sintetiza otros desafíos en la aplicación del enfoque basado en léxico:

- a) Una palabra de sentimiento positivo o negativo puede tener orientaciones o polaridades opuestas en diferentes dominios de aplicación o contextos de oraciones.
- b) Sentencias que contienen términos que no expresan ningún sentimiento. Este fenómeno que ocurre en varios tipos de oraciones (interrogativas y las condicionales).
- c) Oraciones sarcásticas (muy común en las discusiones políticas, lo que hace que minar y obtener los sentimientos en las opiniones políticas sea una tarea muy compleja).
- d) Las sentencias sin términos que presenten criterios de valencia pueden implicar sentimientos u opiniones positivas o negativas desde la perspectiva de sus autores.

Otros enfoques de análisis de sentimientos se basan en aprendizaje de máquina (*machine learning approach*), los cuales aplican algoritmos supervisados y no supervisados (que utilizan análisis lingüísticos más complejos, clasificadores basados en métodos probabilísticos, basados en reglas, métodos lineales y clasificadores de árboles de decisión). La simbiosis de ambos enfoques se denomina enfoque híbrido (LIU, 2015; MEDDEB et al., 2019).

3.4.2 Fase de limpieza de datos

A continuación, se enumeran las diferentes herramientas utilizadas para la minería de datos en el tratamiento de los tuits en este proceso de análisis de sentimientos:

1. Paquete para R *Title Text Mining Package* ('tm') Version 0.7-7 y
2. Paquete dplyr para la construcción del corpus, el filtrado del campo de texto de cada tuits (`<- read.csv(file.choose(), header = T)`) y el filtrado de los tuits (`<- profit %>% filter(isRetuit != T)`), este proceso elimina los retuits del corpus colectado.

Esta etapa de limpieza comprende diversas tareas de pre-procesamiento de los datos:

1. Conversión al formato UTF-8 (este paso elimina el problema de que muchos tuits contienen caracteres no estándares, lo que dificulta su posterior procesamiento).
2. Normalización de casos (mayúsculas y minúsculas) usando la función (*tolower*).
3. Las direcciones web, función (*remove URL*)
4. Tokenización: consiste en dividir los sistemas de texto en términos personales o tokens. Para el corpus en inglés se aplican las siguientes funciones:
 - a) eliminación de las puntuaciones (*remove Numbers function*),
 - b) eliminación de los espacios en blanco (*strip Whitespace function*);
5. Eliminar las palabras vacías (*removeWords, stopwords 'english'*),

3.4.3 Los paquetes y funciones para el análisis de sentimientos

Los paquetes y funciones en R utilizados para el análisis de sentimientos y visualizaciones de datos son:

- Clasificación de los *sentiment scores*: la función *get_nrc_sentiment* obtiene las emociones y valencia a partir de la consulta al NRC Dictionary (Version 0.92) (<https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>). Esta función hace llama al diccionario de sentimientos de NRC para clasificar la presencia de las ocho emociones básicas de Plutchik (1980) que las clasifica desde una perspectiva subjetiva en ira, anticipación, disgusto, miedo, alegría, tristeza, sorpresa, confianza. Plutchik (1980, 16) afirma que las ocho emociones básicas forman cuatro pares opuestos, alegría-tristeza, ira-miedo, confianza-disgusto y anticipación-sorpresa. Tomando como ejemplo el último par anticipación-sorpresa coloca que “son opuestos en el sentido de que uno implica lo impredecible mientras que el otro lo predecible. Estas emociones son detectadas por del dataset de tuits colectados y la correspondiente valencia en sentimientos negativos y positivos.

- *NRC Word-Emotion Association Lexicon* (también conocido como *EmoLex*): es una lista de palabras en inglés y sus asociaciones con las ocho emociones básicas propuestas por

Arnold y Plutchik (1964; PLUTCHIK, 1980, 1984) y la valencia de los sentimientos (negativos y positivos).

- *The Sentiment and Emotion Lexicons* es una colección de léxicos que fue creada en su totalidad por los expertos del Consejo Nacional de Investigación de Canadá. Desarrollada con una amplia gama de aplicaciones, esta colección de léxicos se puede utilizar en una multitud de contextos, como el análisis de sentimientos, el marketing de productos, el comportamiento del consumidor e incluso el análisis de campañas políticas (MOHAMMAD; TURNEY, 2013).

La librería que faculta el uso de *NRC Emotion Lexicon* es *tidytext* y para completar, fue usado el paquete para la visualización de datos: *bar plot*. El modelo de clasificación de sentimientos adoptado es *2-way*, que clasifica los tuits en positivos y negativos (Anexo 5, <https://doi.org/10.6084/m9.figshare.19308236>).

3.5 Análisis de sentimientos para los retuits

Posterior al filtrado de los retuits (`<- profit %>% filter(isRetuit != F)`) y su conversión al formato UTF-8 utilizando la herramienta R, estos son exportados en un fichero csv que es procesado por el software de minería de datos *ORANGE Data Mining* en su versión 3.3.2 para su clasificación en sentimientos positivos y negativos.

Este software fue desarrollado originalmente por científicos de la Universidad de Ljubljana en 1997 utilizando los lenguajes de programación Python para el entorno grafico Qt3 (DEMŠAR et al., 2013). Los componentes gráficos (Widget) de este software proporcionan funciones básicas de PLN que requiere el tratamiento de los retuits para revelar características de la opinión pública relacionada con la prominencia temática y la clasificación de los sentimientos en el corpus analizado.

3.5.1 Flujo de Trabajo

La siguiente figura 4 representa el flujo de trabajo seguido en la minería de datos del tratamiento de los retuits:

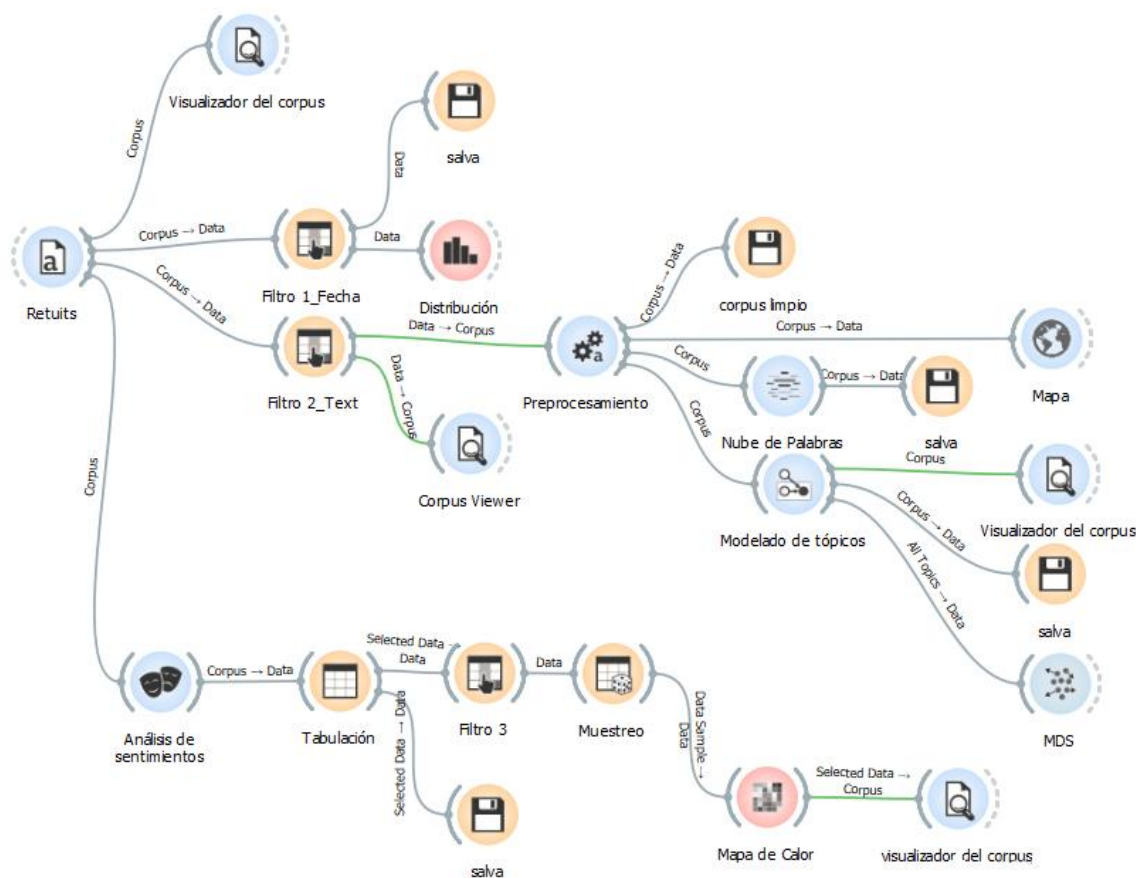


Figura 4. Flujo de Minería de datos de los retuits en #StopHateforProfit
Fuente: Elaboración propia en Orange Data Mining Software

Lo que varía en la primera fase en relación al flujo del NYTimes es el **widget corpus**⁴, pues la recolecta fue realizada externamente en RStudio. Su uso permitió importar el fichero en csv previamente exportado a partir del *dataframe* en R. Se incluyó un widget de visualización del corpus para facilitar el análisis de calidad del proceso de importación, así como de revisiones y análisis del corpus en caso de requerirse. Las variantes en cuanto a los valores procesados se describen a continuación:

En el **widget filtro1** se procede a extraer el campo de fecha de publicación del *dataset* de los retuits (Anexo 6, <https://doi.org/10.6084/m9.figshare.19308254>) para la posterior representación en un gráfico de distribución temporal. En este subproceso se incluye un widget llamado Salva para la exportación de los datos filtrados. Específicamente el **widget filtro 2** extrae el campo de texto de los tuits para su procesamiento.

⁴ Entrada al flujo de minería de datos.

Los widgets de Orange empleados para el **pre-procesamiento** y limpieza de estos datos corresponden a los mismos procesos empleados en el tratamiento de los artículos del NYTimes (Transformación, Normalización, Rango N-gram).

Como salida de este procesamiento se agregaron otros 6 widgets de análisis (bag of words), extracción y clasificación de tópicos (modelado de tópicos), representación (mapa y nube de palabras) y exportación de datos (salva).

En el widget **modelado de tópicos** se aplicó el modelo de tópico Latent Dirichlet Allocation (LDA). Este método probabilístico de modelaje de tópicos (BLEI; AY; MI, 2003) brinda la posibilidad de aplicar técnicas de procesamiento de lenguaje natural y permite la extracción y cálculo de frecuencia de ocurrencia de los 10 tópicos conjunto de términos más abordados en las publicaciones de la campaña, conduciendo a una descubierta de los temas troncales abordados en ella. Su aplicación se justifica pues descubre temas abstractos del corpus basado en 10 grupos de 10 palabras que se encuentran en cada tuit y su frecuencia respectiva. El widget también informa sobre el peso del tema por retuit lo que permite profundizar en la exploración de las temáticas en el corpus del texto. Esto permitirá una mayor comprensión de las narrativas de la opinión pública a partir de la identificación en secuencia de sus principales elementos temáticos en #StopHateforProfit.

Paralelo al flujo de procesos con el corpus para visualizaciones de distribución, representación temática de regiones y países y de modelado de tópicos, se aplica el análisis de sentimientos. Este widget predice los sentimientos presentes en cada tuit del corpus. Se utiliza el léxico de sentimientos multilingües del Data Science Lab que admite varios idiomas, brindando mayor precisión en sus resultados, dadas las características del corpus.

En el **widget tabulación** se observa el sentimiento total de un tuit, donde -1 es el más negativo y 1 el más positivo. La visualización de datos en mapa de calor ocurre posterior a la selección de los resultados a mapear en el **widget filtro 3**, del cual se desprende una función de muestreo intencionado del 70% del corpus para su representación.

El **widget mapa de calor** realiza una visualización de *clustering* basado en una matriz bidireccional. Los valores se representan por color según la paleta de colores

seleccionada por defecto *Blue-Green-Yellow*. Al combinar la variable de clase y los atributos en los ejes x e y, se observa dónde los valores de atributo son más fuertes y dónde son más débiles (sentimientos negativos y positivos), lo que nos permite encontrar características típicas para cada clase. Finalmente se exploran subconjuntos de resultados en el widget **visualizador del corpus**.

A partir del widget **gráfico de escalamiento multidimensional** (MDS, en sus siglas en inglés) (WICKELMAIER, 2003) de Orange se visualizan los tópicos modelados y la representación de su valor de probabilidad de tópico marginal.

Resultados de este subproceso son exportados en el **widget de salva** como fichero en CSV.

4. RESULTADOS DEL ANÁLISIS DE LA CAMPAÑA STOP HATE FOR PROFIT

El presente capítulo pretende esclarecer el alcance y estructura social de esta campaña en los primeros dos meses tras su creación, desde la perspectiva de un Análisis de Redes Sociales y análisis de sentimientos.

4.1 Stop Hate for Profit e interés público

Como se describió en el apartado metodológico, en la recuperación de los tuits emitidos en la campaña se obtuvieron 16 campos de metadatos que describen cada comunicación emitida. Los datos recuperados fueron estructurados en un *dataframe* como un objeto de anotación de JavaScript (JSON), cuyos campos de descripción que permiten aplicar un análisis enriquecido (fecha y horario de publicación, corpus textual del tuit, id del usuario, ubicación, datos sobre el *engagement*, geolocalización, etc.). Este conjunto de datos permite enriquecer el análisis del contexto de los tuits emitidos en la campaña. Como es habitual en datos que proceden de esa fuente, la categorización de tuits recuperados puede realizarse atendiendo a dos conjuntos: orgánicos (también denominados originales) y retuits.

A partir de la recolecta realizada por el hashtag oficial de la campaña #StopHateforProfit se obtuvo un total de 146.771 tuits (incluyendo los retuits) publicados entre el 25 de junio al 6 de agosto del 2020, con excepción de las fechas 12 y 23 de julio en los cuales debido al límite establecido por Twitter a cuentas estándares de 7 días precedentes. La siguiente tabla 2 recoge la clasificación de este conjunto de datos en tuits y retuits:

Tipo de publicación	Total
Tuits	25.480
Retuits	121.291

Tabla 2. Categorías de las publicaciones en la campaña #StopHateforProfit
Fuente: Elaboración propia a partir de datos recolectados

Atendiendo a la fecha de emisión de cada publicación, la distribución de la comunicación generada en esta campaña y desde una perspectiva cuantitativa, permite

describir la dinámica de los mensajes generados por la comunidad participante de la campaña. De esta forma se diseña una línea en el tiempo donde se observa el comportamiento de publicación de la comunidad que construye este discurso. El siguiente gráfico 1 muestra la frecuencia absoluta de los tuits orgánicos emitidos en el periodo comprendido desde el 25 de junio del año 2020 al 6 de agosto de ese mismo año:

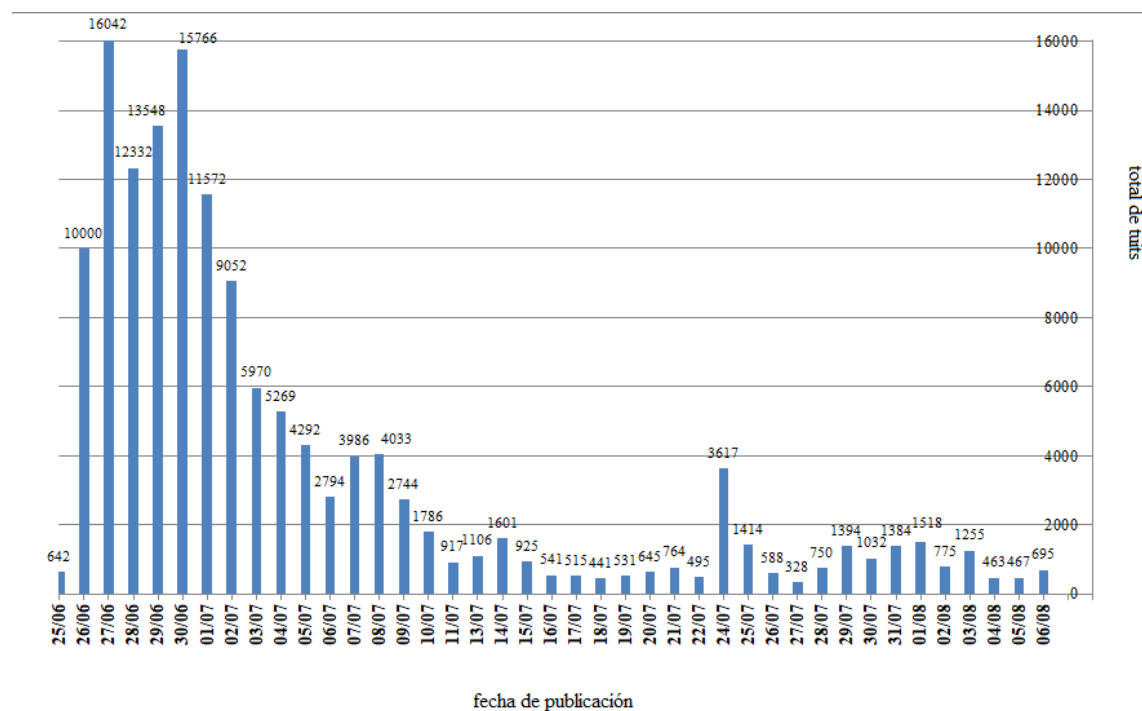


Gráfico 1. Evolución temporal de los tuits orgánicos emitidos #StopHateforProfit: 25/06-06/08/2020

Fuente: Elaboración propia con datos de la investigación en Excel

Los datos del gráfico 1 muestran que la campaña se caracterizó por ser muy productiva cuantitativamente en los días iniciales a su lanzamiento, fueron publicados más de diez mil tuits diarios, siendo el 27 de junio la fecha de mayor auge de participación, desde el punto de vista cuantitativo con aproximadamente 16.042 tuits publicados.

Los datos que muestra el gráfico 2 arrojan los resultados en conjunto de ambos tipos de tuits, los orgánicos y los retuits de estos. A los fines de la investigación interesa analizarlos por separado, para comprender mejor la dinámica en el tiempo del eco generado por la conversación de #StopHateforProfit. El siguiente gráfico 2 recoge esta relación, los tuits

originales publicados sobre la campaña con la línea color azul y los retuits con la línea color rojo entre el 25 de junio del 2020 de los tuits hasta el 6 de agosto.

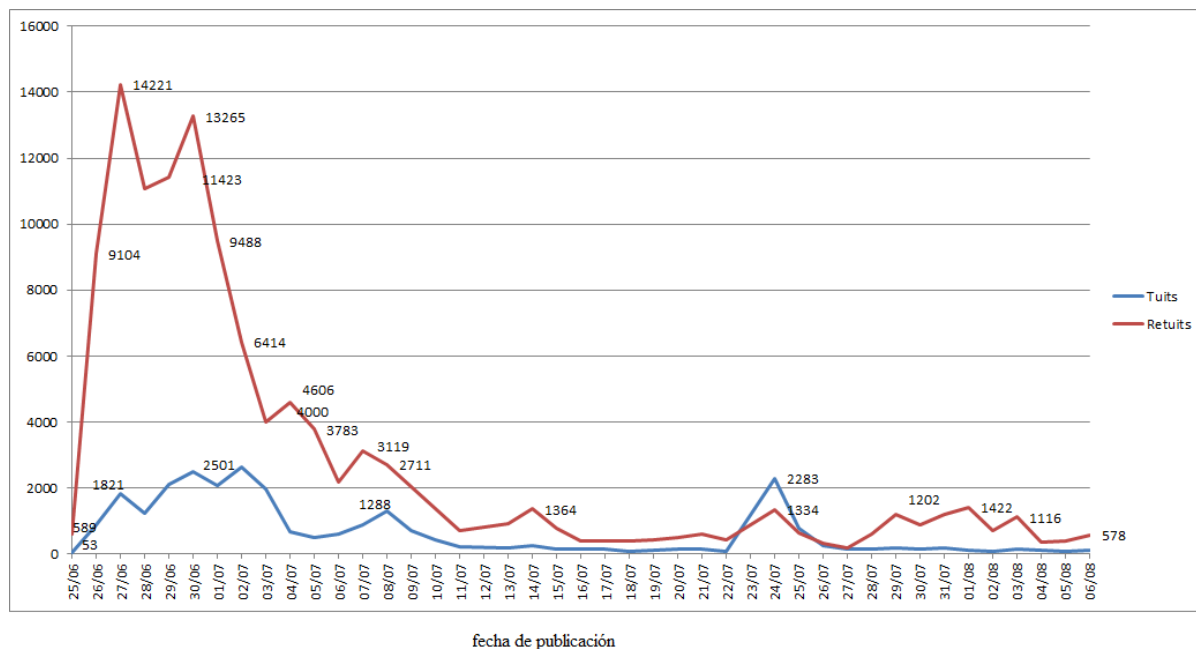


Gráfico 2. Evolución temporal de los tuits y retuits emitidos #StopHateforProfit
Fuente: Elaboración propia con datos de la investigación en Excel

El inicio de la campaña contó con un total de 53 tuits y 589 retuits, cifras que fueron aumentando considerablemente, con mayor proporción en el caso de los retuits. Una característica que destaca es que, desde los inicios de la campaña los tuits tuvieron un alto índice de respuesta. Eso implica el alto eco en la campaña y en este sitio de red social. Como muestra de ello, en su segundo día, los retuits alcanzaron valores superiores a los catorce mil para 1.821 tuits orgánicos.

Durante todo el periodo analizado, que concluyó el 6 de agosto del 2020, los mensajes publicados por la comunidad participante se mantuvo en valores muy próximos al día inicial o en valores más elevados; hay un punto de inflexión hacia un descenso en los valores totales tanto de tuits como de retuits durante los primeros días del mes de julio. Se comprueba que estas cifras aumentan como eco de algunos hitos de la campaña, por ejemplo, en fechas como el viernes 24 de julio con un total de 3.617 tuits, cuando la coalición #StopHateForProfit, *influencers* y otros participantes enviaron un mensaje unificado en sus

redes sociales y otros canales mediáticos, para captar la atención de todos los usuarios de este sitio de red social sobre los objetivos de la campaña.

Inciendiando de nuevo en el alto eco de la campaña, el 27 de junio los tuits emitidos por la coalición de la campaña alcanzaron su mayor valor de retuits, siendo un total de 14.221 difundidos por otros participantes. Seguidamente se muestran algunos de estos mensajes de mayor impacto en la campaña. La mayor cantidad de retuits la obtuvo el mensaje incluido en el tuit que fue emitido por Sacha Baron Cohen el 26 de junio. Como se aprecia en la figura 4, el actor menciona a las compañías que más invierten en publicidad en Facebook y exhorta a unirse al boicot presionando a Facebook para alcanzar los fines de la campaña.



Figura 5. Tuit de mayor difusión en la campaña #StopHateforProfit.
Fuente: Twitter - Sasha Baron Cohen (2020)

Este mensaje fue el de mayor difusión alcanzada y coincidió con ser hasta el 6 de agosto el más favorito (13.764 likes) y retuiteado (5.936 retuits). Sin dudas, la estrategia de mención de las cuentas oficiales de dichas compañías en Twitter ayudó a llamar la atención de sus CEOs con inmediatez y que la comunidad internacional participante hiciera uso de esta funcionalidad de la comunicación en la plataforma para ejercer presión sobre las mismas. En la misma fecha el grupo de activistas digitales que luchan contra el discurso de odio y la desinformación Sleeping Giants publica un tuit confirmando la unión de grandes empresas como Unilever USA a #StopHateforProfit y rápidamente fue retuiteado 1.725 veces y marcado como favorito por 6.076 usuarios.

El 27 de junio en otro tuit (figura 6) la cuenta oficial de Sleeping Giants expone las acciones que realiza la compañía Facebook para frenar que las empresas pausen su publicidad en la plataforma, en vez de desarrollar iniciativas en pos de la lucha contra el discurso de odio difundido en ella. Ese mensaje alcanzó un total de 2.683 retuits y 8.306 likes de otros usuarios de Twitter. En esta misma fecha la NAACP comunicó su agradecimiento al liderazgo de la duquesa y duque de Sussex por reconocer la importancia de la campaña y solidarizarse con su causa; esta noticia inmediatamente tuvo repercusión con un total de 3.668 retuits y 3.668 likes.



Figura 6. Tuit de Sleeping Giants del 27/06/2022.
Fuente: Twitter - Sleeping Giants (2020)

El tercer mayor pico sucede el 29 de junio, durante la primera semana, cuando grandes empresas que publicitan en Facebook oficializan su unión a la campaña. Al analizar este comportamiento se observa que este hito tiene una relación causal con la publicación del revisado tuit de Sasha Baron Cohen. En el contenido del tuit el actor informa sobre la colaboración de compañías como Starbucks, Diageo_News, por mencionar algunas, a #StopHateforProfit. De esta forma se conoce como Facebook pierde esa semana a dos de sus mayores inversionistas en publicidad. Eso hizo que Sasha Baron Cohen aprovechara este logro y como parte de su mensaje exhorta a compañías inversoras como Disney, HomeDepot, ProcterGamble, Walmart, ATT, Cricketnation, sprint, NYTimes y Amazon a unirse a este boicot. Este mensaje alcanzó altos índices de interacción y de difusión, con 8.669 likes y es retuiteado en 2.215 ocasiones.

A inicios del mes de julio se observan valores que superan los 2.000 tuits y más de 9.000 mil retuits. Según un tuit de Sleeping Giants del primero de julio de 2020, se confirmaban 530 anunciantes unidos a la campaña #StopHateForProfit. El 4 de julio, el periodista Kyle Griffin en su cuenta oficial de Twitter anunció el apoyo de más de 400 importantes compañías que retiraron sus anuncios no solo de Facebook sino también de la red social Instagram. Esta publicación fue retuiteada en 1.968 ocasiones y marcada como favorita 6.953 veces por otros usuarios, cuyas interacciones amplifican aún más el alcance de esta noticia en Twitter.

Otro momento clave durante la campaña #StopHateforProfit, que coincide con el aumento de la actividad en Twitter, se dio el 14 de julio de 2020 cuando la coalición de Stop Hate for Profit publicó una carta abierta al CEO de Facebook, Mark Zuckerberg, y a la COO de Facebook, Sheryl Sandberg. La carta recogía la firma de más de 100 grupos laborales, religiosos y de defensa de los derechos civiles. En su mensaje, la carta solicita a los CEO responder ante las demandas de #StopHateForProfit alegando el respaldo por las más de mil empresas que hasta esa fecha habían retirado el pago de su publicidad en la plataforma.

Aunque se aprecia un ascenso de las publicaciones del boicot el día 7 y 14 de julio, como efecto de que nuevas grandes empresas pararon su publicidad en este sitio de red social, se observa un descenso en los mensajes difundidos si se comparan con los valores iniciales alcanzados.

Los tuits de Sleeping Giants, NAACP, Sasha Baron Cohen, periodistas como Carole y Kyle Griffin sin lugar a dudas marcaron una escalada significativa en la difusión y alcance de la campaña a cientos de empresas anunciantes, figuras públicas y miles de usuarios que se sumaron en apoyo de las demandas de #StopHateforProfit contra el discurso del odio, el extremismo y la desinformación.

Las interacciones, ya sean con forma de comentarios, likes o retuits de los participantes a los mensajes de la coalición, organizaciones sin fines de lucro, así como las figuras públicas, ayudaron a difundir la campaña con inmediatez a miles de usuarios en este sitio de red social. Se desconoce en cifras el alcance que tuvo en general la campaña, las métricas en este aspecto no fueron divulgadas por los líderes de la campaña, sin embargo, es

indiscutible la presión ejercida sobre Facebook al ganar el apoyo de más de mil compañías anunciantes. Dada la relevancia y particularidad de los logros de la campaña seguidamente se describe la misma desde una perspectiva de Análisis de Redes Sociales, donde será descrito el rol desempeñado por sus principales líderes y las ventajas estructurales que suponen facilitar la consecución de su éxito.

4.1.1 Alcance público de la campaña #StophateforProfit en Google Trends

Google Trends analiza la frecuencia con la que los temas han aparecido en la Web en escalas temporales y espaciales para así contar con un punto gráfico más alto que identifica la región y el tiempo determinados, donde un valor de 100 es el pico de popularidad del término (GOOGLE, 2022).

Los valores de la figura 6 representan el interés de búsqueda en relación con el punto más alto del gráfico para la Web en general (web, noticias, videos, imágenes), en el alcance global de Internet, es decir, sin límite por región geográfica. En consonancia con los resultados expuestos, se definió el período de tiempo comprendido entre el 6/01/20 y 11/30/20. Los valores del interés del término fueron calculados en una escala de 0 a 100, donde 100 es la ubicación con más popularidad como una fracción del total de búsquedas en esa ubicación y de 0 (países cuyo color aparece en tonalidad gris). Un valor medio en popularidad (50) se alcanza en las regiones representadas con el azul marino. Es importante resaltar que el valor más alto significa la proporción más alta en relación a todas las consultas, no la cifra cuantitativa más alta considerando el recuento de consultas absoluto.

Si el valor de 100 es el pico de popularidad alcanzado por el término en las búsquedas de información en Google, un valor de 50 significa que el término es la mitad de popular y el valor de 0 significa que en la fecha específica Google Trends no obtuvo suficientes datos sobre la búsqueda de sus usuarios por este término.

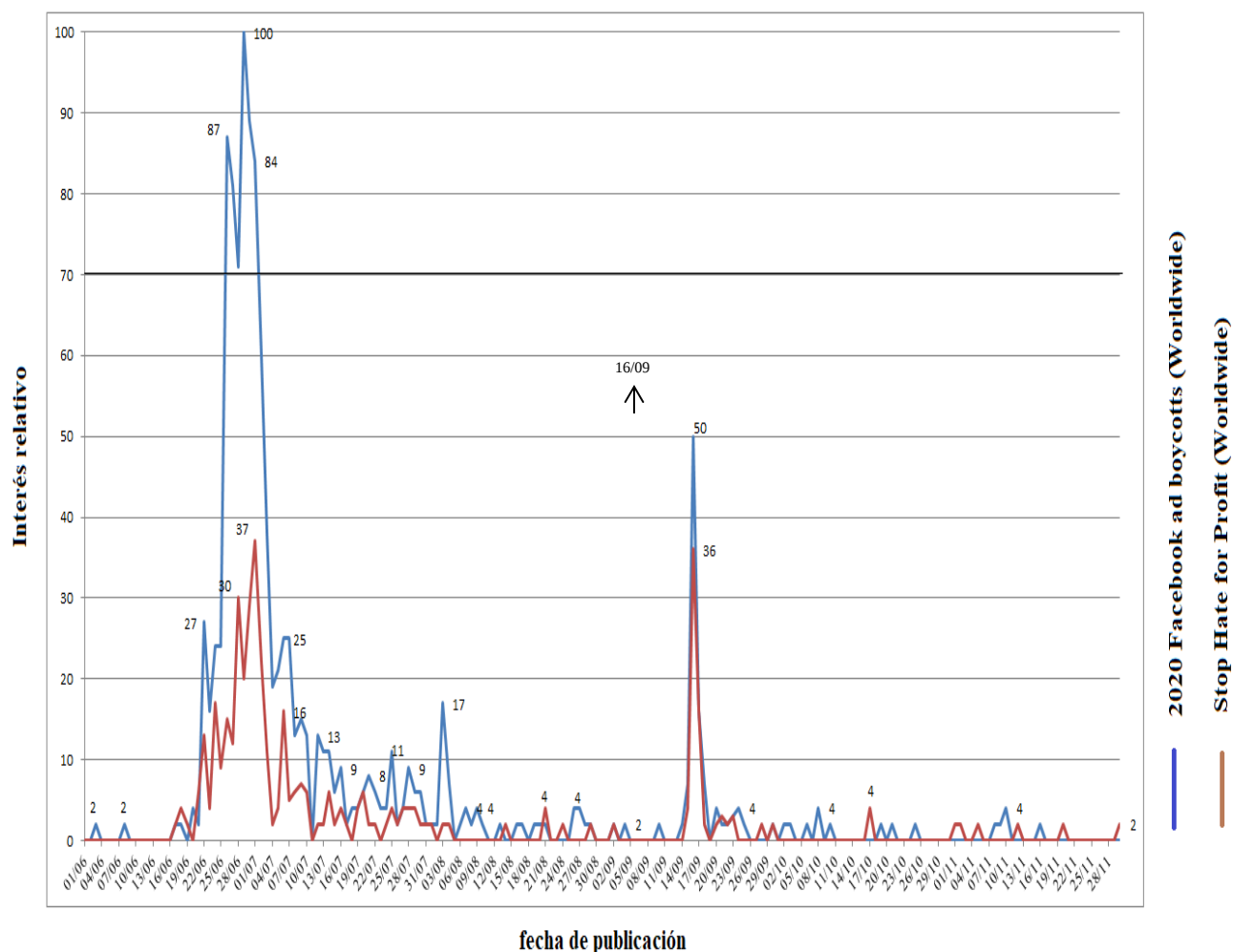


Gráfico 3. Comportamiento temporal del interés y alcance público Web de Stop Hate for Profit y 2020 Facebook ad boycotts.
Fuente: Google Trends (2022)

El análisis de la tendencia en el tiempo del interés público de este fenómeno arrojó varios resultados de relevancia. El gráfico 3 muestra los datos comparativos del término “2020 Facebook ad boycotts” y Stop Hate for Profit. Es visible que desde los inicios de la campaña gana mayor prominencia el primero como tópico usado por los internautas en comparación con el propio nombre oficial de la campaña, el segundo enumerado. Aunque para esa fecha presenta valores muy incipientes, se verifica además que las manifestaciones del interés público se inician el 17 de junio, fecha en que las conversaciones sobre este boicot también comienzan a ganar fuerza en el sitio de red social Twitter.

En el mes de septiembre, específicamente el día 16 se visualizan un aumento en los valores de interés relativo de búsqueda por los términos Stop Hate for Profit (36) y 2020 Facebook ad boycotts (50). Estas estadísticas están relacionadas con tentativa de boicot promovida por actores como Kim Kardashian, entre otros famosos, pero esta vez a la plataforma de Instagram y consistió en congelar la actividad de sus cuentas durante un día por los mensajes de odio.

Google Trends nos muestra, más allá del alcance público del término en una escala temporal, una representación espacial de las necesidades informacionales de los sujetos que realizan las consultas en el buscador general Google (MOURA, 2009). El siguiente mapa incluido en la figura 7 representa que el término Stop Hate for Profit fue más popular en algunas naciones durante el período de tiempo especificado, son las regiones se destacan en el mapa con el color azul intenso. Los valores se calculan en una escala de 0 a 100, donde 100 es la ubicación con la mayor popularidad como fracción del total de búsquedas en ese país, un valor de 50 indica una ubicación que es la mitad de popular, mientras que un valor de 0 (representado por el color gris) indica una ubicación donde no hubo suficientes datos para este término.

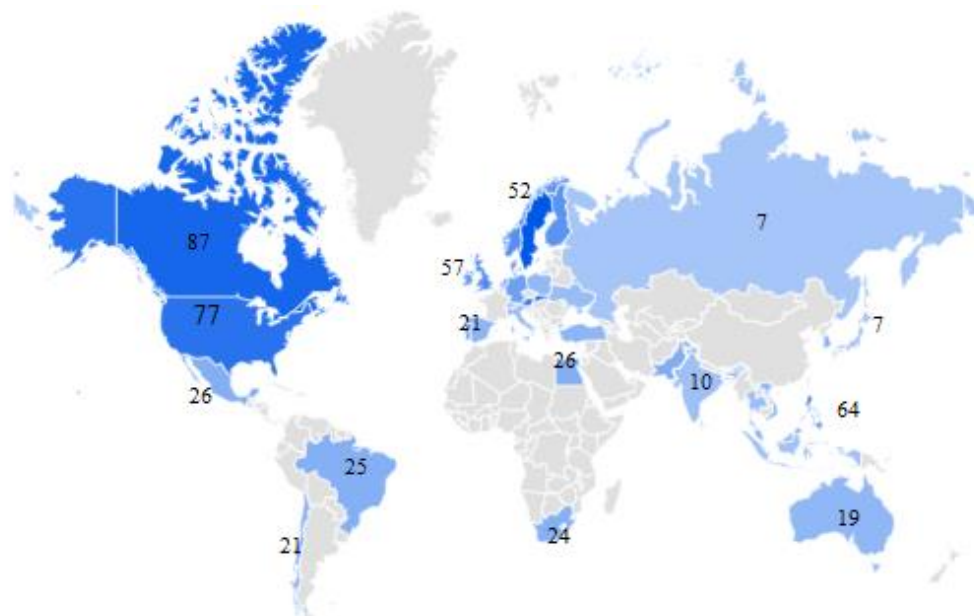


Figura 7. Regiones geográficas de interés y alcance público en Stop Hate for Profit
Fuente: Google Trends (2022).

El interés público en las búsquedas sobre la temática, según su procedencia geográfica, muestra a Suecia con una popularidad máxima 100, le sigue Canadá (87), Eslovaquia (82), Estados Unidos de América (77), Austria (65), Filipinas (64), Reino Unido (57), Noruega (52). El resto de los países identificados presentan valores inferiores a 50 en popularidad. Estos datos sugieren, aunque presentan diferentes valores, una similitud entre esos países sobre el interés en la campaña, lo que denota el alcance internacional de la misma.

Los usuarios que buscaron información sobre la campaña también se interesaron por otros temas, tras cuyo análisis se puede afirmar su adherencia dentro del dominio de la campaña, la medida del interés alcanzada entre dichos temas se relaciona en la siguiente tabla 4:

No.	Temas populares	Puntuación (escala relativa)
1	Hatred	100
2	2020 Facebook ad boycotts	89
3	Profit-Economics	87
4	Meta-Technology Company	10
5	Advertising campaign	8
6	Boycott	7
7	Company	4
8	Advertising	3
9	Hate Speech	1

Tabla 4. Temáticas populares relacionadas al interés público por Stop Hate for Profit.
Fuente: Elaboración propia con datos de Google Trends (2022).

Se observa que *hatred*, término utilizado comúnmente para referirse a odio o rencor, aparece como el tema más popular en las búsquedas de los internautas. Esta puntuación si bien representa un valor de 100 en una escala relativa al término más buscado, indica que el centro del debate coincide con la principal discusión en #StopHateforProfit.

Un elemento interesante es que, como segundo tema más comentado, se encuentra una variante al término oficial que representó la campaña *2020 Facebook ad boycotts*, por tanto, su uso debe ser considerado al analizar este dominio. Este planteamiento puede justificarse también a partir del comportamiento mostrado por los usuarios en sus búsquedas de información sobre la campaña, datos recogidos en el gráfico 3, en este se observa como *2020 Facebook ad boycotts* llega a ser el término más popular en las búsquedas por parte de los usuarios.

La campaña despertó el interés de los internautas por el significado del fenómeno financiero base de su demanda, dígame las implicaciones económicas de un boicot publicitario a la empresa objeto de la demanda realizada por la coalición. Relacionado a esto, se confirma que también fueron solicitadas informaciones sobre boicot, campaña publicitaria y aunque con valores bajos de búsqueda en el caso de las búsquedas *2020 Facebook ad boycott* aparecen términos referentes a las compañías relacionadas con la campaña, como Coca-Cola (3), *The Coca-Cola Company* (3), The North Face (3), Unilever (2), NAACP (2) y Patagonia (1).

Aunque con menor valor de la escala relativa *hate speech* (discurso de odio en su traducción al idioma español) aparece como parte de los temas relacionados en la búsqueda y al validar los datos de consulta de los internautas en relación a ese término en específico, se hace evidente un particular interés por la sociedad en cuanto a su relación con los crímenes de odio, el racismo como una de sus principales manifestaciones, así como la libertad de expresión, la cual coincide con la principal discusión de los estudiosos de este fenómeno en el campo científico; así como las leyes que la regulan.

Por otro lado, los datos de Google Trends muestran que, como parte del comportamiento del interés público, se emplearon otros términos como *stop hate for profit facebook*, *facebook boycott* y *stop hate for profit campaign*, siendo otras tres variantes de términos usados en las consultas a Google y con menos aumento en la frecuencia de consulta en el período de tiempo analizado en relación a los dos principales términos de consultas ya analizados.

4.1.2 Alcance público de la campaña #StopHateforProfit en The New York Times

La trascendencia de Stop Hate for Profit también fue de alcance público en el discurso periodístico observado en el análisis de la cobertura de prensa de esta campaña en el periódico The New York Times, en adelante, NYTimes. En la recuperación de noticias mediante la API de NYTimes se obtuvo un total de siete artículos de las secciones de Opinión, Negocios y Tecnología en el periodo de tiempo del 29/06/2020 al 16/09/2020. La figura 8 muestra una nube de términos con los 369 *tokens* que se obtuvieron tras el procesamiento del lenguaje natural realizado con Orange Data Mining Software para el

contenido de los campos resúmenes, encabezado y párrafo líder a cada registro recuperado del NYTimes:

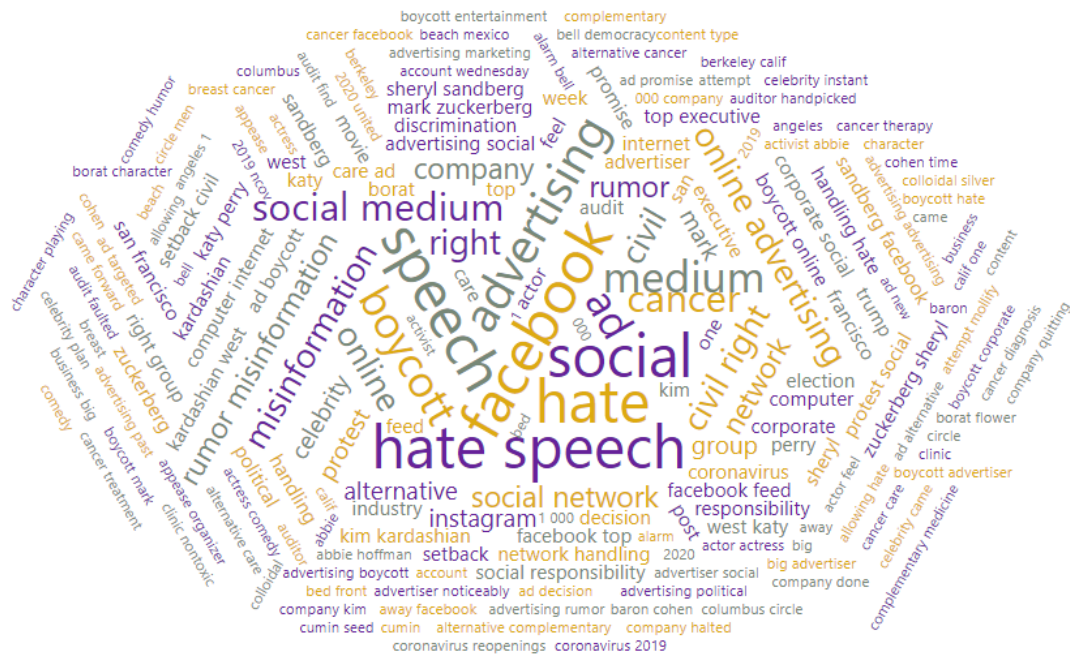


Figura 8. Cobertura temática de los términos sobre #StopHateforProfit en NYTimes.
Fuente: Elaboración propia

El análisis de los términos permite entender varios posicionamientos de la opinión pública sobre este boicot a Facebook. Este análisis se dio en dos momentos, la imagen de la figura 8 representa el resultado de aplicar el algoritmo de bigram en la fase de pre-procesamiento de los datos, con el rango del parámetro ngram a 1-2 obteniéndose los bigram para enriquecer el contexto de análisis de los términos obtenidos de la narrativa periodística del NYTimes sobre la campaña.

En apoyo al análisis visualizado en la figura 8, se completa con la siguiente tabla 5 en la que en la primera columna se posiciona en orden los términos más frecuentes, en la segunda cada uno de ellos, después los pesos. Las columnas cuarta, quinta y sexta, cumplen la misma finalidad que las ya comentadas, pero para los términos posicionados en 11º lugar y

posteriores. Por una cuestión de optimización de espacio, exclusivamente se relacionan los datos de los 20 principales términos prominente en el corpus:

No.	Término	Peso en el corpus
1	Facebook	2.00
2	Social	1.71
3	Speech	1.57
4	Hate	1.43
5	hate speech	1.43
6	Advertising	1.14
7	Boycott	1.00
8	Ad	1.00
9	Médium	1.00
10	Cancer	0.86
11	social médium	0.86
12	Civil	0.71
13	civil right	0.71
14	Misinformation	0.71
15	Online	0.71
16	online advertising	0.71
17	Right	0.71
18	Company	0.57
19	Network	0.57
20	Rumor	0.57

Tabla 5 Términos prominentes sobre #SobreHateforProfit en NYTimes.

Fuente: Elaboración propia con datos de investigación.

Como se esperaba Facebook es el término más frecuente del corpus analizado, le sigue *social*, *speech*, *hate*, los últimos términos como *tokens* independientes y aparece en una quinta posición ambos términos como bigram. Otras temáticas que predominaron fueron *boycott*, *rumor*, *misinformation*, *advertising*, *online advertising*, que parecen representar las

temáticas troncales de la campaña coincidentes, en esencia, a los principales términos asociados a las búsquedas sobre la campaña en Google por los internautas.

El término *cancer* llama la atención en este conjunto, su prominencia en el corpus está vinculada a una noticia de opinión que expone cómo Facebook fomenta el fenómeno de la desinformación médica y pseudociencia tan protegida como viralizada por la publicidad en la plataforma.

El término *audit* se asocia al resultado negativo de la auditoría a las políticas de este sitio de red social que debieran proteger a sus usuarios de publicaciones y anuncios discriminatorios, así como el resultado contraproducente a sus decisiones de dejar las publicaciones de odio y discriminación del expresidente Donald Trump eran "retrocesos significativos para los derechos civiles" (ISSAC, 2020).

Volviendo a considerar la figura 8, en ella también se incluyen términos como el nombre de Mark Zuckerberg y del expresidente de los Estados Unidos de América Donald Trump. Aunque ambos no son destacados en cuanto a su peso, sí parece reflejar cómo la complicada relación del CEO de Facebook con el expresidente de los Estados Unidos de América Donald Trump consiguió ser de alcance público en el periódico y así se argumenta que este elemento se convierte en el centro de la polémica de la campaña para cuestionar si condicionaría, o no, su éxito. Otro debate versa, esta vez desde un enfoque de negocios, en las posibilidades reales de que la salida de los grandes inversores afectara seriamente, o no, a la economía de la compañía, cuando Facebook cuenta con un total de 8 millones de anunciantes (NYTimes, 2020) cuya mayoría son pequeñas empresas.

En un análisis a los términos se observan otros nombres de varias celebridades: Kim Kardashian West, Katy Perry y Mark Ruffalo. Una consulta al *widget* donde se almacenan los registros de las noticias recolectadas permitió conocer su relación a la campaña. Las figuras públicas mencionadas crearon una tentativa de boicot que consistía en congelar la actividad en Instagram, como respuesta a las insuficientes iniciativas de Facebook en pos de la protección de los derechos civiles. Sin embargo esta iniciativa más que recibir amplio apoyo llevó a ser criticada por la opinión pública en Twitter. Si analizamos las fechas en que esta nueva iniciativa se desarrolló y su publicación en el periódico NYTimes, los datos pudieran

indicar el papel de la protesta en Instagram, pos #StopHateforProfit, en el aumento del interés público por la campaña #StopHateforProfit, según indican los datos del Gráfico 3 de Google Trends con el pico alcanzado entre el 15 y 17 de septiembre a escala mundial.

Los resultados hasta aquí analizados permiten diseñar un esbozo de los hitos de #StopHateforProfit en la opinión pública internacional y sus condicionantes contextuales, sin embargo, se precisa de un estudio hacia la estructura social del dominio para entender y describir la comunidad generadora del discurso #StopHateforProfit cuyo papel y dinámicas de los miembros en red condicionaron el desarrollo de la campaña.

4.2 Análisis de Redes Sociales de la campaña #StopHateforProfit

Cuando analizamos #StopHateforProfit con método propio ARS, y utilizando la herramienta Gephi, se pretende comprender la estructura general de la red, describir indicadores como el tamaño de la red, la medida de cuanto recorrido se necesita para que los participantes más distantes en la campaña establezcan vínculos, entre otros. Además, se pretende conocer si la red que construye el uso de #StopHateforProfit está estrechamente conectada, identificar algunos valores atípicos sobre el comportamiento promedio de sus participantes e identificar a aquellos que desempeñaron un papel en la campaña más activo o pasivo. También se busca conocer si la red presenta muchas conexiones entre sus actores, si cuenta con un núcleo de influyentes cohesivo y si está estructurada por agrupaciones claramente definidas.

4.2.1 Actores participantes

La coalición creadora de Stop Hate for Profit al término de julio de 2020 reconoció el apoyo de esos más de 800 firmantes entre empresas, organizaciones defensoras de derechos civiles y sin fines de lucro, figuras públicas como los Duques de Sussex el príncipe Harry y Meghan Markle, organizaciones como la *American Federation of Teachers*, *Black Lives Matter Global Network*, GLAAD, *Guns Down America*, *Matthew Shepard Foundation*, actores reconocidos como Natalie Portman, Isla Fisher, Jamie Foxx y cientos de empresas que se representan en la siguiente figura 9. En verde se destacan los nombres de las empresas, en azul las organizaciones sin fines de lucro y en rojo las personalidades públicas. El tamaño de

las letras responde a la intensidad de la frecuencia con la que cada usuario de Twitter, siempre considerando los datos recopilados desde www.stophateforprofit.org, fue mencionado en los tuits que se generaron durante los primeros días de campaña para la coalición de #StopHateforProfit.



Figura 9. Personalidades, empresas y organizaciones sin fines de lucro firmantes.
Fuente: Elaboración propia a partir de datos públicos en www.stophateforprofit.org

Si bien fueron reconocidos 805 usuarios diferentes, con el método seguido de ARS se identifican un total actores participantes entre los días 25 de junio y 6 de agosto superior, hasta alcanzar los 76.920. La diferencia radica en la fuente de datos empleada para aportar la primera cifra, www.stophateforprofit.org; y para la segunda, Twitter. De ellos el número de emisores de tuits alcanzó un total de 72.814, mientras que el número de mencionados fue 11.727. De ellos se identificaron un total de 11.838 tuits que no establecieron menciones a otros usuarios en sus mensajes y un total de 825 que mencionaron sus propios usuarios en los tuits publicados.

4.2.2 Características de la estructura de red

Las características de la red que generó #StopHateforProfit y con la herramienta Gephi identifica una red compuesta por los ya mencionados 76.920 nodos (usuarios de Twitter participantes) que quedan vinculados por 233.443 aristas (relaciones tanto de mención como de automención).

El grafo para el microdominio #StopHateforProfit analiza una red dirigida. Y en la tabla 6 se recogen las principales métricas que describen la red:

Medida de red	Valor
Número total de nodos	76.920
Número total de aristas	233.443
Promedio de grado	2,38
Densidad	0,048
Densidad con <i>loops</i> permitidos	0,00004
Diámetro de la red	22
Promedio <i>path length</i>	5,445
Coeficiente de agrupación o <i>clustering coefficient</i>	0,00176
Número de triadas	190.380
Modularidad	0,566
Número total de comunidades identificadas	5.556

Tabla 6. Medidas de red de #StopHateforProfit.
Fuente: elaboración propia

Durante los próximos párrafos, se interpretan los valores de las principales medidas. Así, una característica que destaca es el alto valor de grado promedio de la red (2,38), desde la perspectiva del ARS un alto valor en esta medida se interpreta como un alto número de conexiones con otros nodos, comportamiento que se identifica también en la medida de componentes fuertemente conectados (TARJAN, 1972) cuyo algoritmo después de realizar una búsqueda en profundidad identifica un componente fuertemente conexo en la red de 76.233 nodos, cifra que representa el 99% de la misma, lo que confirma una red caracterizada por estar fuertemente conectada. Si la medida del grado promedio de #StopHateforProfit apunta con mayor precisión al nivel de cohesión estructural de su red social, y esta medida alcanza un valor de 2,4 lo que significa que un participante típico de la campaña #StopHateforProfit realiza entre dos y tres menciones a otros participantes durante su actividad en la misma.

Si se retoma la definición de la densidad del grafo que permite poner en relación el nivel de aristas conectadas dentro de una red en relación con el valor total posible, si para esta red es de 0,048, significa que solo ocurre el 4.8% de todas las posibles aristas. A partir del análisis del tamaño de la red se podía inferir el resultado de esta medida, según NOOY, MRVAR, BATAGELJ (2005, 63) la densidad está inversamente relacionada con el tamaño de la red “cuanto mayor es la red social, menor es la densidad porque el número de aristas posibles aumenta rápidamente con el número de nodos”. Este resultado apunta a una propiedad del dominio #StopHateforProfit, que, a pesar del alto número de conexiones establecidas en el grafo en general, éstas no se han completado en su mayoría según el universo de conexiones posibles.

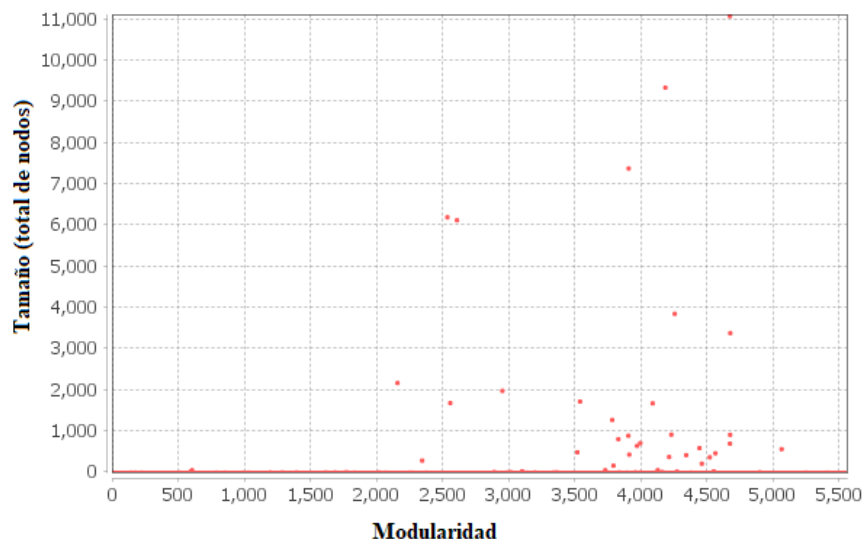
Si se centra la atención en el coeficiente de agrupación de la red en general, este indica de una forma más exacta la densidad del grafo analizada anteriormente. Esta medida en la red general es muy baja (0,00176), lo que significa que aproximadamente sólo el 0,002% de ella está cerrado al estableciendo triadas, más del 99% restante está aún abierto a otras relaciones. Esto permite llevar a la interpretación respecto de que como que en este último caso dos de los tres nodos están conectados, pero falta el tercero, siendo esta medida inversamente proporcional a la medida de centralidad de grado de la red. Según Cherven (2015), si se analiza el comportamiento de esta medida al nivel de los nodos, observaremos que aquellos participantes de nuestra red con pocas conexiones son considerablemente más propensos a tener un alto porcentaje de triadas cerradas, mientras que los coeficientes más bajos provienen de algunos de los miembros de la red conectados más cohesivamente. Este resultado puede reflejar la proximidad o no de los participantes de la campaña durante el periodo de análisis definido, esta medida a nivel de nodo se analizará en el conjunto de líderes de opinión identificados.

Paralelamente, el diámetro del grafo nos permite entender mejor la estructura de la red #StopHateforProfit. Esta medida es de 22, es decir, este es el número máximo de conexiones requeridas para atravesar la red hacia el nodo más alejado y es ciertamente elevado, siendo estos los pasos que se requieren para que el par de nodos más distantes de la red lleguen a alcanzarse entre sí.

Desde la perspectiva de análisis de los actores fue identificada una prevalencia de que el número de actores fuertemente conectados en relación a los débilmente conectados (5.479) mientras que el número de actores fuertemente conectados fue de 76.233.

4.2.2.1 Comunidades en #StopHateforProfit

Entre todas las medidas posibles, la medida de modularidad en **#StopHateforProfit** va a permitir identificar y representar cada subgrafo (las subcomunidades estructurales) por color, siendo una estrategia de visualización más precisa. La métrica de modularidad en la red es de 0,568 y el número total de comunidades identificadas alcanza la cifra de 5.556. Esta distribución por número de nodos se representa en el siguiente gráfico 4.



el criterio de clase de modularidad. La figura 10 visualiza los nodos que han sido objeto de una partición por modularidad de clase usando un *layout* de Fruchterman Reingold. Al centro de la red se vinculan los nodos de la comunidad con mayor valor de grado, el color del nodo representa la subcomunidad a la cual pertenece, siendo los nodos con color ■ parte de la comunidad con modularidad de clase 4.662, ■ nodos de la subcomunidad con modularidad clase 4.174, ■ modularidad clase 3.896, ■ modularidad clase 2.525 y el color ■ agrupa aquellos actores que integran la subcomunidad con modularidad de clase de 2.597.

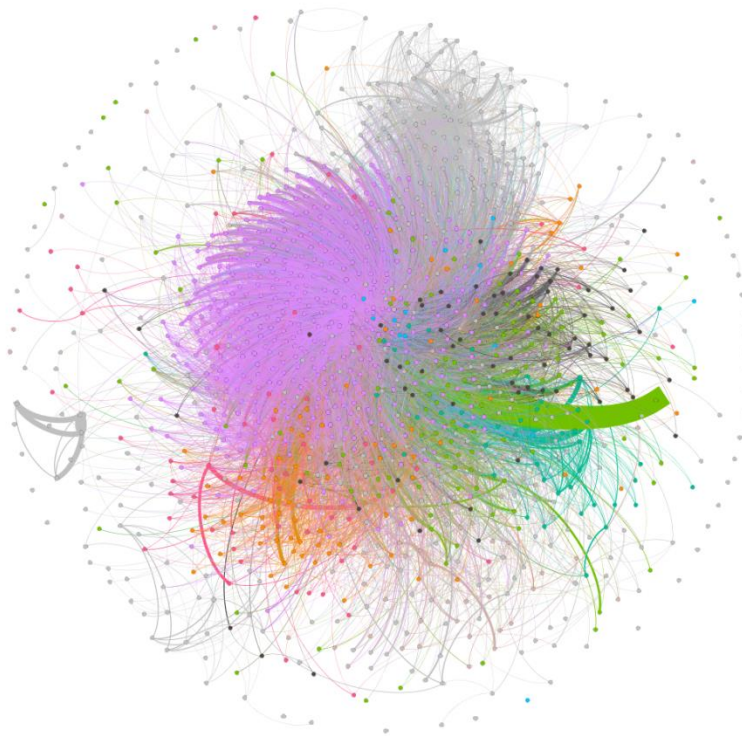


Figura 10. Grafo de las cinco principales comunidades en #StopHateforProfit.
Fuente: Elaboración propia

Considerando el total de miembros de la comunidad se observa en la siguiente tabla 7 que de las cinco comunidades mapeadas en la figura 10, éstas representan el aproximadamente el 50% visible de toda la red con valor de grado igual o mayor de 20.

N comunidad	Clase de modularidad	Nodos	Aristas	% visible de la red
1	4662	11.098	29.794	14,43
2	4174	9.360	17.103	12,17

3	3896	7.397	13.947	9,62
4	2525	6.213	94.66	8,08
5	2597	6.140	83.26	7,98

Tabla 7. Tamaño de las cinco mayores subcomunidades en #StopHateforProfit

Fuente: Elaboración propia

Esta medida revela que el principal subgrafo, según el atributo de la clase de modularidad, está integrado por los usuarios de Twitter que se enumeran slpng_giants, freepress, ColorOfChange, CommonSense, CocaCola, target,barret_williams, Dockers, Ford, Patreon, Verizon, NHMC, LEGO_Group, honda, HP, Senor_Cartmenez, LULAC, adidas, pfizer, Pepsi, VANS_66, HersheyCompany, xp0sefacts, Reebok, dunkindonuts, cvspharmacy, cukey_3, Walgreens, peetscoffee, Chobani, TRULYUNBIASED, Acura, Heineken, JimBeam, beamsuntory, uosdwisrdewohmi, effenvodka, MakersMark, MissScott01, eddiebauer, WhiteCastle, DennysDiner, Earthjustice, rb, adafruit, EJGibney, Marketeer2u, JadeJensen29, FarmAid, VWGroupReddit, behfarj, splcenter, BlueShieldCA, MelissaRyan, Sifill_LDF, Folgers, Jif,smuckers, MilkBone, RideSchwinn, birchHBO,tristanharris, entre otros.

En otro orden, Fresno García, Daly y Sánchez-Cabezudo (2016) afirman que existe un acuerdo en la literatura especializada en ARS que si un nodo presenta valores de mayor centralidad en la red implica una ventaja estructural, rango, poder, influencia o dominio. Las medidas de centralidad, según Freeman, (1979), Newman (2010), Scott (2011) y Cherven (2015), se clasifican en: a) grado, b) intermediación, c) cercanía y d) prestigio o vector propio. A continuación se presentan los análisis del cálculo de las medidas a, b y d, comenzando seguidamente por el análisis de centralidad de grado.

Determinar la centralidad de los nodos de acuerdo con su grado, intermediación, cercanía y prestigio de la red generada a partir de la participación en #StopHateforProfit, permitirá comprender mejor la ventaja estructural, o no, de la posición en la red de alguno de sus actores, así como el rol que jugaron durante la campaña.

4.2.3 Centralidad de grado

Desde una perspectiva sociocéntrica, las redes en general -y sus nodos en particular- son centrales en menor o mayor grado. En este contexto hay que puntualizar que el término ‘centralidad’ es usado para referirnos al nivel de análisis en foco.

4.2.3.1 Actores centrales: grado entrante

Medir el número de conexiones a un nodo refleja cómo de bien o no, está conectado en términos de conexiones directas (FRESNO GARCÍA; DALY; SÁNCHEZ-CABEZUDO, 2016). Entre los usuarios de Twitter que participaron de la conversación en la campaña #StopHateforProfit, la métrica calculada para el conjunto refleja altos valores de grado entrante (*in-degree*) entre 1.000-14.773 para un determinado conjunto de nodos. La figura 11 incluye el grafo que resulta tras aplicar el filtro de acuerdo con el valor de grado entrante de los nodos cuando es mayor que 1.000. El tamaño del nodo es proporcional al valor que alcanza en esa medida.

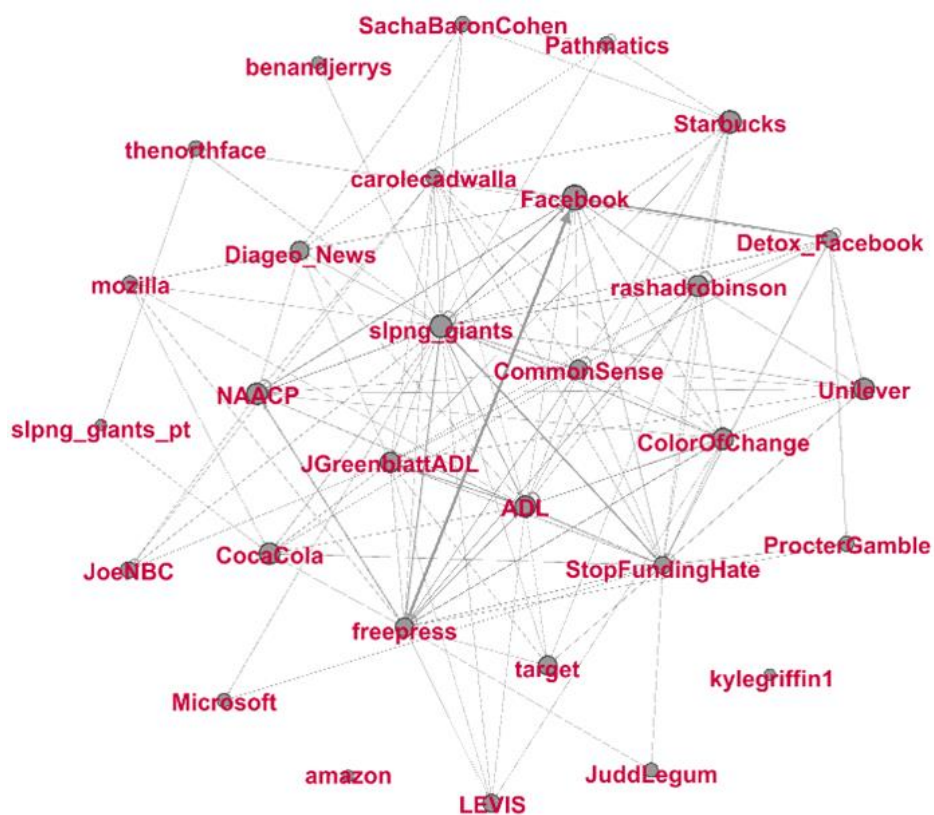


Figura 11. Subgrafo de los nodos con un valor de grado entrante > 1.000.
Fuente: elaboración propia

La intensidad del vínculo entre nodos se observa en el diámetro del arco, en este micro dominio puede observarse que el nodo freepress es quien realiza mayor total de menciones al nodo Facebook dentro de esta red.

Este grafo está formado por las 148 relaciones que vinculan a 30 nodos, lo que representa el 0,04% respecto del total de los nodos que forman la red y el 0,48% de las relaciones entre nodos. Los nodos con la posición más céntrica corresponden a slpng_giants (grado entrante 14.773) y Facebook (grado entrante 11.883). El tercer nodo que recibió mayor valor de grado entrante fue el reconocido actor norteamericano Sacha Baron Cohen, quien recibió un total de 8.506 menciones, así como el CEO de la ADL Jonathan Greenblatt con 3374. Periodistas como Carole Cadwalla, Kyle Griffin y Judd Legum presentan alto valor de grado saliente. La periodista de investigación Carole Cadwalla, quien ganó notoriedad internacional en el año 2018 cuando dio a conocer el escándalo de datos de Facebook-Cambridge Analytica / AggregateIQ, recibió un total de 2.945 menciones.

Como se observa en el núcleo de los 30 actores más populares lo integran miembros de la coalición Stop Hate for Profit: Sleeping Giants, ADL, NAACP, Free Press, Color of Change Common Sense y mozilla, estos ocupan una posición central en la red. A la lista enumerada se les une la compañía de negocio de café Starbucks, que declaró a finales de junio su posible participación en la misma y retirar sus *ads* (anuncios) de Facebook.

Un elemento que destaca del grafo incluido como figura 11 es la ausencia de vínculos creados en la red de dos nodos, uno es el caso notable de Kyle Griffin, escritor y productor en MSNBC del programa de corte político The Last Word, su usuario de Twitter fue mencionado en 1483 ocasiones, sin embargo, se encuentra aislado en el mapa del núcleo de nodos conectados, lo que puede deberse a que dicha mención no aparece vinculada a otras menciones. Paralelamente Amazon, se encuentra en la misma posición aislada de vínculos a los nodos de esta red de relaciones entre los nodos con mayor alto grado entrante. En otro orden de cosas, el nodo correspondiente Judd Legum, periodista, abogado y miembro del personal político estadounidense, alcanza 1.005 de valor de grado entrante, casi en el límite por el que se limitó el núcleo.

Destaca también en este grupo la cuenta oficial de Rashad Robinson, presidente de la organización Color of Change, organización líder en justicia racial cuyo fin es empoderar a las comunidades negras. El presentador de televisión, abogado, comentarista político y expolítico estadounidense Charles Joseph Scarborough, coanfitrión del programa norteamericano *Morning Joe* en la cadena MSNBC, presentan valor de grado entrante 1.255.

Los usuarios de Twitter participantes en #StopHateForProfit hacen un llamado intencionado a marcas y compañías multinacionales como Starbucks (3.712), Unilever (2.684), el canal de noticias de la empresa Diageo_News (2.395), Ben & Jerrys (1.035), LEVIS (1.315), CocaCola (1.301) y Procter & Gamble (1.079), Amazon (1.060) y Microsoft (1.002). Otros usuarios como Pathmatics, plataforma de inteligencia de marketing digital, obtuvo un valor de grado entrante de 2.148 y Detox_Facebook 1.810, en el caso de este usuario, no se pudo comprobar atributos de su perfil debido al carácter privado de la cuenta.

4.2.3.2 Actores centrales: grado saliente

Atendiendo a la medida de grado saliente (*out-degree*), quedó identificado que 66.969 nodos de la red, su 87,06 %, emitió mensajes estableciendo relaciones de mención hacia otros nodos. Este dato se pudiera considerar como un alto índice en el comportamiento de los actores al participar en #StopHateforProfit. Con el objetivo de realizar un análisis más profundo de esta medida de grado, fue extraída una partición de la red cuyo valor de grado saliente es mayor o igual a 50⁵, es decir, que al emitir mensajes a su vez creaban vínculos con una frecuencia de mención mínima a terceros.

La subred que resulta se representa en la figura 12, su estructura se conforma por un total de 93 nodos (0,12% visible) y 531 (0,29% visible) conexiones que se originan desde cada nodo y van hacia otros, del total posible de la red. La intensidad del color azul responde a la frecuencia de las relaciones y el diámetro de los nodos al valor de grado saliente.

⁵ Se selecciona este criterio para analizar la subred dado los intereses de extraer un núcleo condenso de actores más activos.

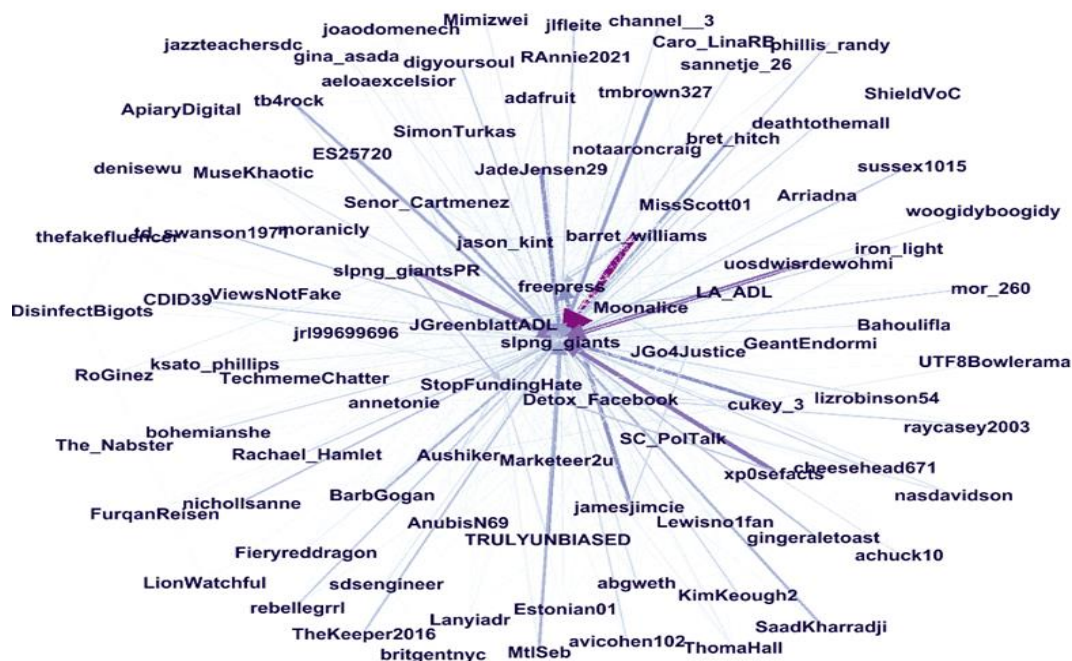


Figura 12. Subgrafo de los nodos con un valor de grado saliente > 50.
Fuente: Elaboración propia

De este subgrafo se debe destacar el núcleo formado por 5 participantes de la campaña con mayor valor de grado saliente. Se identifican en un sentido decreciente a barret_williams (910), seguido por Senor_Cartmenez (416), slpng_giantsPR (387), xp0sefacts (247), y jazzteachersdc (225). La tabla 8 recoge en la primera columna la relación de los 30 nodos con mayor valor de grado saliente y en la segunda, el valor de grado saliente:

Usuario de Twitter	Centralidad de grado saliente
barret_williams	910
Senor_Cartmenez	416
slpng_giantsPR	387
xp0sefacts	247
jazzteachersdc	225
cukey_3	206
TRULYUNBIASED	190
uosdwisrdewohmi	183
slpng_giants	176
jrl99699696	162
MissScott01	160
StopFundingHate	153

JadeJensen29	145
Aushiker	140
BarbGogan	114
Moonalice	109
Freepress	108
MtlSeb	108
Jamesjimcie	106
ShieldVoC	102
Adafruit	101
Caro_LinaRB	94
DisinfectBigots	94
tb4rock	94
deathtothemall	93
tmbrown327	93
Abgweth	92
bret_hitch	92
gina_asada	90
ApiaryDigital	89

Tabla 8. Relación de los 30 nodos con mayor valor de grado saliente.

Fuente: Elaboración propia con datos de la investigación

Partiendo del supuesto de la centralidad grado salientes alcancen altos valores en relación valores del grado de entrante, puede inferirse que, siendo este caso en particular, no se consideran fuentes de información directa; sin embargo, estos nodos pueden proporcionar rutas al conjunto de otros actores centrales en la campaña, sirviendo como posibles agregadores de información para otros participantes de la red.

Analizando estos resultados en el contexto de Twitter, los datos sugieren que el comportamiento de los usuarios en #StopHateforProfit no estaban twitteando con una mínima asiduidad para mencionar a terceros durante el periodo en que se analiza la campaña. Sin embargo, a pesar de la gran actividad del núcleo de nodos, este no puede ser considerado un valor alto para toda la red.

Otros participantes no mencionaron con la misma frecuencia durante el período de la campaña. Los usuarios que participaron, pero no establecieron relaciones de mención a otros usuarios, fue un total 9.951 nodos. Revisando su centralidad y como cabe esperar, todos ellos presentan un valor de grado saliente igual a 0. Se aporta una muestra del conjunto de ellos

con forma de nube de palabras de la figura 13. El criterio de selección de esta muestra es, de una parte, su ocurrencia en la lista de firmantes de la campaña, divulgada en su sitio web oficial y de otra, por ser compañías y figuras públicas reconocidas internacionalmente.

Figura 13. Nodos en #StopHateforProfit con grado saliente = 0.
Fuente: Elaboración propia con <https://wordart.com/>

4.2.3.3 Subred de los actores más influyentes: centralidad de grado

De entre todos los actores de la red (76.920), los 100 actores más influyentes de StopHateforProfit presentan un rango de grado entre 265-14949, representando el 0.13% de la red general, los cuales se vinculan por 821 conexiones, el 0.45% de los vínculos posibles. La figura 14 muestra la subred que generan los actores más influyentes.

La interpretación de la figura 14 pasa por considerar que a mayor valor de grado el nodo se encuentra al centro de la figura, siendo los nodos periféricos los de menor valor dentro de este conjunto seleccionado. El tamaño del nodo se define por el valor del indicador *in-degree*, es decir, a mayor tamaño mayor popularidad del actor (menciones recibidas). El valor cromático es ordinal y presenta el valor *out-degree* del actor.

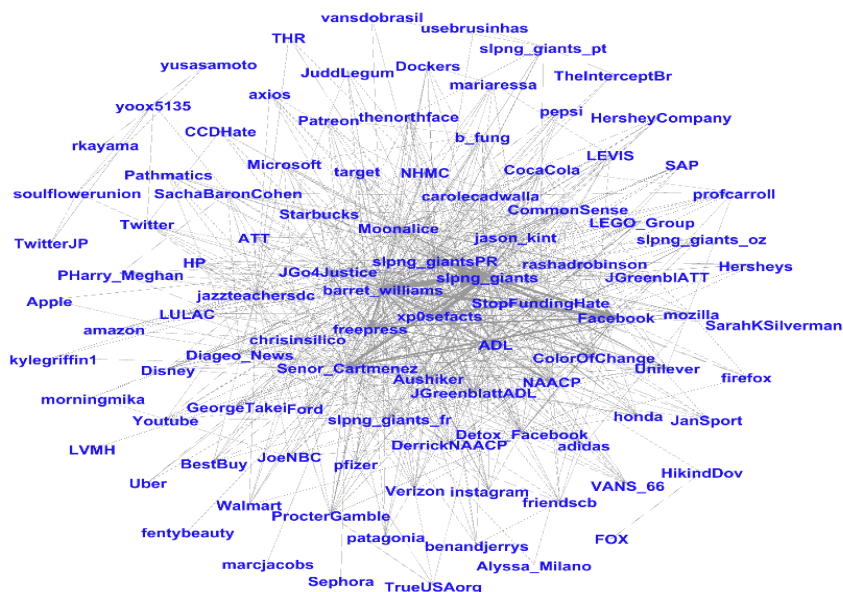


Figura 14. Red de los 100 actores más influyentes en #StopHateforProfit.
Fuente: Elaboración propia

Y las medidas de la red incluida en la anterior figura 14 se recogen en la tabla 9:

Medida de red	Valor
Promedio de grado	8,210
Densidad	0,083
Diámetro de la red	7
Promedio o <i>path length</i>	2,21
Coeficiente de agrupación o <i>clustering coefficient</i>	0,326
Número de triadas	190.380
Número de componentes débilmente conectados	2
Número de componentes fuertemente conectados	70
Modularidad	0,075
Número de comunidades	7

Tabla 9. Medidas de la red formada por los nodos más influyente #StopHateforProfit.
Fuente: Elaboración propia

Considerando la figura 14 se puede apreciar que, de entre los participantes con mayor valor de grado, 7 nodos coinciden con la coalición creadora de la campaña: slpng_giants, ADL, NAACP, freepress, ColorOfChange, CommonSense y mozilla.

Un análisis de las medidas de grado entrante y saliente de este núcleo de 100 participantes denota que si bien existen un conjunto de compañías que resultaron muy populares en la campaña por el valor de grado entrante no se encuentran entre las cuentas activas que crean vínculos de mención en la comunidad participante. Tal es el caso de Starbucks, Unilever, Amazon, Disney, UBER, Walmart, Verizon, Fox, ATT, Dockers, la propia compañía Twitter, entre otras grandes empresas que publicitan en Facebook.

Como se observa en el conjunto de los 100 actores más influyentes únicamente Sleeping Giants forma parte de ambos núcleos con mayores valores, cumpliendo doble un rol y teniendo una posición destacada en la estructura de la red que conforma la comunidad participante de Stop Hate for Profit. Otra figura influyente que se destaca en la campaña, con valores de grado saliente mayor e igual a 100 son freepress, StopFundingHate, Moonalice, barret_williams y slpng_giantsPR; aunque con menores valores de *out-degree* se encuentra Carole Jane Cadwalladr, periodista de investigación y escritora británica para The Observer que fue reconocida internacionalmente en 2018 cuando expuso el escándalo de *Facebook-Cambridge Analytica*. Otros nodos que pertenecen a este grupo, pero que presentan valor de grado entrante 0, no recibieron menciones, fueron cukey_3, BarbGogan y el usuario de Twitter jrl99699696. En el caso de los influyentes con valor 0 de grado saliente son los que se representan en la siguiente nube de palabras de la figura 15.



Figura 15. Actores de #StopHateforProfit con valor de grado saliente = 0.

Como se esperaba, Facebook es un nodo con un rol muy específico en la red, su papel se limitó a ser nodo destino de las menciones realizadas en las conversaciones de StopHateforProfit, por lo que llega a ser el segundo nodo más popular de la campaña.

4.2.4 Centralidad de intermediación

La perspectiva de la centralidad de intermediación se fundamenta en identificar a aquellos actores que ejercen el papel de puente entre actores de una red por la que circulan noticias, información, percepciones, rumores o falsedades es un factor crítico para cualquier organización.

Su aplicación en la campaña busca comprender la influencia relativa de los nodos individuales dentro de la red en términos de cuáles son altamente influyentes en conectar otras regiones de nodos de la red. La siguiente figura 16 muestra una subred, cuya topología corresponde a los nodos que presentan mayores medidas de centralidad de intermediación entre el rango de 1-150, normalizados a valores de entre 0 y 1. El diámetro de la red resalta los mayores valores obtenidos y dada la totalidad de nodos de la red origen, la topología de la red también cumple con incluir sólo aquellos nodos con mayor valor de grado en el rango 200-14949 y la intensidad del color verde representa el mayor valor, nodos con menores valores de grado la intensidad cromática tiende a aproximarse al color blanco. El algoritmo usado para el *layout* fue Force Atlas 2 con ajuste de la escala a 200 entre otras opciones de configuración para evitar el solapamiento de los nodos y obtener una mejor definición de este grafo.

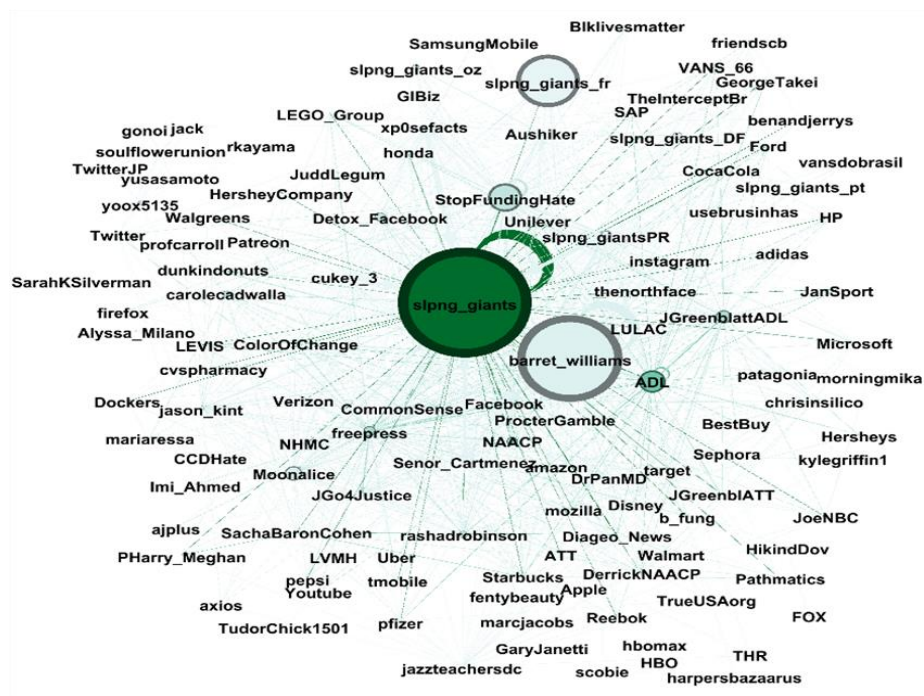


Figura 16. Red de centralidad de intermediación.

Fuente: Elaboración propia en Gephi.

Como se observa en el mapa de la figura 16 coincide con que las mayores medidas de centralidad de intermediación la presentan nodos como los recogidos en la siguiente tabla 10:

Usuario de Twitter	Centralidad normalizada de intermediación
slpng_giants	0,03950335247234424
barret_williams	0,031068589394270558
slpng_giants_fr	0,018784796234679747
StopFundingHate	0,010000679280102592
ADL	0,007805797394572568
Moonalice	0,004993160613934969
JGreenblattADL	0,00449438608666436
Freepress	0,003398918981913474
Detox_Facebook	0,002970429128424218
slpng_giantsPR	0,0028479604794427464

Tabla 10. Relación de nodos con mayor valor de intermediación.

Fuente: Elaboración propia

Este método proporciona una definición diferente de centralidad, estos nodos al contar con el mayor puntaje de centralidad de intermediación son los que ofrecen la ruta más directa entre *clusters* de nodos, que de otro modo estarían desconectados en la red. Pudieran

considerarse puentes en la topología de la red visto que juegan un papel único en la comprensión de cómo las influencias de los mensajes fluyen dentro de la misma. Estos resultados sugieren con qué frecuencia los participantes de la campaña se encuentran en el camino más corto entre otros dos nodos.

Es probable, que un miembro con una puntuación de centralidad de intermediación muy alta pueda alcanzar valores muy bajos en otras medidas de centralidad. Esto dependerá en gran medida de la estructura de la red y es mucho más probable que sea cierto en redes con niveles de agrupamiento más altos. Este supuesto será analizado en los siguientes epígrafes.

Cuando los nodos están altamente conectados a otros nodos con altos niveles de influencia, el resultado será un alto nivel de centralidad de *eigenvector* o vector propio. Al aplicar esta medida se identifica una partición de 2.843 nodos con altos mayor valor de vector propio de la red 6.6, diez de los líderes de opinión se listan a continuación. En la primera columna de la tabla 11, la identificación del actor en Twitter; en la segunda, el valor de centralidad de grado entrante; en la tercera, el valor de centralidad de grado saliente; en la cuarta, el valor de grado; y, la última, el valor de vector propio. El orden en el que se exponen atiende al criterio de grado entrante:

Actor	Grado entrante	Grado saliente	Grado	Eigenvector
slpng_giants	14.773	176	14.949	1
Facebook	11.883	0	11.883	0,8.1321
SachaBaronCohen	8.506	3	8.509	0,5.7284
ADL	6.818	47	6.865	0,4.7418
NAACP	4.841	15	4.856	0,3.4341
Starbucks	3.712	0	3.712	0,2.6804
freepress	3.379	108	3.487	0,2.4094
JGreenblattADL	3.374	63	3.437	0,2.4058
ColorOfChange	3.133	5	3.138	0,2.2842
CommonSense	2.963	10	2.973	0,2.1292

Tabla 11 Actores con mayor valor de grado de vector propio.

Fuente: elaboración propia

Esta medida se basa en que la centralidad de un nodo depende de la centralidad de sus vecinos (prominencia). Así, el poder y el status de los miembros de la coalición slpng_giants, DL, NAACP, freepress y CommonSense es indiscutible; además del liderazgo de SachaBaronCohen y Greenblatt. La connotación de poder procede del *status* de los vecinos

de un determinado actor (*alters*). El análisis de los datos recogidos en la última columna de la tabla 11, confirma la calidad de la centralidad del nodo Facebook considerando la centralidad los nodos a los que está conectado: slpng_giants, ADL, NAACP, freepress y CommonSense.

Una vez identificado el núcleo de participantes influyentes y su rol en la estructura de la red interesa analizar las temáticas que definieron el discurso del dominio #StopHateforProfit en el siguiente epígrafe.

4.3 Análisis temático de la campaña #StopHateForProfit

Para identificar los temas discutidos y las emociones implícitas en la narrativa de los usuarios que participaron en la campaña tuiteando el hashtag #StopHateForProfit, se emplearon varios métodos y técnicas de procesamiento del lenguaje natural, de identificación de sentimientos y de modelado de tópicos.

En un primer momento se analizaron los términos con mayor adherencia al debate en StopHateforProfit, lo cual fue posible a partir de las funciones en R (para el procesamiento de los tuits) y Python (para el procesamiento de los retuits) que miden la frecuencia de las palabras en los tuits (figura 17) y retuits (figura 18). En las siguientes nubes de palabras la frecuencia del término en ambos corpus procesados está representada por el tamaño de la palabra. El color rojo se escoge para dar destaque a los términos con mayores valores de frecuencia obtenidos.



Los términos con mayores valores de ocurrencia son el hashtag oficial de la campaña y Facebook; en el caso del conteo final de Facebook se observó en los tuits (fig 17) que este fue mencionado tanto como hashtag, como palabra del corpus o como mención de la cuenta en Twitter, en total aparece en 12423 ocasiones. Como se esperaba el comportamiento de este mismo término se amplificó al ser retuiteado. Los participantes en general lo mencionaron en un total de 76832 veces (figura 18).

En relación a nombres propios, si bien Mark Zuckerberg fue la persona más comentada en los tuits como texto o como mención de cuenta de twitter, este resultado muda en el conjunto de los retuits donde Sasha Baron Cohen llega a ser la figura pública central de las discusiones de #StopHateforProfit con un total de 9832 ocurrencias en el corpus. Este resultado apoya la discusión del epígrafe anterior donde se identificó al actor como la figura pública más influyente y de impacto al ser también el usuario de los tuis más retuiteados y que recibió mayor cantidad de likes por otros usuarios de Twitter.

Al observar la figura 17 se aprecia que prevalecen términos de acción, como *join*, *act*, *stand*, empleados con el claro propósito de hacer un llamamiento a las compañías a unirse a la campaña parando la publicidad en la plataforma, e término *loud* (ruidoso, en su traducción al español) destaca en cuanto a frecuencia de ocurrencia y está relacionado a esa

estrategia seguida por la coalición de generar ruido por parte de la opinión pública y llamar la atención en las redes a las empresas y CEOs de Facebook.

Interesa destacar que si bien la coalición no incluyó como parte de su narrativa en la campaña mensajes que incitaran a los usuarios o empresas a eliminar la cuenta de Facebook, sino a pausar su publicidad, curiosamente aparecen hashtags emitidos por otros participantes con esa intencionalidad, como son *deletefacebook* (479 ocurrencias) y *deletefacebookknow* (96), los que fueron difundidos en retuits 225 y 16 veces respectivamente. Aunque no son cuantitativamente representativos de la totalidad de términos, este resultado sugiere cuáles fueron las reacciones de la opinión pública en relación a Facebook, por su parte el termino *pause* aparece con 656 de frecuencia total en los tuits y empleado un total de 8611 veces en la difusión de dichos tuits.

Los temas estaban conectados con el tema principal de la campaña contra el discurso del odio, sin embargo, este término aparece con menor frecuencia (419 veces) en relación a los 5 principales del *dataset* de tuits y en retuits se emplea 376 veces. Un resultado similar arrojó el análisis del comportamiento de las consultas en Google donde los internautas al buscar información sobre la campaña mostraban una mayor adherencia a temas como *advertising*, *boycott ads*, para informarse sobre la misma.

En cuanto al tema sobre el fenómeno del discurso del odio en la plataforma, destacan por un lado términos asociados al racismo, la intolerancia, la presencia de la supremacía blanca en la plataforma como *racism* (597 tuits/3010 retuits), *bigotry* (382 en tuits/783 en retuits), *white supremacy* (14 tuits/80retuits) y *misinformation* (271 tuits/1585 retuits) y por otro lado un conjunto de términos relacionados a los movimientos activistas que combaten con fuerza esos fenómenos nocivos a la democracia, como son el movimiento black lives matters (141tuits/568retuits) y el hashtag de forma abreviada #BLM (48 tuits/ 20 retuits), también destaca términos sobre el fenómeno de la desinformación (271 tuits/1585 retuits).

El dominio StopHateforProfit se caracteriza por incluir debates con foco en determinadas regiones geográficas, para representar esta dimensión de análisis se procesa el dataset de los tuits y retuits, así se obtiene mayor cobertura y representatividad. El resultado de esta clasificación, se obtuvo empleando el widget *Document Map* de Orange. En la figura

19 se muestra el resultado del análisis de los retuits. Las zonas mencionadas aparecen con destaque en color rojo, la intensidad cromática representa la frecuencia de ocurrencia tuits transmitidos por sus participantes.

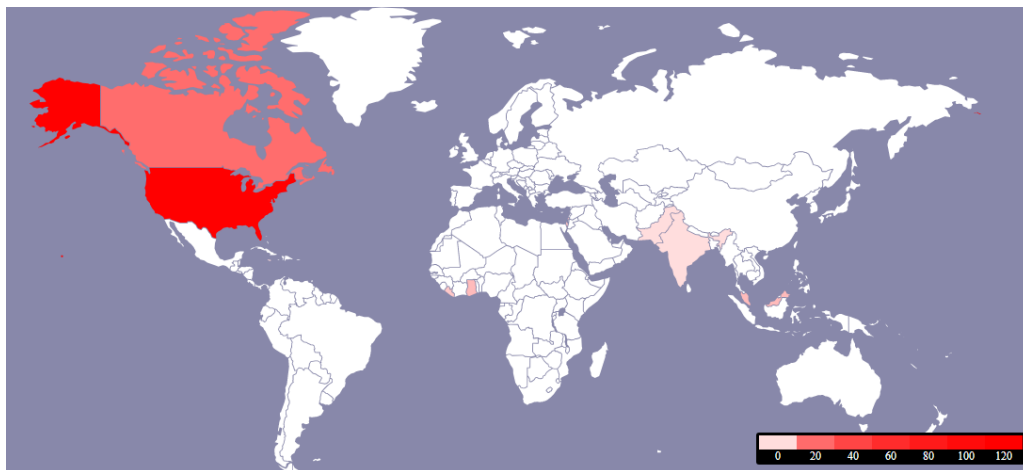


Figura 19 Regiones que fueron temas de debate en #StopHateforProfit (retuits)
Fuente: Mapa generado por el widget de Orange con datos de la investigación.

Las regiones geográficas más abordadas se encuadran en varias regiones de los Estados Unidos de América (116 menciones), Canadá (19), Ghana y Malasia (3), Liberia (2), Reino unido, India y Paquistán (1). En la región norteamericana se centra principalmente en el estado de Ohio (93).

Una posible hipótesis de los datos procedentes de los Estados Unidos de América es el aumento de las preocupaciones entre sus ciudadanos por el discurso del odio a partir del aumento de crímenes de odio reportados por la Oficina Federal de Investigaciones (FBI) (ADL, 2020).

Un análisis de contenido de los tuits que incluyen el término Canadá muestra la trascendencia de la campaña el 1ro de julio del 2020 en este contexto, celebración del Día de la independencia de Canadá. En este día festivo fue noticia y obtuvo muchos retuits un post de la cuenta oficial en Twitter de Globalnews.ca que los cinco bancos más grandes de ese país, Scotiabank, RBC, CIBC, BMO y TD, se unieron al boicot internacional a Facebook por la preocupación sobre las manifestaciones del racismo, la violencia y la desinformación en la plataforma (CANADIAN PRESS, 2020). También destacan los anuncios de la participación a parar de publicitar en el sitio de Facebook durante el mes de julio grandes marcas como

Lululemon Athletica y Mountain Equipment Co-op en apoyo a la campaña #StopHateForProfit.

Este análisis sugiere alguno de los temas centrales pero no llega a revelar las temáticas implícitas en la narrativa predominante de la campaña y como método no llega a ser validado para el modelado de tópicos, donde métodos probabilísticos serían más precisos en descubrirlos. Para validar y profundizar en el análisis del contexto de empleo de dichos términos en la campaña fue empleado el algoritmo LDA que es un método que facilita advertir tópicos latentes que no están explícitos mediante términos (BLEI; NG; JORDAN, 2003; BLEI; LAFFERTY, 2006). Estos modelos se denominan "modelos temáticos" porque los patrones descubiertos a menudo reflejan los temas subyacentes que se combinaron para formar los documentos. A estos fines resulta útil a las Ciencias de la Información pues las inferencias a nivel de documento beneficiarían la recuperación de información, clasificación y navegación por los temas descubiertos. La suposición básica bajo LDA es que los documentos se ven como una mezcla probabilística de temas latentes; un tema tiene una distribución de probabilidad sobre las palabras y cada documento se modela sobre la base de un modelo de *bag of words*.

La siguiente tabla 12 muestra los resultados de empleo de este método en el conjunto de datos de retuits, ya que al incluir mayor cantidad de datos beneficia descubrimiento de tales temáticas. Cada tópico está compuesto por k términos, en esta investigación se asigna valor 10. El orden jerárquico de k sigue el criterio de su relevancia en el corpus analizado.

Tópicos	Términos
1	slpng_giants, stophateforprofit, confirmed, facebook, july, campaign, remains, advertising, joined, caller
2	stophateforprofit, medium, detox_facebook, social, stopfundinghate, said, american, month, major, directly
3	facebook, platform, hate, outlet, continue, inaction, user, adl, disinformation, sit
4	new, one, conspiracy, theory, say, movement, responsibility, hate, billion, worry
5	ad, facebook, company, 11, today, boycott, join, stop, first, sign
6	advertiser, racism, hatred, sky, next, month, let, hearing, demand, sort
7	piece, past, daily, week, group, right, stophateforprofit, civil, ceo, stay
8	jgreenblattadl, spending, news, f, joining, ending, bomb, wall, ceasing, stophateforprofit

Tópicos	Términos
9	mark, zuckerberg, facebook, published, stophateforprofit, result, cocacola, effort, july, responsible
10	hate, white, supremacist, facebook, time, naacp, adl, stand, stop, freepress

Tabla 12. Relación de los diez tópicos latentes principales que describen el debate de #StophateforProfit

Fuente: elaboración propia con datos extraídos del reporte del widget *Topic modeling* del Orange

Para medir la calidad de este resultado se usó la métrica de evaluación coherencia de tópicos (*Coherence Topic*, término en inglés), (MIMNO et al., 2011), de esta forma se identifican los tres tópicos más apropiados que representan la Campaña #StopHateforProfit.

La métrica de coherencia de tópicos presentó un valor global de 0.49259, valor medido en un rango entre 0 y 1 (EKINCI; OMURCA, 2019) que refleja el valor de coherencia semántica de los temas individuales extraídos. En relación a los valores de coherencia, un resultado próximo a cero indica baja coherencia en relación a valores con proximidad al valor 1, siendo en este caso de alta calidad su coherencia

A partir del widget gráfico de escalamiento multidimensional (MDS, en sus siglas en inglés) (WICKELMAIER, 2003) de Orange se visualizan los tópicos modelados y la representación de su valor de probabilidad de tópico marginal. La siguiente figura 20 representa esta relación por tópicos descubiertos a partir del LDA, representados por puntos (nodos) en el MDS, donde el color y tamaño del nodo es una función de la probabilidad de tema marginal (MTP, en sus siglas en inglés) para cada tema extraído del corpus de tuits representan el atributo del valor de probabilidad de tópico marginal, el cual se incluye como rango de valores en la leyenda de la esquina inferior derecha de la figura. Cuanto mayor sea el tamaño del nodo, más fuerte será el tema representado por los términos en el corpus.

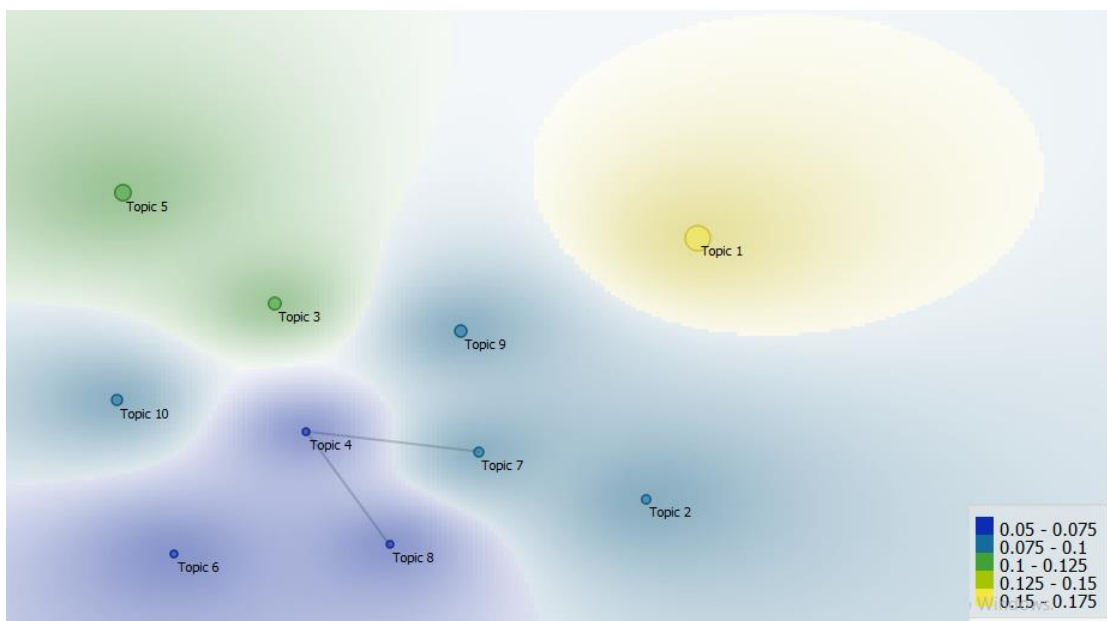


Figura 20. Mapa MDS mostrando la probabilidad de t pico marginal de los t picos descubiertos en #StopHateforProfit utilizando LDA.

Fuente: visualizaci n generada en el widget MDS de Orange con datos de la investigaci n.

La visualizaci n de los t picos LDA con MDS muestra que el t pico 1 tiene la probabilidad marginal m s alta de 0.171879, seguido del t pico cinco con valor 0.120859 y el t pico tres con valor 0.102541. Estos resultados sugieren la fortaleza de representaci n del t pico en el corpus.

Con valor medio se encuentran los t picos 9 (0.096191), el t pico 10 (0.089522), t pico 7 (0.0846564) t pico 2 (0.080945). Los t picos 4 (0.072077), 6 (0.725202) y 8 (0.0709327) presentan probabilidad marginal m s baja en el corpus analizado.

El t pico 1 resulta el tema m s fuertemente representado en la camp a #StopHateforProfit. En la siguiente figura 21 se muestra una descripci n detallada de los tokens, tags y el texto de uno de los tuits con gran adherencia al mismo:

Topic 1: 0.838292
 Topic 2: 0
 Topic 3: 0
 Topic 4: 0
 Topic 5: 0
 Topic 6: 0
 Topic 7: 0
 Topic 8: 0.0950368
 Topic 9: 0
 Topic 10: 0
 text: RT @slpng_giants: **CONFIRMED**: BOOM! @verizon has joined the #StopHateForProfit campaign and will pause their advertising on Facebook platfor...
 Tokens & Tags: slpng_giants confirmed boom verizon joined stophateforprofit campaign pause advertising facebook platfor

Figura 21. Ejemplo de retuit con mayor adherencia al tópico 1
 Fuente: reporte de Orange con datos de la investigación

Se advierte analizando el texto del tuit que este tópico refiere a la inclusión de la compañía Verizon en la campaña #StopHateforProfit, cuyo emisor de origen fue la cuenta oficial de Sleeping Giants. Si retomamos el debate inicial del epígrafe donde se analizan los hitos según fecha de publicación, se menciona que este tuit fue una de las discusiones más difundidas. Este resultado confirma uno de los elementos ya discutidos sobre los comportamientos de divulgación mensajes y la campaña que es, la estrecha relación temática de los temas troncales con los hitos de inclusión de las compañías firmantes a favor de su misión en contra el discurso del odio y la desinformación que se difunde en la plataforma de Facebook.

Conocidas las temáticas centrales de #StopHateforProfit, interesa conocer cuáles fueron las emociones y sentimientos que representaron su debate, lo cual es discutido en el siguiente subepígrafe 4.6.

4.4 Análisis de sentimientos en #StopHateforProfit

El análisis de emociones fue aplicado en dos momentos, en el procesamiento del dataset de los 25480 tuits recolectados fue empleando el lenguaje R, que se basa en el NRC Emotion Lexicon descrito en el epígrafe de procedimientos metodológicos y el lenguaje Python mediante el software de mineración Orange para el procesamiento y visualización de los retuits. Previo a ejecutarse la función que mide la presencia de las emociones en el texto,

fue necesario aplicar un control de calidad del pre-procesamiento de los tuits, a fin de garantizar mayor precisión en los resultados del método.

Antes de profundizar en las categorías de las 8 emociones conviene saber de forma general cuál fue el proceso de análisis llevado a cabo por la función ‘syuzhet’ para extraer sentimientos a partir de su consulta al diccionario NRC y los resultados obtenidos.

Fueron identificados varios idiomas en el corpus de los datos recuperados de Twitter por tanto se precisó aplicar varios procesos de *stopwords* según el idioma, a saber: ‘*english*’, ‘*spanish*’, ‘*portuguese*’, ‘*german*’, ‘*italian*’, ‘*dutch*’, ‘*french*’ y ‘*finnish*’. En este proceso la revisión de calidad de la limpieza se aplicó con la función *inspect*, con la función *removeWords* y *stripWhitespace*, se eliminaron palabras y espacios extras en el corpus que precisaron ser subsanados en vistas a obtener mayor calidad del método de análisis de sentimientos. El siguiente tramo de código permitió sustituir algunos *strings* donde las funciones de limpieza no funcionaron:

```
corpus <- tm_map(corpus, gsub,  
                  pattern = '????T',  
                  replacement = "")  
inspect(corpus[5000:6000])
```

Tabla 13. Script de R con la función ‘reemplazar’ para sustituir términos en el corpus.

R procesó el texto y lo transformó en un vector de caracteres, para analizarlos de forma individual. Finalmente en la figura 21 se representa gráficamente el resultado de este proceso.

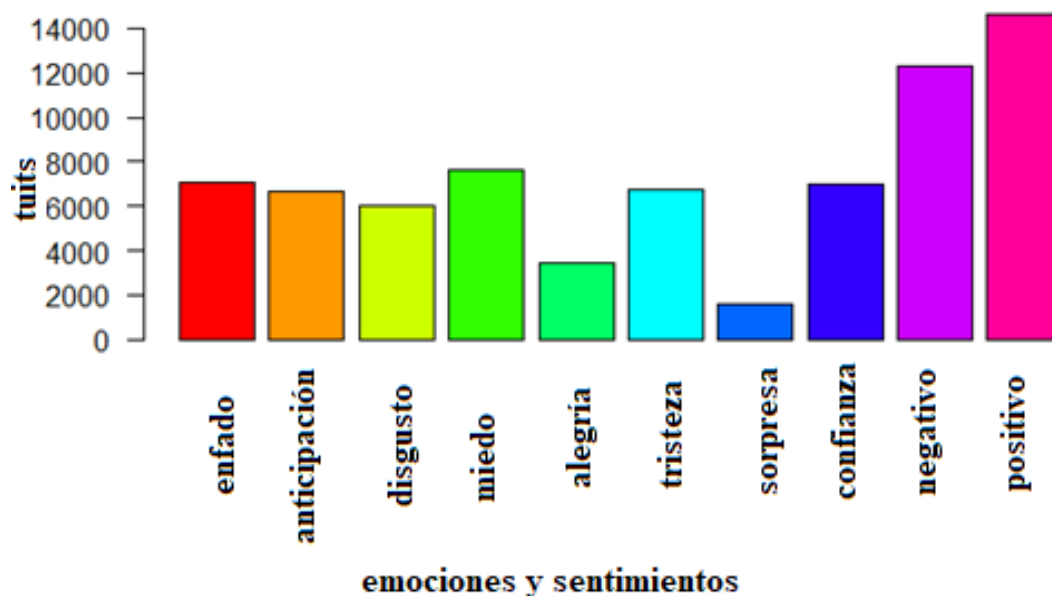


Figura 22. Distribución de sentimientos y emociones para #StopHateforProfit en 2020
Fuente: obtenido de la minería de datos en R con datos de la investigación.

Según la valencia de los sentimientos, la campaña #StopHateforProfit se caracterizó por una representatividad de tuits donde prevalecen sentimientos positivos sobre los negativos, pues son más de 14 mil tuits que lo expresan. A pesar de ser un resultado relevante, alrededor de 12 mil tuits indican sentimientos negativos en la narrativa de la comunidad activa de la campaña. La siguiente tabla 14 muestra la distribución de los 30 principales tuits cuyo vocabulario es predominantemente positivo. Los valores numéricos representan la ocurrencia en el texto de palabras con carga emocional de enfado, anticipación, disgusto, miedo, alegría, tristeza, sorpresa y confianza.

ID	enfado	anticipación	disgusto	miedo	alegría	tristeza	sorpresa	confianza	(-)	(+)
15562	0	3	0	0	2	0	1	3	0	6
5394	2	1	1	0	1	0	0	2	1	5
14384	1	1	1	1	1	1	0	2	1	5
17596	0	1	1	1	2	0	0	2	1	5
15518	2	4	0	0	5	2	2	3	1	5
13547	1	3	0	0	3	0	1	2	0	5
14306	0	2	0	1	2	0	1	4	0	5
13256	0	0	0	0	2	0	0	4	0	5
24791	0	1	0	0	2	0	1	3	1	5

ID	enfado	anticipación	disgusto	miedo	alegría	tristeza	sorpresa	confianza	(-)	(+)
5138	0	1	0	0	1	0	0	3	0	5
8048	0	1	0	0	2	0	0	2	0	5
18357	0	1	0	0	1	0	0	1	0	5
24065	0	1	0	0	1	0	0	1	0	5
2931	2	1	2	1	0	1	1	1	2	4
22867	2	0	2	2	1	2	0	1	2	4
23777	1	2	2	1	1	0	0	3	3	4
9033	3	1	1	2	2	1	1	3	2	4
1567	2	1	1	0	2	0	2	3	1	4
9500	2	1	1	1	1	1	0	2	2	4
16387	2	2	1	0	0	1	0	2	1	4
6722	2	1	1	2	2	1	0	1	2	4
13137	2	2	1	2	1	2	0	0	2	4
19729	1	2	1	2	1	1	1	3	2	4
20192	1	0	1	1	2	1	0	2	2	4
6968	1	2	1	0	1	0	1	1	0	4
9910	1	1	1	1	1	0	1	1	1	4
14388	1	1	1	1	1	1	0	1	1	4
2555	1	1	1	0	1	0	0	1	2	4
20663	1	0	1	1	0	1	0	0	1	4
18844	0	1	1	1	0	0	0	2	2	4

Tabla 14. Clasificación de emociones y sentimientos de 30 principales tuits con mayor peso positivo.
Fuente: elaboración propia a partir de datos extraídos con RStudio.

Para entender cómo es identificada la presencia de valor emocional en los tuits, la tabla 15 muestra los cinco principales tuits con mayor valor de sentimientos positivos y en la 16 se analizan las emociones evocadas por algunos términos del propio tuit.

ID	Corpus del tuit
5138	facebook <i>agree hire</i> csuite level executive deep civil <i>rights</i> expertise evaluates produc..
13256	frutabites believes <i>freedom equality</i> justice well humbly follow footsteps brands <i>admire</i> such
24791	youve got love innocent <i>great</i> example responsible business using influence good csr
8048	samansco stand <i>love</i> opportunity equality truth stand <i>acceptance</i> stand with
18357	oh fun tidbits can learn measure quickanswers pick trend see people think question

Tabla 15. Cinco principales tuits con mayor ocurrencia de vocabulario positivo.

Fuente: elaboración propia a partir de datos extraídos con RStudio.

ID	enfado	anticipación	disgusto	miedo	alegría	tristeza	sorpresa	confianza	(-)	(+)
Agree	0	0	0	0	0	0	0	0	0	1
hire	0	1	0	0	1	0	0	1	0	1
Freedom	0	0	0	0	1	0	0	1	0	1
Admire	0	0	0	0	0	0	0	1	0	1
Responsible	0	0	0	0	0	0	0	1	0	1
Acceptance	0	0	0	0	0	0	0	1	0	1
fun	0	1	0	0	0	0	0	1	0	1

Tabla 16. Clasificación de emociones y sentimientos de vocabularios extraídos de los cinco tuits con mayor con mayor valor positivo.

Fuente: elaboración propia a partir de datos extraídos con RStudio.

Como se observa en la tabla 16 la ocurrencia de emociones como anticipación no afecta que se clasifique como un tuit negativo, por la predominancia de términos con carga emocional de alegría, sorpresa y confianza.

Emociones asociadas a sentimientos negativos también caracterizaron los mensajes difundidos en #StopHateforProfit. Las siguientes tablas 17 y 18 explicitan la presencia de este tipo de sentimiento y emociones evocadas a partir de los términos que estructuran el tuit.

ID tuit	Corpus del tuit
3672	totally despise mark zuckerberg facebook just <i>vacuous greedy</i>
3896	please keep <i>pressure</i> join boycottfacebook movement force psychopath zuckerberg erase <i>hate</i> bigot

Tabla 17. Principales tuits con mayor ocurrencia de vocabulario negativo.

Fuente: elaboración propia a partir de datos extraídos con RStudio.

ID	enfado	anticipación	disgusto	miedo	alegría	tristeza	sorpresa	confianza	(-)	(+)
greedy	0	0	1	0	0	0	0	0	1	0
vacuous	0	0	1	0	0	0	0	0	1	0
idiot	0	0	1	0	0	0	0	0	1	0
devil	1	1	1	1	0	1	0	0	1	0
depression	0	0	0	0	0	1	0	0	1	0
nasty	1	0	1	1	0	1	0	0	1	0
hate	1	0	1	1	0	1	0	0	1	0
militia	1	0	0	1	0	1	0	0	1	0
holocaust	1	0	1	1	0	1	0	0	1	0
anxiety	1	1	0	1	0	1	0	0	1	0

Tabla 18. Clasificación de emociones y sentimientos de vocabularios extraídos de los cinco tuits con mayor con mayor peso negativo. Fuente: elaboración propia a partir de datos extraídos con RStudio.

Como se observa en los datos de las cuatro anteriores tablas, el *NRC Emotion Lexicon* detecta si un término dentro de un mensaje transmite un sentimiento positivo o negativo, y tiene en cuenta que el mismo término puede expresar diferentes sentimientos en diferentes contextos, para ello, el algoritmo está entrenado para detectar la presencia de términos negativos que condicionen este resultado y la posición de la palabra frase analizada en la oración. Si se analiza esta relación de sentimientos negativos en el corpus de tuits se aprecia incluso el uso de términos muy ofensivos que denotan enfado hacia el CEO de la compañía de Facebook. La presencia predominante de emociones de enfado también se acompaña de una carga de emociones que denotan miedo, como es el caso del tuit con ID 16688 de la tabla 17.

En resumen, se observa en los tuis analizados que hay una presencia positiva más fuerte que negativa en la opinión del público participante. El miedo activo en la opinión pública tuvo un papel fundamental en la prevención y anticipación de las consecuencias negativas del fenómeno de difusión del discurso del odio y desinformación en Facebook. El sentimiento de miedo como respuesta al discurso de odio, presente en forma de tuits con el hashtag #StopHateforProfit, puede interpretarse como una manifestación activa del miedo como respuesta social adaptativa y una herramienta para convencer sobre los males derivados de este discurso dañino a la democracia de la sociedad.

Para entender cómo se comportó esta dimensión de análisis de sentimientos en la difusión de mensajes se mapeó esta clasificación de sentimientos en los retuits. El siguiente mapa de calor de la figura 23 muestra esta relación. El mapa de calor de la figura es un método gráfico que nos visualizar valores de atributos en una matriz bidireccional con las variables numéricas del peso de los sentimientos presentes en los mensajes sobre la campaña #StopHateforProfit. Los valores se representan por color según la paleta de colores seleccionada (tonalidades entre color blanco, amarillo y azul). Al combinar la variable de clase y los atributos en los ejes x e y, vemos dónde los valores de atributo positivo o negativo son más fuertes y dónde más débiles, lo que nos permite encontrar características típicas para

cada clase de sentimientos. Los retuits son agrupados atendiendo a la similitud jerárquica en distancias euclidianas.

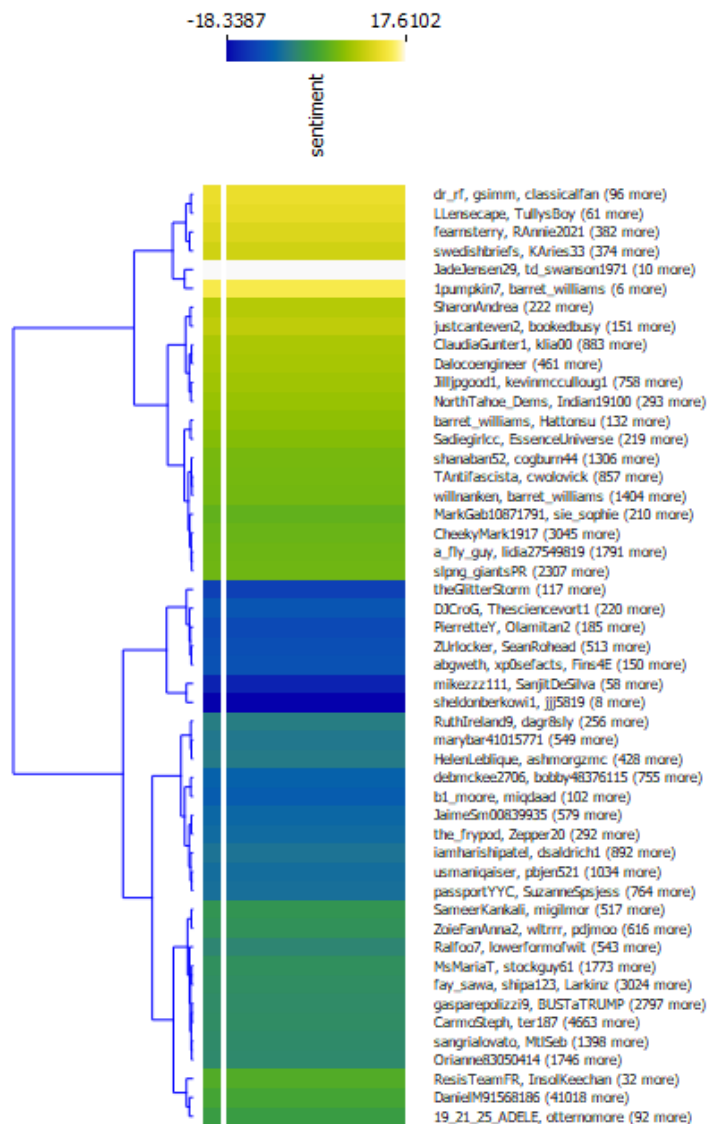


Figura 23. Mapa de calor de la clasificación de sentimientos en los retuits.

Fuente: imagen generada en el widget mapa de calor de Orange.

La representación de los sentimientos en este mapa de calor permite entender con detalle la medida de distribución por clusters el peso de los sentimientos negativos y positivos en #StopHateforProfit, a la vez que conecta a sus emisores. El resultado del análisis de

sentimiento son cuatro columnas también con puntaje de neutral, y un puntaje compuesto que agrega los valores anteriores en un único número. En el mapa de calor se pueden apreciar valores positivos (representados con color blanco) estos representan mensajes positivos, mientras que los negativos (azul) representan los negativos. El widget nos permite seleccionar un segmento del cluster y analizar los retuit de origen de los sentimientos mapeados. En la siguiente figura se muestran cuatro ejemplos del cluster de mayor peso en cuanto a la presencia de vocabulario con sentimientos negativos.

sentiment:	-17.8571
screenName:	sheldonberkowi1
text:	RT @DrPanMD: Unfortunately, online threats are just trolling. Anti-vaxxers use violent rhetoric online to incite res world violence. #Stop...
sentiment:	-18.1818
screenName:	jjj5819
text:	RT @IHateTr63686024: #StopHateForProfit #Boycott these RACIST COMPANIES
sentiment:	-20
screenName:	MegaphoneTexas
text:	RT @roshan18: Facebook has failed to curb toxic content on its platform — including misinformation, violent threats, and voter suppression...

Figura 24. Ejemplo de visualización de retuits con mayor puntuación negativa del mapa de calor.
Fuente: imagen generada en el widget mapa de calor de Orange.

Estos resultados sugieren que, si bien en los mensajes en forma de tuits originales de la campaña prevalecen sentimientos positivos, su difusión hacia toda la red por otros participantes de la campaña tiende a ser mayor cuando se trata de mensajes con mayor peso de vocabulario negativo.

CONCLUSIONES

Stop Hate for Profit, es una campaña que alentó a las empresas con fines de lucro a pausar la publicidad en Facebook durante el mes de julio y creció rápidamente a amplia escala internacional a solo días de ser lanzada.

Los principales delineamientos teóricos del dominio de conocimiento del ‘discurso de odio’ lo muestran como un debate científico no próximo a acabar en las disímiles áreas de las ciencias, donde cada vez más se establecen diálogos interdisciplinarios para su abordaje. Es un fenómeno sistemático y nocivo que atenta contra los principios de las sociedades democráticas.

El análisis de dominio como método neurálgico de la investigación contribuyó al entendimiento del dominio #StopHateforProfit a través del activismo. El Análisis de Redes Sociales permitió contribuir al estudio del dominio de la campaña #StopHateforProfit, describiendo su estructura social y las temáticas de la conversación generada por los participantes, así como los mensajes que abrieron mayor impacto entre los participantes.

Se confirmó la estrecha relación existente entre las temáticas del interés público y los tópicos centrales de la cobertura periodística internacional. La campaña no solo despertó el interés de los internautas por el significado de la campaña, sus demandas y el fenómeno del discurso del odio, sino también sobre el significado del fenómeno financiero base de su demanda, dígame las implicaciones del boicot publicitario a Facebook. La minería de datos de las noticias arrojó esta relación causal con las principales preocupaciones del interés público sobre las compañías relacionadas con la campaña como Starbuck y Unilever.

La campaña a solo un mes de lanzada en Twitter generó un gran éxito en la pausa publicitaria en el tiempo demandado, pues más de 1,000 empresas pausaron las inversiones publicitarias en la plataforma. Entre ellos se incluyeron las grandes empresas Adidas, Best Buy, Starbucks, Unilever, Coca-Cola, Verizon y Clorox, algunas incluso declararon mantener esta decisión por más de un año. La participación de las marcas más emblemáticas del mundo en Stop Hate for Profit presionó a Facebook a hacer una serie de concesiones a las históricas demandas de las organizaciones de defensa de los derechos civiles.

Los hitos de la campaña fueron detectados por medio del ARS, advirtiendo que tuvo el apoyo e influencia de pequeñas empresas, organizaciones sin fines de lucro y figuras públicas que, mediante un vínculo fuerte apelaron a la responsabilidad social y conciencia del liderazgo de Facebook en prevenir la propagación de la desinformación en anuncios políticos, entre otras demandas. Los participantes de la campaña formaron comunidades cohesivas independientemente de su geolocalización, sino a partir de sus intereses de defensa de los derechos civiles y las temáticas que los vinculaban. Si bien los influyentes más activos que dominaron las conversaciones eran desproporcionadamente usuarios comunes de Twitter, la campaña movilizó a influencers, grandes y pequeñas empresas, organizaciones sin fines de lucro, defensoras de los derechos civiles, políticos, y periodistas en general que desempeñaron un papel de "puente" entre la coalición y la comunidad participante.

Los usuarios se comunicaron sobre la campaña más allá de los temas originales sobre el discurso del odio también sobre su relación con la desinformación e influencia en las campañas presidenciales, de las desigualdades sociales y raciales.

Los tuits de los influyentes centrales Sleeping Giants, NAACP, Sasha Baron Cohen, periodistas como Carole y Kyle Griffin sin lugar a dudas marcaron una escalada significativa en la difusión y alcance de la campaña a cientos de empresas anunciantes, figuras públicas y miles de usuarios que se sumaron en apoyo de las demandas de #StopHateforProfit contra el discurso del odio, el extremismo y la desinformación. Las interacciones, ya sea por comentarios, likes y retuits, del público en general participante, a estos mensajes de la coalición, organizaciones sin fines de lucro, así como las figuras públicas, ayudaron a difundir la campaña con inmediatez a miles de usuarios en este sitio de red social.

Aunque se desconocen las cifras del alcance que tuvo en general la campaña, las métricas brindadas por el ARS develan la relación de las características del conocimiento sobre el dominio #StopHateforProfit, ejercido sobre Facebook que conllevaron a ganar el apoyo de más de mil compañías anunciantes y formar un movimiento activista de escala internacional.

Los resultados ilustran cómo la narrativa general en Twitter en torno a la campaña no solo promulga mensajes positivos para generar conciencia a las principales compañías que

publicitan en Facebook sobre el discurso del odio y la desinformación como fenómenos dañinos a la democracia, también articula demandas de cambio en Facebook contra el discurso del odio y la desinformación desde una narrativa positiva, sino también la opinión pública hizo uso de vocabulario cargados de sentimientos negativos para con prevalencia de emociones de miedo para alertar sobre las consecuencias que este fenómeno causa a la sociedad.

El fenómeno social e informacional generado Stop Hate for Profit constituye en sí un ejemplo de movimiento activista contra el discurso del odio del cual los estudiosos de la información a partir de su estudio como dominio empleando una perspectiva de ARS pueden ayudar no solo a entender el perfil de los participantes, y su papel jugado en la estructura de red social construida a partir de los vínculos establecidos sino debe apoyarse también en la clasificación de la narrativa según la presencia de sentimientos positivos y negativos, para entender a cabalidad el comportamiento de los flujos informacionales que se dio entre las estructuras sociales de la comunidad creada. Todo ello facilitará propuestas de herramientas basadas en los principios de la Organización de la Información y del Conocimiento que tributen al diseño de estrategias de campañas de activismo social contra el fenómeno del discurso del odio y la desinformación.

REFERENCIAS BIBLIOGRÁFICAS

- ANDEREGG, W. R. L.; GOLDSMITH, G. R. Public interest in climate change over the past decade and the effects of the ‘climategate’ media event. **Environmental Research Letters**, v. 9, n. 5, p. 054005, 2014. Disponible en: <https://doi.org/10.1088/1748-9326/9/5/054005>
- ANDERSEN, N. Å. **Discursive analytical strategies**: Understanding Foucault, Koselleck, Laclau, Luhmann. Policy Press, 2003.
- ANTI-DEFAMATION LEAGUE. **Demand Facebook Stop Hate for Profit**. 2020. Disponible en: <https://www.stophateforprofit.org/demand-change>
- ARNOLD, M. B.; PLUTCHIK, R. The Emotions: Facts, Theories and a New Model. **The American Journal of Psychology**, v. 77, n. 3, p. 518, 1964.
- BENNETT, W. L.; SEGERBERG, A. The Logic of connective action: Digital media and the personalization of contentious politics. **Information, Communication & Society**, v. 15, n. 5, p. 739–768, Jun. 2012. Disponible en: <https://doi.org/10.1080/1369118X.2012.670661>
- BLEI, D. M.; LAFFERTY, J. D. Dynamic topic models. En: **Proceedings of the 23rd international conference on Machine learning**, 113-120, 2006.
- BLEI, D. M.; NG, ANDREW, Y.; JORDAN, M.I. Latent dirichlet allocation. **Journal of machine Learning research**, v. 3, n. Jan, p. 993-1022, 2003.
- BLEICH, E. Freedom of Expression versus Racist Hate Speech: Explaining Differences Between High Court Regulations in the USA and Europe. **Journal of Ethnic and Migration Studies**, v.40, n. 2, p. 283-300, 2014. Disponible en: <https://doi.org/10.1080/1369183x.2013.851476>
- BLONDEL, V. D. et al. Fast unfolding of communities in large networks. **Journal of Statistical Mechanics: Theory and Experiment**, v. 2008, n. 10, p. P10008, 9 out. 2008.

- BOYD, D. M.; ELLISON, N. B. Social network sites: Definition, history, and scholarship. **Journal of computer-mediated Communication**, v.13, n.1, p.210-230, 2007. Disponible en: <https://doi.org/10.1111/j.1083-6101.2007.00393.x>
- BRANDES, U. A faster algorithm for betweenness centrality. **The Journal of Mathematical Sociology**, v. 25, n. 2, p. 163–177. 2001.
- BROWN, A. **Hate Speech Law: A Philosophical Examination**. Routledge: (2015)
- BROWN, A. The “Who?” Question in the Hate Speech Debate: Part 1: Consistency, Practical, and Formal Approaches. **Canadian Journal of Law & Jurisprudence**, v. 29, n. 2, p. 275–320, ago. 2016.
- BROWN, A. What is hate speech? Part 1: The myth of hate. **Law and Philosophy**, v. 36, n. 4, p.419-468. 2017. Disponible en: <https://doi.org/10.1007/s10982-017-9297-1>
- BRUGGER, W. Proibição ou proteção do discurso do ódio? Algumas observações sobre o direito alemão e o americano. **Direito Público**, v.4, n.15, p.117-136. 2007.
- BURNAP, P.; WILLIAMS, M. L. Cyber Hate Speech on Twitter: An Application of Machine Classification and Statistical Modeling for Policy and Decision Making. **Policy and Internet**, v.7, n.2, p.223-24, 2015. Disponible en: <https://doi.org/10.1002/poi3.85>
- BUYSE, A. Dangerous expressions: the echo, violence and free speech. **International & Comparative Law Quarterly**, v.63, n.2, p.491-503, 2014. Disponible en: <https://doi.org/10.1017/S0020589314000104>
- CABO ISASI, A.; GARCÍA JUANATEY, A. El discurso del odio en las redes sociales: un estado de la cuestión. P.38 2016. Disponible en: https://ajuntament.barcelona.cat/bcnvsodi/wp-content/uploads/2015/03/Informe_discurso-del-odio_ES.pdf
- CAMMAERTS, B. Radical pluralism and free speech in online public spaces: The case of North Belgian extreme right discourses. **International Journal of Cultural Studies**, v.12, n.6, p. 555-575, 2009. Disponible en: [10.1177/1367877909342479](https://doi.org/10.1177/1367877909342479)
- CAPOZZI, A. T. E; PATTI, V.; RUFFO, G.; BOSCO, C. A Data Viz Platform as a Support to Study, Analyze and Understand the Hate Speech Phenomenon. Proceedings of the

- 2nd International Conference on Web Studies. Anales: WS.2 2018. New York, NY, USA: Association for Computing Machinery, 3 out. 2018. Disponível em: <https://doi.org/10.1145/3240431.3240437>
- CAPPI, J. Internet, big data e discurso de ódio: reflexões sobre as dinâmicas de interação no Twitter e os novos ambientes de debate político. 2017. 204 f. Tese (Doutorado em Comunicação e Semiótica) - Programa de Estudos Pós-Graduados em Comunicação e Semiótica, Pontifícia Universidade Católica de São Paulo, São Paulo, 2017. <https://tede2.pucsp.br/bitstream/handle/20644/2/Juliano%20Cappi.pdf>
- CHAPP, C. B. God Talk: **Experimenting with the Religious Causes of Public Opinion**. By Paul A. Djupe and Brian R. Calfano. Philadelphia, PA: Temple University Press, 2013. 276 pp. Politics and Religion, v. 8, n. 1, p. 188–190, 2015. Disponível em: <https://doi.org/10.1017/S1755048314000625>
- CHARAUDEAU, Patrick. Discurso y Contexto. **Discurso & Sociedad**, v. 3, n. 2, p. 253-279 2009.
- CHEN, C.; IBEKWE-SANJUAN, F.; HOU, J. The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis. **Journal of the American Society for Information Science and Technology**, v. 61, n. 7, p. 1386–1409, 2010. Disponível em: <https://doi.org/10.1002/asi.21309>
- CHETTY, N.; SREEJITH, A. Hate speech review in the context of online social networks. **Aggression and violent behavior**, v.40, p.108-118. 2018.
- CITRON, D. K.; NORTON, H. Intermediaries and hate speech: fostering digital citizenship for our information age. **Boston University Law Review**, v.91, n.4, p.1435-1484, 2011. Disponível em: <https://ssrn.com/abstract=1764004>
- COHEN, S. B. These are the companies that spend the most \$ on Facebook ads. Will you join the boycott and tell Facebook to stop spreading racism and hate?!@ProcterGamble @Walmart @Microsoft @Target @amazon @HomeDepot @Diageo_News @nytimes @dominos @BestBuy #StopHateForProfit. 26 jun 2020. Twitter: @SashaBaronCohen. Disponível em:

<https://twitter.com/sachabaroncohen/status/1276612542735974402> Acceso en: 27 jun 2020.

COHEN-ALMAGOR, R. Taking North American White Supremacist Groups Seriously: The Scope and Challenge of Hate Speech on the Internet. **International Journal for Crime Justice and Social Democracy**, v.7, n.2, p.38-57, 2018. Disponible en: <https://doi.org/10.5204/ijcjsd.v7i2.517>

COHEN-ALMAGOR, R. **Confronting the Internet's dark side: moral and social responsibility on the free highway**. New York, NY : Washington, DC: Cambridge University Press ; Woodrow Wilson Center Press, 2015.

COLLINS, P. H. On violence, intersectionality and transversal politics. **Ethnic and Racial Studies**, v.40, n.9, p.1460-1473, 2017. Disponible en: <https://doi.org/10.1080/01419870.2017.1317827>

DAMUS, MA; ACUÑA, GN. Aproximación al Análisis de Dominio (AD) desde la investigación en Bibliotecología y Ciencia de la Información. **E-Ciencias de la Información**, v.9, n.2, p 47-67, 2019. Disponible en: <https://doi.org/10.15517/eci.v9i2.37497>

FRESNO GARCÍA, M.; DALY, A. J.; SEGADO SÁNCHEZ-CABEZUDO, S. Identificando a los nuevos influyentes en tiempos de Internet: medios sociales y análisis de redes sociales / Identifying the new Influencers in the Internet Era: Social Media and Social Network Analysis. **Revista Española de Investigaciones Sociológicas**, 2016.

DEMSAR, J; CURK, T; ERJAVEC, A; GORUP C; HOCEVAR, T; MILUTINOVIC, M; MOZINA, M; POLAJNAR, M; TOPLAK, M; STARIC, A; STAJDOHAR, M; UMEK, L; ZAGAR, L; ZBONTAR, J; ZITNIK, M; ZUPAN, B. Orange: Data Mining Toolbox in Python, **Journal of Machine Learning Research** 14(Aug), p. 2349–2353, 2013.

DEY, A. K. Understanding and using context. **Personal and ubiquitous computing**, v. 5, n. 1, p. 4-7, 2001. Disponible en: <https://doi.org/10.1007/s007790170019>

DI FELICE, M. Redes sociais digitais, epistemologias reticulares e a crise do antropomorfismo social. *Revista USP*, no 92, p. 6-19, 2012.

- DJUPE, P.; CALFANO, B. **God talk**: Experimenting with the religious causes of public opinion. Philadelphia, PA: Temple University Press, 2013.
- DO, H.H.; PRASAD, P.W.C.; MAAG, A; ALSADOON, A. Deep learning for aspect-based sentiment analysis: a comparative review. **Expert Systems with Applications**, n. 118, p.272-299, 2019.
- DOBROVOLSKYI, H.; KEBERLE, N. **Collecting the Seminal Scientific Abstracts with Topic Modelling, Snowball Sampling and Citation Analysis**. En: Ermolayev, V., Suárez-Figueroa, M. C., Ławrynowicz, A., Palma, R., Yakovyna, V., Mayr, H. C., Nikitchenko, M.; Spivakovsky, A. (Eds.): ICT in Education, Research, and Industrial Applications. Proc. 14th Int. Conf. ICTERI 2018. Volume I: Main Conference. Kyiv, Ukraine, May 14-17, 2018, CEUR-WS.org, vol. 2105, p. 179-192, 2018. Disponible en: <http://ceur-ws.org/Vol-2105>
- EKINCI, E.; OMURCA, I.S. Concept-LDA: Incorporating BabelFy into LDA for aspect extraction. **Journal of Information Science**, v. 46, n. 3, p. 406–418, jun. 2020.
- ERIKSON, R. S.; TEDIN, K. L. **American public opinion**: Its origins, content and impact. Routledge, 2015.
- ESQUIVEL ALONSO, Y. El discurso del odio en la jurisprudencia del Tribunal Europeo de Derechos Humanos. **Cuestiones Constitucionales Revista Mexicana de Derecho Constitucional**, june, 2016. ISSN 2448-4881. Disponible en: <http://dx.doi.org/10.22201/ijj.24484881e.2016.35.10491>. Acceso: 31 oct. 2019
- EVERETT, M.; BORGATTI, S. P. Ego network betweenness. **Social networks**, v.27, n.1, p.31-38, 2005. Disponible en: <https://doi.org/10.1016/j.socnet.2004.11.007>
- FISCHER, A.; HALPERIN, E.; CANETTI, D.; JASINI, A. Why we hate. **Emotion Review**, v. 10, n.4, p.309-320, 2018.
- FORTUNA, P.; NUNES, S. A survey on automatic detection of hate speech in text. **ACM Computing Surveys (CSUR)**, v.51, n.4, p.1-30. 2018

- FRIED, A. The Forgotten Lindsay Rogers and the Development of American Political Science. **American Political Science Review**, v. 100, n. 4, p. 555, 2006. Disponível en: <https://doi.org/10.1017/S0003055406062423>
- GLYNN, C. J., HERBST, S., LINDEMAN, M., O'KEEFE, G. J., SHAPIRO, R. Y. **Public Opinion**. New York: Routledge, p. 419. 2018.
- GOUVEIA, F.C. Introdução ao Gephi. En: Fialho, J. **Redes sociais: ¿Cómo comprendê-las?** Uma introdução à análise de redes sociais. 1ed. Lisboa: Edições Sílabo.
- GREENE, A. R.; SIMPSON, R. M. Tolerating hate in the name of democracy. **The Modern Law Review**, v.80, n.4, p746–765. 2017. Disponível en: <https://doi.org/10.1111/1468-2230.12283>
- GU, L.; KROPOTOV, V.; YAROCHKIN, F. The fake news machine: How propagandists abuse the internet and manipulate the public. **Trendslab**. 2017. Disponível en: https://documents.trendmicro.com/assets/white_papers/wp-fake-news-machine-how-propagandists-abuse-the-internet.pdf
- HICKS, D.; WANG, J. The New York Times as a resource for mode 2. **Science, technology, & human values**, v. 38, n. 6, p. 851-877, 2013. Disponível en: <https://doi.org/10.1177/0162243913497806>
- HJØRLAND, B. Domain analysis in information science: eleven approaches-traditional as well as innovative. **Journal of Documentation**, v. 58, n. 4, p. 422-462, 2002. Disponível en: <https://doi.org/10.1108/00220410210431136>
- HJØRLAND, B. Domain analysis. **Knowledge Organization**, v. 44, n.6, p. 436-464, 2017. Disponível en: <https://www.nomos-elibrary.de/10.5771/0943-7444-2017-6-436/domain-analysis-jahrgang-44-2017-heft-6>
- HJØRLAND, B.; ALBRECHTSEN, H An analysis of some trends in classification research. **KO Knowledge Organization**, v.26, n.3, p. 131-139, 1999.
- HJØRLAND, B.; ALBRECHTSEN, H. Toward a new horizon in Information Science: domain-analysis. **Journal of the American Society for Information Science**, v. 6, n.

- 6, p. 400-425, 1995. Disponible en: [https://doi.org/10.1002/\(SICI\)1097-4571\(199507\)46:6<400::AID-ASI2>3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1097-4571(199507)46:6<400::AID-ASI2>3.0.CO;2-Y)
- HU, M.; LIU, B. Mining opinion features in customer reviews. **AAAI**, v. 4, n. 4, p. 755-760, 2004. Disponible en: <https://www.aaai.org/Papers/AAAI/2004/AAAI04-119.pdf>
- JIMÉNEZ-ZAFRA, S. M.; MARTÍN-VALDIVIA, M. T.; MARTÍNEZ-CÁMARA, E.; UREÑA-LÓPEZ, L. A. Combining resources to improve unsupervised sentiment analysis at aspect-level. **Journal of Information Science**, v.42, n.2, p. 213-229, 2016.
- JINDAL, N.; LIU, B., July. Mining comparative sentences and relations. En: **AAAI**, v. 22, n. 13311336, p. 9, 2006.
- JUN SP; YOO HS; CHOI S. Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. **Technological forecasting and social change**, v. 130, p. 69-87, 2018. Disponible en: <https://doi.org/10.1016/j.techfore.2017.11.009>
- KAPRALSKI, S. The Evolution of Anti-Gypsyism in Poland: From Ritual Scapegoat to Surrogate Victims to Racial Hate Speech? **Polish Sociological Review**, n.193, p. 101-117, 2016. Disponible en: <https://www.jstor.org/stable/44113928>
- KURSUNCU, U.; GAUR, M.; LOKALA, U.; THIRUNARAYAN, K.; SHETH, A.; ARPINAR, I. B. Predictive analysis on Twitter: Techniques and applications. In **Emerging research challenges and opportunities in computational social network analysis and mining**, Springer, Cham, p. 67-104, 2019.
- LIPPMANN, W. **Public opinion**. New York: The Free Press Publishers, p. 271, 1966.
- LIU, B. **Sentiment analysis: Mining opinions, sentiments, and emotions**. Cambridge university press.2015
- MADELIN, R. The evolving social responsibilities of internet corporate actors: Pointers past and present. **Philosophy & Technology**, v.24, n.4, p.455–461. 2011.
- MAITRA, I. Silencing speech. **Canadian Journal of Philosophy**, v.39, n. 2, p.309-338. 2009
- MARQUES, T. The Expression of Hate in Hate Speech. **6º Congresso da Sociedade Brasileira de Filosofia Analítica**, (5 de setembro de 2020). 2020.

- MARTÍNEZ-ÁVILA, D. Problems and Characteristics of Foucauldian Discourse Analysis as a Research Method. En: **Cultural Frames of Knowledge**, eds., Richard P. Smiraglia; Hur-Li Lee. Würzburg, Germany: Ergon. p. 99-110, 2012.
- McCAIN, R. Reusable software component construction-A product-oriented paradigm. En: **5th Computers in Aerospace Conference**, p. 5068, 1985. Disponible en: <https://doi.org/10.2514/6.1985-5068>
- MCCALLUM, M. L.; BURY, G. W. Google search patterns suggest declining interest in the environment. **Biodiversity and Conservation**, v. 22, n. 6–7, p. 1355–1367, Jun. 2013. Disponible en: <https://doi.org/10.1007/s10531-013-0476-6>
- MCLAREN, N.; SHANBHOGUE, R. Using Internet Search Data as Economic Indicators. SSRN **Electronic Journal**, 2011. Disponible en: <https://dx.doi.org/10.2139/ssrn.1865276>
- MCQUAIL, D. **Media performance**: Mass communication and the public interest. London. 1992.
- MEDHAT, W., HASSAN, A. AND KORASHY, H. Sentiment analysis algorithms and applications: A survey. **Ain Shams engineering journal**, v.5, n.4, p.1093-1113, 2014. Disponible en: <http://dx.doi.org/10.1016/j.asej.2014.04.011>
- MEYER-PFLUG, S.R. Liberdade de expressão e discurso de ódio. **Revista dos Tribunais**, 2009.
- MIMNO, D.; WALLACH, H.; TALLEY, E.; LEENDERS, M.; MCCALLUM, A. Optimizing semantic coherence in topic models. En: **Proceedings of the 2011 conference on empirical methods in natural language processing**, p. 262-272, 2011. Disponible en: <https://aclanthology.org/D11-1024.pdf>
- MINKOV, D., 2002. Trading System Applied to Large Mutual Fund Company. En: **Empirical Science of Financial Fluctuations**, p. 275-286. Springer, Tokyo.
- MOHAMMAD, S. M., TURNEY, P. D. Crowdsourcing a word–emotion association lexicon. **Computational Intelligence**, v. 29, n. 3, p. 436-465, 2013. Disponible en: <http://www.saifmohammad.com/WebDocs/NRCemotionlexicon.pdf>

- MOURA, M.A. Informação, ferramentas ontológicas e redes sociais ad hoc: a interoperabilidade na construção de tesouros e ontologias. **Informação & Sociedade: Estudos**, v. 19, n. 1. 2009.
- NEIGHBORS, J.M The Draco approach to constructing software from reusable components. **IEEE Transactions on Software Engineering**, n.5, p.564-574, 1984. <https://doi.org/10.1109/TSE.1984.5010280>
- NEIGHBORS, J.M. **Software construction using components**. University of California, Irvine; 1980.
- Newbury Park: Sage Publications. MEDDEB, I.; LAVANDIER, C.; KOTZINOS. Using Twitter Streams for Opinion Mining: A Case Study on Airport Noise. En: **International Workshop on Information Search, Integration, and Personalization**, p 145-160, 2019 Springer, Cham. Disponible en: https://doi.org/10.1007/978-3-030-44900-1_10
- NOELLE-NEUMANN, E. The Theory of Public Opinion: The Concept of the Spiral of Silence. **Annals of the International Communication Association**, v. 14, n. 1, p. 256–287, Jan. 1991. Disponible en: <https://doi.org/10.1080/23808985.1991.11678790>
- NYTimes. Facebook Fails to Appease Organizers of Ad Boycott. Julio 7, 2020. <https://www.nytimes.com/2020/07/07/technology/facebook-ad-boycott-civil-rights.html>
- ORGANIZACIÓN DE NACIONES UNIDAS. La estrategia y plan de acción de las naciones unidas para la lucha contra el discurso de odio. 2019. Disponible en: [https://www.un.org/en/genocideprevention/documents/advising-and-mobilizing/Action_plan_on_hate_speech ES.pdf](https://www.un.org/en/genocideprevention/documents/advising-and-mobilizing/Action_plan_on_hate_speech_ES.pdf)
- OVALLE-PERANDONES, M. A., OLMEDA-GÓMEZ, C., PERIANES-RODRÍGUEZ, A. ARS y redes de colaboración científica. En: M. Fresno García, M. P. Marqués-Sánchez & D. Sánchez-Paunero (Eds.). **Conectados por redes sociales**: introducción al análisis de redes sociales y casos prácticos. Barcelona: Editorial UOC, p. 219-245, 2014.

- OVALLE-PERANDONES, M.A; MARQUES-SANCHEZ, P.; BENITEZ-ANDRADE, J.A; FIALHO, J. Fundamentos teóricos das redes de Modo 2 e casos aplicados de redes bimodales na saúde. p.135-161. En: Fialho, J. **Redes sociais ¿Cómo comprendê-las?** Uma introdução à análise de redes sociais. 1ed. Lisboa: Edições Sílabo, 2020.
- PALLIS, G.; ZEINALIPOUR-YAZTI, D.; DIKAIKOS, M. D. Online social networks: status and trends. New directions in web data management 1. En: Vakali, A., & Jain, L. C. (Eds.). *New Directions in Web Data Management 1* (Vol. 331). Springer, p. 213-234, 2011.
- PANG, B.; LEE L.; VAITHYANATHAN, S. Thumbs Up? Sentiment Classification Using Machine Learning Techniques. En: **Proceedings of Conference on Empirical Methods in Natural Language Processing, (EMNLP-2002)**, 2002.
- PAWAR, K.K.; SHRISHRIMAL, P.P.; DESHMUKH, R.R. Twitter sentiment analysis: A review. **International Journal of Scientific & Engineering Research**, v.6, n.4, p.957-964, 2015. Disponível en: https://www.researchgate.net/profile/Kishori-Pawar/publication/277554643_Twitter_Sentiment_Analysis_A_Review/links/556c64a008aec22683049811/Twitter-Sentiment-Analysis-A-Review.pdf
- PERSSON, O.; DANELL, R.; SCHNEIDER, J. W. How to use Bibexcel for various types of bibliometric analysis. Celebrating scholarly communication studies: A Festschrift for Olle Persson at his 60th Birthday, v. 5, p. 9-24, 2009. Disponível en: <https://portal.research.lu.se/ws/files/5902071/1458992.pdf>
- PLUTCHIK, R. Emotion: A psychoevolutionary synthesis. New York: Harper & Row, 1980.
- PLUTCHIK, R. Emotions: A general psychoevolutionary theory. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 197-219). Hillsdale, NJ: Erlbaum. (1984).
- PRABOWO, R.; THELWALL, M. Sentiment analysis: A combined approach. **Journal of Informetrics**, v. 3, n. 2, p. 143-157, 2009.
- PRICE, D. D. S. A general theory of bibliometric and other cumulative advantage processes. **Journal of the American Society for Information Science**, v. 27, n. 5, p. 292-306, 1976. Disponível en: <https://doi.org/10.1002/asi.4630270505>

- PRICE, D. J. Citations of literature. *Acta cytologica*, v. 13, n. 10, p. 544-544, 1969.
- PRICE, D. J. S. Little science, big science. Columbia University Press, 1963
- PRIETO-DIAZ, R. Domain analysis: An introduction. **ACM SIGSOFT Software Engineering Notes**. v. 15, n. 2, p.47-54, 1990. Disponible en: <https://doi.org/10.1145/382296.382703>
- PRIETO-DIAZ, R. Reuse as a new paradigm for software development. En: **Systematic Reuse: Issues in Initiating and Improving a Reuse Program**, p. 1-13, 1996. Springer, London. Disponible en: https://doi.org/10.1007/978-1-4471-1484-0_1
- PUERTA-DIAZ, M.; OVALLE-PERANDONES, M. A.; MARTÍNEZ-ÁVILA, D. Patrones y tendencias emergentes de la estructura científica internacional en el dominio “discurso del odio”. **Revista Ibero-Americana de Ciência da Informação**, v. 13, n. 3, 2020, p. 978-963. Disponible en: [10.26512/rici.v13.n3.2020.33017](https://doi.org/10.26512/rici.v13.n3.2020.33017)
- RAVI, K.; RAVI, V. A survey on opinion mining and sentiment analysis: tasks, approaches and applications. **Knowledge-Based Systems**, n. 89, p. 14-46, 2015. DOI: 10.1016/j.knosys.2015.06.015
- REVENGA SÁNCHEZ, M. Libertad de expresión y discursos del odio. Alcalá de Henares, Spain: Editorial Universidad de Alcalá, p. 190, 2015. Disponible en: <https://elibro.net/es/ereader/universidadcomlutense/44373?page=13> .
- RODRÍGUEZ-IZQUIERDO SERRANO, M. El Discurso Del Odio A Través De Internet. En: Revenga Sánchez, M., Coord. **Libertad De Expresión Y Discursos Del Odio**. Madrid: Universidad De Alcalá De Henares, 2015.
- ROMERO QUESADA, MA. El análisis de dominio y sus presupuestos teóricos metodológicos. *Bibliotecas. Anales de investigación*, v. 8, n. 8-9, p. 228-38, 2013. Disponible en: <http://eprints.rclis.org/25302/1/19-Reflexi%C3%B3n%2006.pdf>
- ROYZMAN, E. B.; MCCAULEY, C.; ROZIN, P. From Plato to Putnam: Four Ways to Think About Hate. En: R. J. Sternberg (Ed.). *The psychology of hate*, p 3-35. Washington, D.C.: APA Books, 2005.

- SCHÄFER, G.; LEIVAS, P. G. C.; SANTOS, R. H. Discurso de ódio: da abordagem conceitual ao discurso parlamentar. **Revista de Informação Legislativa**, v. 52, n.207, p. 143-158, jul./set. 2015. Disponível em: https://www12.senado.leg.br/ril/edicoes/52/207/ril_v52_n207_p143
- SCHEUFLE, D. A.; MOY, P. Twenty-five years of the spiral of silence: A conceptual review and empirical outlook. **International journal of public opinion research**, v. 12, n. 1, p. 3-28, 2000. Disponível em: <https://doi.org/10.1093/ijpor/12.1.3>
- SEGLOW, J. Hate speech, dignity and self-respect. **Ethical Theory and Moral Practice**, v. 19, n. 5, p.1103-1116, 2016.
- SELLARS, A. **Defining Hate Speech**. Berkman Klein Center Research Publication No. 2016-20, Boston Univ. School of Law, Public Law Research Paper No. 16-48, 2016. Disponível em: <http://dx.doi.org/10.2139/ssrn.2882244>
- SHANG, S.; SHI, M.; SHANG, W.; HONG, Z. Research on public opinion based on Big Data. In: **IEEE/ACIS 14TH International Conference on Computer and Information Science (ICIS)**, Jun. 2015. 2015 IEEE/ACIS 14th International Conference on Computer and Information Science (ICIS). Las Vegas, NV, USA: IEEE, Jun. 2015. p. 559-562. Disponível em: <https://doi.org/10.1109/ICIS.2015.7166655>
- SILVA, A E F A; PRADO, J W; ALCANTARA, V C; TONELLI, D F; PEREIRA, J R Public opinion: bibliometric analysis for the systematization of trends. **HOLOS**, v. 34, n 4, p. 2-30, 2018. Disponível em: <http://dx.doi.org/10.15628/holos.2018.6905>
- SILVA, A. K. A. da. A dinâmica das redes sociais e as redes de coautoria. **Perspectivas em Gestão & Conhecimento**, n. 4, p.27-47, 2014. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/pgc/article/view/21275>
- SLEEPING GIANTS. Hearing that Facebook execs have been working the phones to stop companies from pausing their advertising instead of spending time thinking about changing their hate & disinfo policies to bring them back. This should tell you everything about their priorities. #StopHateForProfit. 27 jun 2020.Twitter:

@SashaBaronCohen.

Disponível

en:

https://twitter.com/slpng_giants/status/1277016099599470592

SMIRAGLIA, R. P. **Domain analysis for knowledge organization: tools for ontology extraction.** Chandos Publishing, 2015.

SMIRAGLIA, R. P. **The elements of knowledge organization.** Springer, 2014.

SONNENWALD, D. H. Scientific Collaboration. **Annual Review of Information Science and Technology**, v. 42, n. 1, p.643 681, 2008.

SPINAK. E. Indicadores Cienciométricos. **Ciência da Informação**, Brasília, v. 27, n. 2, p. 141-148, 1998.

STURGES, P. Limits to freedom of expression? The problem of blasphemy. **IFLA Journal**, v. 41, n. 2, p. 112-119, 2015. <https://doi.org/10.1177%2F0340035215584778>

TABOADA, M.; BROOKE, J.; TOFILOSKI, M.; VOLL, K.; STEDE, M., 2011. Lexicon-based methods for sentiment analysis. *Computational linguistics*, v. 37, n. 2, p. 267-307. DOI: 10.1162/COLI_a_00049

TADDEO, M.; FLORIDI, L. **The moral responsibilities of online service providers.** En: The responsibilities of online service providers p. 13-42. Springer, Cham. 2017.

TAN, C., LEE, L., TANG, J., JIANG, L., ZHOU, M. AND LI, P. User-level sentiment analysis incorporating social networks. En: **Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining**, p. 1397-1405. August 2011.

TARJAN, R. Depth-First Search and Linear Graph Algorithms. **SIAM Journal on Computing**, v. 1, n. 2, p. 146–160, 1 jun. 1972.

THE CANADIAN PRESS. Canada's 5 big banks join anti-hate advertising boycott of Facebook. 2020. Disponível em: <https://www.cbc.ca/news/business/facebook-canada-big-banks-advertising-boycott-1.5634395>

TONTODIMAMMA, A.; NISSI, E.; SARRA, A.; FONTANELLA, L. Thirty years of research into hate speech: topics of interest and their evolution. **Scientometrics**, v.

- 126, n. 1, p. 57-179. 2021. Disponible en:
<https://link.springer.com/article/10.1007/s11192-020-03737-6>
- VAN DIJK, T. A. El discurso como interacción social. V. 302483. Editorial Gedisa, 2019.
- VAN DIJK, T.A. Discurso y racismo. Persona y sociedad, p. 191-205, 2001.
- VELDHUIS, C. B.; DRABBLE, L.; RIGGLE, E. D. B.; WOOTTON, A. R.; HUGHES, T. L.
 I Fear for My Safety, but Want to Show Bravery for Others: Violence and
 Discrimination Concerns Among Transgender and Gender-Nonconforming
 Individuals After the 2016 Presidential Election. **Violence and Gender**, v. 5, n.1,
 p.26-36, 2018. Disponible en: <https://doi.org/10.1089/vio.2017.0032>
- VOSEN, S; SCHMIDT T. Forecasting private consumption: survey-based indicators vs.
 Google Trends. **J Forecast**, n. 30, 2011. Disponible en:
<https://doi.org/10.1002/for.1213>
- WALDRON, J. **The harm in hate speech**. Harvard University Press, 2012.
- WASSERMAN, S.; FAUST, K. **Social network analysis: Methods and
 applications**. Cambridge University Press, 1994. Disponible en:
<https://doi.org/10.1017/CBO9780511815478>
- WHITE, H. D. Authors as Citers over Time. **Journal of the American Society for
 Information Science and Technology**, v. 52, n. 2, p .87–108, 2001.
- WICKELMAIER, F. **An Introduction to MDS**. Sound Quality Research Unit, Aalborg
 University. 2003.