

Patrones y tendencias emergentes de la estructura científica internacional en el dominio “discurso del odio”

Mirelys Puerta-Díaz

Universidade Estadual Paulista Júlio de Mesquita Filho, Faculdade de Filosofia e Ciências, Marília/SP, Brasil
<https://orcid.org/0000-0002-2312-2540>
mirelys.puerta@unesp.br

María-Antonia Ovalle-Perandones

Universidad Complutense de Madrid, Facultad de Ciencias de la Documentación, Madrid, España
<https://orcid.org/0000-0002-6149-4724>
maovalle@ucm.es

Daniel Martínez-Ávila

Universidad Carlos III de Madrid, Departamento de Biblioteconomía y Documentación, Madrid, España
<https://orcid.org/0000-0003-2236-553X>
daniel.martinez@uc3m.es

DOI: <https://doi.org/10.26512/rici.v13.n3.33017>

Recibido/Recibido/Received: 2020-07-02

Aceitado/Aceptado/Accepted: 2020-08-05

ARTIGOS

Resumen: El objetivo de esta investigación es identificar los patrones y tendencias en la estructura científica internacional sobre discursos de odio. Revela la red de colaboración científica, la estructura de cocitación, las áreas de conocimiento con las que está vinculado el tema y las materias que marcan las tendencias en este dominio. Se recuperaron los 411 artículos científicos que contienen la expresión “hate speech” en los campos título, resumen y palabras clave de la Web of Science Core Collection, periodo 2009-2018. Los programas Bibexcel versión 2017 y Pajek para el análisis y visualización de las redes de cocitación y colaboración; y Latent Dirichlet Allocation como técnica de procesamiento de lenguaje natural en los campos títulos, resúmenes y palabras claves. Los resultados muestran un significativo aumento de publicaciones a partir de 2013 y un pico en 2018. Se revelaron los autores, países y fuentes más productivas, así como las redes de coautoría y cocitación de esta producción. El Análisis de Redes Sociales mostró que a pesar de que la presencia de discursos del odio en aún incipiente en la literatura científica existe un fuerte núcleo teórico sobre la temática que está siendo frecuentemente citado por la comunidad académica internacional.

Palabras clave: Discurso del odio. Estudios métricos. Estructura científica.

Padrões emergentes e tendências da estrutura científica internacional no domínio "discurso do ódio"

Resumo: O objetivo desta pesquisa é identificar padrões e tendências da estrutura científica internacional sobre discurso de ódio. Revela a rede de colaboração científica, a estrutura de cocitação, as áreas de conhecimento com as quais o tópico está vinculado e os assuntos que definem as tendências nesse domínio. Foram recuperados 441 artigos contendo a expressão "hate speech" nos campos de título, resumo e palavras-chave da Coleção Principal da Web of Science (WoS), na janela temporária 2009. As redes de colaboração e cocitação foram modeladas usando Bibexcel versão 2017, assim como Pajek na análise e visualização. O *Latent Dirichlet Allocation* foi empregado como uma técnica de processamento de linguagem natural aos textos dos campos de título, resumo e palavras-chave. Os resultados revelam um aumento significativo nas publicações a partir de 2013 e um pico no ano 2018. A Análise das Redes Sociais (ARS) mostrou que, embora a presença do discurso de ódio ainda seja incipiente na literatura científica, há um forte núcleo teórico sobre o assunto que é frequentemente citado pela comunidade acadêmica internacional. Esse núcleo de obras é

reconhecido como altamente relevante nas abordagens teóricas, conceituais e metodológicas do discurso de ódio.

Palavras-chave: Discurso do ódio. Estudos métricos. Estrutura científica. Processamento de linguagem natural.

Emerging patterns and trends in the international scientific structure of the “hate speech” domain

Abstract: The aim of this study is to identify the patterns and trends in the international scientific structure on hate speech. It revealed the network of scientific collaboration, the structure of co-citations, the areas of knowledge linked to the topic, and the subjects that define the trends in the domain. We retrieved 441 articles that include the expression "hate speech" in the title, abstract or keywords field from the Web of Science Core Collection for the period 2009-2018. The collaboration and co-citation networks were modeled using Bibexcel version 2017 and Pajek for the analysis and visualization. Latent Dirichlet Allocation was used as a natural language processing technique for the title, abstract and keywords fields. The results show a significant increase in publications since 2013 and a peak in 2018. The Social Network Analysis showed that despite the incipient presence of hate speech in the scientific literature, there is a prominent theoretical core of publications on the topic which is frequently cited by the international scientific community. There is also a core of works that is recognized as highly relevant in the theoretical, conceptual and methodological approaches to the study of hate speech.

Keywords: Hate speech. Metric studies. Scientific structure.

1 Introducción

Los avances en Internet y las tecnologías de redes sociales han propiciado el desarrollo de un nuevo ambiente de actuación de la sociedad; su potencial democratizador constituye un catalizador en la dinámica comunicativa e informacional de los actores sociales. Dadas sus características, estas tecnologías resultan herramientas eficaces de expresión y socialización de opiniones y son potenciadas a partir del anonimato que ofrecen sus plataformas, como es el caso de las redes sociales, las cuales tienen una implicación directa en el ejercicio de la libertad de expresión. Sin embargo, en estos medios de expresión también existe 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, pues a menudo estas plataformas son usadas para ejercer la libertad de manifestar odio hacia el otro.

Bertoni (2007) señala que los discursos de odio pueden definirse tanto por su intención como por sus objetivos, los cuales, históricamente, no han tenido límites temporales, espaciales o sociales ya que han sido empleados por una amplia gama de actores.

Según Revenga Sánchez (2015), “discursos del odio” es una expresión en español firmemente asentada en la literatura académica que es una traducción literal de “*hate speech*”. 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.

El término se emplea con el claro propósito de alejarse de cuestiones penales y constitucionales, motivado por cuestiones étnicas, raciales, religiosas, violencia, hostilidad, entre otras, según la clasificación propuesta por Esquivel Alonso (2016). Según esta autora, el mayor dilema en los debates contemporáneos sobre discurso del odio se encuentra en limitar o no la libertad de expresión de quienes utilizan las expresiones de odio.

Explicitar discursos de odio se convierte entonces en una práctica social amplificadora, una tendencia alarmante que está siendo utilizada como una vía eficiente para alcanzar popularidad instantánea en las redes sociales sin mucho esfuerzo (CHETTY; ALATHUR, 2018). Dado que este tipo de comunicación prejuiciosa puede ser extremadamente perjudicial para la sociedad, por causas diversas como el inminente peligro de generar diversos actos de violencia o la vulneración de derechos humanos, la complejidad de la identificación y demostración de su intencionalidad, así como la reciente y creciente necesidad de su detección automatizada, los aspectos relacionados con los discursos de odio en la web están recibiendo cada vez más la atención de la comunidad científica internacional (lo que ha llevado a un rápido crecimiento de la producción científica sobre el tema en el último año). Considerando lo expuesto, el objetivo de la presente investigación es identificar los patrones y tendencias de la estructura científica internacional sobre discursos de odio, revelando la red de colaboración científica, la estructura de cocitación, las áreas de conocimiento con las que está vinculado y las materias que marcan las tendencias en este dominio.

2 Discurso del odio: una revisión de la literatura

El discurso de odio ha sido teorizado desde varias áreas disciplinares como la jurisprudencia, la filosofía, la comunicación, la lingüística, las ciencias políticas y, más recientemente, la ciencia de la computación y la ciencia de la información.

En particular, la regulación legislativa sobre el discurso del odio ha enfrentado diversos problemas debido a 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” (el cual es el principal término asociado a los debates teóricos sobre el tema). Aunque la libertad de expresión propicia la creación y pluralidad de la opinión pública y el debate que en ella se gesta, su abuso genera serios conflictos con otros derechos sociales (ESQUIVEL ALONSO, 2016). Por ello, en la literatura científica aparecen estrechamente relacionados ambos términos y se encuentran asociados además a expresiones como “libertad del discurso”, “ética del discurso”, “crímenes de odio” y otros.

En la literatura se puede encontrar dos vertientes relacionadas con 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). Brow (2015) afirma que el estudio de discursos de odio debe centrarse en la trilogía forma-contenido-contexto en que se manifiesta. Es en este sentido en el que adoptar la perspectiva inferencial de las categorías contextuales en que ocurren dichas manifestaciones cobra especial importancia. Este abordaje permite determinar el grado de daño que determinadas expresiones, símbolos y manifestaciones pueden llegar a ocasionar, las cuales, según Cabo Isasi y García Juanatey (2016), pueden oscilar desde el daño emocional hasta la provocación de escaladas de violencia masiva.

Según Esquivel Alonso (2016), pueden encontrarse expresiones de odio por motivos étnicos y raciales (en relación al color de piel, la pertenencia a un grupo étnico e incluso la nacionalidad), religiosos (relacionados con los históricos choques entre los diferentes credos como por ejemplo entre cristianos y musulmanes fanáticos, así como con ateos), haciendo apología del delito, violencia y hostilidad (en la emisión de mensajes ofensivos, ultrajantes o que denoten desprecio por una etnia, grupo o sector poblacional determinado), o en el discurso negacionista del Holocausto (que cuestiona o niega la realidad del genocidio cometido por los nazis durante la II Guerra Mundial). Estas expresiones de odio comprenden toda expresión, símbolos y manifestaciones considerados ofensivos por cierta comunidad o persona, incluyendo también la destrucción pública de imágenes religiosas o símbolos patrios, así como las imágenes que promueven la sumisión y subordinación de la mujer de manera vejatoria, comúnmente presentes en los discursos machistas, sexistas o en la pornografía.

Históricamente se reconoce que las peores expresiones de repudio se han iniciado desde la propagación de un discurso de apatía o indiferencia hacia grupos minoritarios como aquellos señalados por Esquivel Alonso (2016). Con el objetivo de eliminar o minimizar dichas manifestaciones, diversas iniciativas no jurídicas que incluyen medidas legales, constitucionales y sociales han surgido para prevenir los peligros inherentes que este fenómeno ocasiona. Diversos instrumentos (códigos penales, legislaciones administrativas y otros mecanismos punitivos) señalan que las expresiones que promuevan, inciten o justifiquen el odio racial, la xenofobia, el antisemitismo, la violencia de género deben ser sancionadas penalmente. Por otro lado, también se ha promovido la empatía y el apoyo a los grupos que son víctimas del discurso del odio bajo el principio de que dichas manifestaciones de odio son dañinas para la convivencia en sociedad. Puede decirse que en la sociedad se ha intentado hacer prevalecer las políticas públicas que alientan una solución no penal como modelo no sancionatorio en el combate de discursos de odio.

3 Métodos y procedimientos

La presente investigación sigue un doble abordaje metodológico cuanti-cualitativo, el

primero propio de los estudios métricos y el segundo siguiendo el paradigma de análisis de dominio en la organización del conocimiento (HJØRLAND; ALBRECHTSEN, 1995; HJØRLAND, 2002, 2017; SMIRAGLIA, 2015). La obtención de datos se realizó tras la búsqueda y recuperación de los registros bibliográficos con la expresión “hate speech” en los campos título, resumen y palabras clave del recurso *Web of Science* (WoS) Core Collection, para el tipo documental artículo y limitando el periodo temporal a 2009-2018.

Posteriormente se modelaron y visualizaron redes con el propósito de mapear la estructura intelectual del dominio según la información mostrada en los 411 registros en fecha de 8 de noviembre de 2019. Este proceso de extracción de información se realizó utilizando el software Bibexcel versión 2017 (PERSSON, 2009). El *software* Pajek fue empleado para el análisis y visualización de las redes de cocitación y colaboración. Como resultado se espera una base para una definición más precisa del dominio que lo permita ampliar con el método *snowball* (DOBROVOLSKYI; KEBERLE, 2018).

Los indicadores bibliométricos empleados son indicadores de producción (obtenidos a partir del cálculo del número de publicaciones) y de relación (en el caso del análisis de coautoría y cocitación) que son utilizados para elaborar la red de interacción que existe en el campo científico sobre discursos de odio entre agentes científicos implicados (SONNENWALD, 2008).

Se aplicó el método probabilístico de modelaje de tópicos *Latent Dirichlet Allocation* o LDA (BLEIR; NG; JORDAN, 2003), como técnica de procesamiento de lenguaje natural del contenido de los campos títulos, resúmenes y palabras claves para la extracción y cálculo de frecuencia de ocurrencia de las palabras abordadas en las publicaciones. Con el propósito de implementar el LDA se utilizó el lenguaje Python y su biblioteca Gensim. Posteriormente se aplicaron los procedimientos de tratamiento de datos de tokenización (que elimina números, signos de puntuación y letras individuales presentes en el corpus analizado), eliminación de palabras vacías¹ sin aparente significado, y finalmente aplicación del algoritmo de *n_gram*. Es importante destacar que como fuente de datos se consideraron 356 resúmenes de los 411 documentos de la muestra debido a la ausencia de resumen significativo en 55 documentos.

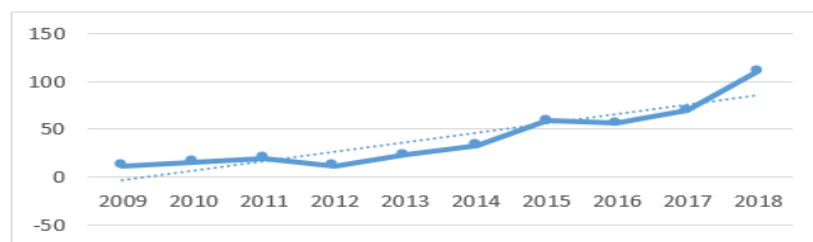
La visualización gráfica del resultado de este proceso requirió de un previo tratamiento de formato de texto para su migración a formato *csv* (*comma-separated values*) para finalmente ser exportado a la herramienta de visualización *online* disponible en el sitio <https://wordart.com/>.

4 Resultados

¹ Término introducido por Hans Peter Luhn, científico informático y experto en información, uno de los pioneros en recuperación de información, se le atribuye la acuñación en 1958 de la locución inglesa *stop words* y el uso del concepto en su diseño.

El acumulativo recuperado de WoS devolvió un conjunto de 411 artículos científicos los cuales no presentan duplicidad de registros, constatado mediante observación directa. La evolución temporal de las publicaciones recuperadas de la base de datos WoS, representada en el gráfico 1, visualiza los totales de publicaciones anuales cada vez mayores.

Gráfico 1. Evolución temporal de publicaciones entre el 2009-2018



Fuente: Elaboración propia con datos colectados desde WoS.

A partir del año 2013 se verifica un crecimiento ininterrumpido y elevado de la literatura científica en el dominio. Se destaca el año 2018 como el más productivo en relación a los estudios de discursos de odio. Este crecimiento acompaña las preocupaciones a escala internacional por combatir estas manifestaciones que generan crímenes de odio a la vez que reflejan una respuesta científica y social a los tumultuosos acontecimientos políticos en diversas esferas.

De un total de 411 registros analizados, fueron identificados 181 autores que contribuyeron con al menos una publicación. De esos autores apenas 21 presentan más de una publicación cada uno, lo que revela la dispersión que caracteriza los abordajes emergentes no consolidados dentro de determinados dominios del conocimiento.

La identificación de la élite investigadora se basó en la ley de Price, según la cual "el número de productores prolíficos es igual a la raíz cuadrada del número total de autores" (PRICE, 1976, p. 3). Según esta fórmula, los investigadores de $\sqrt{556}$ corresponden a aproximadamente 20 autores más productivos, de los cuales se destacan los siguientes: Brown, A. (6), Cohen-Almagor, R. (5), Bleich, E. (5), Boromisza-Habashi, D. (4) y Gelber, K. (4). El resto de los autores que conforman la élite de investigadores contribuyeron con al menos 2 publicaciones.

Entre los tres principales países líderes, en este grupo, destacan en producción universidades de Estados Unidos (15), Inglaterra (14) y Australia (9), seguidas de instituciones de educación superior de países europeos como Italia y Dinamarca, y la República Popular de China para el continente asiático.

En lo que respecta a las fuentes de estas publicaciones, del total de 298 revistas científicas identificadas, 232 revistas presentan solamente una publicación sobre discursos de odio, lo que

representa un 77,9 % del total de revistas identificadas. Estos datos indican que dichas fuentes presentan poca especialización temática sobre el dominio en estudio. En este sentido. Destacan solamente cinco revistas con más de 4 publicaciones sobre discursos de odio: Journal of Ethnic and Migration Studies (8), International Journal for Crime Justice and Social Democracy (6), Annales-Anali Za Istrske in Mediteranske Studije-Series Historia Et Sociologia (6) y Review of Faith & International Affairs (6) y Philosophia (5).

4.1 Nube de palabras de los términos incluidos en los artículos “hate speech” y su frecuencia absoluta de inclusión (2009-2018)

La aplicación del LDA identificó 15.492 términos diferentes, de los cuales 14.868 no están representados en la nube de palabras (figura 1). En todos esos casos su frecuencia de aparición es de 9 a 1. El término “hate speech” también fue excluido en la visualización para no alterar el peso específico de otros términos de este dominio. En esta única visualización con forma de nube de palabras (OLMEDA, 2014), el tamaño de cada término es proporcional a la frecuencia absoluta con la que ha sido utilizado. La gama cromática no guarda relación con ninguna característica implícita en los términos.

Figura 1. *Cloudwords* de los 624 términos incluidos en los campos de los título, resumen y palabras clave, de los 411 trabajos en el dominio “hate speech” (excepto 14.870 términos).



Fuente: Elaboración propia con <https://wordart.com/>

Destacan por su valor de frecuencia los términos online (192) e internet (110) que denotan el interés de la comunidad científica por los discursos de odio que se propagan en el ambiente digital. Este aspecto se refleja en los 78 artículos que abordan la temática y se enfocan en describir este fenómeno manifestado en redes sociales, específicamente Twitter. De forma análoga, la frecuencia de co-ocurrencia del término *political* (138) muestra el enfoque para analizar

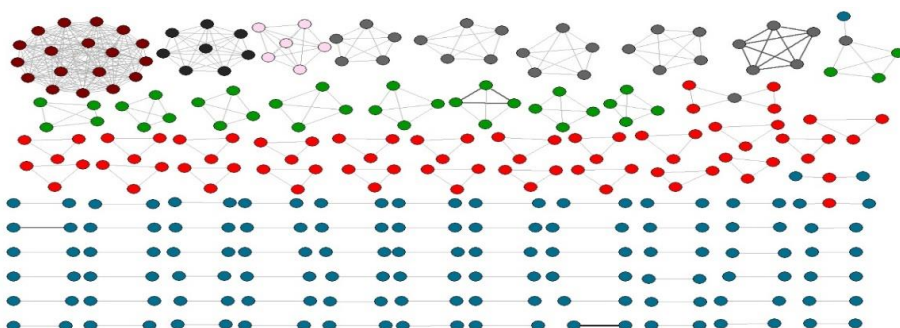
dichas expresiones en la narrativa de partidos políticos, utilizados durante sus campañas y las elecciones presidenciales. Las frecuencias de los términos *freedom expression* (114), *freedom* (104), *violence* (100) y *free_speech* (96) ratifican la afirmación de Esquivel Alonso (2016) de que la principal polémica en estos estudios reside en la definición de las fronteras de los derechos civiles de libertad de expresión, sobre los cuales mayormente se amparan los discursos de odio, y los derechos humanos respecto a la integridad de las personas y a su identidad religiosa de género, política y otros.

4.2 Red de coautoría

Entre las diferentes formas de medir la colaboración entre científicos, la coautoría es un aspecto esencial ya que se basa en la constancia presente en las publicaciones científicas. Su análisis desde la perspectiva del Análisis de Redes Sociales (ARS)² se fundamenta en la existencia de las relaciones entre autores. La principal característica de las relaciones de coautoría es que son recíprocas, de tal manera que si el autor A firmó en coautoría un determinado documento con el autor B, entonces el autor B también es coautor del denominado autor A (OVALLE-PERANDONES; OLMEDA-GÓMEZ; PERIANES-RODRÍGUEZ, 2014).

De los 411 artículos extraídos de la fuente de datos, 295 fueron publicados por autores que no colaboraron con ningún otro autor (no hubo coautoría). La proporción de esas cantidades permite afirmar que en los artículos internacionales publicados en el dominio del discurso del odio es habitual el desarrollo de investigaciones y la publicación de los resultados de manera predominantemente individual. La tasa de coautoría para el dominio es de 1,5 autores, valor muy próximo a la unidad, aspecto que confirma esa tendencia.

Figura 2. Red de coautoría de los artículos en el dominio “hate speech”. Fuente: Elaboración propia con Web of Science (2009-2018)



La red de coautoría del dominio “hate speech” (figura 2) muestra que, cuando los autores

² En inglés *Social Network Analysis* (SNA).

no trabajan de manera individual (lo que en ARS se conoce como nodos aislados o *isolated*), la red resultante es una red estructuralmente formada por muchos componentes (107). Estos nodos aislados no se incluyen en la representación, ya que en la red se incluyen los autores que al menos en una ocasión publicaron un trabajo en coautoría. La red quedaría así formada por un total de 301 autores (nodos) diferentes relacionados por 442 vínculos. Por otro lado, puede comprobarse que las relaciones que sí ocurren están representadas con tono gris claro, lo que significa que casi siempre la relación ocurre una sola vez, sinónimo de un único documento en coautoría. Son 15 las relaciones que ocurren en más de una ocasión, es decir, en más de un artículo, las cuales se visualizan con un tono más oscuro. Se puede apreciar que es una red con muy poca densidad de relaciones (0,009).

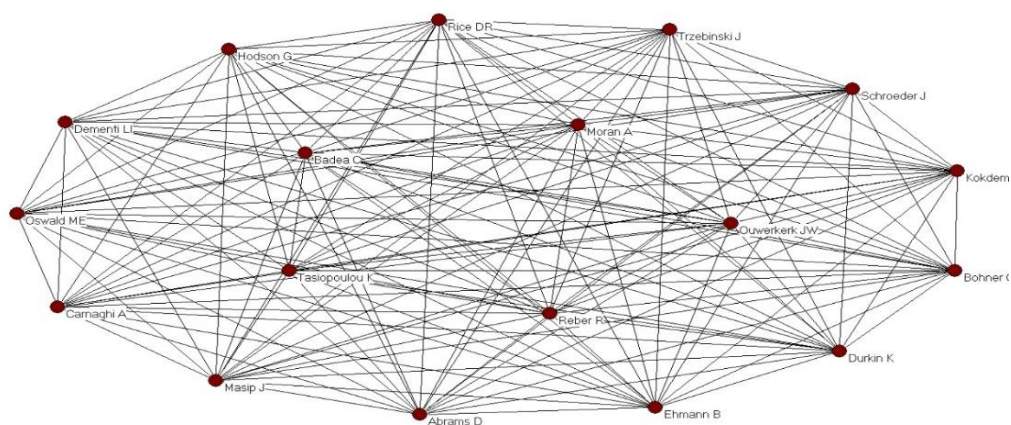
Hay que considerar también que los componentes en un grafo son partes internamente conectadas pero desconectadas a su vez de otras partes formando subestructuras (Hanneman y Riddler, 2005). La presente red es por lo tanto una red muy desconectada. El color de cada nodo se corresponde con la centralidad de grado en el grafo: color azul (grado 1), rojo (2), verde (3), gris (4), rosa (5), negro (7) y granate (17 y componente principal). Este componente principal es el que se visualiza en la figura 2 en la parte superior izquierda.

Como en el presente grafo están excluidos los nodos aislados, es decir, aquellos autores no participantes en una relación de coautoría, la densidad es moderadamente superior y el grado medio asciende a 2,9. La distancia entre los nodos es aleatoria y no responde a ninguna medida de redes. El algoritmo con el que se visualiza el grafo de la figura 2 es Kamada-Kawai, componentes separados.

El grafo del componente principal está formado por el mayor número de nodos o coautores y puede visualizarse en la figura 3. En un dominio con una tasa de coautoría tan reducida, resulta importante prestar atención a una ruptura en ese valor medio de 1,5 autores hasta 17. Este caso se corresponde con el artículo de título “What Did You Just Call Me? European and American Ratings of the Valence of Ethnophaulisms” publicado en el Journal of Language and Social Sociology en 2010.

Figura 3. Componente principal de red de coautoría de los artículos en el dominio “hate

speech”. Fuente: Elaboración propia con Web of Science (2009-2018)



4.3 Red de cocitación

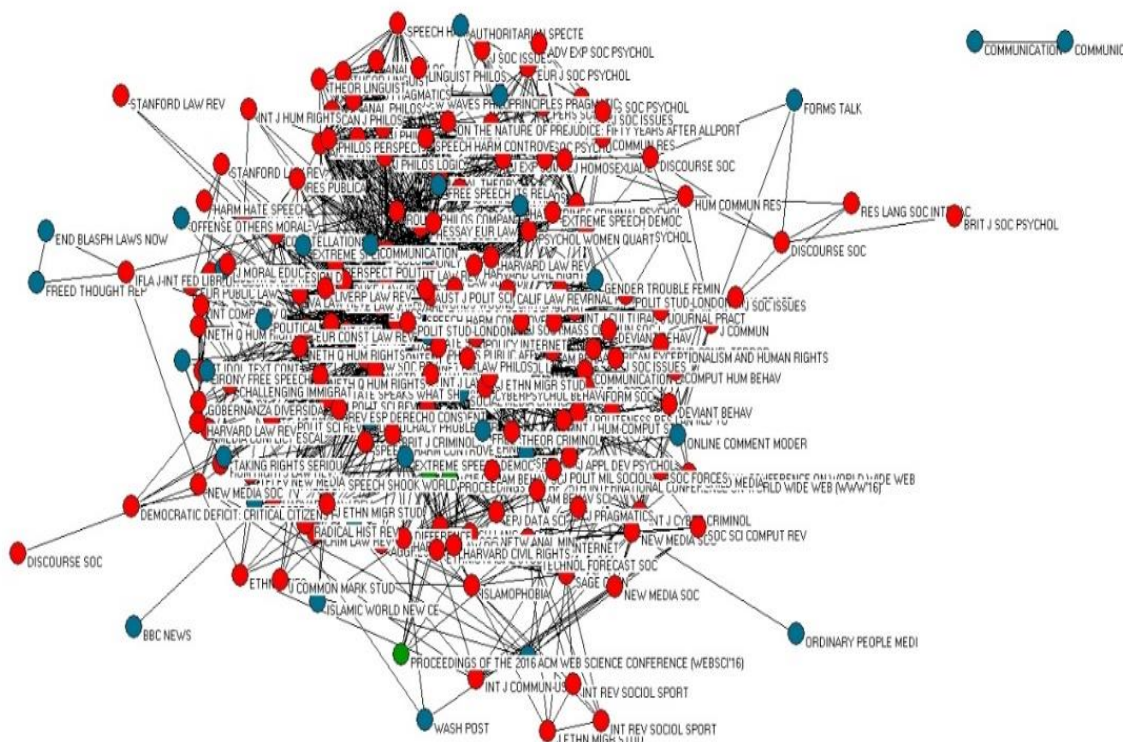
En contraposición con las relaciones de coautoría recíprocas están las relaciones de citación, las cuales tienen la característica de ser no recíprocas. En el contexto de la citación lo más probable es que un documento A cite a un documento B, pero el documento B no cite al documento A. Algunos aspectos que afectan las prácticas de citación son el tiempo en que transcurre para realizar las citas desde un determinado documento y el periodo que éste tarda en ser citado (en disciplinas, conocido como obsolescencia).

En ese mismo contexto de la literatura científica, habría una tercera posibilidad que corresponde con las relaciones de cocitación, las cuales van más allá de la presencia de reciprocidad en las relaciones o su asimetría. El análisis de cocitas se basa en la idea de que entre dos o más documentos que son citados conjuntamente por un tercer trabajo posterior, existe, al menos desde la perspectiva del trabajo citante, una similitud temática. A esto debe añadirse que cuanto mayor sea la frecuencia de cocitación, mayor será la afinidad entre ellos. La intensidad de esta relación vendrá dada por la cantidad de documentos citantes que tienen el mismo par de documentos en sus referencias. Si se acepta que los documentos altamente citados representan los conceptos, métodos o experimentos claves en un dominio, los patrones de cocitación pueden ser usados para identificar y visualizar las relaciones entre estos conceptos fundamentales (Small, 1973).

En la figura 4 se incluye la red de relaciones entre los trabajos cocitados en los 411 artículos del dominio “hate speech”. Siendo la referencia de esos documentos citados conjuntamente la unidad de análisis se visualiza que cuanto más cocitados son dos trabajos, más cerca aparecerán en el grafo. De este modo, los nodos son los trabajos citados que se vinculan por su relación de cocitación, mientras que la intensidad con la que eso ocurre se representa por el inverso de la distancia (o similitud) entre ellos. El color en esta ocasión refleja el tipo documental

de los documentos cocitados. Esta cualidad relativa al tipo documental se ha obtenido tras considerar la forma en la que los trabajos fueron citados y los elementos de sus referencias, por lo que existe la posibilidad de que puntualmente pudiera existir algún tipo erróneamente descrito o identificado. Para cada nodo se ha utilizado el color rojo para representar los artículos publicados en revistas científicas, azul para los libros y verde los para trabajos publicados en actas de congresos.

Figura 4. Red de cocitación en el dominio “hate speech” con títulos fuente. Fuente: Elaboración propia con Web of Science (2009-2018)



La pareja de cocitación más frecuente es [Anonymous], 1938, COMMUNICATION con el trabajo similar [Anonymous] 1939, COMMUNICATION.

Dado los altos valores de frecuencia de este par cabe destacar que éstos se concentran en un solo artículo científico publicado en la revista *Patterns of Prejudice* por David Goodman en el 2015 que se titula “Before hate speech: Charles Coughlin, free speech and listeners' rights”. Los pares cocitados corresponden a cartas, de diferente autoría, dirigidas a la Comisión Federal de Comunicaciones de Estados Unidos (*Federal Communications Commission* o FCC) de los Estados Unidos, con el objetivo de denunciar ciertos tipos de expresiones impropias en los medios de comunicación del país (de radio para el caso específico del artículo).

La red de cocitación identifica los trabajos centrales como los más altamente cocitados, Por lo tanto, ese debería ser considerado el núcleo teórico más central del dominio.

4.4 Distribución de los autores más citados

Los autores que fueron citados en 7 o más ocasiones en las referencias incluidas en los 411 trabajos son (según orden alfabético): Anonymous³, Allport G, Altman A, Austin John L, Baker C, Bleich E, Brison S, Brown A, Burnap P, Butler J, Cammaerts B, Citron D, Citron Danielle K, Cohen-Almagor R, Delgado R, Djuric N, Erjavec K, Farror S, Foxman A, Gagliardone I, Gelber K, Gerstenfeld P, Greenberg J, Hare I, Herz M, Hom C, Keane D, Langton R, Lawrence C, Lee E, Leets L, Leets L, Lillian D, Mackinnon C, Maitra I, Matsuda M, Mcnamee L, Npa (National Police A), Parekh B, Perry B, Post R, Tsesis A, Van Dijk T, Waldron J, Walker S, Weber A, Yong C.

El trabajo más citado pertenece a Jeremy Waldron titulado “The harm in hate speech” con un total 45 citas. Waldron estudia lo que se percibe como la discusión más polémica en cuanto a discursos de odio: el punto entre los límites entre la libertad de expresión y la regulación de discursos de odio, aspecto que describe de manera sintética este dominio de conocimiento.

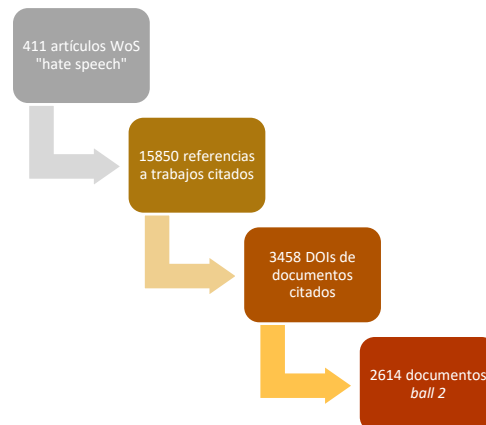
4.5 Snowball

Considerando las 15.850 referencias citadas en los 411 artículos “hate speech”, se llegó a la conclusión de que la forma de elaborar las referencias puede ser muy diferente además de que puede acumular diferentes tipos de fallos. En el contexto documental, los identificadores únicos como DOI (*Digital Object Identifier*) son una solución ante ese tipo de problemas. Por esta razón, se del conjunto de referencias citadas se extrajeron 3.450 DOIs (quedando así definido el conjunto de documentos *seed* utilizando en el snowball) para identificarse después la segunda oleada de análisis utilizando la misma fuente de información (Web of Science, Core Collection). Tras realizar este proceso, se recuperaron 2.614 documentos. La diferencia entre los documentos buscados y los recuperados puede encontrarse en DOIs erróneos o que no están controlados en la base de datos (el proceso queda esquematizado en la figura 5). A diferencia de otros trabajos similares, aquí no

³ Anonymous refleja la participación de una masa anónima ciudadana (aún con nombres individuales) que no es académica ni tiene afiliación institucional. Esta aparición de un autor “Anonymous” de estas características puede corresponder con el momento histórico del artículo que refleja la transición de una discusión sobre el tema en el ámbito académico del Derecho a una emergencia de la discusión en la sociedad en la que se da voz a los ciudadanos y que posteriormente fue reconocido por la Sociología. Por otro lado, el formato establecido por las normas de la revista citante para el tipo documental comunicaciones personales hace que los algoritmos de extracción de datos de WoS consideren el campo de autoría sin valor, lo que conlleva que sea un término asignado automáticamente para representar a los autores originales de las cartas enviadas al FCC.

se estableció la red de citación manteniendo el vínculo de la segunda oleada con los documentos *seed* (LECY; BEATTY, 2001) y sí se analizaron en su contexto global. Esta decisión metodológica se fundamentó en el poco uso de los identificadores para los documentos *seed* publicados al inicio del período.

Figura 5. Proceso para la obtención de la oleada 2



Algunas características descriptivas de la oleada 2 muestran las revistas más importantes de los documentos incluidos en esta fase, denotando las principales áreas de conocimiento desde las cuales este dominio es estudiado: Journal of Personality and Social Psychology (113), Personality and Social Psychology Bulletin (42), New Media Society (41), Discourse Society (34), Journal of Social Issues (26), American Political Science Review (22), Journal of Communication (23), American Psychologist (22), European Journal of Social Psychology (22), Harvard Law Review (22), Journal of Experimental Social Psychology (22), Yale Law Journal (22), Journal of Computer Mediated Communication (19), Disability Society (18), Ethics (18), Political Studies (17), Journal of Ethic and Migration (16), Journal of Applied Social Psychology (15), American Sociological Review (14), British Journal of Social Psychology (14), Psychological Bulletin (14), Journal of Homosexuality (14) Columbia Law Review (13), Computer in Human Behavior (13) y Ethic and Racial Studies (13).

Entre los autores con más documentos en esta segunda oleada destacan: Devine PG (10), Mullen B (10), Hatzenbuehler ML (9), Sidanius J (9), Crandall CS (8), Dovidio JF (8), Fiske ST (8), Hayes AF (8), Herek GM (8), Judd CM (8), Anderson (7), Augoustinos M (7), Cleland J (7), Gibson JL (7), Granger Da(7), Jost JT (7), Ullman SE (7), Burnap P (6), Cohen-Almagor R (6), Conan G (6), Duckitt J (6), Ellison Nb (6), Henry PJ(6) y Kawakami K (6).

5 Consideraciones finales

Los patrones y tendencias de la estructura científica internacional sobre discursos de odio

delinean un dominio que en términos científicos está creciendo en lo que respecta a publicaciones científicas. Su red de colaboración científica revela un dominio con una reducida tasa de coautoría no conservada en el efecto del componente principal de la red de coautoría. La estructura de cocitación revela el papel importante que en este dominio ha tenido en un momento histórico un tipo documental ajeno a los canales de publicación científico, como son las cartas dirigidas a la Comisión Federal de las Comunicaciones americana. En la actualidad, el dominio se vincula con áreas de conocimiento relativas a comunidades online e Internet, tal como se desprende del uso reiterado que se hace de estos términos. Las materias que marcan las tendencias en este dominio están relacionadas con la libertad de expresión y la regulación jurídica de discursos de odio. La principal aportación del método propio del ARS, nos permite afirmar que a pesar de que la presencia de discursos del odio es incipiente en la literatura científica, existe un importante núcleo teórico sobre la temática que está siendo frecuentemente citado en la comunidad académica internacional. Este núcleo de trabajos es reconocido como altamente relevante en los abordajes teóricos, conceptuales y metodológicos sobre discursos de odio y puede empezar a considerarse clásico y/o fundamental en esta reciente temática.

Agradecimientos

Esta investigación se encuadra dentro de los proyectos “*Understanding opinion and language dynamics using massive data*” financiado por el programa Trans-Atlantic Platform for Social Sciences and Humanities, “EMPATÍA-CM. protección integral de las víctimas de violencia de género Mediante comPutación AfecTiva multimodal” (Y2018/TCS-5046) financiado por la CAM – Consejería de Educación e Investigación e “Identificación de sesgos de género en inteligencia artificial. Discursos tecnológico, científico y mediático” (PDI2019-106695RB-100/AEI/10.13039/501100011033) financiado por el Ministerio de Ciencia e Innovación. El presente trabajo fue realizado con apoyo de la Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Referencias

BLEI, D. M.; NG, A. Y.; JORDAN, M. I. Latent Dirichlet allocation. **Journal of machine Learning research**, v. 3, p. 993-1022, Jan 2003.

BROWN, A. **Hate Speech Law**. London: Routledge, 2015.

CHETTY, N.; ALATHUR, S. Hate speech review in the context of online social networks. **Aggression and violent behavior**, v. 40, p. 108-118, 2018.

DOBROVOLSKYI, H.; KEBERLE, N. On Convergence of Controlled Snowball Sampling for Scientific Abstracts Collection. En: 14th International Conference on Information and

Communication Technologies in Education, Research, and Industrial Applications (ICTERI), p. 179-192, 2018. Disponible en: <http://ceur-ws.org/Vol-2105/10000179.pdf>. Acceso: 10 oct. 2019.

ESQUIVEL ALONSO, Y. El discurso del odio en la jurisprudencia del Tribunal Europeo de Derechos Humanos. **Cuestiones constitucionales**, n. 35, p. 3-44, 2016. Disponible en: <http://dx.doi.org/10.22201/ijj.24484881e.2016.35.10491>. Acceso: 1 nov. 2019.

HANNEMAN, R.A.; RIDDLE, M. Components. In: **Introduction to social network methods**. Riverside, CA: University of California, Riverside, 2005. Disponible en: <https://faculty.ucr.edu/~hanneman/nettext/C11 Cliques.html#Components> Acceso: 10 oct. 2019.

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.

HJØRLAND, B. Domain analysis. **Knowledge Organization**, v. 44, n. 6, p. 436-464, 2017.

HJØRLAND, B.; ALBRECHTSEN, H. Toward a new horizon in information science: Domain-analysis. **Journal of the American Society for Information Science**, v. 46, n. 6, p. 400-425, 1995.

ISASI, A. C.; JUANATEY, A. G. El discurso del odio en las redes sociales: Un estado de la cuestión. **Ajuntament de Barcelona progress report**. 2017.

LECY, J D.; BEATTY, K. E. **Representative literature reviews using constrained snowball sampling and citation network analysis** 2012. Disponible en: <http://dx.doi.org/10.2139/ssrn.1992601> Acceso: 11 oct. 2019

OLMEDA-GÓMEZ, C. Visualización de información. **El profesional de la información**, v. 23, n. 3, p. 213-219, 2014.

OVALLE-PERANDONES, M. A.; OLMEDA-GÓMEZ, C.; PERIANES-RODRÍGUEZ, A. ARS y redes de colaboración científica. Em: DEL FRESNO, M. P.; MARQUÉS-SÁNCHEZ; SÁNCHEZ-PAUNERO, D. (Eds.). **Conectados por redes sociales: introducción al análisis de redes sociales y casos prácticos**. Barcelona: Editorial UOC, p. 219-245, 2014.

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.

REVENGA SÁNCHEZ, M. **Libertad de expresión y discursos del odio**. Madrid: Universidad de Alcalá, 2015.

SMALL, H. Co-citation in the scientific literature: A new measure of the relationship between two documents **Journal of the American Society for Information Science**, v. 24, n. 4, p. 265-269, 1973.

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

SONNENWALD, D. H. Scientific collaboration. **Annual review of information science and technology**, v. 41, n. 1, p. 643-681, 2007.

STURGES, P. Limits to freedom of expression? The problem of blasphemy. **IFLA Journal**, v. 41, n. 2, p. 112-119, 2015.

WASSERMAN, S.; FAUST, K. **Social network analysis: Methods and applications** (Vol. 8). Cambridge University Press. 1994.