

Investigación del comportamiento de menores y jóvenes en las redes sociales mediante técnicas de Social Big Data

Research into the behaviour of minors and young people on social networks using Social Big Data techniques



Rebeca Suárez-Álvarez. Profesora ayudante doctora en la Universidad Rey Juan Carlos (Departamento de Comunicación y Sociología). Doctora en Comunicación Social por la Universidad CEU San Pablo. Licenciada en Periodismo por la Universidad Complutense de Madrid (UCM). Posee dos másteres: uno en radio por la Universidad San Pablo CEU y otro en Comunicación de Instituciones Públicas y Políticas en la UCM. Sus principales líneas de investigación son la comunicación y los públicos vulnerables (concretamente, menores), medios digitales, alfabetización mediática y competencia digital. Ha participado en diferentes proyectos de investigación y en la actualidad, es investigadora en el proyecto “Nuevos escenarios de vulnerabilidad digital: alfabetización mediática para una sociedad inclusiva” (PROVULDIG-2CM) (ref. H2019/HUM5775), financiado por la Comunidad Madrid y el Fondo Social Europeo. Miembro del grupo de investigación sobre Comunicación, Sociedad y Cultura (GICOMSOC).

Universidad Rey Juan Carlos, Madrid, España
rebeca.suarez@urjc.es
ORCID: 0000-0002-0102-4472



Antonio García-Jiménez. Catedrático de Periodismo en la Universidad Rey Juan Carlos (Departamento de Comunicación y Sociología). Doctor en Ciencias de la Información. Ex Decano de la Facultad de Ciencias de la Comunicación de la Universidad Rey Juan Carlos (2008-2014). Ha sido también director del Máster en Comunicación y Problemas Socioculturales (2015-2018). Profesor en el grado de Periodismo y en varios másteres de Comunicación (Medios Sociales; Periodismo de Datos). Investigador principal del grupo de investigación sobre comunicación, sociedad y cultura (GICOMSOC). Ha dirigido o participado en 19 proyectos de investigación relacionados con los usos y riesgos digitales para los adolescentes y es autor de un gran número de publicaciones de impacto, tanto nacionales como internacionales. Es miembro de ECREA, IAMCR y AE+IC.

Universidad Rey Juan Carlos, Madrid, España
antonio.garcia@urjc.es
ORCID: 0000-0002-8423-9486

Recibido: 15/09/2020 - Aceptado: 07/04/2021

Resumen:

El objetivo es identificar la producción científica sobre el comportamiento y consumo comunicativo de los menores y los jóvenes en internet utilizando una metodología Social Big Data (SBD). Mediante una revisión sistemática se han identificado 58 documentos académicos publicados entre 2010-2020 (mayo). Se compendian las dimensiones más investigadas, los países con

Received: 15/09/2020 - Accepted: 07/04/2021

Abstract:

The aim of this study is to identify the academic research related to the behaviour and communication consumption of minors and young people on the internet through the use of a Social Big Data (SBD) methodology. A systematic review has identified 58 scholarly works published between 2010 and 2020 (May). This article has summarised the most widely studied aspect of the issue, as well as

Cómo citar este artículo:

Suárez-Álvarez, R. y García-Jiménez, A. (2021). Investigación del comportamiento de menores y jóvenes en las redes sociales mediante técnicas de Social Big Data. *Doxa Comunicación*, 32, pp. 95-113.

<https://doi.org/10.31921/doxacom.n32a5>

mayor producción científica, el perfil académico de las revistas, las técnicas de investigación basadas en SBD y los hallazgos más relevantes. Las principales conclusiones son que los científicos están utilizando el SBD para, mediante gran cantidad de datos analizados en tiempo real, conocer los usos y efectos de las acciones de los adolescentes en la red así como para crear algoritmos que posibiliten la identificación de tendencias de riesgo adolescente. Se confirma que la producción científica es escasa en revistas de ciencias sociales y la necesaria asociación y coautoría de comunicólogos y sociólogos con científicos de perfil técnico para lograr mayor comprensión de la realidad e incrementar las publicaciones en ciencias sociales.

Palabras clave:

Social Big Data; adolescentes; jóvenes; datificación; redes sociales.

the countries that have conducted the largest amount of research, the academic profile of the journals, research techniques based on SBD, and the most relevant findings. The main conclusions of the study are that scientists currently use SBD to understand the uses and effects of adolescent online activity by analysing large amounts of data in real time, and also to create algorithms that make it possible to identify trends in adolescent risk. It has been confirmed that there is a scarce amount of academic research related to this topic in social science journals, and there is also a shortage of an association and co-authorship between communicologists and sociologists, on the one hand, and scientists with a technical background on the other hand, which is necessary in order to achieve a greater understanding of the situation and increase the number of publications related to this issue in the area of social science.

Keywords:

Social Big Data; young people; adolescents; datafication; social networks.

1. Introducción

Este artículo presenta una investigación sistemática de las investigaciones publicadas sobre el comportamiento de los menores en la red y que empleen técnicas de Big Data (BD). Según el objetivo de la investigación, las técnicas BD se pueden dividir en descriptivas, predictivas y prescriptivas. Las descriptivas pretenden entender la realidad mediante técnicas como las tasas de variación, tablas de frecuencias, test A/B, análisis factorial y de *cluster*, etc. Las predictivas anticiparse a los acontecimientos (series temporales, técnicas de regresión, redes neuronales, *Machine learning* y *deep learning* y algoritmos de *boosting* como XGBoost). Y las prescriptivas buscan una recomendación mediante la identificación de reglas de causa/efecto o algoritmos de optimización (métodos de probabilidad condicionada, técnicas de regresiones, reglas de asociación, método de Montecarlo y simulación estocástica, algoritmos genéticos y técnicas de optimización espacial).

Para poder aplicarlas, los científicos precisan de desarrollos tecnológicos específicos para lograr convertir los millones de datos en información como son: *Cassandra*, *Extract, transform, and load*, *Hadoop*, *HBase*, *MapReduce*, *Linguistic inquiry and word count (LIWC)*, *Hamlet*, *WordStat*, *QDAMiner* o *Python* para el análisis automático del lenguaje, entre otros (Pérez, 2015; Joyanes, 2016; Arcila-Calderón, Barbosa-Caro & Cabezuelo-Lorenzo, 2016; Hernández-Leal, Duque-Méndez & Moreno-Cadavid, 2017).

El objetivo de esta investigación es profundizar en este tipo de investigación y determinar su potencial incorporación al estudio de los comportamientos y consumos comunicativos de jóvenes en internet, para complementar el conocimiento que tenemos de este fenómeno.

2. Estado de la cuestión

2.1. Orientación del Big Data

Tal y como apuntan Pérez (2016) y Schwab (2017), la sociedad está viviendo la “cuarta revolución industrial” o “revolución digital” en la que las innovaciones tecnológicas y los dispositivos digitales están modificando el paradigma social. Esta revolución se caracteriza por el incremento exponencial de la cantidad de datos digitales que generan los individuos en tiempo real de forma automática, rutinaria y a través de diversos dispositivos (Batty, 2013) y que son los conocidos como Big Data (BD). Paredes-Moreno (2015) señala que el ámbito de actuación del BD está relacionado con la información que no puede ser analizada adecuadamente por otros medios y Arcila-Calderón, Barbosa-Caro y Cabezuelo-Lorenzo (2016) lo describen en términos de grandes flujos y cantidades de información, tanto estructurados como no estructurados, y a los que se aplican instrumentos y métodos computacionales con vistas a extraer conocimiento.

2.2. Big Data en el paradigma científico de las ciencias sociales

El análisis de los grandes datos surge como nuevo paradigma de comunicación que va más allá de ser considerado como un factor puramente tecnológico, sino que adquiere una dimensión social, política y económica (Malvicino & Yogue, 2014) que lleva a considerar a la tecnología digital, no como un subproducto de las relaciones sociales, sino tomando en consideración su papel social y comunicativo (Tapia, 2018) en la codificación de los datos.

Desde esta perspectiva, surge el “cuarto paradigma científico” propuesto por Hey, Tansley y Tolle (2009) en el que las TIC y los científicos convergen dando lugar al descubrimiento de un nuevo arquetipo de investigación basado en una ciencia basada en datos masivos (Bell, Hey & Szalay, 2009) por medio de la captura de información y de técnicas como la minería de datos impensables hace tan solo una década.

Este planteamiento da origen al debate sobre los límites y la interrelación entre el método, el objeto de estudio y el dato (Tapia, 2018). Será necesario replantearse las preguntas clave sobre la construcción del conocimiento y los procesos de investigación social para lograr categorizar la realidad (Boyd & Crawford, 2012). Manovich (2011) plantea cómo el BD modifica el método de investigación en ciencias sociales y humanidades en la medida en que el mundo se vuelve más digital. Así, serán necesarias nuevas técnicas para investigar, analizar y entender un gran volumen de datos cuyo impacto se puede medir en términos cualitativos y cuantitativos (Mayer-Schönberger & Cukier, 2013) y que según Cerezo (2015), está dando lugar a una ciencia relevante en el presente y el futuro. Qin (2014) identifica la actualidad con la “era Big Data”, un momento en el que está adquiriendo relevancia en la investigación científica presentando implicaciones en la propia teoría social. En esta línea, Mayer-Schönberger y Cukier (2013) afirman que el BD no solo provee de datos masivos sino que está provocando tres modificaciones en el paradigma de la investigación en ciencias sociales: mayor disponibilidad y acceso de los datos, mayor aceptación de los niveles de imprecisión y desorden en los datos y el incremento de la posibilidad de centrarse más en las correlaciones en vez de buscar la causalidad.

2.3. *Social Big Data*

El incremento masivo de datos revela la necesidad de que las metodologías de investigación evolucionen para lograr que los datos se conviertan en conocimiento que se transfiera a la sociedad. De la posibilidad de datificar la realidad social online surge el Big Data Social (Social Big Data, SBD) (Manovich, 2011). Se centra en el estudio de los contenidos públicos, así como de las expresiones de gusto, estados o corrientes de opinión, temas tratados, descripción de perfiles, o las interacciones entre personas, entre otros aspectos (Mayer-Schönberger & Cukier, 2013). A juicio de Russell (2013), el SBD ofrece la posibilidad de capturar los detalles de la comunicación online a través de los dispositivos móviles y brinda múltiples posibilidades para el análisis de datos almacenados en sitios web.

Es producto de la confluencia de tres grandes áreas: los medios sociales, el análisis de datos y los datos masivos (Bello-Organ, Jung & Camacho, 2016) y su fusión está generando una nueva disciplina en la que se genera conocimiento a raíz de procesar y analizar la información y los datos que se derivan de las redes sociales (Gualda & Rebollo, 2020). Con el SBD se logra analizar y descubrir el conocimiento que albergan las aplicaciones y los servicios web que no se limitan a los mensajes de texto sino también a sonidos, imágenes y vídeos (Kumar, Sangwan & Nayyar, 2020).

2.4. *Posiciones críticas al uso de Big Data en ciencias sociales*

Esta visión tecno-optimista de la aplicación de BD en la investigación también ofrece dudas a los científicos del área de ciencias sociales. López (2018) señala que la promesa de una mejora del conocimiento y de un cambio de paradigma está apoyada por la posición tecnoneopositivista promovida desde Silicon Valley e impulsada por la creciente capacidad de procesamiento de datos actuales. Por su parte, Boyd y Crawford (2012) y Taylor-Sakyi (2016), que no ponen en duda los beneficios de la gestión de los grandes datos, manifiestan que es necesario interrogar críticamente los supuestos y sesgos del BD. En el fondo, estaríamos ante una mitología creada en el supuesto de que los grandes datos son una suerte de modelo de inteligencia superior que garantiza resultados más precisos y objetivos (Boyd & Crawford, 2012; Martínez-Martínez & Lara-Navarra, 2014).

Asimismo, Pybus, Coté y Blanke (2015) cuestionan la fe en la datificación de la sociedad teorizada por Mayer-Schönberger y Cukier (2013) puesto que consideran que adolece de “resolución de verosimilitud”. Ponen en duda que el análisis de datos a través de un algoritmo produzca necesariamente información válida. Por otra parte, van Dijck (2014) apunta a los problemas ontológicos y epistemológicos que se encuentran en el trabajo con datos, puesto que su recogida, tratamiento y análisis no es un proceso neutral. Y O’Neil (2013) advierte que en el análisis de los datos se debe tener en cuenta quién los controla, cómo los estructura y con qué propósito.

Del mismo modo se debe apreciar que la recopilación de datos a gran escala debe ser cuidadosa con la representatividad de los datos puesto que pueden dar lugar a decisiones erróneas (UN Global Pulse, 2018; Uman, 2018), lo que no es ajeno a la perspectiva de la ciencia social computacional (Törnberg & Törnberg, 2018). Además, se encuentran los potenciales riesgos relacionados con el daño a la privacidad e intimidad, la libertad de expresión y el derecho a la información (Arellano, 2014; Shuijing, 2015).

2.5. Investigación sobre niños, adolescentes y jóvenes en internet

En este trabajo, el SBD se aplica como punto de partida para estudiar el uso que hacen los menores de Internet y de las redes sociales. Aunque se tiene en consideración el uso y consumo por parte de esta parte de la población, el interés de este trabajo supera el enfoque comunicológico y social. Es muy amplia la literatura sobre los niños, adolescentes y jóvenes en internet. La investigación desarrollada en los últimos años tienen al proyecto EU Kids Online (<http://www.lse.ac.uk/media-and-communications/research/research-projects/eu-kids-online>) y a su directora Sonia Livingstone (Livingstone, Mascheroni & Staksrud, 2018) como máximos exponentes de los estudios centrados en adolescentes, junto a otras muchas iniciativas de carácter nacional e internacional. Las técnicas que habitualmente se han estado empleando son las encuestas y los grupos de discusión/*focus group*. Además de los riesgos y los hábitos de uso y consumo, se abordan cuestiones como el control parental, la educación en la era digital, o las políticas vinculadas, entre otras (Jiménez, Garmendia & Casado, 2018).

Se trata, a su vez, de un objeto de estudio en el que participan otros dominios como el psicológico (Alsehaima & Alanazi, 2018), el de la salud (Falzone, et al., 2017) o la educación (Giménez, Luengo & Bartrina, 2017), con una clara profusión de literatura en los últimos años. Al mismo tiempo es un campo que no solo permite, sino que requiere aproximaciones multidisciplinares, y que ha abierto la oportunidad de ámbitos mixtos de estudio como es el caso de la educomunicación (Aparici & García, 2016; Nupairoj, 2016).

En este artículo, se estudian las características de las publicaciones existentes sobre el uso y el comportamiento de los menores y jóvenes en internet y las redes sociales que aplican una metodología que se puede vincular con el SBD. Se analizan las aportaciones de esta literatura académica y se busca descubrir la potencial contribución metodológica del SBD al estudio del comportamiento social y comunicacional de este sector de la población en internet.

3. Metodología

Para la investigación se aplica el análisis bibliométrico (Díaz-Campo, 2015) que estudia el flujo de conocimiento dentro de un campo específico durante un período de tiempo (Leung, Sun & Bai, 2017) a través del análisis estadístico cuantitativo de la literatura académica existente y de las citas obtenidas (van Eck & Waltman, 2014). La idea es identificar y analizar la literatura existente en tres fases: búsqueda de la literatura, extracción de los datos más relevantes y análisis de los contenidos abordados por la bibliografía estudiada (Ramírez-Montoya & García-Peñalvo, 2018). Para ello, esta investigación presente responder a las siguientes preguntas:

P1) ¿Cuál es la cantidad de estudios que tratan el objeto de estudio? ¿En qué publicaciones y en qué fecha aparecen?

P2) ¿Qué conceptos y dimensiones de investigación son las más estudiadas?

P3) ¿Cuál es el dominio de conocimiento de las publicaciones que se dedican al objeto de estudio?

P4) ¿Qué técnicas basadas en SBD han aplicado y cuáles son las principales muestras analizadas?

P5) ¿Cuáles son las principales aportaciones de las investigaciones?

La búsqueda documental se ha realizado en las bases de datos WoS, Scopus, IEEE Xplore y en el buscador Google Scholar de todos los artículos científicos publicados entre 2010 y 2020 (mayo), en inglés y en español, en revistas científicas (artículo), informes, libros (capítulos) y tesis como fuentes de investigación.

La estrategia documental recoge los siguientes términos de búsqueda y sus combinaciones con los operadores booleanos (y/o): “joven*”, “niños”, “menor*”, “youth”, “child*”, “childhood”, “adolesc*” y “social big data”. Los documentos sin conclusiones científicas se han desestimado y, a través del hipertexto, se aplicó metodología de investigación “bola de nieve” para localizar más documentos con las referencias de búsqueda seleccionadas en las mismas bases de datos (Wohlin, 2014). Las variables de registro con las que se ha trabajado son:

- Referencia completa del documento: título, año, autor/es, revista/congreso, *keywords*, *abstract* y número de citas.
- Autores: nombres, país e institución.
- Investigación: hipótesis de la investigación, objetivos, variables estudiadas.
- Resultados: hallazgos, limitaciones y recomendaciones de los autores.

Asimismo, se ha acudido a la base de datos *SCImago journal Rank* para acceder a los descriptores de las áreas y categorías que aborda cada publicación así como para conocer el *quarter* en el que se encuentra posicionada cada una. Se utilizó el programa *Publish or Perish* (<http://www.harzing.com/pop.htm>) y el programa Excel del paquete Office de Word para la clasificación e identificación de la información así como para registrar cada elemento de la muestra.

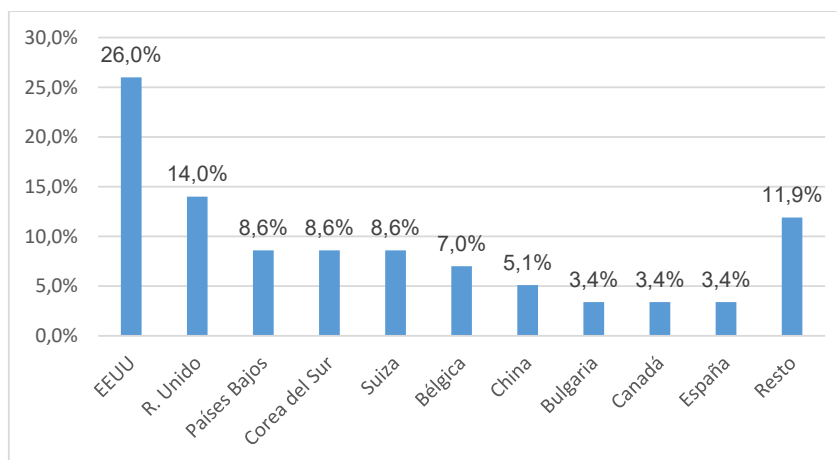
Tras la obtención de resultados, se procedió a la necesaria depuración de resultados en la que se realizó una revisión preliminar en la que se eliminaron falsos positivos como duplicados así como en la medida en que dichas búsquedas no abordan el objeto de estudio, se centran en corpus teóricos o técnicas de aplicación, no ofrecen texto completo, no abordan temáticas relacionadas con adolescentes, revisiones bibliográficas y datos que no coincidan con los propósitos de la investigación. Posteriormente, se procedió a la lectura de todos los artículos para comprobar su objeto de estudio y corroborar la metodología y las muestras seleccionadas con el fin de analizar sus contenidos para dar respuesta a las preguntas planteadas.

4. Resultados

4.1. Número de publicaciones, fuentes y fecha

El resultado final del proceso de criba bibliográfica ofrece 58 (N) resultados que componen la muestra. El 63,8% (n=37) son artículos en revistas, el 12% (n=7) son capítulos de libros, el 10,3% congresos (n=6) y tesis e informes 7% (n=4) cada uno. Destacan 3 revistas por su citación: *Yonsei Medical Journal* con 57 citas (2014), *Telecommunications Policy* con 46 citas (2015) y *Journal of Adolescent Health* con 29 citas (2016). Estados Unidos es el país con mayor producción científica (26%; n=15), seguido de Reino Unido (14%; n=8) y Países Bajos, Corea del Sur y Suiza con un 8,6% (n=5) respectivamente. Bruselas alcanza el 7% (n=4), China el 5,1% (n=3) y Bulgaria, Canadá y España con 3,4% (n=2) cada uno. El resto – Australia, Birmania, Finlandia, India, Nueva Zelanda, Singapur y Vietnam- publican 1,7% (n=1) (ver figura 1).

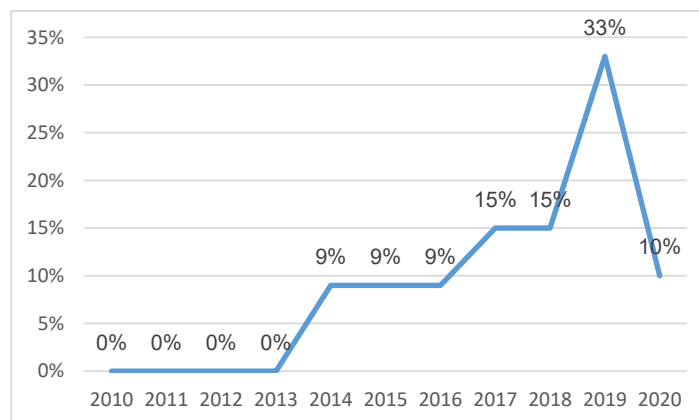
Figura 1. Países con mayor producción científica sobre SBD y menores



Fuente: elaboración propia

En relación a la publicación por años (ver figura 2), se observa la tendencia temporal hacia un mayor interés científico para estudiar cuestiones relativas al SBD centrado en menores pasando del 18% (n=10) artículos entre 2010 y 2015 al 82% (n=48) entre 2016 a 2020.

Figura 2. Evolución de la producción científica tras la depuración del número de artículos (2010-2020)



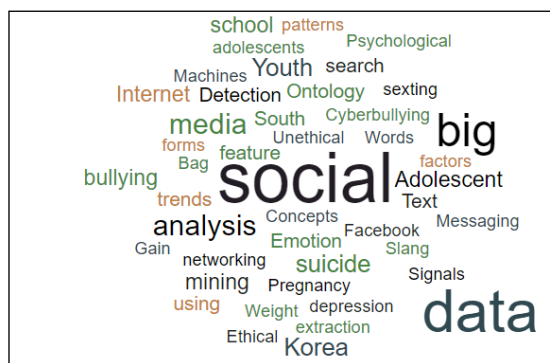
Fuente: elaboración propia

Entre los autores con más publicaciones científicas se encuentran las investigadoras Tae Min Song (*Korea Institute for Health and Social Affairs*) y Juyoung Song (*University of West Georgia*, EE.UU.) que han escrito el 18,6% (n=11) de los artículos como coautoras junto con otras científicas de Corea del Sur o de Estados Unidos, siendo su investigación “*Psychological and social factors affecting Internet searches on suicide in Korea: a big data analysis of Google search trends*” (2014) la más citada (n=57) y “*Data mining of web-based documents on social networking sites that included suicide-related words among Korean adolescents*” la tercera con más citas (n=29). De los artículos científicos publicados en revistas científicas el 67,5% (n=25) aparecen en la base de datos *Scimago Journal & Country Rank*. De ellas más de la mitad (59,5%, n=22) son Q1 y Q2 y 8,1% (n=3) Q3. El resto no aparecen indexadas en esta base de datos.

4.2. Conceptos y dimensiones estudiadas

El universo de conceptos incluidos en el título y en las palabras clave de los documentos registrados demuestra el acierto de la elección de los términos para la búsqueda documental y muestra los aspectos relevantes y significativos de esta temática (ver figura 3).

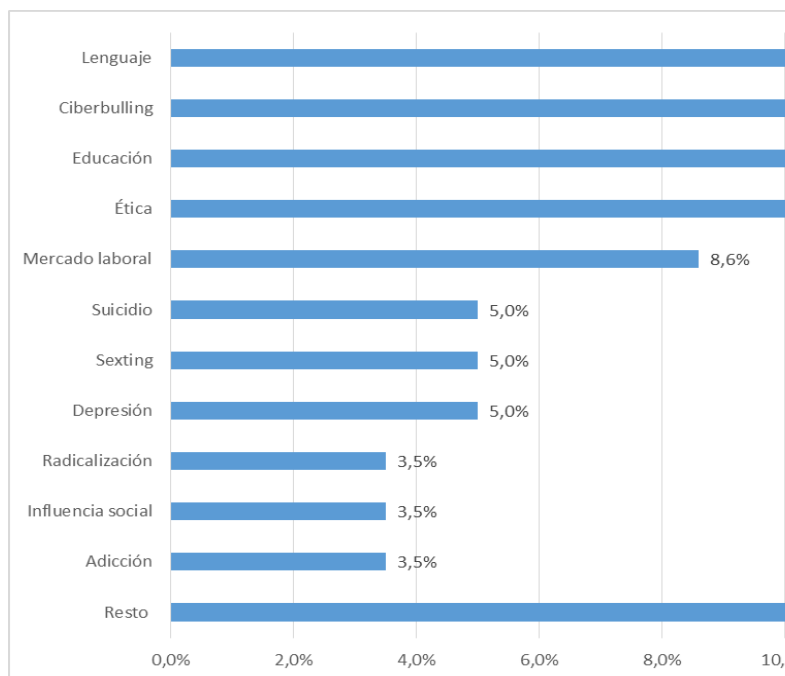
Figura 3. Conceptos clave recurrentes en la investigación



Fuente: realizado con Word it out. Elaboración propia

En relación a las dimensiones identificadas en la literatura científica con temáticas relacionadas con los menores y de sus propio interés (Fondo de las Naciones Unidas para la Infancia [UNICEF], 2019) investigadas con SBD destacan los estudios centrados en el análisis del lenguaje y opiniones en redes sociales (19%; n=11), el acoso escolar online (15,3%; n=9), las prácticas vulnerables que invaden la privacidad de los adolescentes y las políticas educativas (10,3%; n=6).

El resto de los registros (ver figura 4) abordan temas tan dispares como la situación y previsiones del mercado laboral con un 8,6% (n=5), depresión, sexting y suicidio adolescente con un 5% (n=3) cada uno e influencia social de las redes sociales y radicalización juvenil (3,5%; n=2). Le siguen con un 1,7% (n=1) cada uno (identificados como “resto” en el gráfico) delincuencia juvenil, deporte y ocio, embarazo adolescente, medio ambiente, móviles, moda y privacidad.

Figura 4. Dimensiones abordadas en la investigación SBD y menores

Fuente: elaboración propia

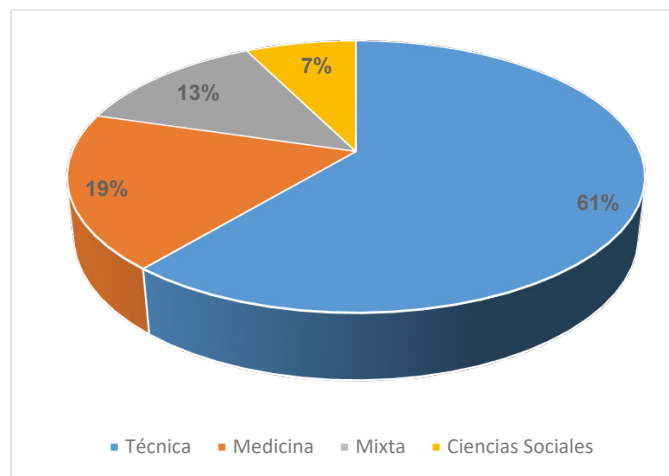
4.3. Dominio de conocimiento de los estudios

Otro de los aspectos a estudiar en la presente investigación bibliográfica es conocer el dominio de conocimiento al que pertenecen las publicaciones. Para ello se han segmentado las publicaciones en 4 categorías que se corresponden con el perfil de las revistas que componen la muestra:

- Técnicas: abordan cuestiones como computación, ingeniería e informática.
- Ciencias sociales: que se componen de 13 principales ramas: Antropología, Historia, Ciencia del derecho, Economía, Sociología, Psicología, Lingüística, Semiología, Arqueología, Demografía, Ecología humana, Pedagogía y Relaciones internacionales (Centro de Estudios Cervantinos, 2020).
- Mixtas, que combina la categorías de revistas técnicas y revistas de ciencias sociales.
- Medicina.

En esta clasificación no se recogen los 4 informes de la Unión Europea que no ofrecen sesgo temático. Del restante 93% (n=54), el 61% (n=33) son publicaciones técnicas, médicas el 19% (n=10), de las denominadas mixtas que integran contenidos y categorías técnicas y de ciencias sociales 13% (n=7) y de revistas de ciencias sociales 7% (n=4) (ver figura 5).

Figura 5. Dominio de conocimiento de las publicaciones



Fuente: elaboración propia

4.4. Técnicas de SBD aplicadas y muestras

Las técnicas aplicadas varían según el objeto de estudio y los autores no expresan que existan diferencias entre las técnicas SBD y BD. Los investigadores realizan sus estudios aplicando técnicas estadísticas como el análisis de frecuencias para detectar la importancia de una o varias palabras en la red, diferentes test estadísticos correlacionales, se han aplicado diversos algoritmos para la detección de tendencias, y también se ha encontrado el uso del análisis estadístico bivariable, el análisis de *cluster*, los análisis temporales y ecuaciones estructurales, entre otras técnicas. Asimismo, también se observa la existencia de estudios que hacen uso del vocabulario utilizado en las redes sociales para aplicar el *sentiment analysis* y también el *machine learning*, lo que permite el análisis y codificación de textos a gran escala. En otros casos se emplea el *data mining*, *text mining* o bien *opinion mining* para identificar y analizar los comentarios, post y mensajes que los jóvenes dejan en las redes sociales.

Gracias al uso del SBD las muestras con las que trabajan los investigadores son muy elevadas. Entre ellos destacan, Kim, et al., (2019) que, para comprender las experiencias actuales y futuras de la intimidación de adolescente analizan 436.508 documentos web de 279 canales online entre 2013 a 2017 en los que identifican 177 términos relativos a formas específicas de acoso escolar. A su vez, Song et al., (2016) revisan más de 2.35 billones de post durante 2 años de 163 *websites* y

consolidan 99.693 registros de menores concernientes al suicidio juvenil. Han, *et al.*, (2019) estudian 2.400 términos en tres categorías en relación al *cyberbullying* y Song, *et al.*, (2014) recopilan documentos web de documentos de 227 canales online de los que extraen 435.563 casos relacionados con el acoso cibernético e identifican 118 términos como causas del *cyberbullying*. Finalmente, Matsumoto, Yoshida y Kita (2019) crean un corpus propio para la investigación de tuits basados en argot juvenil cuando se comunican en redes sociales. Logran crear listas y vectores de expresiones de emoción de hasta 30.000 elementos.

4.5 .Principales propuestas y aportaciones

Para este apartado nos hemos centrado en los estudios que de modo específico han controlado el corpus analizado en el sector de la población que es relevante para esta investigación. De todos los registros encontrados, 16 estudian y trabajan con muestras específicas de adolescentes o se refieren específicamente a ellos cuando elaboran y desarrollan sus corpus en los temas/dimensiones que se recogen a continuación.

4.5.1. Detección del *cyberbullying*

Kim, et al., (2019), mediante sistemas de reconocimiento automático del lenguaje en el contexto de las redes sociales, señalan que los problemas de acoso online no se limitan a la escuela, sino que se ha extendido a otros entornos como las comunidades de adolescentes “*fandom*”, que reúnen a jóvenes aficionados a algún pasatiempo o afición. En este sentido, recomiendan al gobierno sistemas de control de modos emergentes de las diferentes formas de acoso. También Han, et al., (2019) analizan el acoso escolar como fenómeno social gracias al análisis de los términos utilizados en las redes sociales por los menores. Coincide con el estudio de Kim et al. (2019) que apuesta por crear políticas de seguridad de los datos y de protección de la privacidad de los individuos investigados al utilizar SBD.

Song, at al., (2014) estudian los factores de riesgo para cada tipo de acoso online con el objetivo de generar automáticamente un modelo predictivo con el que explicar las variables que afectan a este comportamiento juvenil. También sugieren que se deberían instaurar programas de educación para padres con vistas a ayudar a los menores ante estas situaciones. Tarwani, Chorasias & Shukla (2017), que extraen las opiniones de internautas en *Facebook*, *MySpace* y *Twitter* mediante técnicas de aprendizaje automático, ofrecen la posibilidad de realizar una detección instantánea de los mensajes de intimidación en las redes sociales y proponen la solución al problema del ciberacoso basándose en la identificación de los términos intimidatorios.

Por su parte, Lara y Giancinto (2017) trabajan con modelos estadísticos para la detección de expresiones ofensivas así como con un sistema que captura estas comunicaciones individuales. La clave se sitúa en la generación de una aplicación que reconoce comentarios y declaraciones con connotaciones negativas. Su principal limitación es la interpretación del análisis contextual de las conversaciones, lo que puede conducir a falsos positivos, especialmente cuando el lenguaje puede ser sarcástico o grosero, pero no ofensivo, en una situación específica.

4.5.2. *Sexting como práctica virtual adolescente*

El *sexting* es un fenómeno emergente que atrae una atención significativa por su potencial para generar daños en los menores (Song & Song, 2015). Song, Song y Lee (2018) aplicando SBD en las comunicaciones de menores de 18 años en Twitter, concluyen que los adolescentes sextean a través de mensajes de texto con teléfonos inteligentes con la esperanza de obtener una mayor atención de los amigos. Indican que el asesoramiento experto y el fomento de la integridad de los menores suponen factores que ayudan a minimizar los riesgos del *sexting*. Con respecto al uso del SBD, afirman que, si bien esta metodología es útil y efectiva, presenta dificultades como el seguimiento en tiempo real y el análisis de los datos individuales. Consideran clave la continuación de la investigación con esta tecnología sobre los factores de riesgo relacionados con esta práctica.

4.5.3. *El riesgo de suicidio en los jóvenes*

Esta dimensión en las investigaciones con SBD adquiere especial relevancia en Corea del Sur puesto que las tasas de suicidio juvenil en ese país es una de las más altas del mundo (Song, *et al.*, 2016) y las búsquedas de términos relacionados con el suicidio en Google en este país crece constantemente desde 2007.

Song *et al.*, (2014) señalan que la investigación en tiempo real de los grandes datos resulta útil para detectar las tendencias suicidas. En concreto, estos autores crean un modelo estadístico a raíz del análisis de la relación entre el número de búsquedas en Google de las palabras clave *stress, drinking, exercise, suicide*, y por otra parte, la detección de los riesgos del suicidio adolescente. Determinan que un mayor número de búsquedas de las palabras estrés y consumo de alcohol y el número real de suicidios resulta significativa. Son conscientes que utilización y aplicación de SBD para este tipo de investigaciones se encuentran en las primeras etapas y aconsejan que, para mejorar esta situación y poder establecer planes de prevención del suicidio a nivel nacional, se deben crear programas de capacitación para profesionales en la gestión de grandes datos, mejorar las infraestructuras e implementar servicios *cloud computing*.

Por su parte, Song, *et al.*, (2016) confirman que los adolescentes expresan el estrés, las emociones depresivas y sus pensamientos suicidas en el ciberespacio y que lo comunican con otros internautas. Entre los factores que se asocian con una mayor actividad en la búsqueda de palabras relacionadas con el suicidio adolescente destaca la presión académica, la baja imagen corporal, ser víctima de acoso, la preocupación por alguna enfermedad, la baja tasa de empleo, el índice de precios de alquiler más alto y la intimidación. Para reducirlo señalan que es necesario desarrollar un sistema de monitoreo y respuesta de búsqueda de palabras relacionadas con el suicidio en tiempo real en redes sociales mediante robots y mensajería instantánea.

4.5.4. *El lenguaje adolescente en las redes sociales*

Entre los artículos que estudian específicamente la conversación de adolescentes que se desarrolla en redes sociales encontramos a Matsumoto, *et al.*, (2019) y Matsumoto, Yoshida y Kita (2019) que se centran en conocer el vocabulario y expresiones cortas que los jóvenes usan en redes sociales y que no están recogidas en los diccionarios. Trabajan para diseñar modelos a partir de grupos de documentos que incluyen estas expresiones juveniles en Twitter y proponen un método de clasificación basado en la codificación de sus características. Con este modelo de agrupación de argot basado en datos

han logrado identificar 40 términos y expresiones cortas de la manera de expresarse los jóvenes en redes sociales. En concreto, Matsumoto, *et al.*, (2019) trabajan en el reconocimiento de emociones y crean un método que denominan “*bag of concepts*” con vistas a elaborar un corpus emocional a través de la recolección de textos de *weblogs* que incluyen estas nuevas palabras. Señalan que la mayoría de las palabras del argot juvenil que expresan sus sentimientos son demasiado ambiguas para ser registradas en los diccionarios.

4.5.5. Otras dimensiones de investigación

Existen otras dimensiones que ocupan también las investigaciones con SBD. A juicio de Jung, Park y Song (2016, 2017) las redes sociales contienen ingente información sobre los sentimientos, pensamientos, intereses y patrones de comportamiento de los menores que se pueden aprovechar para conocer sus tendencias hacia la depresión. En su investigación de 2017 señalan que la ontología que desarrollan resulta especialmente novedosa con respecto a análogos estudios por tres principales razones: incluyen factores de depresión adolescente antes no identificadas relacionadas, no solo con las características individuales, sino también con factores medioambientales como la familia, la escuela y la comunidad, así como el bajo rendimiento escolar, la delincuencia y el absentismo escolar. Además, consideran situaciones de riesgo, síntomas, diagnósticos, intervenciones preventivas y tratamientos de la depresión adolescente y, en tercer lugar, tienen en cuenta la jerga, las palabras de moda y los neologismos utilizados por los adolescentes, lo que hace que el análisis con SBD resulte particularmente adecuado para analizar la información que generan los adolescentes en las redes sociales.

En cuanto a las prácticas que vulneran la privacidad de los adolescentes cuando están conectados, Montgomery (2015) destaca que una de las claves del modelo de negocio de Facebook es que ha logrado datificar las relaciones sociales a través del sistema de análisis de datos “*The Social Graph*”. Esto le permite estudiar y controlar todos los movimientos que se producen en su red, elemento clave para las diferentes operaciones comerciales. Esta vigilancia comercial omnipresente no tiene precedentes en su alcance y penetración en la vida de los jóvenes. Al mismo tiempo, entiende necesario incrementar las políticas públicas para proteger a los menores de la intromisión de Facebook en sus vidas y en sus datos así como implantar leyes de privacidad para salvaguardar los excesos del BD.

Finalmente, DeJonckheere, *et al.*, (2019) integran varias técnicas de investigación con jóvenes embarazadas: una encuesta, el correspondiente análisis de mensajes de texto y el *social data mining* con procesamiento natural del lenguaje. Su objetivo es detectar las creencias y las normas sociales de estas jóvenes sobre el aumento de peso durante el periodo de embarazo. La parte del estudio que emplea SBD se orienta a las cuentas de *Facebook*, *Twitter* e *Instagram* de la muestra de jóvenes seleccionadas en 2 momentos precisos: durante sus visitas al centro médico en el primer y tercer trimestre. Por medio del análisis de los mensajes en redes sociales y de imágenes publicadas se estudiaron las creencias y declaraciones públicas sobre el embarazo, el aumento de peso, la imagen corporal y la alimentación en los meses en que las mujeres estaban embarazadas.

5. Conclusiones y discusión

Este trabajo se ha aproximado a las publicaciones sobre el comportamiento online de los menores y jóvenes en las que se han empleado técnicas vinculadas con el SBD. El análisis sobre este objeto de estudio se muestra en constante crecimiento si atendemos al número de estudios académicos encontrados, sus fechas de publicación y al volumen de citación. Los científicos están trabajando para conocer, con el análisis de grandes datos sociales, qué están haciendo los menores en la red o para crear algoritmos que permitan la identificación de tendencias de riesgo que afectan a la adolescencia.

Las dimensiones que más atención han recogido son el reconocimiento de las emociones y expresiones en las redes sociales, el *ciberbullying*, la monitorización e implantación de políticas educativas, el estudio de prácticas poco éticas que interfieren en la vida virtual de los adolescentes como la manipulación virtual, el suicidio o la depresión juvenil. Otros temas detectados son la delincuencia juvenil, la percepción del embarazo adolescente, el *sexting*, los contactos peligrosos en redes sociales, el abuso sexual a menores en Internet, la comercialización de los datos privados así como la arbitrariedad con que utilizan las distintas redes sociales la privacidad de los menores.

Se registran autores especializados en la materia que están trabajando para explotar datos de las redes sociales sobre asuntos que afectan directamente a los menores o los jóvenes. Por medio del SBD, los científicos pueden investigar con grandes cantidades de datos en tiempo real, con muestras de hasta 2.35 billones de post (Song, *et al.*, 2016) y crear algoritmos y modelos automatizados que detectan los comportamientos y sentimientos de los menores, con la capacidad de identificar tendencias.

Destaca que solo el 6,8% (n=4) de las publicaciones sean en revistas de ciencias sociales y el 13,5% (n=8) en mixtas. El resto se publican en las áreas de investigación técnicas de Informática, Computación e Ingeniería y en revistas médicas. Poniendo en valor las visiones más tecnófilas y sistemáticas que aporta la muestra estudiada de un asunto tan complejo como el comportamiento online de los seres humanos ante las multipantallas, queda pendiente la fusión y entendimiento con el conocimiento con otras áreas de conocimiento, tales como la Comunicación, la Sociología o la Educación.

En este trabajo se ha optado por una búsqueda centrada en el SBD. No obstante, se constata cierta discordancia para denominar los estudios y las técnicas relativas a este ámbito. La mayoría de los autores titulan sus investigaciones e incluyen en sus palabras clave el término SBD o BD cuando están investigando las redes sociales y emplean las técnicas de investigación como el *data mining*, *machine learning*, el análisis semántico o análisis de sentimiento en redes sociales. Pero lo hacen en realidad sin especificar que están aplicando una metodología concreta SBD o BD.

Sospechamos un incremento en el número de estudios que aborden la vida virtual de los menores con estas técnicas de investigación puesto que todavía se manifiesta una brecha entre su potencial y su realización. Ciertamente, el SBD solo es una forma de aproximación (Malvicino & Yoguel, 2014) y los datos por sí solos no suponen una solución mágica para el análisis de la realidad social. No se debe caer en el reduccionismo cuantitativo (Sánchez-Bayón, 2017) y abandonar los aspectos culturales, entre otros, a la hora de investigar los movimientos sociales. Los científicos deben acercarse a los grandes datos con cautela para garantizar el progreso social (Kosinski & Behrend, 2017) y, sin duda, los científicos seguirán siendo necesarios para crear la teoría y proporcionar la legitimidad que la ciencia necesita como entidad social normativa (Sætra, 2018).

Esta investigación presenta dos limitaciones. Por una parte, no incorpora todas las posibles bases de datos a las que se puede acudir para localizar literatura académica en este ámbito. Por otra, la estrategia de búsqueda por etiquetas siempre puede no cubrir algún aspecto que pudiera ser de interés.

Financiación

Este trabajo ha recibido financiación de los proyectos “Nuevos escenarios de vulnerabilidad digital: alfabetización mediática para una sociedad inclusiva” (PROVULDIG-2CM) (ref. H2019/HUM5775), financiado por la CAM y el Fondo Social Europeo y del Proyecto del plan nacional I+D+i “Redes sociales, adolescentes y jóvenes: convergencia de medios y cultura digital” (CSO2016-74980-C2-2-R).

6. Referencias bibliográficas

- Alsehaima, A. O., & Alanazi, A. A. (2018). Psychological and social risks to children of using the internet: Literature Review. *Journal of Child Adolescent Behaviour*, 6(380), 2. <http://dx.doi.org/10.4172/2375-4494.1000380>
- Aparici, R., & García Matilla, A. (2016). ¿Qué ha ocurrido con la educación en comunicación en los últimos 35 años?: pensar el futuro. *Espacios en Blanco. Revista de Educación (Serie Indagaciones)*. <https://bit.ly/2Twn6ko>
- Arcila-Calderón, C., Barbosa-Caro, E., & Cabezuelo-Lorenzo, F. (2016). Técnicas BD: análisis de textos a gran escala para la investigación científica y periodística. *El profesional de la información*, 25(4), 623-631. <https://doi.org/10.3145/epi.2016.jul.12>
- Arellano Toledo, W. (2014). Gobierno abierto y privacidad: la problemática del Big data y el cómputo en la nube. *Virtualis*, 5(10), 34-59. <https://bit.ly/2FADr3y>
- Batty, M. (2013). Big Data, smart cities and city planning. *Dialogues in Human Geography*, 3(3), 274-279. <http://dx.doi.org/10.1177/2043820613513390>
- Bell, G., Hey, T. & Szalay, A. (2009). Beyond the data deluge. *Science*, 323, 1297-1298. <http://dx.doi.org/10.1126/science.1170411>.
- Bello-Orgaz, G., Jung, J. J., & Camacho, D. (2016). Social big data: Recent achievements and new challenges. *Information Fusion*, 28, 45-59. <https://doi.org/10.1016/j.inffus.2015.08.005>
- Boyd, D. & Crawford, K. (2012). Critical Questions for Big Data. Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication & Society*, 15(5), 662-679. <https://doi.org/10.1080/1369118X.2012.678878>
- Centro de Estudios Cervantinos (2020). *Cuáles son las ramas de las ciencias sociales*. <https://www.centroestudioscervantinos.es/ramas-de-las-ciencias-sociales/>
- Cerezo Gilarranz, J. (2015). Big Data, la nueva ciencia del siglo XXI. En: Pérez-Hernández y Torra (eds). *Big Data. Revista de Ciencias y Humanidades de la Fundación Ramón Areces*, 14, 7-11. <https://bit.ly/3hsHTyA>

DeJonckheere, M., Nichols, L. P., Vydiswaran, V. V., Zhao, X., Collins-Thompson, K., Resnicow, K., & Chang, T. (2019). Using text messaging, social media, and interviews to understand what pregnant youth think about weight gain during pregnancy. *JMIR formative research*, 3(2): e11397. <https://doi.org/10.2196/11397>

Díaz-Campo, J. (2015). Análisis bibliométrico de las tesis doctorales sobre Ética de los Medios de Comunicación presentadas en España (1979-2013). *Doxa Comunicación*, 65-88. <https://doi.org/10.31921/doxacom.n20a3>

Falzone, A. E., Brindis, C. D., Chren, M. M., Junn, A., Pagoto, S., Wehner, M., & Linos, E. (2017). Teens, tweets, and tanning beds: rethinking the use of social media for skin cancer prevention. *American journal of preventive medicine*, 53(3), S86-S94. <https://doi.org/10.1016/j.amepre.2017.04.027>

Fondo de las Naciones Unidas para la Infancia (UNICEF, 2019). *Barómetro de Opinión de Infancia y Adolescencia 2019*. <https://www.unicef.es/publicacion/que-opinan-los-ninos-y-las-ninas>

Giménez, A. M., Luengo, J. A., & Bartrina, M. J. (2017). ¿Qué hacen los menores en internet? Usos de las TIC, estrategias de supervisión parental y exposición a riesgos. *Electronic Journal of Research in Education Psychology*, 15(3), 533-552. <http://dx.doi.org/10.14204/ejrep.43.16123>

Gualda, E., & Rebollo, C. (2020). Big Data y Twitter para el estudio de procesos migratorios: Métodos, técnicas de investigación y software. *Empiria. Revista de metodología de ciencias sociales*, 46, 147-177. <https://doi.org/10.5944/empiria.46.2020.26970>

Han, Y., Kim, H., Song, J., & Song, T. M. (2019). Ontology development of school bullying for social big data collection and analysis. *The Journal of the Korea Contents Association*, 19(6), 10-23. <https://doi.org/10.3390/ijerph16142596>

Hernández-Leal, E. J., Duque-Méndez, N. D., & Moreno-Cadavid, J. (2017). Big Data: una exploración de investigaciones, tecnologías y casos de aplicación. *TecnoLógicas*, 20(39), 17-24. <http://www.scielo.org.co/pdf/teclo/v20n39/v20n39a02.pdf>

Hey, T., Tansley, S., & Tolle, K. (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Washington: Microsoft Research.

Jiménez, E., Garmendia, M., & Casado, M. A. (2018). *Entre selfies y whatsapps. Oportunidades y riesgos para la infancia y la adolescencia conectada*. Barcelona: Gedisa.

Joyanes Aguilar, L. (2016). *Big Data, Análisis de grandes volúmenes de datos en organizaciones*. Alfaomega Grupo Editor.

Jung, H., Park, HA., & Song, TM. (2016). Development and Evaluation of an Adolescents' Depression Ontology for an Analysing Social Data. En: *SERMEUS, et al., (Eds.) Nursing Informatics 2016: EHealth for All: Every Level Collaboration-From Project to Realization*, 225-442. Amsterdam: IOS Press BW.. <https://doi.org/10.3233/978-1-61499-658-3-442>

Jung, H., Park, HA., & Song, TM. (2017). Ontology-based approach to social data sentiment analysis: detection of adolescent depression signals. *Journal of medical internet research*, 19(7), e259. <https://doi.org/10.2196/jmir.7452>

- Kim, H., Han, Y., Song, J., & Song, T. M. (2019). Application of social big data to identify trends of school bullying forms in South Korea. *International journal of environmental research and public health*, 16(14), 2596. <https://doi.org/10.3390/ijerph16142596>
- Kosinski, M., & Behrend, T. (2017). Editorial overview: Big data in the behavioral sciences. *Current Opinion in Behavioral Sciences*, 18, iv-vi. <https://doi.org/10.1016/j.cobeha.2017.11.007>
- Kumar A., Sangwan S.R., Nayyar A. (2020) Multimedia Social Big Data: Mining. In: *Tanwar S., Tyagi S., Kumar N. (eds) Multimedia Big Data Computing for IoT Applications*. Intelligent Systems Reference Library, 163. Springer, Singapore. https://doi.org/10.1007/978-981-13-8759-3_11
- Lara Palma, A. M., & Giacinto, R. (2017). Ethical Machines against Unethical Computer-mediated Social Interactions. *Journal of Information and Management*, 37(2), 22-38. https://doi.org/10.20627/jsim.37.2_22
- Leung, X.Y., Sun, J., & Bai, B. (2017). Bibliometrics of social media research: A co-citation and co-word analysis. *International Journal of Hospitality Management*, 66, 35-45. <https://doi.org/10.1016/j.ijhm.2017.06.012>
- Livingstone, S., Mascheroni, G., & Staksrud, E. (2018). European research on children's internet use: Assessing the past and anticipating the future. *New Media & Society*, 20(3), 1103-1122 <https://doi.org/10.1177/1461444816685930>
- López Cantos, F. J. (2018). *Cultura visual y conocimiento científico: Comunicación transmedia de la ciencia en la era Big Data*. Barcelona: Editorial UOC.
- Malvicino, F., & Yoguel, G. (2014). Big Data. Avances recientes a nivel internacional y perspectivas para el desarrollo local. *Documento de Trabajo. Centro Interdisciplinario de Estudios en Ciencia Tecnología e Innovación (CIECTI-MinCyT)*. Buenos Aires. <https://bit.ly/3cFeizD>
- Manovich, L. (2011). Trending: The promises and the challenges of big social data. En *Gold (eds) Debates in the digital humanities*, 460-475. London: University Minnesota Press.
- Martínez-Martínez, S., & Lara-Navarra, P. (2014). El big data transforma la interpretación de los medios sociales. *El profesional de la información*, 23(6), 575-581. <https://doi.org/10.3145/epi.2014.nov.03>
- Matsumoto, K., Ren, E., Matsuoka, M., Yoshida, M., & Kita, K. (2019). Slang feature extraction by analysing topic change on social media. *CAAI Transactions on Intelligence Technology*, 4(1), 64-71. <https://doi.org/10.1049/trit.2018.1060>
- Matsumoto, K., Yoshida, M., & Kita, K. (2019). Emotion Recognition for Japanese Short Sentences Including Slangs Based on Bag of Concepts Feature Trained by Large Web Text. *Current Analysis on Instrumentation and Control*, 2019(2), 9-18. <https://bit.ly/2XeP6dk>
- Mayer-Schönberger, V. & Cukier, K. (2013). *Big Data. La revolución de los datos masivos*. London: Turner.
- Montgomery, K. C. (2015). Youth and surveillance in the Facebook era: Policy interventions and social implications. *Telecommunications Policy*, 39(9), 771-786. <http://dx.doi.org/10.1016/j.telpol.2014.12.006>

Nupairoj, N. (2016). El ecosistema de la alfabetización mediática: Un enfoque integral y sistemático para divulgar la educomunicación. *Comunicar*, 24(49), 29-37. <https://doi.org/10.3916/C49-2016-03>

O'Neil C. (2013). *The rise of big data, big brother*. <https://bit.ly/3eR7Jvi>

Paredes-Moreno, A. (2015). Big Data: Estado de la cuestión. *International Journal of Information Systems and Software Engineering for Big Companies (IJISEBC)*, 2(1), 38-59.

Pérez, M. (2015). *BIG DATA-Técnicas, herramientas y aplicaciones*. Alfaomega Grupo Editor.

Pérez, M. J. (2016). Davos y la cuarta revolución industrial. *Nueva Revista*, 157. <https://bit.ly/3wIUlP>

Pybus, J., Coté, M., & Blanke, T. (2015). Hacking the social life of big data. *Big Data & Society*, 2(2), 1-10. <https://doi.org/10.1177/2053951715616649>

Qin, S. J. (2014). Process data analytics in the era of big data. *AIChE Journal*, 60(9), 3092-3100. <https://doi.org/10.1002/aic.14523>

Ramírez-Montoya, M., & García-Peñalvo, F. (2018). Co-creation and open innovation: Systematic literature review. *Comunicar. Media Education Research Journal*, 54, 9-18. <https://doi.org/10.3916/C54-2018-01>

Russell, M. A. (2013). *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More*. Sebastopol: O'Reilly Media, Inc.

Sætra, H. S. (2018). Science as a vocation in the era of big data: The philosophy of science behind big data and humanity's continued part in science. *Integrative Psychological and Behavioral Science*, 52, 508-522. <https://doi.org/10.1007/s12124-018-9447-5>

Sánchez-Bayón, A. (2017). Apuntes para una teoría crítica humanista y su praxis económico-empresarial en la posglobalización. *Miscelánea Comillas. Revista de Ciencias Humanas y Sociales*, 75(147), 305-329. <https://bit.ly/2RtY188>

Schwab, K. (2017). *The fourth industrial revolution*. London: Currency.

Shuijing, H. (2015). Big Data Mining: Essential Technologies and Concerns. In *Proceedings of the 2015 Sixth International Conference on Digital Manufacturing and Automation*. 88-92. <https://bit.ly/3dA77dd>

Song, J., Song, T-M., & Lee, J. R. (2018). Stay alert: Forecasting the risks of sexting in Korea using social big data. *Computers in Human Behavior*, 81, 294-302. <https://doi.org/10.1016/j.chb.2017.12.035>

Song, J., Song, T-M., Seo, D. C., & Jin, J. H. (2016). Data mining of web-based documents on social networking sites that included suicide-related words among Korean adolescents. *Journal of Adolescent Health*, 59(6), 668-673. <http://dx.doi.org/10.1016/j.jadohealth.2016.07.025>

Song, T., & Song, J. (2015). *Social Big Data Analysis and Utilization Methodologies-With Special Reference to Forecasting the dangers of sexting in Korea using social big data*. <https://bit.ly/2Zbpq2d>

- Song, T., Song, J., An, J. Y., & Woo, J. M. (2014). *Social Risk Factor Prediction Utilizing Social Big. Kihasa*. <https://bit.ly/3c4sq6D>
- Tapia Nava, E. (2018). *El uso del Big Data en los estudios de opinión pública*. Instituto Belisario Domínguez Dirección General de Análisis Legislativo. <https://bit.ly/2z7HUb4>
- Tarwani, N., Chorasias, U., & Shukla, P. K. (2017). Survey of Cyberbullying Detection on Social Media Big-Data. *International Journal of Advanced Research in Computer Science*, 8(5), 831-835.
- Taylor-Sakya, K. (2016). *Big data: Understanding big data*. arXiv preprint arXiv:1601.04602.
- Törnberg, P., & Törnberg, A. (2018). The limits of computation: A philosophical critique of contemporary Big Data research. *Big Data & Society*, 5(2), 1-12. <https://doi.org/10.1177/2053951718811843>
- Uman, I. (2018). Big Data y Memoria Digital: claves para su exploración e investigación desde las ciencias sociales. *Avatares de la Comunicación y la Cultura*, 15. <https://bit.ly/3gRtOeR>
- UN Global Pulse, IAPP. (2018). Building Ethics into Privacy Frameworks for Big Data and AI. <https://bit.ly/3sYaySp>
- van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & society*, 12(2), 197-208. <https://doi.org/10.24908/ss.v12i2.4776>
- van Eck, N. J., & Waltman, L. (2014). Visualizing bibliometric networks. In: *Ding Y., Rousseau R., Wolfram D. (eds) Measuring Scholarly Impact*, 285-320. Springer, Cham. https://doi.org/10.1007/978-3-319-10377-8_13
- Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: *Proceedings of the 18th international conference on evaluation and assessment in software engineering*, 38, 1-10. <https://doi.org/10.1145/2601248.2601268>

