



Big Data y Business Intelligence en Twitter e Instagram para la inclusión digital

Big Data and Business Intelligence on Twitter and Instagram for digital inclusion

- Carlos Barroso-Moreno. Investigador Pre-doctoral, Departamento de Estudios Educativos, Universidad Complutense de Madrid (España) (carbarro@ucm.es) (<https://orcid.org/0000-0002-1609-2267>)
- Dra. Laura Rayon-Rumayor. Profesora Titular, Departamento de Estudios Educativos, Universidad Complutense de Madrid (España) (larayon@ucm.es) (<https://orcid.org/0000-0003-0339-8221>)
- Dr. Antonio Bautista García-Vera. Catedrático, Departamento de Estudios Educativos, Universidad Complutense de Madrid (España) (bautista@edu.ucm.es) (<https://orcid.org/0000-0002-5194-7419>)

RESUMEN

Las redes sociales pueden contribuir a una sociedad inclusiva, pero también son espacios de comunicación asimétricos y polarizados. Ello requiere de un profesorado competente para la construcción de una ciudadanía digital crítica. Este artículo tiene un doble objetivo, presentar las técnicas «Web Scraping» y «Text Analytics» como herramientas que definen competencias digitales docentes, e indagar sobre qué publicaciones, en Twitter e Instagram, son más virales en relación con educación, discapacidad e inclusión. Se analizaron 48.991 publicaciones en español e inglés, correspondientes al periodo del 13 de octubre de 2021 al 1 de mayo de 2022. Se seleccionaron las 100 más virales, e identificaron las correlaciones entre el sentimiento, género e influencia asociado al contenido, su espacio temporal y geográfico. Los resultados evidencian que los grupos de influencia económica y política son los más virales, relegando a un segundo plano a las organizaciones sin ánimo de lucro o particulares con difusión altruista; solo en los días internacionales se invierte esta tendencia. Los «bots» no interfieren para imponer mensajes; son los algoritmos de inteligencia artificial los que opacan contenido reivindicativo y humanístico. Las personas más influyentes tienen una prevalencia de género masculino asociadas a cuentas institucionales de ámbito político. Se concluye que las herramientas de «Big Data» y «Business Intelligence» ayudan al profesorado a analizar temas educativos y sociales relevantes, y a adquirir una ética colectiva frente a los nuevos retos educativos.

ABSTRACT

Social media can contribute to an inclusive society, but they are also asymmetrical and polarised communication spaces. This requires competent teachers to build critical digital citizenship. The aim of this article is twofold: to present Web Scraping and Text Analytics as tools that define teachers' digital competences, and to investigate which posts on Twitter and Instagram are most viral in relation to education, disability and inclusion. A total of 48,991 publications in Spanish and English were analysed, corresponding to the period from 13 October 2021 to 1 May 2022. The 100 most viral posts were selected, and correlations were identified between the sentiment, gender and influence associated with the content, its temporal and geographic space. The results show that economic and political influence groups are the most viral, relegating non-profit organisations or individuals with altruistic outreach to second place; only on international days is this trend reversed. Bots do not interfere to impose messages; it is artificial intelligence algorithms that overshadow vindictive and humanistic content. The most influential people are predominantly male, associated with institutional accounts in the political sphere. It is concluded that Big Data and Business Intelligence tools help teachers to analyse relevant educational and social issues, and to acquire a collective ethic in the face of new educational challenges.

PALABRAS CLAVE | KEYWORDS

Análisis de redes sociales, Big Data, educación, discapacidad, inclusión digital, grupos de influencia.
Social network analysis, Big Data, education, disability, digital inclusion, influence groups.



1. Introducción y estado de la cuestión

Informes actuales (We are social, 2021; Ditrencia, 2020) señalan el incremento exponencial de usuarios conectados a las redes sociales (RRSS) a nivel mundial, siendo estos no solo personas individuales, sino también grupos y medios profesionales e institucionales que impactan en la construcción de la realidad y en cómo se comparten significados (Dellwing, 2021; Del-Fresno-García, 2014, 2019; Ladogina et al., 2020).

En cuanto a los estudios sobre las funciones, roles y relaciones entre los miembros que las integran (Awidi et al., 2019; Brunner et al., 2019; De-Groot et al., 2022; Grace et al., 2019; Tuzel & Hobbs, 2017; White & Forrester-Jones, 2020), nos preocupa de forma especial el enfoque que enfatiza las RRSS como espacios asimétricos de comunicación. Concretamente, Barberá (2015), Barberá et al. (2015) y Brady et al. (2019) señalan que los usuarios comparten mensajes que representan creencias, opiniones y valores que respaldan y siguen a perfiles en los que confían ideológicamente. Otro factor relevante que condiciona la comunicación es el lenguaje moral-emocional que los líderes políticos emplean y su efecto de contagio moral, así como la asimetría en la comunicación entre los creadores de contenidos y sus potenciales seguidores, según el contenido de los mensajes. Una asimetría amplificada por los «bots» (Robles et al., 2022). No hay duda del potencial de Twitter para que determinados usuarios y grupos, como las élites políticas y corporaciones, lleguen a grandes audiencias, potenciales votantes y consumidores, a través de vínculos directos e indirectos de influencia que definen la interacción en esta red, un factor predictivo de la influencia e impacto social fuera de la misma (Brady et al., 2019). En este contexto, adquieren especial relevancia los discursos de odio que están generando una importante polarización entre las personas en función de su ideología, con efectos en el «sesgo de percepción selectiva» que favorece la valoración positiva del mensaje de los emisores con los que existe afinidad ideológica y el rechazo de los discursos de ideología contraria. Un importante factor predictivo para el comportamiento ante determinados colectivos fuera de línea, que son representados mediante mensajes excluyentes y antidemocráticos en las RRSS (Ortega-Sánchez et al., 2021).

Para afrontar esta preocupación, entendemos que el profesorado tiene un papel clave en la construcción de una ciudadanía digital con capacidad de participar de un modo informado y responsable, y contribuir así a una sociedad en red democrática e inclusiva (Bautista, 2021; Carlsson, 2019; Ortega-Sánchez et al., 2021). Como evidencian Tuzel y Hoobs (2017: 64), el uso de redes sociales para una ciudadanía intercultural requiere que el profesorado tenga una «apreciación sólida de las asimetrías y desigualdades inherentes a los flujos de información», y una comprensión del funcionamiento de estas plataformas digitales como espacios de divulgación, amplificación de ideas y movilización de acciones para colectivos y personas en desigualdad de derechos, como aquellas que tienen una discapacidad, con el fin de difundir y criticar ideas, así como divulgar sus logros en pro de una sociedad más inclusiva (Hemsley et al., 2018). Entonces, ¿cómo se han de trabajar estas competencias digitales que preparen al profesorado para una educación inclusiva? ¿Qué herramientas y procedimientos le ayudarán en la compleja y hermosa tarea de conocer los significados, creencias y actitudes que circulan por redes sociales con miles y miles de participantes? Entendemos que una forma adecuada de indagar y generar ese conocimiento que debe tener el profesorado para promover relaciones que lleven a crear sentimientos de inclusión y pertenencia del alumnado al grupo de referencia, es mediante el rastreo de redes sociales con técnicas y herramientas como las de «Web Scrapping» y «Text Analytics» que presentamos en este artículo.

Ante las anteriores preocupaciones y propuestas derivadas de la revisión del estado de la cuestión, este artículo tiene una doble finalidad. Una, ejemplificar en un contexto de investigación el uso de estas herramientas que conforma una de las competencias digitales docentes para alentar el debate, e inspirar y alumbrar evidencias del valor de las técnicas señaladas en el análisis de las RRSS y la visualización interactiva de datos. La otra, responder a tres cuestiones/hipótesis sobre los procesos de relación asimétrica que ayuden a entender mejor los procesos de remodelación y relegación de estos temas en las RRSS, un conocimiento útil para definir el contenido de la competencia digital docente, concretamente:

- H1: La influencia de los grupos de poder económico y político junto con un interés de las plataformas digitales remodelan las temáticas asociadas a la educación, discapacidad e inclusión en función de sus propios intereses.

- H2: Los algoritmos de inteligencia artificial relegan a un segundo o tercer plano las ideas minoritarias o difusiones altruistas.
- H3: Las publicaciones con sentimientos negativos están asociadas a grupos y líderes políticos, frente a las publicaciones positivas vinculadas a asociaciones y creadores de contenido individuales en defensa de los derechos humanos de las personas con discapacidad, sin influencia por género.

Una aportación innovadora del artículo es el acceso a los datos y resultados obtenidos, que se ponen a disposición de investigadores, profesorado y otros profesionales para que puedan realizar una lectura dinámica e interactiva de los mismos.

2. Material y métodos

En este artículo analizamos qué publicaciones en Twitter e Instagram son más virales en relación con la educación y la discapacidad e inclusión y, en este sentido, nos permitirán identificar mejor las correlaciones entre el contenido, sentimiento, género, influencia, y su espacio temporal y geográfico. La base de datos se compone de publicaciones descargadas de ambas redes sociales sobre contenido de educación relacionado con la inclusión o discapacidad durante 200 días. Se aplican técnicas de «Social Big Data Analysis», como técnicas de «Web Scrapping» para extraer la información y analizarla con algoritmos de «Big Data» y «Text Mining». Posteriormente, la representación de los resultados emplea la herramienta de negocio Power BI, permitiendo interactuar de forma dinámica a los lectores.

2.1. Flujo de procesos en la metodología

El conjunto de procesos de la investigación, desde el origen de los datos hasta la representación gráfica, conceptualmente se referencia como la herramienta «Social Networks Tools», la cual está representada en la Figura 1. Los procesos se dividen en dos grandes bloques, los delimitados en color azul son modificables por el investigador como es el origen de los datos o las palabras clave empleadas. El bloque delimitado con líneas discontinuas rojas referencia a técnicas de «Web Scrapping», algoritmos de «Big Data», «Data Mining» y parte de la herramienta de «Business Intelligence», propias del grupo de investigación con propiedad intelectual, en la cual los interesados no pueden realizar modificaciones ni disponer de acceso total.



El primer proceso corresponde a la elección de las redes sociales con posibilidad de extracción de los datos como Twitter, Instagram, YouTube, Tik Tok o Facebook. En este mismo punto, se seleccionan las palabras clave necesarias para considerar el almacenamiento de la información. El segundo proceso aplica técnicas de «Web Scrapping» para la extracción de datos en las RRSS con Python, que literalmente significa arañar de la web (Mitchell, 2018). Posteriormente, con la información obtenida se aplican técnicas de «Big Data» para procesar la gran cantidad de datos recopilada, además, se emplea minería de datos para

analizar textos de las publicaciones. A continuación, se incluyen variables específicas de la investigación sobre el perfil de comunicación como género, actividad profesional, geolocalización y espacio temporal. El proceso siguiente es la carga de los datos en el servicio de análisis de datos de Microsoft Power BI con el fin de proporcionar visualizaciones interactivas para que los investigadores y el profesorado puedan generar sus propios informes y paneles (Becker & Gould, 2019). Por último, el usuario interpreta los gráficos generados para la comprensión de la temática abordada, en busca de un uso científico y/o educativo con un impacto real en la sociedad. Los procesos de Extraer, Transformar y Cargar (ETL) permiten mezclar datos de múltiples fuentes, en nuestro caso particular, se extraen datos de Twitter e Instagram, se transforman mediante «script» de Python con una serie de requisitos y se cargan en un «data warehouse» para canalizar los datos de manera homogénea y analizarlos mediante diferentes algoritmos.

2.2. Extracción de datos

Los datos de Twitter se descargan de la API de Twitter mediante la librería «tweepy» y los datos de Instagram mediante algoritmos privados desarrollados específicamente para la investigación, en ambos casos, se recopila información en tiempo real con el lenguaje de programación de Python. Dada la necesidad de capturar todo el tráfico en tiempo real, a diferencia de la mayoría de los estudios previos, empleamos publicaciones que tengan las palabras clave demandadas, no nos centramos en lo que dice un grupo determinado de perfiles para evitar sesgos. Esta situación conlleva otra implicación, la captura de datos se inicia en el momento de la activación de dichas palabras, no se recopilan datos con carácter retroactivo.

La base de datos generada se compone de las publicaciones en el periodo del 13 de octubre de 2021 al 1 de mayo de 2022, con un total de 48.991 publicaciones en la escena hispanohablante y anglosajona. El requisito es contener una serie de palabras clave independientemente del usuario que las difunda. La palabra clave común empleada es educación, y al menos una palabra de discapacidad o inclusión, así como palabras derivadas de las mismas. Se incluyen, de manera adicional, las palabras sin tilde debido a la viralidad de algunas noticias con dicha ortografía. Por ello, toda publicación dispone de las palabras educación o «education» junto con una o varias palabras de discapacidad, inclusión, «disability» o «inclusion» en la red social de Twitter o Instagram. La elección de la palabra educación junto con discapacidad o inclusión es motivada porque determinadas publicaciones que abordan la discapacidad no hacen referencia a este término, sino a la idea de la inclusión, además, la inclusión sí refleja diferentes sensibilidades sociales entorno a la temática abordada. Las publicaciones son actualizadas con toda la información necesaria una semana posterior a la publicación, esta situación no afecta significativamente a los datos porque las publicaciones tienen impacto significativo los primeros días de la difusión, comprobado de manera empírica. De hecho, el lector puede corroborar en posteriores apartados cómo el número de «likes, followers, followings o comment» no aumenta de forma relevante, por consiguiente, dicha situación no influye en el análisis posterior.

2.3. Transformación de datos

Una vez almacenadas en bruto las publicaciones, se procedió a realizar la minería de datos con algoritmos de Procesamiento del Lenguaje Natural de las librerías NLTK y Scikit-Learn de Python, una rama de la inteligencia artificial para determinar la interacción entre ordenadores y ser humano (Cheng & Tsai, 2019). En este punto, se diferencian dos transformaciones: manual y automática. Las transformaciones automáticas se aplican a las 48.991 publicaciones de la investigación; las transformaciones manuales se aplican por el equipo investigador a las 100 publicaciones de mayor interés con clasificaciones particulares para la temática abordada. Todas las variables disponibles en la base de datos se encuentran detalladas por grupo, tipo, categoría y ejemplo (Tabla 1).

2.3.1. Automática

La aplicación del análisis del texto o «text mining» comienza con el proceso de «tokenización» del contenido, que permite separar las palabras por los espacios que componen la frase. A continuación, se eliminan las palabras denominadas «stopwords» compuestas por preposiciones, determinantes o palabras particulares, entre otras. Esta separación y limpieza del texto permite analizar la frecuencia de repetición,

nubes de palabras etc. En lo concerniente al análisis de sentimientos se aplica el diccionario de Liu (2010) para detectar el contenido positivo, negativo o neutro de las palabras, aportando un valor final a la frase en su conjunto. Entre las limitaciones de este tipo de diccionarios se encuentran la invisibilidad o la clasificación errónea de las ironías o juegos de palabras. En referencia a la detección de temáticas se agrupan por los «hashtags» para el contenido de Instagram, en función de las palabras más frecuentes se clasifica los tweets para agruparse en alguno de los «hashtags». Atendiendo a la detección de «bots» se emplea la API de Botometer para los perfiles de Twitter y extraer más de 1.200 características como patrones de actividad, lenguaje, sentimientos, estructura social o amigos, asignando un 1 si es un «bot» o un 0 si la cuenta es real. Se establecen criterios elevados para considerarlo un «bot» y no incluir usuarios reales en dicha categoría, por ello, se establece una validación cruzada para el «Area Under the ROC Curve» (AUC) de 0,99.

2.3.2. Manual

Para contestar a las hipótesis planteadas es necesario generar variables concretas, las cuales no son posible automatizar. El equipo de investigación se distribuye el etiquetado con criterios homogéneos para garantizar una correcta clasificación. Esto exige un análisis minucioso de los perfiles de usuario, basado en el rastreo de información de difusión pública en otras redes o blogs personales para determinar variables como: género, ámbito, profesión, multimodalidad o ubicaciones estimadas.

2.4. Carga de datos

Para analizar los datos una vez se dispone del «data mark», denominado educación, con todas las variables limpias se cargan los datos al programa Power BI. El objetivo es analizar los datos e indicadores para contrastar hipótesis a través de diferentes análisis de datos de tipo visual. Algunos de los gráficos abordan estadística simple y compleja, incluyen mapas (geográficos o de calor) sobre las ubicaciones de los emisores más virales (Arcila-Calderón et al., 2022), correlación estadística para analizar si existe una relación fuerte o débil, grafos multidimensionales, series temporales o nubes de palabras entre otras.

Tabla 1. Descripción de las variables, tipo y categorías del Data Mart educación en RRSS				
Grupo de variables	Variables	Categorías	Tipo	Ejemplos
Identificador	Identificador	-	N Numérica	02186348225, 31186048213...
	Etiqueta	3	C Categórica	T asociado al Top 10 V asociado a viral 10 al 50 P asociado a publicaciones 50 a 100
	Nombre del perfil	-	T Texto	Mzavalagc
Datos transaccionales	Nombre real	-	T Texto	Margarita Zavala
	Origen	-	U URL	www.instagram.com/p/CZKkN4lp13G/
Temporal	Dominio	2	C Categórica	Twitter o Instagram
	Día	200	F Fecha	2022-01-24
Contenido	Hora	1.440	F Fecha	21:32
	Título	-	T Texto	La inclusión real
	Contenido	-	T Texto	La inclusión se logra con...
Viralidad	Hashtags	-	T Texto	#educacion #education #inclusión
	Retweets (twitter)	-	N Numérica	1957, 34...
	Comment (Instagram)	-	N Numérica	33.971, 4.664...
	Likes	-	N Numérica	139.103, 83.400...
	Reproducciones (Video)	-	N Numérica	826, 453...
	Following	-	N Numérica	1.596.670, 320...
Comportamiento	Follower	-	N Numérica	6.068, 234...
	Publicaciones (Instagram)	-	N Numérica	6.068, 234...
	Sentimiento	3	C Categórica	Positivo, neutro y negativo
Ámbito	Profesión	>10	C Categórica	Político, creador digital, docente...
	Ámbito	3	C Categórica	1. Institucional 2. Grupal 3. Individual
Multimodalidad	Subámbito	9	C Categórica	11. Local, 12. Supranacional (política) 13. Medios de comunicación 21. Academias, clínicas, o empresas 22. Asociaciones sin ánimo de lucro 23. Divulgadores de ciencia 31. Profesional 32. Profesional con discapacidad 33. Particular
	Modal	2	C Categórica	Monomodal y multimodal
Demográfico	Submultimodal	4	C Categórica	Texto y video; texto e imagen; texto audio y audio e imagen
	Género	3	C Categórica	Masculino, femenino y neutro
	Ubicación de la persona estimada	195	C Categoría	Estados Unidos, España

3. Análisis y resultados

La publicación de los datos se encuentra en la herramienta de Power Bi, debiéndose abrir preferiblemente con el navegador Microsoft Edge desde un ordenador (<https://bit.ly/3z9wDU6>), o con el Código QR, presentado en la Figura 2.



Extraemos una captura inicial de las publicaciones (N=48.991), que se distribuyen en Twitter con un 59,38% (N=29.095) e Instagram con un 40,61% (N= 19.896). La mayor difusión se produce en Twitter, sin embargo, la de mayor virilidad en función del número de «likes» y comentarios en Instagram.

Los resultados al analizar las publicaciones más virales, sentimiento y el tipo de perfil asociado evidencian cinco ideas principales que se presentan en la Figura 3. (i) Las publicaciones con sentimientos polarizados son las más virales, en concreto las positivas, de las 10 más virales 9 son positivas. (ii) Las 25 publicaciones más virales son campañas de grupos de poder relevantes, orquestadas por multinacionales, políticos, «influencer» o creadores digitales con interés económico, mayoritariamente. Prueba de ello son las comunicaciones más virales [T1] (<https://bit.ly/3z5bEBO>), [T4] y [V11] relacionadas con McDonald's, ubicadas en cuentas que no reivindican ningún aspecto social o educativo en relación con la discapacidad e inclusión, publicadas en la misma fecha, con la misma imagen y contenido de texto. Para orquestar una campaña de tanto impacto se requieren cuentas de miles de seguidores, aunque tenga una tasa de difusión baja por cada seguidor el impacto total es muy elevado. (iii) Tenemos que relegarnos a la posición 6 [T6] y 28 [V28] por número de «likes» para encontrar un espacio de comunicación social.

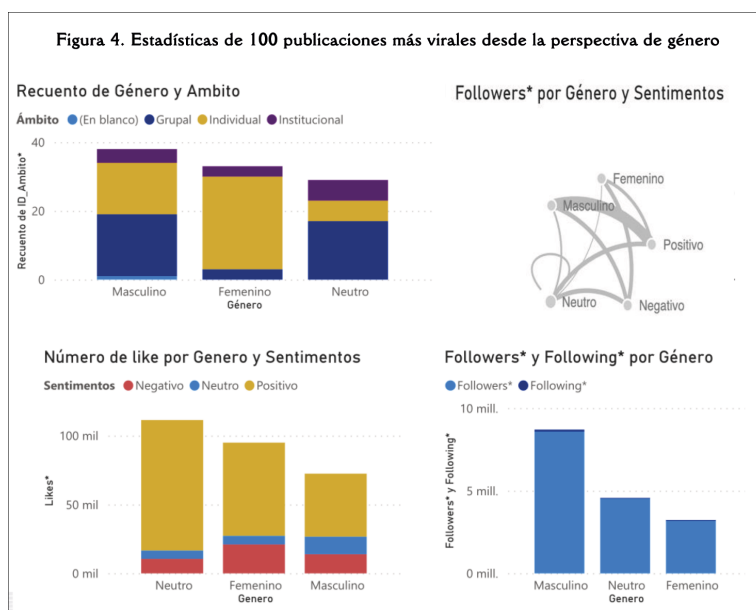
No obstante, el [T6] se define como un espacio para aprender de diversidad y LGBTQ+, sin embargo, cierto contenido exclusivo requiere un desembolso económico (<https://bit.ly/3N0Vv4G>). Hay que descender a la posición 34 [T34] para encontrar la Fundación «Solo Faltas Tú», una Organización No Gubernamental (ONG) sin ánimo de lucro para la concienciación sobre el espectro autista (<https://bit.ly/3t2Gpno>). (iv) Los casos de denuncia de situaciones de exclusión son casi inexistentes y no disponen de recorrido en las redes sociales, debido principalmente a la denuncia desde redes con poco número de seguidores, y un sentimiento neutro e institucional de las denuncias. (v) Mayoritariamente, lo más difundido tiene carácter positivo o negativo, pero los mensajes neutros no son tan difundidos. A la luz de estos resultados podemos afirmar que se respalda la hipótesis (H1), determinados grupos y líderes políticos remodelan las temáticas analizadas para sus propios intereses, relegando a un segundo plano a colectivos y actores individuales con intereses altruistas.

Los resultados al analizar las 100 publicaciones más virales desde la perspectiva de género se representan en la Figura 4, se distribuyen del siguiente modo: el 38% hombres, 33% mujeres y el 29% neutro. Los resultados significativos son los siguientes: (i) Las cuentas de género masculino disponen de más seguidores y sus comentarios son más broncos (negativos) que las del género femenino, estas últimas quedan asociadas a comentarios positivos.

En datos, el género masculino tiene «followers» 8.582.921, «following» 118.981; el género femenino tiene «followers» 4.523.345 y «following» 43.307. (ii) Las cuentas de género femenino tienen menos seguidores, pero más «likes». (iii) Las cuentas de instituciones difunden más neutralidad o mensajes positivos. (iv) El género masculino es el que más cuentas institucionales dispone en RRSS. Por ello, la hipótesis (H3) se confirma solo en parte, los sentimientos negativos están vinculados a grupos de influencia y líderes políticos, pero, dichas cuentas son de género masculino lo cual evidencia una influencia por género.



En lo concerniente a la Figura 4, aparece la correlación entre las variables: género, ámbito, sentimientos y multimodalidad. Los resultados son los siguientes: (i) A través de los nodos observamos como una relación (grosor grueso) entre los grupos institucionales y el género masculino. El género femenino tiene menor presencia en el ámbito institucional, aparece ligado a perfiles individuales (grosor fino). (ii) Nicolás Maduro tiene 4 millones de seguidores, un caso paradigmático que contribuye a la asociación entre las variables masculino e institucional, como sucede con otros líderes políticos. La publicación asociada a este caso [P80] recopila la publicación (<https://bit.ly/3N0QOro>), destaca por la baja ratio de difusión con 824 «likes» en relación con su elevado número de seguidores. (iii) La mayoría son mensajes multimodales (texto e imagen) y son escasas las publicaciones monomodales.



En lo referente a la detección de «bots», con una probabilidad superior al 98%, se corresponde con cuentas automatizadas, se detectaron 23 cuentas. Esto supone menos de un 0,07% del total de perfiles correspondientes a 89 publicaciones que fueron eliminadas. Dichas publicaciones cuentan con temáticas enfocadas a temas políticos con una carga de sentimientos negativa a favor o en contra de determinadas ideologías y partidos políticos en periodo de campaña electoral. Otro aspecto preocupante, como sucede en todas las RRSS, son las «face news». Es el caso de la publicación [P60], asociada a la cuenta «adhd_understood» de Instagram con un número significativo de seguidores, cuya propietaria

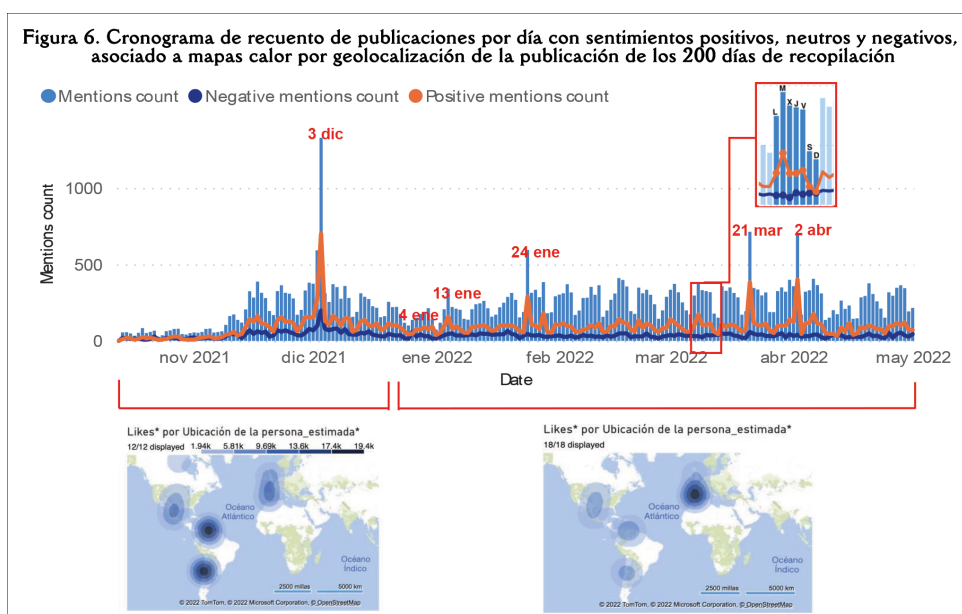
es Ms. Donna Giachino, que se define como médico especialista en TDAH (<https://bit.ly/3z5yIFV>). Sin embargo, el Colegio de Médicos en Vancouver en 2019 tuvo que pronunciarse para indicar que no estaba colegiada y, por ello, no podía ejercer de profesional del habla y el lenguaje (<https://bit.ly/3wTC7Rt>).

La Figura 5 ejemplifica los resultados presentados hasta el momento, mediante dos publicaciones en el contexto de la última campaña electoral en Chile, publicadas en un espacio temporal próximo. Muestran cómo el género masculino frente al femenino tiene una diferencia cuantitativa de «followers» 560.321 y «following» 80.002, en línea con los resultados (i) de género sobre un mayor número de «followers» y «following» en el género masculino. Otro aspecto relevante de las publicaciones analizadas es el número de «likes», cuya publicación de José Antonio Kast [V32] dispone de 3.019 «likes» y 1.265 «retweets» frente a los 3.854 «likes» y 3.681 «retweets» de Claudia Aldana [V17]. Es decir, con un menor número de seguidores en género femenino la publicación alcanza un impacto social mayor en Twitter. En este punto, se ejemplifica la idea secundaria de género (ii), son las mujeres las que tienen seguidores más fieles en la interacción con las publicaciones. Si observamos el número de visualizaciones del video [V17] se sitúa en 85.693 visualizaciones, una cifra mayor a su número de seguidores. Estos resultados evidencian que son los propios algoritmos de las plataformas digitales los que sugieren al resto de usuarios las publicaciones, aunque no sean seguidores del perfil, motivado por la polarización del sentimiento asociado a las campañas electorales que conduce a una fidelidad segmentada. Esta situación evidencia que los algoritmos generan «eco chambers». Es decir, a los usuarios se les presentan sugerencias por sus intereses temáticos y afinidad ideológica de manera recurrente, este procedimiento se repite escuchando solo la información que las plataformas digitales consideran afines al usuario. Esta manipulación de difusión de información nos conduce a los sesgos cognitivos de los usuarios analizados en la herramienta de Power BI.



Las plataformas digitales tienen un papel relevante en el ámbito educativo y social, porque son un altavoz de los días internacionales cuya finalidad reivindicativa busca sensibilizar y concienciar sobre problemáticas sociales. Como se evidencia en la Figura 6, el recuento de las publicaciones por día en función de los sentimientos arroja los siguientes resultados. Se aprecian una serie de días con mayor número de publicaciones que se traducen en la gráfica en los picos enumerados. Si atendemos a la de mayor magnitud, corresponde al día 3 de diciembre 2021, el Día Internacional de las personas con discapacidad, con 1.329 publicaciones de las cuales 702 y 199 son de sentimientos positivos y negativos, respectivamente. Si analizamos las publicaciones de dicho día, la publicación más viral es positiva (<https://bit.ly/3GwET20>) [T7], asociada a un perfil fuera de los grupos de influencia de intereses políticos y económicos tradicionales, mencionados más arriba. La creadora de dicha publicación es una persona con Síndrome de Down, María José Paiz Arias, conocida en las RRSS como Majo. Su publicación dispone de 7.034 «likes» y 55.463 reproducciones, cuyo título conceptualiza la idea reivindicativa de los días internacionales analizados: «la inclusión se logra con menos etiquetas y más acción». Sin embargo, líderes políticos como la Senadora de la República de México, Presidenta de la Comisión de Derechos de

la Niñez y de la Adolescencia con más de 100.000 seguidores, apenas recibe 13 «likes» y 6 «retweets» (<https://bit.ly/3wYaCVJ>). Es decir, el poder viral de los grupos de influencia frente a colectivos minoritarios se invierte, a diferencia del resto de días del año. Otros días destacables que evidencian la inversión de la tendencia son el 4 de enero Día Mundial del Braille, el 13 de enero Día Mundial de la Lucha contra la Depresión, el 26 de enero Día Mundial de la Educación Ambiental y el 21 de marzo Día Mundial del Síndrome de Down. En referencia al Día Mundial de la Educación, el 24 de enero de 2022, sucede la situación descrita con anterioridad, los líderes políticos publican, pero quedan relegados a un segundo plano independientemente de sus miles o millones de seguidores, como sucede con la cuenta Oficial del Gobierno de España con 778.858 «followers», que apenas recibe 74 «likes» (<https://bit.ly/3IUDyIV>). Otro día significativo, de relevancia para esta investigación, porque no contiene directamente las palabras clave determinadas para el estudio, es el 2 de abril Día Mundial de Concienciación sobre el Autismo, donde la publicación más viral tiene un componente claramente altruista de un perfil individual, un padre con un hijo con Trastorno del Espectro Autista (TEA): <https://bit.ly/3tqQY3R> [P95].



En referencia a la frecuencia de publicaciones destacan los días laborables como espacios temporales de mayor intensidad, pero los sábados y los domingos para la temática analizada se reducen las publicaciones de forma drástica, las causas se desconocen por el momento. No obstante, cabe mencionar que, si el día internacional es festivo, se mantiene posicionado como altavoz de reivindicación, debido al alto número de publicaciones. Por lo tanto, se evidencia que: (i) las fechas internacionales dan visibilidad a las reivindicaciones sociales, invisibles el resto del año a los ojos de los usuarios de las RRSS, con el fin de sensibilizar, orientar y reivindicar la causa social. (ii) Los perfiles involucrados en la reivindicación social son los más virales, relegando a un segundo plano a los grupos de influencia, aunque estos grupos conmemoran estos días con publicaciones. Por lo tanto, la hipótesis (H1) planteada de prevalencia dominante de los grupos de influencia dispone de una excepción en los días internacionales.

En lo concerniente a la Figura 6 se representa un cronograma, asociado a mapas de calor por ubicación de los perfiles más virales que muestra los siguientes resultados: la división cronológica para generar dichos mapas es motivada para argumentar la hipótesis H1, donde los intereses de los grupos de influencia se superponen sobre los grupos altruistas. Si observamos los mapas de calor generados por las geolocalizaciones de las publicaciones, en las fechas correspondientes a 2021, destaca Argentina y Chile. El motivo es porque el 12 de noviembre fue la primera vuelta de las elecciones en Chile y el 14 de noviembre las elecciones en Argentina. En el año 2022 el foco de calor se centra en Europa, donde tienen lugar distintos eventos electorales, en Portugal, Francia y Alemania. Por tanto, (i) también los eventos de

interés político opacan reivindicaciones sociales y contenido altruista.

Si atendemos a los 100 hashtags (#) de Instagram más mencionados en toda la horquilla temporal, representados en la Figura 6, las primeras posiciones las establecen las palabras clave de la investigación, como es obvio. Los siguientes hashtags, en la posición octava, se encuentra la palabra autismo, #autism (N=2.117) y #autismo (N=1.962). Según la Organización Mundial de la Salud (OMS), se estima una prevalencia media del autismo en 1/160 personas, por lo tanto, es lógica la presencia del tema. En otros hashtags se refleja #adhd (21) N=913, #dyslexia (23) N=876, #tea (29) N=801, #downsyndrome (35) N=733, o #asperger (86) N=457. Además, destacan hashtags relacionados con la cultura, valores y el respeto como #diversidad (10) N=1927, #cultura (19) N=1024, #teacher (31) N=773, #educacionespecial N=762. Otros hashtags que reflejan la sintonía de las publicaciones más sociales en el top 100 son #diversity, #igualdad, #valores, #equidad, #respeto, #educacioninclusiva, #accesibilidad. (i) Alineado con la H1 de los grupos de influencia entre los hashtags más repetidos se encuentra #influencer, relacionados con creadores digitales. (ii) En línea con la filosofía de los profesionales de la educación juegan un papel fundamental en dichos asuntos, reflejado en los siguientes hashtags: #profesores, #docentes, #educacionprimaria, #primaria, #neuroeducación, #estudiantes, #colegio, #maestro, #bibliotecas. En relación con las profesiones de los perfiles más virales, (i) la que mayor número de «likes» recibe se vincula a los creadores digitales [T1], es algo lógico ya que su fin es conseguir impacto en las RRSS. A diferencia de una asociación que se encuentra oculta en estas RRSS cuyo fin es sensibilizar y colaborar con la causa que los mueve [T34], pero dados los mensajes neutros y los contenidos que publican quedan relegados a un segundo plano debido a un «algoritmo de cristal» en las redes sociales, es decir, son visibles para cualquier usuario que los busque explícitamente, pero no por sugerencias de las plataformas digitales.

4. Discusión y conclusiones

Los grupos de influencia tienen un papel relevante como actores que estarían fagocitando la influencia de otros actores con intereses más altruistas y humanistas. Es una cuestión importante si tenemos en cuenta que las temáticas analizadas afectan a colectivos y personas muchas veces en desigualdad, que necesitan visibilizar sus reivindicaciones, experiencias y logros conseguidos. Las RRSS como espacios en los que se intermedia entre percepción y realidad, con efectos en las formas de entenderla y de lo que es o no importante para el debate colectivo público, juega a favor de grupos y actores con intereses políticos (Del-Fresno-García, 2019). En este sentido, la función de las RRSS como espacios de afinidad, participación y colaboración, como evidencian los trabajos de Grace et al. (2019) y De-Groot et al. (2022), en este estudio y para la temática objeto análisis, no ha sido confirmada. En cuanto al papel de las RRSS como oportunidades para que los usuarios interactúen y compartan información en unas estructuras relacionales de apoyo y seguimiento influyentes, tampoco han sido validadas en esta investigación. Las dos únicas formas de revertir el poder viral de los grupos de influencia son dos: cambio en los algoritmos de las plataformas digitales, y usuarios concienciados en las RRSS. Para la primera posibilidad, se necesitarían políticas públicas estatales y supraestatales que promueven alguna ley para establecer cierto control sobre los grupos de influencia, con la paradoja, como se ha demostrado en este estudio, que son los actores de mayor influencia en las redes. Para la segunda posibilidad, se requiere un profesorado con competencia digital. Los resultados obtenidos señalan la necesidad de concienciar a la ciudadanía digital con una visión crítica para comprender cómo los algoritmos imponen los mensajes de determinados actores, frente a otros colectivos interesados en visibilizar situaciones vinculadas a la discapacidad, y reivindicar su valor educativo y social desde referentes inclusivos. Algunos de los 100 hashtags (#) de Instagram más virales confirman estos intereses. Los docentes deben hacer comprender a las nuevas generaciones la asimetría en la comunicación, tal y como la representan Barberá et al. (2015) y Brady et al. (2019). Ello les permitirá reconocer que determinadas publicaciones, y los perfiles a ellas asociadas, definen valores humanos y sociales inestimables. En relación con esto, es importante que el profesorado ayude a las nuevas generaciones a sensibilizarse y tomar conciencia progresiva sobre los riesgos de caer en los filtros burbujas o «eco chambers». Ello evitaría un dominio como factor de desestabilización de las democracias a nivel mundial, ante la desinformación promovida por los grupos de influencia según sus intereses, que

define otro desafío importante para el profesorado (Ortega-Sánchez et al., 2021). En referencia al estudio Robles et al. (2022) sobre los «bots» y sentimientos negativos asociados a los mismos como combustible para polarización política, no tiene presencia en la temática social y educativa objeto de análisis. Esto puede ser debido a la sensibilización ciudadana que no asumirá y aceptará la confrontación explícita, como sí sucede con temáticas sobre inmigración, economía, seguridad, política, o sanidad. Temáticas propensas a la «bots» que polarizan las publicaciones, y con ello a los ciudadanos. La relegación del poder de influencia de los grupos influyentes en los días internacionales avalaría esta conclusión. Otros resultados que validarían estas conclusiones, es que, aunque el género masculino tiene mayor número de seguidores que el femenino, son las mujeres las que más «likes» cosechan. Lo que podría estar vinculado a los sentimientos positivos que prevalecen en sus publicaciones para el contenido analizado. Estos resultados indicarían otra tarea importante para el profesorado, enseñar a las nuevas generaciones a publicar en las RRSS desde actitudes constructivas, alejadas de emociones negativas (Arcila-Calderón, 2022).

El aspecto innovador del uso de la herramienta de «Business Intelligence» justifica que sea considerada como contenido de la competencia digital del profesorado para la educación en los próximos años. El impacto científico-técnico supone un cambio radical en el modo de conceptualizar y utilizar las RRSS como espacios de comunicación y construcción de discursos y acciones en el campo de la educación inclusiva y la discapacidad. Entre otras razones, porque nuestro análisis deja de contemplar las RRSS como herramientas para el consumo acrítico e irreflexivo de discursos y de contactos superficiales, para constituirse en valiosos medios que soportan relaciones y contenidos de valor educativo y social. Otro hecho importante que avala el carácter innovador de las herramientas y técnicas aplicadas reside en los estudios de caso que hemos iniciado, porque permitirán sistematizar el valor y función de la representación multimodal para comunicar y promover concienciación, debates y narrativas que generen acciones contra-hegemónicas. Otra razón reside en el impacto metodológico inherente al estudio definido por las técnicas de análisis que llevaremos a cabo. Ello contribuirá a definir en el debate científico sobre la comunicación y el pensamiento humano el valor del trabajo interdisciplinar en dos ámbitos, el de la Tecnología Educativa y el de las Telecomunicaciones.

Para finalizar, señalar que el proceso de análisis realizado con herramientas de «Big Data» y «Business Intelligence» es en sí mismo una competencia digital en espacios de relación inclusiva en instituciones educativas, que ayudará al profesorado a comprender la estructura relacional de las RRSS y saber llegar a significados sobre el contenido de los intercambios que hacen los participantes en dichas redes. Esto supone que tal competencia docente contemple saber identificar los perfiles más influyentes, indagar sobre el sentido que se da a los temas y los comportamientos, no siempre visibles (Del-Fresno-García, 2014), vinculados con los mensajes en RRSS y el tipo de perfiles asociados a ellos. Este conocimiento llevará a maestros y maestras a adquirir una ética colectiva, para lo cual, a veces, será necesario desaprender ideas y creencias construidas en la historia de cada uno de los participantes en sus contextos de vida. Solo así, se podrá construir una moralidad compartida desde la razón, y no solo desde las emociones en las relaciones contemporáneas, mestizas y diversas que caracterizan a los centros educativos que buscan situaciones inclusivas de enseñanza. Esta exigencia en la práctica docente conlleva la necesidad de asumir la dimensión indagadora-innovadora sobre las RRSS, de naturaleza crítica, como uno de los objetivos de la competencia digital del profesorado (Bautista, 2021).

Contribución de Autores

Idea, C.B.; Revisión de literatura (estado del arte), C.B., L.R.; Metodología, C.B.; Análisis de datos, C.B., L.R.; Resultados, C.B, L.R, A.B; Discusión y conclusiones, C.B, L.R, A.B.; Redacción (borrador original), C.B, L.R, A.B.; Revisiones finales, C.B, L.R, A.B; Diseño del Proyecto y patrocinios, C.B, L.R, A.B.

Apoyos

Acción financiada por la Comunidad de Madrid a través del Convenio Plurianual con la Universidad Complutense de Madrid en su línea Programa de Excelencia para el profesorado universitario, en el marco del V PRICIT (Plan Regional de Investigación Científica e Innovación Tecnológica).

Referencias

- Arcila-Calderón, C., Sánchez-Holgado, P., Quintana-Moreno, C., Amores, J., & Blanco-Herrero, D. (2022). Hate speech and social acceptance of migrants in Europe: Analysis of tweets with geolocation. [Discurso de odio y aceptación social hacia migrantes en Europa: Análisis de tuits con geolocalización]. *Comunicar*, 71, 21-35. <https://doi.org/10.3916/C71-2022-02>
- Awidi, I.T., Paynter, M., & Vujosevic, T. (2019). Facebook group in the learning design of a higher education course: An analysis of factors influencing positive learning experience for students. *Computers & Education*, 129, 106-121. <https://doi.org/10.1016/j.compedu.2018.10.018>
- Barberá, P. (2015). Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data. *Political Analysis*, 23(1), 76-91. <https://doi.org/10.1093/pan/mpu011>
- Barberá, P., Jost, J.T., Nagler, J., Tucker, J.A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, 26(10), 1531-1542. <https://doi.org/10.1177/0956797615594620>
- Bautista, A. (2021). Functional resignification and technological innovation as a digital teaching competence. *IEEE. Revista Iberoamericana de Tecnologías del Aprendizaje*, 16(1), 93-99. <https://doi.org/10.1109/RITA.2021.3052679>
- Becker, L.T., & Gould, E.M. (2019). Microsoft power BI: Extending excel to manipulate, analyze, and visualize diverse data. *Serials Review*, 45(3), 184-188. <https://doi.org/10.1080/00987913.2019.1644891>
- Brady, W.V.J., Willis, J.A., Burkart, D., Jost, J.T., & Van-Bavel, J.J. (2019). An ideological asymmetry in the diffusion of moralized content on social media among political leaders. *Journal of Experimental Psychology: General*, 148(10), 1802-1813. <https://doi.org/10.1037/xge0000532>
- Brunner, M., Palmer, S., Togher, L., & Hemsley, B. (2019). 'I kind of figured it out': The views and experiences of people with traumatic brain injury in using social media-self-determination for participation and inclusion online. *International Journal of Language & Communication Disorders*, 54(2), 221-233. <https://doi.org/10.1111/1460-6984.12405>
- Carlsson (Ed.) (2019). *Understanding media and information literacy (MIL) in the digital age. A question of democracy*. University of Gothenburg. <https://bit.ly/3xVwCsZ>
- Cheng, L.C., & Tsai, S.L. (2019). Deep learning for automated sentiment analysis of social media. In F. Spezzano, W. Chen, & X. Xiao (Eds.), *Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining* (pp. 1001-1004). Association for Computing Machinery. <https://doi.org/10.1145/3341161.3344821>
- De-Groot, R., Kaal, L.H., & Wouter, P.S. (2022). The online lives of adolescents with mild or borderline intellectual disabilities in the Netherlands: Care staff knowledge and perceptions. *Journal of Intellectual & Developmental Disability*. <https://doi.org/10.3109/13668250.2021.2004635>
- Del-Fresno-García, M. (2014). Haciendo visible lo invisible: Visualización de la estructura de las relaciones en red en Twitter por medio del análisis de redes sociales. *Profesional de la Información*, 23, 246-252. <https://doi.org/10.3145/epi.2014.may.04>
- Del-Fresno-García, M. (2019). Desórdenes informativos: sobreexpuestos e infrainformados en la era de la posverdad. *Profesional de la Información*, 28(3), 1-11. <https://doi.org/10.3145/epi.2019.may.02>
- Dellwing, M. (2021). *The continuities of Twitter strategies and algorithmic terror. Society for the study of symbolic interaction*. Sage. <https://doi.org/10.1002/symb.585>
- Ditrendia (Ed.) (2009). *Móvil en España y en el Mundo*. Ditrendia Digital Marketing Trends. <https://bit.ly/3MNim3j>
- Grace, E., Raghavendra, P., Mcmillan, J., & Gunson, J. (2019). Exploring participation experiences of youth who use AAC in social media settings: Impact of an e-mentoring intervention. *Augmentative and Alternative Communication*, 35(2), 132-141. <https://doi.org/10.1080/07434618.2018.1557250>
- Hemsley, B., Dann, D., Palmer, S., Allan, M., & Balandin, S. (2018). Using Twitter to access the human right of communication for people who use Augmentative and Alternative Communication. *International Journal of Speech-Language Pathology*, 20(1), 50-58. <https://doi.org/10.1080/17549507.2017.1413137>
- Ladogina, A., Samoylenko, I., Golovina, V., Razina, N., & Petushkova, E. (2020). Communication effectiveness in social networks of leading universities. *Diálogo*, 43, 35-50. <https://doi.org/10.18316/dialogo.v0i43.6497>
- Liu, B. (2010). Sentiment analysis and subjectivity. In N. Indurkha, & F. J. Damerau (Eds.), *Handbook of Natural Language Processing*. Chapman and Hall/CRC. <https://doi.org/10.1201/9781420085938>
- Mitchell, R. (2018). *Web scraping with Python: Collecting more data from the modern web*. O'Reilly Media: O'Reilly Media, Inc. <https://bit.ly/3zw5nOn>
- Ortega-Sánchez, D., Blanch, J.P., Quintana, J.I., Cal, E., Fuente-Anuncibay, C., & Hate, R. (2021). Speech, Emotions, and Gender Identities: A Study of Social Narratives on Twitter with Trainee Teachers. *International Journal of Environmental Research and Public Health*, 18(8), 4055-4055. <https://doi.org/10.3390/ijerph18084055>
- Robles, J., Guevara, J., Casas-Mas, B., & Gómez, D. (2022). When negativity is the fuel. Bots and Political Polarization in the COVID-19 debate. [Cuando la negatividad es el combustible. Bots y polarización política en el debate sobre el COVID-19]. *Comunicar*, 71, 63-75. <https://doi.org/10.3916/C71-2022-05>
- Tuzel, S., & Hobbs, R. (2017). The use of social media and popular culture to advance cross-cultural understanding. [El uso de las redes sociales y la cultura popular para una mejor comprensión intercultural]. *Comunicar*, 51, 63-72. <https://doi.org/10.3916/C51-2017-06>
- We are social (Ed.) (2021). *Digital 2021. The latest insights into the 'state of digital'*. We Are Social Ltd. <https://bit.ly/3POLhJV>
- White, P., & Forrester-Jones, R. (2020). Valuing e-inclusion: Social media and the social networks of adolescents with intellectual disability. *Journal of Intellectual Disabilities*, 24(3), 381-397. <https://doi.org/10.1177/1744629518821240>