

¿Callejón sin salida?: El problema muestral en el estudio de las redes sociales digitales¹

Dead end? The sampling problem in the study of digital social networks

Beco sem saída? O problema de amostragem no estudo das redes sociais digitais

DOI: <https://doi.org/10.32870/cys.v2024.8580>

DANIEL PATTIER²

<https://orcid.org/0000-0003-3426-922X>

Los objetivos de este trabajo son profundizar en la problemática muestral en el estudio de las redes sociales digitales y mostrar las posibilidades y soluciones que aporta la literatura para hacer frente a este problema. A través de una metodología teórico-reflexiva se indaga sobre el tema, destacando y analizando las opciones procedimentales utilizadas en la literatura. Se plantean, además, algunas mejoras metodológicas que pueden contribuir al desarrollo de investigaciones más fiables sobre la temática.

PALABRAS CLAVE: Comunicación digital, redes sociales, metodología de investigación, tamaño de la muestra.

The objectives of this work are to delve into the sampling problem in the study of digital social networks and to show the possibilities and solutions that the literature has provided to deal with this problem. Through a theoretical-reflexive methodology, the subject is investigated highlighting and analyzing the procedural options used by the literature. In addition, some methodological improvements are proposed that can contribute to the development of more reliable research on the subject.

KEYWORDS: Digital communication, social networks, research methodology, sample size.

Os objetivos deste trabalho são aprofundar os problemas exemplares no estudo das redes sociais digitais e mostrar as possibilidades e soluções que a literatura oferece para resolver este problema. Através de uma metodologia teórico-reflexiva, o tema é investigado, destacando e analisando as opções processuais utilizadas na literatura. Além disso, são propostas algumas melhorias metodológicas que podem contribuir para o desenvolvimento de pesquisas mais confiáveis sobre o tema.

PALAVRAS-CHAVE: Comunicação digital, redes sociais, metodologia de pesquisa, tamanho de amostra.

Cómo citar este artículo:

Pattier, D. (2024). ¿Callejón sin salida?: El problema muestral en el estudio de las redes sociales digitales. *Comunicación y Sociedad*, e8580. <https://doi.org/10.32870/cys.v2024.8580>

¹ Trabajo financiado por el Ministerio de Ciencia e Innovación de España a través del proyecto I+D+i “El imperativo de la innovación educativa: análisis de su recepción y articulación en el sistema educativo español (IMP-NOVA)” con referencia PID2022-138878NA-I00.

² Universidad Complutense de Madrid, España.
dpattier@ucm.es

Fecha de recepción: 16/02/23. Aceptación: 12/09/23. Publicado: 14/02/24.

INTRODUCCIÓN

Las redes sociales, como plataformas digitales de comunicación global, ponen en contacto a un gran número de usuarios en todo el mundo. Potencian la creación de entornos donde los usuarios crean sus propios perfiles con la finalidad de comunicarse con otros, expresarse, ver o comentar contenido creado por otros usuarios, o compartir su propio contenido (Phua et al., 2016). Son entornos dinámicos ya que los usuarios se unen, utilizan y abandonan las comunidades según sus intereses a lo largo del tiempo (Mahmoudi et al., 2020).

Las redes sociales han transformado la manera en que nos comunicamos, aprendemos, hablamos o interactuamos con los demás. La accesibilidad ofrecida por los dispositivos tecnológicos ha generado un uso habitual y frecuente de todo tipo de páginas web y redes sociales. Según el *Digital 2022 Global Overview Report* (Kemp, 2022), que analiza el uso digital a nivel mundial, cada año que pasa aumenta el tiempo que las personas destinamos a la utilización de dispositivos tecnológicos conectados a Internet. En el año 2022 se situaba prácticamente en una media de siete horas diarias. De este modo, si tenemos en cuenta que una persona suele dormir entre siete y ocho horas por día, podemos afirmar que, cuando estamos despiertos, pasamos más del 40% de horas conectados.

Este informe analiza en profundidad el uso de redes sociales digitales a nivel mundial, y destaca que un 95.2 % de las personas entre 16 y 64 años las han utilizado en el último mes. Además, es relevante que el usuario medio utilice 7.5 redes sociales diferentes al mes, lo que demuestra una variada utilización de este tipo de plataformas digitales. Por último, es significativo que cuatro de las diez páginas web más empleadas por la audiencia sean redes sociales: YouTube (segunda posición), Facebook (tercera posición), Instagram (sexta posición) y Twitter (novena posición).

Este uso creciente de las redes sociodigitales ha propiciado que la literatura se haya volcado en la investigación de páginas, perfiles, cuentas y canales pertenecientes a un gran número de ellas para obtener datos relativos a numerosas áreas del saber. Así, podemos encontrar investigaciones sobre la temática en el ámbito de la psiquiatría (Brooks

et al., 2022), la atención social (Fernández-Peña et al., 2022), la actividad física (Prochnow & Patterson, 2022), la política y la economía (Faciroli et al., 2022), la educación (Marcelo & Marcelo, 2021), la sociología (Tobón Berrio et al., 2022), la comunicación (Fernández et al., 2021) o la medicina (Kumar et al., 2021), entre otros.

El estudio de redes sociales digitales

La literatura se ha acercado a la temática de las redes sociales desde innumerables puntos de vista, ya sea meramente desde los datos cuantitativos, hasta una profundización en el análisis de redes del discurso (DNA por sus siglas en inglés), de redes semánticas o, incluso, de redes socioecológicas (Tindall et al., 2022).

Sin embargo, en general, las investigaciones abordan la temática de las redes sociales desde dos propuestas: una enfocada en las relaciones reportadas por un individuo focal y otra considerando una red completa a través de la delimitación de la población con algún criterio específico (Vatrapu et al., 2016). Además, el carácter dinámico de este tipo de entornos impone la necesidad de fragmentar la red en algunas ventanas de tiempo. Como señalan Mahmoudi et al. (2018), muchos estudios solamente han utilizado una sola ventana de tiempo como información indivisible, lo que puede provocar una interpretación incorrecta de los datos.

Es importante señalar, además, la discusión actual entre el desarrollo de investigaciones que emplean dos tipos de procesos en la utilización de los datos. Por un lado, el uso del *big data*, donde se manejan enormes cantidades de datos a través de programas específicos de análisis de una manera cuantitativa. En estos estudios los problemas pueden surgir debido a factores de heterogeneidad, de acumulación de ruido, de falsas correlaciones y a la endogeneidad (Hernández-Leal et al., 2017). Por otro lado, el *soft data* se basa en una perspectiva cualitativa a través de opiniones e interpretaciones, lo que puede producir un desvío subjetivo de los resultados o conclusiones de la investigación.

Asimismo, siguiendo a Tindall et al. (2022), es reseñable la existencia de formas estructurales específicas en la arquitectura de las redes sociales donde entran en juego varios procesos que se condicionan simultáneamente, como son la reciprocidad, la homofilia, la transitividad o el apego preferencial. Como señala su investigación, si tenemos en

cuenta los modelos generativos para datos de redes transversales, podemos destacar la utilización en la actualidad de modelos de gráficos aleatorios exponenciales (ERGMs en su terminología en inglés).

Además, a la hora de recolectar los datos de las redes sociales o de páginas web, la literatura señala dos formas fiables de hacerlo: *web scraping* y solicitudes a las interfaces de programación de aplicaciones (API en su nomenclatura en inglés). Este proceso puede ser complejo dependiendo de la investigación que se esté llevando a cabo (Tindall et al., 2022). Sin embargo, la falta de información sobre los marcos de muestreo (Rocheva et al., 2022) o la enorme cantidad de datos que generan los usuarios a tiempo real hace que el muestreo sea ineludible y que se desarrollen alternativas metodológicas como el *social mining* (Piña-García et al., 2016). De ahí la importancia de hacer un muestreo con una intención clara y de saber acotar la muestra eficientemente debido a la imposibilidad de uso de la población en la mayoría de los procesos investigativos (Pihlaja, 2022).

Esta utilización de cierta parte muestral puede ocasionar un error de muestreo cuyos resultados no representen aquellos que se obtendrían mediante el análisis de toda la población. La literatura señala ciertos tipos de errores de muestreo. En primer lugar, el error específico de la población, que ocurre cuando no se delimita adecuadamente el total de la población de la investigación. En segundo lugar, el error de selección, que puede generarse cuando la muestra es autoseleccionada o cuando participan solamente aquellas personas interesadas en la temática. En tercer lugar, el error de marco de muestra, producido por una selección incorrecta de la parte de la población a analizar. En cuarto lugar, el error de falta de respuesta, cuando la investigación no consigue obtener respuestas útiles por parte de los participantes en ella.

Como señala la literatura, es importante que la investigación en redes sociales consiga afrontar el reto de profundizar en un campo tan cambiante como los medios digitales (Jones et al., 2015). Debido a ello, es importante mencionar la aportación de Vásquez (2022), que ofrece una mirada poliédrica defendiendo la eficiencia de diferentes metodologías de investigación para acercarse a la realidad de las redes sociales y del discurso digital tanto desde una perspectiva cuantitativa como cualitativa.

Por último, son reseñables diversas investigaciones que se han realizado sobre la temática de la privacidad y de la ética investigativa en la búsqueda, selección y utilización de datos en las investigaciones sobre redes sociales (Tindall et al., 2022). Existen factores sensibles que tienen que ver con la privacidad, con el procesamiento de los datos, con el cumplimiento de la normativa legal y con la confidencialidad (Zhang et al., 2010). En este aspecto, es notable la aportación de Wang et al. (2022) quienes presentan un método para preservar la privacidad en el tratamiento de datos mediante dos tareas: detección y preservación. En la tarea de detección se identifican por separado las imágenes y las leyendas para ubicar los objetos específicos. En la tarea de preservación se utilizan métodos de ocultamiento o reemplazo para manejar las imágenes y leyendas detectadas y garantizar al mismo tiempo la utilidad de los recursos.

Por último, no podemos olvidar la posibilidad de un comportamiento sesgado y malicioso por parte de los usuarios, lo que infiere una problemática añadida a las investigaciones actuales (Sirisala et al., 2022).

Objetivos y preguntas de investigación

En este trabajo estableceremos una discusión metodológica frente a un problema que surge en los estudios de comunicación digital, concretamente en investigaciones de redes sociodigitales. Así, los dos objetivos de este artículo son los siguientes: 1) profundizar en la problemática muestral en el estudio de las redes sociales digitales, y 2) mostrar las posibilidades y soluciones que aporta la literatura para hacer frente a esta problemática.

Las preguntas de investigación son las siguientes: ¿Qué problemas metodológicos existen cuando se diseña la muestra en los estudios de redes sociales? ¿Cómo abordan las investigaciones estos problemas para obtener datos y resultados?

Para responder a estas preguntas de investigación mencionaremos la metodología empleada. Más adelante, se expondrá el problema muestral en el estudio de las redes sociales digitales y se indagará sobre las diversas soluciones aportadas por la literatura hasta la fecha. Por último, se desarrollará un apartado de discusión para finalizar con las conclusiones.

METODOLOGÍA

Este trabajo plantea una metodología teórico-reflexiva con la finalidad de profundizar en las relaciones esenciales de problematicidad de diversos fenómenos procedimentales en el ámbito de la investigación en comunicación (Del Sol Fabregat et al., 2017). Debido a la similitud en las dificultades que surgen en el estudio de las redes sociales digitales, se ha mantenido una visión generalista que ofrece una visión global y proporciona un marco de entendimiento de la cuestión.

Los planteamientos desarrollados en este escrito surgen de un estudio amplio del estado de la cuestión, que se ha expuesto con anterioridad, y que serán reforzados, además, con ejemplos concretos de investigaciones que utilizan cada una de las opciones metodológicas expuestas para tratar de solventar la problematicidad indagada por este trabajo.

El problema muestral

Las investigaciones realizadas sobre redes sociodigitales en las diversas áreas disciplinares se deparan con un problema metodológico inherente a la naturaleza propia de estas páginas y plataformas tecnológicas: el establecimiento de la muestra.

Dentro de una investigación, la muestra es un aspecto fundamental ya que es la base que proporciona los datos que serán interpretados para corroborar o desmentir la o las hipótesis o responder a las diversas preguntas de investigación. La utilización de muestras con déficits de calidad metodológica puede generar la obtención de resultados no generalizables o incluso erróneos (Mucha-Hospinal et al., 2021). De ahí la importancia de establecer procesos metodológicos rigurosos y de calidad que aporten datos de valor a la discusión global sobre la temática de estudio (Vásquez, 2022).

La problemática muestral en las investigaciones sobre las redes sociales recae en dos factores que afectan directamente a dicho aspecto metodológico. En primer lugar, las redes sociales están en continuo cambio, ya que dependen de los usuarios, quienes suben contenidos o interactúan en su día a día generando millones de nuevos contenidos en la red (Jones et al., 2015; Mahmoudi et al., 2020). Como

podemos entrever, se genera una problemática en la delimitación de la muestra. ¿Qué perfiles investigar? ¿Durante cuánto tiempo? ¿Qué tipo de contenidos tener en cuenta y cuáles no? Estas son algunas de las preguntas que toda investigación sobre redes sociales debe hacerse para establecer un marco metodológico riguroso (Vásquez, 2022).

En segundo lugar, existe una problemática en cuanto al establecimiento de la muestra, ya que no existen listados oficiales por temáticas o áreas disciplinares que contemplen todos los perfiles, cuentas o canales de una red social. Debemos entender que las redes sociales digitales no están creadas ni pensadas de una manera metodológica estructurada, sino que se basan en algoritmos de posicionamiento que buscan obtener la mayor interacción y seguimiento de la audiencia. De ahí que la idoneidad o calidad de los contenidos no dependa de su posicionamiento en la plataforma (Burgos et al., 2020; Osman et al., 2022).

De este modo, cuando los investigadores pretenden estudiar, por ejemplo, los influencers de divulgación científica presentes en Twitch (Buitrago & Torres-Ortiz, 2022), dicha plataforma no genera un listado oficial de todos los perfiles teniendo en cuenta que estén destinados a la educación informal y, más concretamente, a la divulgación científica. Esta problemática surge en todas y cada una de las redes sociales disponibles hasta la fecha y en todas las áreas disciplinares.

Además, es importante resaltar la dificultad de categorizar, por un lado, algunos canales o perfiles que se dedican a varias áreas o temáticas y, por otro lado, que se encuentran entre dos categorías y que, dependiendo de la interpretación del propio creador de contenido o de los investigadores, pudiera ser agrupado en una u otra. Por ejemplo, un canal de YouTube destinado a subir canciones educativas podría categorizarse bajo la agrupación de “Formación” o de “Música”.

Al no tener la posibilidad de acceso a un listado oficial propio de la plataforma que indique cuáles son los canales o perfiles destinados al objeto de estudio de la investigación, se genera una problemática en cuanto a la delimitación muestral y, por lo tanto, al corazón de la metodología de las diversas investigaciones que se realizan sobre las redes sociales digitales (Pihlaja, 2022).

SOLUCIONES APORTADAS POR LA LITERATURA

Como hemos señalado anteriormente, existe una problemática metodológica a la hora de establecer la muestra de estudio en las redes sociales propiciada por dos factores: el continuo cambio y la actualización de contenidos, y la inexistencia de listados públicos categorizados oficialmente en las diversas plataformas digitales.

A continuación, expondremos las diversas soluciones aportadas por la literatura a lo largo de estos años que han intentado solventar esta problemática metodológica utilizando diversas decisiones desde una perspectiva creativa y con ánimo de una mejora en la calidad de las investigaciones (Flores-Márquez & González Reyes, 2021).

Estudio de caso

En primer lugar, diversos estudios han profundizado en la temática analizando algunos perfiles o canales de redes sociales utilizando una metodología de estudio de caso. La delimitación de la muestra, por tanto, no sufre de la problemática indicada anteriormente, ya que se centra en uno o varios casos concretos.

Se observa en la literatura que las investigaciones que han utilizado este método lo han hecho por conveniencia y debido a la facilidad de acceso a la muestra. Así, simplemente analizando los datos públicos que ofrecen las diversas cuentas, se puede realizar un estudio de caso sobre alguna temática en concreto. Por ejemplo, Martorell y Serra Folch (2018) analizaron la presencia de las ONGs en Instagram realizando un estudio de caso de la organización Proactiva Open Arms. Sus conclusiones muestran la evolución de la actividad de la ONG en la red social aumentando la frecuencia, la regularidad y el uso de imágenes que generan más captación por parte de la audiencia, lo que generó un crecimiento del perfil de Instagram.

Por otro lado, consiguiendo el permiso de un canal o perfil de redes sociales no solamente se tiene acceso a los datos que pueden ser observados y cotejados de manera pública en la red social, sino que también entrarán en juego algunos datos que las diversas plataformas ofrecen de manera privada a los creadores de contenido o usuarios de los perfiles de las redes sociales. En este sentido, algunos investigadores han ana-

lizado los datos públicos y privados de sus propios perfiles o canales, ya que tienen acceso ilimitado a todos los datos que ofrece la plataforma. Por ejemplo, Saurabh y Gautam (2019) demostraron, a través del análisis de su propio canal de YouTube, que el posicionamiento de los videos y la cantidad de visualizaciones siguen la distribución Zipf para videos educativos. Además, observaron una fuerte correlación entre la ubicación geográfica de la audiencia y la ubicación de la industria que atiende el canal.

Sin embargo, al tratarse de estudios de caso de tan solo uno o varios perfiles o cuentas, los datos son limitados en número y es difícil poder generalizar los resultados y conclusiones aportadas por las investigaciones. La concordancia de resultados con otros estudios de caso de otros canales o perfiles de redes sociales podrían indicar una posible generalización, pero, al tratarse de una muestra muy limitada, podrían generarse fácilmente resultados contrapuestos que habría que dilucidar en la discusión global sobre la temática.

Utilización de la muestra de otras investigaciones

En segundo lugar, una solución para evitar la problemática de la delimitación de la muestra en el estudio de redes sociales es la utilización de muestras establecidas por otras investigaciones. Un artículo que ha pasado un filtro para ser publicado en una revista posee cierto rigor que podemos aceptar a la hora de establecer la muestra para el estudio de redes sociales. De esta manera, el establecimiento de la muestra es sencillo porque es obtenido directamente de otra investigación.

Es la opción utilizada, por ejemplo, por Buitrago y Torres-Ortiz (2022) en su investigación sobre los influencers de ciencia en Twitch, que establece una serie de criterios para filtrar un listado elaborado por Zaragoza y Roca Marín (2020) concerniente a canales divulgativos de YouTube de España. Esta opción les valió para concluir que Twitch es un vehículo eficiente para captar la atención de la audiencia joven hacia temas científicos y evidenciar el compromiso de los creadores de contenido en esta plataforma por un discurso basado en el rigor científico.

Incluso algunas investigaciones toman, en parte, listados provenientes de fuentes menos fiables como páginas web que no muestran la metodología utilizada, pero que proponen una serie de perfiles o

canales sobre diversas temáticas de estudio. Es el caso, por ejemplo, de Feijoo Fernández y Fernández-Gómez (2021) quienes utilizaron el listado ofrecido por una página web para configurar parte de la muestra de su estudio sobre niños y niñas influyentes en YouTube e Instagram durante el periodo de la pandemia por Covid-19. Dicha investigación concluye la utilización por parte de estos creadores de contenido de temáticas relacionadas con la pandemia de manera frecuente, implementando historias guionizadas basadas en rutinas, recomendaciones y hábitos saludables.

Teniendo en cuenta esto, es importante señalar que esta opción para solventar la problemática de la delimitación de la muestra en el estudio de las redes sociales hace depender la metodología de la investigación en las bases metodológicas de otra. Por tanto, si la metodología de establecimiento de la muestra de la investigación que se toma como base no ofrece un rigor adecuado, todos aquellos estudios cuyas muestras provengan directamente de esta no alcanzarán ese rigor necesario en el ámbito académico y científico.

Delimitación de una pequeña muestra

En tercer lugar, existe la posibilidad de utilizar una muestra relativamente pequeña para analizar los perfiles o canales de las redes sociales y obtener datos de valor para la discusión global sobre la temática.

El establecimiento de dicha muestra puede realizarse de manera aleatoria, por conveniencia, o utilizando algún tipo de proceso sistemático que conduzca a la delimitación de una muestra que podemos considerar relativamente pequeña.

Es el caso, por ejemplo, de la investigación de Beltrán-Pellicer et al. (2018), quienes para delimitar la muestra de estudio de videos educativos de matemáticas en YouTube propuso la utilización de la palabra clave “distribución proporcional” en el buscador general de YouTube, para filtrar luego el listado con los videos que superan las mil visualizaciones y por la temática concreta que desarrollaban en el contenido. De esta manera, utilizaron para su investigación un total de 31 videos cuyo análisis les llevó a concluir que estos tenían una débil idoneidad epistémica y que no estaba relacionada con el nivel de popularidad y de posicionamiento en la plataforma.

Otro ejemplo de la utilización de esta opción es la implementada por Eslen-Ziya (2022) quien analiza un total de 1 295 tuits obtenidos en una semana de manera aleatoria sobre la temática del cambio climático. Esta investigación evidencia la presencia de perfiles que se muestran escépticos acerca de que el gobierno de su país tome una postura más activa hacia el cambio climático.

Como puede observarse en estos dos ejemplos que nos pueden servir de referencia, esta opción metodológica para resolver el problema de la delimitación de la muestra en el estudio de las redes sociales produce investigaciones que se valen del fácil acceso público a ciertos datos y contenidos de las plataformas digitales para establecer una pequeña muestra por conveniencia o afinidad. Sin embargo, es importante reseñar que las muestras más limitadas en número tienen una dificultad para generalizar resultados y conclusiones. Por lo tanto, la obtención de datos globales se ve limitada por el escaso número de la muestra.

Delimitación de una gran muestra

En cuarto lugar, otra solución a la problemática que analiza este artículo es la delimitación de una gran muestra de perfiles o canales de redes sociales digitales para su estudio. Al igual que en la delimitación de pequeñas muestras, la metodología empleada puede tender más hacia la conveniencia o hacia un proceso más sistemático, pero el resultado final es el establecimiento de una muestra que se puede considerar de gran tamaño.

La imposibilidad de conocer la totalidad de los perfiles o canales de las redes sociales relativos a una temática o a un área disciplinar hace que, cuanto mayor sea la muestra, más se acercará a la totalidad, y los resultados podrán ser más generalizables y utilizados en una discusión global (Pihlaja, 2022; Rocheva et al., 2022). El problema de esta posibilidad es la mayor dificultad en la obtención de la muestra, ya que se necesita emplear un tiempo significativamente superior en la búsqueda de canales o perfiles concernientes a la temática investigada.

Por ejemplo, la investigación de Pattier (2021) reconoce que la búsqueda de la muestra de canales educativos de YouTube siguiendo una serie de parámetros y pautas duró más de dos años. Como podemos apreciar, es un tiempo significativo que confiere una mayor dificultad

en cuanto a dicho aspecto metodológico. Así, esa investigación consiguió elaborar un listado de los 204 mejores canales educativos de YouTube España, una muestra muy por encima de la mayoría de las investigaciones sobre la temática. Dicho estudio evidenció la existencia de una brecha de género en el colectivo de los edutubers con una presencia significativamente mayor de hombres que de mujeres como creadores de contenido audiovisual en la plataforma.

Como podemos entrever, los resultados y conclusiones alcanzados por una investigación que utiliza una muestra de 204 canales educativos de YouTube son más generalizables que otras de la misma temática, pero con una muestra más pequeña (De Souza & Costa, 2021; Dos Santos Galvão, 2019; López et al., 2020; Moreira Queiroga & Spyer Dulci, 2019; Zaragoza & Roca Marín, 2020).

Uso de lectores o compiladores web de redes sociales

En quinto lugar, se ha generado en los últimos años una posibilidad antes no desarrollada que se apoya en el uso de páginas o plataformas ajenas a la red social que se pretende analizar para el establecimiento de muestras de redes sociales digitales con una categorización y una búsqueda sistemática aportada por las propias páginas.

Es el caso de los lectores o compiladores web de redes sociales, tales como Social Blade, Noxinfluencer o Feedspot, que impulsan esta posibilidad para la búsqueda de muestras de perfiles o canales en redes sociales digitales. Señalando una serie de categorías en la web o utilizando diversas palabras clave, se obtiene un listado que podría ser la base muestral de una investigación. Es una metodología replicable, aunque se debe tener en cuenta que las redes sociales varían continuamente (Jones et al., 2015) y, por lo tanto, los resultados que utilizan las mismas categorías de análisis podrían variar si son buscados en momentos diferentes (Mahmoudi et al., 2020).

Esta solución es utilizada, por ejemplo, por Ojeda-Serna y García-Ruiz (2022), quienes emplearon diversas palabras clave en la herramienta Social Blade para determinar los museos con el mayor número de suscriptores en la plataforma YouTube y Noxinfluencer para buscar youtubers sobre la temática. Su trabajo discute la necesidad de considerar las estrategias que utilizan los youtubers por parte de los centros

de educación superior y de los museos para mejorar su eficacia como divulgadores científicos. Asimismo, dicha propuesta es utilizada por Vizcaíno-Verdú y Contreras Pulido (2019), quienes, implementando la herramienta Feedspot, que tiene en cuenta el impacto y la medición de sistemas algorítmicos de tendencia, consiguieron elaborar un listado de canales de YouTube que consideran las diversas áreas del saber.

Sin embargo, es importante señalar que la categorización que proponen este tipo de herramientas es deficiente en cuanto a la metodología investigativa. Las categorías son demasiado simples y no existe la posibilidad de generar nuevas categorías o divisiones que profundicen algunas de las temáticas de estudio. Además, estas páginas que recopilan información de perfiles y cuentas de redes sociales tienen en cuenta los datos proporcionados por los propios creadores de contenido y no pasa ningún filtro humano, sino que son leídos por sistemas algorítmicos. Por ello, los perfiles que no estipulan la categoría de su contenido o que puedan ser englobados en varias categorías podrían quedar fuera del listado que ofrece la página. Por tanto, y aunque metodológicamente esta solución se acerca a un proceso de establecimiento de la muestra replicable y de rigor, se sigue sin obtener la totalidad posible de la población que se pretende analizar (Pihlaja, 2022).

Como se puede comprobar, siguiendo algunos de los ejemplos planteados en este trabajo, existen algunos canales de YouTube creados antes de 2019 que aparecen en la investigación de Pattier (2021) que utilizan la vía del establecimiento de una gran muestra mediante búsquedas sistemáticas de todo tipo, pero no aparecen en la de Vizcaíno-Verdú y Contreras Pulido (2019), quienes implementan la opción de la herramienta Feedspot, lo que muestra el déficit de esta última posible solución al problema del establecimiento de la muestra en el estudio de las redes sociales digitales.

De este modo, como hemos analizado, se han desarrollado a lo largo de la literatura cinco posibilidades para solventar dicho problema, como muestra la Tabla 1.

TABLA 1
VENTAJAS E INCONVENIENTES DE LAS SOLUCIONES UTILIZADAS
POR LA LITERATURA ANTE EL PROBLEMA DE LA DELIMITACIÓN DE LA
MUESTRA EN EL ESTUDIO DE REDES SOCIALES DIGITALES

Solución	Ventajas	Inconvenientes
Estudio de caso	Facilidad de acceso. Elección por conveniencia. Posibilidad de obtención de datos privados.	Dificultad de generalización de resultados.
Utilización de la muestra de otras investigaciones	Facilidad de establecimiento de la muestra.	Dependencia del rigor metodológico de otra investigación.
Delimitación de una pequeña muestra	Facilidad de búsqueda de la muestra. Elección por conveniencia.	Dificultad de generalización de resultados.
Delimitación de una gran muestra	Viabilidad de generalización de resultados.	Dificultad de búsqueda de la muestra. Empleo de mayor tiempo en la búsqueda.
Uso de lectores o compiladores web de redes sociales	Facilidad de búsqueda de la muestra.	Categorización deficiente. Continua actualización de los lectores o compiladores. Posibles errores de generalización.

Fuente: Elaboración propia.

DISCUSIÓN

El uso generalizado de las redes sociales entre la población general (Kemp, 2022) ha propiciado el desarrollo de numerosas investigaciones

sobre un número considerable de temáticas y áreas del saber (Brooks et al., 2022; Faciroli et al., 2022; Fernández et al., 2021; Fernández-Peña et al., 2022; Kumar et al., 2021; Marcelo & Marcelo, 2021; Prochnow & Patterson, 2022; Tobón Berrio et al., 2022).

El establecimiento de la muestra, dentro del proceso metodológico de dichos estudios, es fundamental para obtener bases fiables de resultados que fomenten el conocimiento que tenemos acerca de las diversas temáticas de estudio (Vásquez, 2022). De ahí la importancia de analizar y reflexionar acerca del rigor y las posibilidades que ofrecen las diversas soluciones que se han dado a lo largo del tiempo para solventar esta problemática dentro de las investigaciones sobre las redes sociales.

No se trata de analizar cuál es mejor y cuál es peor. Como podemos observar en la Tabla 1, todas las propuestas tienen ventajas e inconvenientes ya que, en la actualidad, no tenemos las herramientas necesarias para obtener la muestra completa de perfiles o canales sobre una temática en concreto.

Es más, dependiendo de las preguntas de investigación y de los propósitos de los investigadores, puede que el estudio de caso o el establecimiento de pequeñas muestras consigan ofrecer los resultados necesarios para responder a estas cuestiones y continuar avanzando el conocimiento en la temática (Beltrán-Pellicer et al., 2018; Eslen-Ziya, 2022; Martorell & Serra Folch, 2018; Saurabh & Gautam, 2019). Es por ello que es relevante reflexionar y acotar la finalidad de la investigación (Pihlaja, 2022).

La problematicidad surgiría a la hora de generalizar resultados, donde las investigaciones con mayores muestras podrían ofrecer resultados de mayor rigor en este sentido (Pattier, 2021; Vizcaíno-Verdú & Pulido, 2019), ya sean establecidas con búsquedas en la propia red social o a través de la utilización de lectores o compiladores web de redes sociales. En los últimos años se ha originado una tendencia a la utilización de la inteligencia artificial y de la minería de datos para intentar inferir patrones en grandes volúmenes de conjuntos de datos, lo que podría beneficiar la interpretación de los resultados a gran escala, pero estos datos provienen de la utilización de una muestra donde desconocemos la población total (Hassan & Behadili, 2022; Jones et al., 2015; Pihlaja, 2022).

Sin duda, la posibilidad de establecer un proceso sistemático y replicable de búsqueda de muestras para estudios de redes sociales digitales es un aspecto a valorar y, por tanto, la utilización de lectores o compiladores web podría acercarse a la solución del problema, aunque, como ya hemos anunciado, habría que realizar una serie de cambios que planteamos a continuación.

En primer lugar, la categorización de búsqueda debería ser mayor. Para solventar esta dificultad, las redes sociales digitales tendrían que ofrecer la posibilidad a los creadores de contenido de categorizar el tipo de recurso que suben de una manera más concreta. Por ejemplo, YouTube podría ofrecer la posibilidad de que los creadores de contenido no solamente indiquen que se trata de canales destinados a formación, sino también indicar el área del saber. Así, podríamos obtener datos más fiables, por ejemplo, de los canales de YouTube destinados a la divulgación de la ciencia (Buitrago & Torres-Ortiz, 2022; Zaragoza & Roca Marín, 2020).

En segundo lugar, y como hemos indicado con anterioridad, existe la posibilidad de que los perfiles o canales de redes sociales no compartan la categorización como información para que la usen los lectores o compiladores web, o que puedan ser ubicados en categorías diferentes a las que indica el usuario. En este sentido, y siguiendo el planteamiento que hacen Osman et al. (2022) en cuanto al posicionamiento en redes sociales digitales, podría proponerse la participación de expertos que emitan juicios sobre los diversos perfiles y canales introduciendo estos datos de evaluación en el algoritmo de clasificación. De esta manera, los listados ofrecidos por las páginas que ofrecen herramientas como lectores o compiladores web de redes sociales pasarían un filtro metodológico que mejoraría considerablemente los resultados actuales.

Así, el problema en cuanto al establecimiento de la muestra en los estudios de redes sociales ha sido enfrentado por las investigaciones a lo largo de la literatura utilizando diversas técnicas, y debemos reconocer que, al día de hoy, no se ha llegado a solucionar. Por ello, es destacable la propuesta planteada en este trabajo que mejoraría considerablemente la utilización de lectores o compiladores web de redes sociales tecnológicas.

Actualmente parece que nos encontramos ante un callejón sin salida, ya que no existe la posibilidad real de obtener listados completos y rigurosos sobre diversas temáticas o áreas del saber en las redes sociales digitales. Sin embargo, y como hemos podido comprobar en este trabajo, la investigación va buscando formas creativas de mejorar este problema metodológico y solventar, al menos en parte, esta dificultad (Flores-Márquez & González Reyes, 2021). La mirada ahora está puesta en el futuro, donde tal vez se abra dicha posibilidad a través de la mejora de alguna de las opciones utilizadas por la investigación sobre redes sociales a lo largo de los años o de la aparición de una nueva posibilidad que entre en escena debido a los avances en innovación tecnológica, donde se posiciona como un elemento a tener en cuenta el uso de la inteligencia artificial.

CONCLUSIONES

Se puede afirmar que existe un problema metodológico en cuanto al establecimiento de la muestra en los estudios de las redes sociales digitales que utilizan perfiles, cuentas o canales. A día de hoy no existe la posibilidad de obtención de listados por temáticas o áreas disciplinares con el rigor adecuado y, por lo tanto, las investigaciones que se han adentrado en este tipo de estudios de redes sociales han tenido que barajar diversas opciones metodológicas, que se han categorizado en este artículo.

En primer lugar se encuentra el estudio de caso, bastante usado por la literatura debido a la facilidad de acceso a uno o varios perfiles o canales y a la posibilidad de elegir dicho estudio de caso por conveniencia. Aunque aporta datos de valor, es importante señalar la dificultad de generalización de resultados inherente a este planteamiento metodológico.

En segundo lugar está la utilización de la muestra de otras investigaciones, lo que conlleva una facilidad del establecimiento de la muestra debido a que ya está disponible en las publicaciones de referencia. Sin embargo, esta opción metodológica hace depender el rigor metodológico del proceso de obtención de la muestra de dichas publicaciones.

En tercer lugar, la delimitación de una pequeña muestra. La imposibilidad de establecer mecanismos de búsqueda con herramientas

de compilación web de redes sociales antes de que estas existiesen, junto con la inexistencia de listados oficiales fiables, han impulsado esta posibilidad metodológica de obtener una pequeña muestra de una manera sencilla y por conveniencia. Pero, precisamente por ello, existe una limitación en la generalización de resultados.

En cuarto lugar, la delimitación de una gran muestra que genera una viabilidad de generalización de resultados. Cuanto más grande sea la muestra más se evitará caer en desviaciones propias de la utilización de pequeñas muestras. No obstante, se debe considerar la dificultad de búsqueda y la necesidad de destinar un mayor tiempo en el proceso de indagación.

En quinto lugar se ubica el uso de lectores o compiladores web de redes sociales que propone una posibilidad replicable y sistemática de obtener muestras de una manera sencilla. A pesar de ello, es importante señalar que, al día de hoy, la categorización de las redes sociales digitales y de estas herramientas es deficiente. Existe, además, una continua actualización de los contenidos, lo que provoca una necesaria estipulación del momento en el que se ha realizado la búsqueda. Y, por último, y teniendo en cuenta estas dos cosas, no podemos olvidar que los listados ofrecidos en estas páginas no son ni lo completos ni lo precisos que una investigación académica pretende. Por ello proponemos dos posibles mejoras de esta posibilidad metodológica: una mayor categorización en los perfiles y cuentas de las redes sociales digitales, y la participación de expertos que filtren dichos listados para ofrecer datos más rigurosos.

La limitación de este estudio recae en la ingente cantidad de investigaciones sobre redes sociales que se han desarrollado a lo largo de la literatura. Nuestra categorización de las posibilidades y planteamientos que se han realizado para intentar solventar la problemática del establecimiento de la muestra en dichas investigaciones puede ser completada teniendo en cuenta algún nuevo aspecto no considerado en este artículo. Como prospectiva de este trabajo, se propone la realización de entrevistas a investigadores e investigadoras que han utilizado durante su carrera académica varias de las posibilidades categorizadas en este artículo. Nadie mejor que ellos para profundizar, todavía más si cabe, en las ventajas y desventajas de cada una de las opciones metodológicas.

Referencias bibliográficas

- Beltrán-Pellicer, P., Giacomone, M. B. & Navarro, M. B. (2018). On-line educational videos according to specific didactics: The case of Mathematics. *Culture and Education*, 30(4), 633-662. <https://doi.org/10.1080/11356405.2018.1524651>
- Brooks, H., Devereux-Fitzgerald, A., Richmond, L., Bee, P., Lovell, K., Caton, N., Cherry, M. G., Edwards, B. M., Downs, J., Bush, L., Vasilev, I. & Young, B. (2022). Assessing the effectiveness of social network interventions for adults with a diagnosis of mental health problems: a systematic review and narrative synthesis of impact. *Social Psychiatry and Psychiatric Epidemiology*, 57(5), 907-925. <https://doi.org/10.1007/s00127-022-02242-w>
- Buitrago, A. & Torres-Ortiz, L. (2022). Influencers de ciencia en Twitch. Divulgación científica a través de vídeo-streaming en tiempos de COVID-19. *Teknokultura*, 19(2), 165-176. <https://doi.org/10.5209/tekn.77941>
- Burgos, M., Beltrán-Pellicer, P. & Godino, J. D. (2020). La cuestión de la idoneidad de los vídeos educativos de matemáticas. *Revista Española de Pedagogía*, 78(275), 27-50. <https://doi.org/10.22550/rep78-1-2020-07>
- Del Sol Fabregat, L. A., Tejada Castañeda, E. & Mirabal Díaz, J. M. (2017). Los métodos teóricos: una necesidad de conocimiento en la investigación científico-pedagógica. *Edumecentro*, 9(4), 250-253. <https://revedumecentro.sld.cu/index.php/edumc/article/view/934>
- De Souza, S. T. & Costa, P. H. (2021). Educommunicative culture and new narrative creations. The use of video and storytelling language in an educative context. *Indagatio Didactica*, 13(3), 215-232. <https://doi.org/10.34624/id.v13i3.25539>
- Dos Santos Galvão, N. M. (2019). The YouTube as a study tool for the sufficiency test in accounting. *Revista EDaPECI*, 19(2), 76-84. <https://periodicos.ufs.br/edapeci/article/view/10738>
- Eslen-Ziya, H. (2022). Humour and sarcasm: expressions of global warming on Twitter. *Humanities and Social Sciences Communications*, 9(1), 1-8. <https://doi.org/10.1057/s41599-022-01236-y>
- Facioli, J., Freguglia, R. D. S., Sirqueira, T. F. M., Vieira, M. D. T. (2022). Social networks effects on outcomes of government pro-

- grams: a systematic review. *Brazilian Journal of Political Economy*, 42(1), 222-243. <https://doi.org/10.1590/0101-31572022-3056>
- Feijoo Fernández, B. & Fernández-Gómez, E. (2021). Niños y niñas influyentes en YouTube e Instagram: contenidos y presencia de marcas durante el confinamiento. *Cuadernos.info*, (49), 300-328. <https://doi.org/10.7764/cdi.49.27309>
- Fernández, V. F., Suárez, M. Á. S. & Machas, R. C. (2021). Communication in museums through social media during the pandemic: unveiling new opportunities for interaction. *Fonseca, Journal of Communication*, 23, 129-149. <https://doi.org/10.14201/fjc202123129149>
- Fernández-Peña, R., Orvalle-Perandones, M. A., Marqués-Sánchez, P., Ortego-Maté, C. & Serrano-Fuentes, N. (2022). The use of social network analysis in social support and care: a systematic scoping review protocol. *Systematic Reviews*, 11(1), 1-6. <https://doi.org/10.1186/s13643-021-01876-2>
- Flores-Márquez, D. & González Reyes, R. (2021). En busca de coordenadas metodológicas para estudiar la cultura digital. En D. Flores-Márquez & R. González Reyes (Coords.), *La imaginación metodológica. Coordenadas, rutas y apuestas para el estudio de la cultura digital* (pp. 15-23). Tintable.
- Hassan, F. & Behadili, S. F. (2022). Modeling Social Networks using Data Mining Approaches-Review. *Iraqi Journal of Science*, 63(3), 1313-1338. <https://doi.org/10.24996/ijis.2022.63.3.35>
- 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), 15-38. <https://doi.org/10.22430/22565337.685>
- Jones, R., Chik, A. & Hafner, C. (Eds.). (2015). *Discourse Analysis and Digital Practices*. Routledge.
- Kemp, S. (2022). *Digital 2022 Global Overview Report*. <https://bit.ly/3OVJRYYP>
- Kumar, N., Oles, W., Howell, B., Janmohamed, K., Lee, S. T., Funaro, M. C., O'Connor, P. G. & Alexander, M. (2021). The role of social network support in treatment outcomes for medication for

- opioid use disorder: A systematic review. *Journal of Substance Abuse Treatment*, 127, 1-9. <https://doi.org/10.1016/j.jsat.2021.108367>
- López, L., Maza-Córdoba, J. & Tusa, F. (2020). Educar en el contexto digital: el reto de ser edutuber. *Revista Ibérica de Sistemas e Tecnologías de Informação*, (E25), 188-200. <https://bit.ly/3qnyHBS>
- Mahmoudi, A., Yaakub, M. R. & Abu Bakar, A. (2018). A new method to discretize time to identify the milestones of online social networks. *Social Network Analysis and Mining*, 8(1), 1-20. <https://doi.org/10.1007/s13278-018-0511-4>
- Mahmoudi, A., Yaakub, M. R. & Abu Bakar, A. (2020). A new real-time link prediction method based on user community changes in online social networks. *The Computer Journal*, 63(3), 448-459. <https://doi.org/10.1093/comjnl/bxz050>
- Marcelo, C. & Marcelo, P. (2021). Educational influencers on Twitter. Analysis of hashtags and relationship structure. *Comunicar*, 68, 73-83. <https://doi.org/10.3916/C68-2021-06>
- Martorell, C. & Serra Folch, C. (2018). La presencia de las ONG en Instagram: estudio de caso de Proactiva Open Arms. *Tripodos*, 43, 83-97. <http://hdl.handle.net/20.500.12328/1363>
- Moreira Queiroga, T. & Spyer Dulci, T. M. (2019). “Teachers you-tubers”: Analysis of three channels of YouTube dedicated to the teaching of History. *Escritas do Tempo*, 1(1), 4-29. <https://doi.org/10.47694/issn.2674-7758.v1.i1.2019.0429>
- Mucha-Hospinal, L. F., Chamorro-Mejía, R., Oseda-Lazo, M. E. & Alania-Contreras, R. D. (2021). Evaluación de procedimientos empleados para determinar la población y muestra en trabajos de investigación de posgrado. *Desafíos*, 12(1), 44-51. <https://doi.org/10.37711/desafios.2021.12.1.253>
- Ojeda-Serna, V. & García-Ruiz, R. (2022). Divulgación científica en YouTube en Latinoamérica. Estudio de Casos de universidades, museos y YouTubers. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias*, 19(2), 2204-2204. https://doi.org/10.25267/rev_eureka_ensen_divulg_cienc.2022.v19.i2.2204
- Osman, W., Mohamed, F., Elhassan, M. & Shoufan, A. (2022). Is YouTube a reliable source of health-related information? A sys-

- tematic review. *BMC Medical Education*, 22(1), 1-12. <https://doi.org/10.1186/s12909-022-03446-z>
- Pattier, D. (2021). The gender gap among EduTubers and the factors significantly influencing it. *Journal of New Approaches in Educational Research*, 10(2), 313-329. <https://doi.org/10.7821/naer.2021.7.732>
- Phua, J., Jin, S. V. & Kim, J. J. (2016). Gratifications of Using Facebook, Twitter, Instagram, or Snapchat to Follow Brands: The Moderating Effect of Social Comparison, Trust, Tie Strength, and Network Homophily on. *Telematics and Informatics*, 34(1), 412-424. <https://doi.org/10.1016/j.tele.2016.06.004>
- Pihlaja, S. (2022). Data sampling and digital discourse. En C. Vásquez (Ed.), *Research methods for digital discourse analysis* (pp. 55-71). Bloomsbury Publishing.
- Piña-García, C. A., Gershenson, C. & Siqueiros-García, J. M. (2016). Towards a standard sampling methodology on online social networks: collecting global trends on Twitter. *Applied Network Science*, 1(3), 1-19. <https://doi.org/10.1007/s41109-016-0004-1>
- Prochnow, T. & Patterson, M. S. (2022). Assessing Social Network Influences on Adult Physical Activity Using Social Network Analysis: A Systematic Review. *American Journal of Health Promotion*, 36(3), 537-558. <https://doi.org/10.1177/08901171211060701>
- Rocheva, A., Varshaver, E. & Ivanova, N. (2022). Targeting on Social Networking Sites as Sampling Strategy for Online Migrant Surveys: The Challenge of Biases and Search for Possible Solutions. En S. Pöttschke & S. Rinken (Eds.), *Migration Research in a Digitized World*. IMISCOE Research Series, Springer, Cham. https://doi.org/10.1007/978-3-031-01319-5_3
- Saurabh, S. & Gautam, S. (2019). Modelling and statistical analysis of YouTube's educational videos: A channel Owner's perspective. *Computers & Education*, 128, 145-158. <https://doi.org/10.1016/j.compedu.2018.09.003>
- Sirisala, N., Yarava, A., Reddy, Y. P. & Poola, V. (2022). A novel trust recommendation model in online social networks using soft computing methods. *Concurrency and Computation: Practice and Experience*, 34(22), e7153. <https://doi.org/10.1002/cpe.7153>
- Tindall, D., McLevey, J., Koop-Monteiro, Y. & Graham, A. (2022). Big data, computational social science, and other recent innovations in

- social network analysis. *Canadian Review of Sociology/Revue Canadienne de Sociologie*, 59(2), 271-288. <https://doi.org/10.1111/cars.12377>
- Tobón Berrio, L. E., Sabatier, C., Palacio Sañudo, J. E. & Navarro, O. (2021). The social representation of the rights of children: legality built by fathers and mothers. *Revista Colombiana de Sociología*, 44(2), 243-265. <https://doi.org/10.15446/rcs.v44n2.87666>
- Vásquez, C. (Ed.). (2022). *Research methods for digital discourse analysis*. Bloomsbury Publishing.
- Vatrapu, R., Mukkamala, R. R., Hussain, A. & Flesch, B. (2016). Social set analysis: A set theoretical approach to big data analytics. *IEEE Access*, 4, 2542-2571. <https://doi.org/10.1109/access.2016.2559584>
- Vizcaíno-Verdú, A. & Contreras Pulido, P. (2019). From the classroom to the universal online school: Thematic dimensions on YouTube channels. *HAMUT'AY*, 6(3), 12-25. <http://dx.doi.org/10.21503/hamu.v6i3.1841>
- Wang, C., Tianqing, Z., Xiong, P., Ren, W. & Choo, K. K. R. (2022). A privacy preservation method for multiple-source unstructured data in online social networks. *Computers & Security*, 113, 102574. <https://doi.org/10.1016/j.cose.2021.102574>
- Zaragoza, J. C. & Roca Marín, D. (2020). El movimiento youtuber en la divulgación científica española. *Revista Prisma Social*, (31), 212-238. <https://revistaprismasocial.es/article/view/3942>
- Zhang, C., Sun, J., Zhu, X. & Fang, Y. (2010). Privacy and security for online social networks: challenges and opportunities. *IEEE Network*, 24(4), 13-18. <https://doi.org/10.1109/mnet.2010.5510913>

DANIEL PATTIER

Es doctor en educación y profesor ayudante en el Departamento de Estudios Educativos de la Facultad de Educación, Centro de Formación del Profesorado de la Universidad Complutense de Madrid. Su principal línea de investigación se centra en el estudio de la utilización de las redes sociales en el ámbito educativo. Es autor de diversos artículos científicos publicados en revistas indexadas.