

De *Twitter* a X. El riesgo de la dependencia de fuentes privadas para la investigación

From *Twitter* to X. The risk of reliance on private sources for research

Teresa Piñeiro-Otero

Piñeiro-Otero, Teresa (2023). "De *Twitter* a X. El riesgo de la dependencia de fuentes privadas para la investigación". *Anuario ThinkEPI*, v. 17, e17a46.

<https://doi.org/10.3145/thinkepi.2023.e17a46>

Publicado en *IweTel* el 15 de diciembre de 2023

Teresa Piñeiro-Otero

<https://orcid.org/0000-0001-6414-2700>

<https://www.directorioexit.info/ficha2086>

Universidade da Coruña

Facultad de Ciencias da Comunicación

Campus de Elviña

15074 A Coruña, España

teresa.pineiro@udc.es



Resumen: La investigación académica está profundamente arraigada en su contexto social, económico, político y cultural. En un momento de expansión y diversificación de la esfera digital las plataformas sociales, en especial *Twitter*, suponen herramientas clave para la investigación comunicativa. La evolución de la API de *Twitter* ha facilitado significativamente la recopilación de grandes datos para la investigación comunicativa. Sin embargo, la reciente limitación en el acceso gratuito de dicha herramienta plantea, con el cambio de las políticas corporativas de X, desafíos importantes que exigen repensar la dependencia de fuentes privadas, así como el desarrollo de estrategias adaptativas en investigación.

Palabras clave: *Twitter*; API; Investigación; Dependencia de fuentes; Data.

Abstract: Academic research is deeply rooted in its social, economic, political, and cultural contexts. At a time of expansion and diversification of the digital sphere, social platforms, especially *Twitter*, are key tools for communication research. The evolution of the *Twitter* API has significantly facilitated the collection of big data for communication research. However, the recent limitation of free access to this tool poses, with the change in X's corporate policies, important challenges that require rethinking dependence on private sources as well as the development of adaptive research strategies.

Keywords: *Twitter*; API; Research; Source dependence; Data.

1. *Twitter*¹ como fuente de información para la investigación comunicativa

La investigación no es una actividad aislada; está profundamente arraigada en un contexto social, económico, político y cultural en el que se produce.

Esta interconexión se ha hecho más evidente en la era digital. El auge de las plataformas sociales, la multiplicación de los flujos de información –y desinformación–, la mediamorfosis, la convergencia de medios y contenidos, o las transformaciones en su acceso y consumo han tenido proyección en los intereses de la comunidad investigadora.

Basta con tomar una instantánea de la producción científica para comprobar el cambio en la fisonomía de la investigación comunicativa, de sus principales temas y enfoques, en la última década.

La comunidad investigadora también ha sido sensible al impacto de movimientos como el 15M o #MeToo, de campañas políticas como el Brexit, del auge de los populismos, la crisis climática y, por supuesto, de emergencias de salud pública. Así, la pandemia por el Covid-19 suscitó una producción científica sin precedentes, reflejo de la diversidad y urgencia de los desafíos planteados. En este contexto requieren especial atención las plataformas sociales como espacios de interacción, soporte o manifestación de muchos de los fenómenos analizados.

Los social media han transformado radicalmente el ámbito de la comunicación. El incremento en el número y tipo de actores implicados en la conversación digital permite evadir los marcos y representaciones impuestos por los medios y emisores clásicos para dar lugar a un nuevo entorno abierto, descentralizado y horizontal –al menos en apariencia.

Esta expansión de la esfera pública a las plataformas sociales presenta numerosas potencialidades para la comunicación política y el activismo, pero también innumerables retos: filtros burbuja, cámaras eco, *bots* y *trolls*, radicalización algorítmica, *astroturfing*... La comunicación política digital en su sentido más amplio –desde la política institucional, partidista o electoral al activismo y la participación de la ciudadanía– ha vivido un importante impulso por y gracias a *Twitter*, convirtiéndose en una de las líneas con mayor pujanza en la actual investigación comunicativa.

Desde 2008 *Twitter* se ha consolidado como herramienta esencial para la investigación académica de dinámicas sociales, movimientos políticos o tendencias de opinión pública. Su estudio ha evolucionado de la presencia y enfoques tecnodeterministas hacia aproximaciones críticas, a realidades más complejas que han mostrado fisuras en su reflejo de la esfera pública, como su carácter predictor (**Campos-Domínguez**, 2017).

Aun cuando *Facebook* presenta una comunidad usuaria mayor, *Twitter* ha acaparado la atención de la comunidad científica. Para **Casero-Ripollés** (2018) esta atención responde a:

- enfoque noticioso e informativo de la plataforma –tres de cada cuatro usuarios en España la consideran de utilidad para seguir la actualidad política (*Twitter*, 2019)–;
- su concentración de personalidades de la política y el periodismo, que sitúan a la plataforma en el epicentro de la influencia política;
- su percepción como termómetro de la opinión pública.

El carácter público de *Twitter*, la estructura de organización y consumo de información, y la accesibilidad de los datos han situado a la plataforma de *microblogging* como base de un importante corpus de investigaciones.

Las estrategias de acceso y recuperación de datos de dichos trabajos basculan de la rudimentaria recolección de *tweets* a través de consultas en la propia plataforma (palabras clave, cuentas, *hashtags*...) al acceso a través de interfaces de programación de aplicaciones (API²) bien de forma directa, bien empleando servicios de terceros (*twittonomy*, *tweetreach*...).

Desde el lanzamiento de su primera API –en 2006– *Twitter* ha previsto el acceso a diversos datos de perfiles e interacciones que han facilitado el estudio de la influencia y patrones de comportamiento en la plataforma, haciendo visibles los lazos e interrelaciones entre nodos (**Martínez-Rolán; Piñero-Otero**, 2017). Estos grandes datos han permitido entender mejor la formación de la opinión pública, la articulación de discursos de odio o la propagación de información-desinformación, entre otros.

La progresión de versiones de esta API ha influenciado significativamente la recopilación de datos de *Twitter* por la comunidad investigadora, pero también por profesionales y activistas.

Tabla 1. Principales versiones de la API de *Twitter*. Elaboración propia a partir de **Congosto, Basanta-Val y Sánchez Fernández** (2017).

Año	Versión de la API	Principales cambios/características
2006	API inicial	Acceso básico a tweets y perfiles de usuarios; posibilidad de publicar <i>tweets</i> .
2010	API v1	Mejoras en la autenticación; incorporación de nuevas funciones como la búsqueda de <i>tweets</i> y límites de tasa más estructurados.
2013	API v1.1	Introducción de la autenticación OAuth 2.0; límites de tasa más estrictos y eliminación de ciertas funcionalidades no autenticadas. En actualizaciones posteriores se introducirían nuevas características como los <i>endpoint</i> para datos de audiencia y compromiso o mejoras en la recuperación de datos históricos.
2017	API Premium	Acceso a datos más detallados e históricos, mejores capacidades de búsqueda, planes pagados para un acceso más amplio.
2020	API v2	Introducción de nuevas funciones como el acceso a conversaciones completas, mejoras en la búsqueda y en la transmisión de datos, y un enfoque más modular.
2021	API Académica	Acceso extendido a datos históricos para investigadores académicos, mayor capacidad para el análisis de grandes conjuntos de datos y el seguimiento de tendencias a largo plazo.

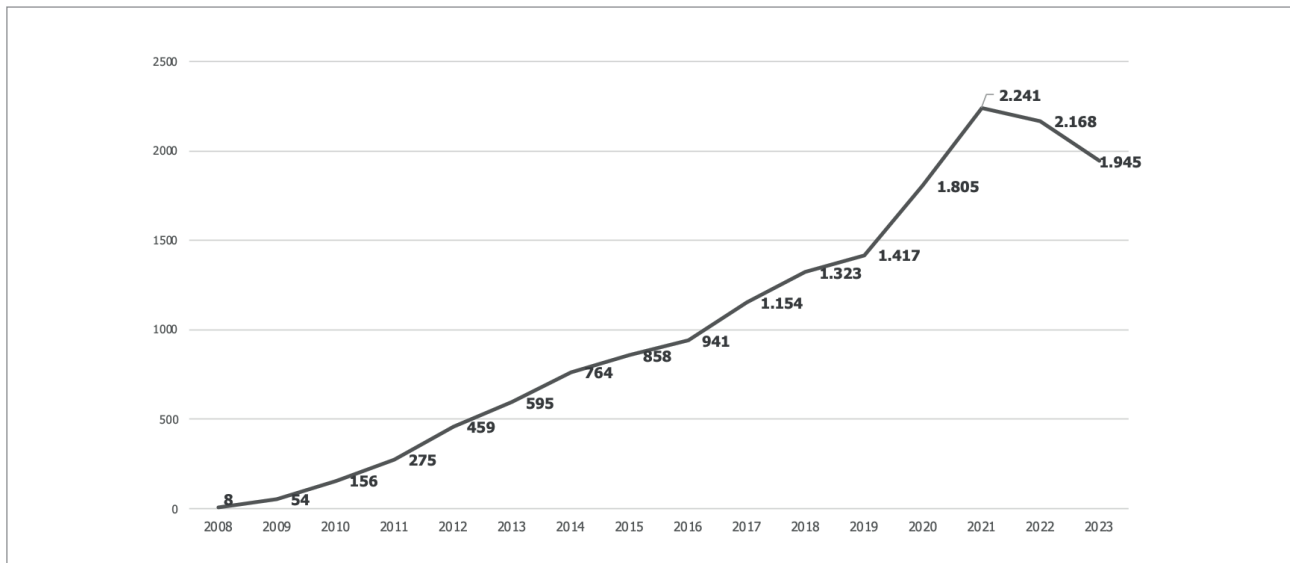


Figura 1. Evolución de la investigación comunicativa basada en *Twitter*. *Dimensions* categoría Communication & Media.

En enero de 2021 *Twitter* lanzó su API Académica. Esta API, diseñada específicamente para apoyar la investigación, proveía acceso a un conjunto detallado de datos:

- históricos desde 2006;
- *tweets* en tiempo real y archivados;
- *tweets* completos –texto y multimedia–;
- metatados;
- *engagement* (*retweets*, favoritos, menciones...);
- filtros avanzados de búsqueda que proveían un acceso a un conjunto de datos sociales masivo y en constante evolución, de gran valor para el estudio de diversos fenómenos comunicativos.

Aun cuando el acceso a la información presentaba limitaciones como la recuperación de *tweets* borrados o eliminados –por ejemplo los mensajes tóxicos contra mujeres periodistas y activistas suprimidos por la aplicación *Harrassment Manager* de Jigsaw– o metadatos vinculados al momento de recogida (no de emisión), el desarrollo de la API académica ofrecía un acceso sin parangón a los datos de *Twitter* (visibles y no visibles) que hacía presagiar un salto cualitativo en la investigación, más su impacto fue limitado.

2. El inicio y fin de una era

En octubre de 2022 Elon Musk compró *Twitter*. Esta adquisición marcó un punto inflexión en la gestión de la plataforma, tanto en su funcionamiento interno como en la apertura de la información y su relación con la Academia. El cambio en el liderazgo de *Twitter* suscitó un intenso debate sobre el equilibrio entre la libertad de expresión, la prevención de la radicalización y la propagación de desinformación, aún hoy inconcluso.

En febrero de 2023 *Twitter* anunció el fin del soporte gratuito de su API para proveer un acceso básico de pago. Este paso afectó a la investigación académica, pero también al quehacer de periodistas, verificadores de noticias y activistas. El importante volumen de datos disponibles a través de la API de *Twitter* proveía una información de gran valor para investigar y rastrear discursos de odio, desinformación, interacción entre personas usuarias, movimientos y tendencias, entre otros.

La decisión de la plataforma social de acotar el acceso a la información ha marcado y marcará el devenir de la investigación, especialmente en comunicación digital. La misma que había impulsado.

Desde 2020 *Dimensions* ha publicado un total de 35.194 estudios con datos de *Twitter*, 10.286 de los cuáles señalan específicamente el uso de la API.

Tras el anuncio del cierre, el acceso a los datos se fue acotando poco a poco, lastrando el devenir de investigaciones y otros proyectos. Un ejemplo fue *t-hoarder R*, un conjunto de herramientas elaborado por la investigadora Mariluz Congosto para facilitar el análisis y visualización de la información

“*Twitter* se ha consolidado como herramienta esencial para la investigación de dinámicas sociales, movimientos políticos o tendencias de opinión pública”

a la comunidad investigadora sin conocimientos de programación, subido a *GitHub* en febrero de 2023 (**Congosto, 2023**).

No ha sido el único.

En julio de 2022, presentamos un proyecto³ a la convocatoria de subvenciones investigación del *Instituto de las Mujeres*, para mapear las desigualdades de género en la comunicación política en *Twitter*.

En un momento en el que la API Académica permitía la recuperación de 10 millones de *tweets* mensuales, planteamos un estudio de la conversación en torno a 297 representantes en el *Congreso de los Diputados* con presencia en *Twitter*.

Planteábamos desarrollar un algoritmo para automatizar la detección de discurso de odio en *Twitter* en la línea de *Automatic Misogyny Identification* de **Fersini, Nozza y Rosso (2008)**, además de un análisis de redes sociales (ARS) para visualizar el grado de cohesión e integración de cada nodo en la conversación social.

Su puesta en marcha coincidió con el cierre de la API por lo que se tuvieron que emplear técnicas de raspado web (*web scraping*) para la extracción de datos, con un máximo de 300 *tweets* diarios y 40 por perfil. Otras restricciones fueron el acceso únicamente a las publicaciones emitidas por representantes políticos –no a las respuestas– y la imposibilidad de recuperar contenidos multimedia habituales en la conversación online (imágenes, vídeos, memes...).

Solo un dato: en el intervalo del 28 de mayo al 23 de julio (entre las elecciones autonómicas-municipales y generales) se recuperaron 25.978 publicaciones de 297 representantes políticos. En un estudio previo sobre veinte perfiles políticos –mujeres y varones– se recuperaron más de 421.000 *tweets* (entre publicaciones y respuestas a las publicaciones) para el mismo período en 2021 (**Martínez-Rolán et al. 2021**).

En un 2023 marcado por las citas electorales no resulta complejo evaluar el coste de oportunidad perdida, máxime teniendo en cuenta que los datos recabados ofrecen una perspectiva jerárquica, unidireccional –sin presencia de la ciudadanía– de *Twitter*.

Frente al acceso y democratización de datos propugnado por *Twitter*, una de las plataformas más activa en el intercambio de datos con la comunidad investigadora, X ha vuelto a recuperar para las élites el acceso a la información. No es una frase hecha. El plan de precios de su API parte de 42.000\$ mensuales para acceder a 50 millones de *tweets* (**Pfeffer et al., 2023**, sitúan las publicaciones de X en 350 millones diarias).

“El desarrollo de la API académica ofrecía un acceso sin parangón a los datos de *Twitter* que hacía presagiar un salto cualitativo en la investigación”



Figuras 2 y 3. Reacciones ante el cierre definitivo de la API académica y la V1.1. @SMLabTO y @congosto.

Musk ha justificado este cierre como parte de su lucha contra la manipulación de opinión, ante la falta de verificación de la API. La realidad es que deja a oscuras a la comunidad científica, activistas y *factcheckers* para desenmarañar esta manipulación.

“Frente al acceso y democratización de datos propugnada por *Twitter*, *X* ha vuelto a recuperar para las élites el acceso a la información.”

3. Consideraciones finales

La evolución de *Twitter* a *X* subraya un desafío crucial en la investigación contemporánea: el complejo equilibrio entre la dependencia de datos controlados por entidades privadas y la integridad y accesibilidad del conocimiento científico.

En su análisis de la investigación sobre información política y redes sociales, **Casero-Ripollés** (2018) incidía en la elevada dependencia de *Twitter* y señalaba tres riesgos derivados:

- exceso de atención que obvie innovaciones transformadoras en otras redes sociales;
- el intento de extrapolar sus hallazgos a otras plataformas;
- que la orientación a las campañas electorales enmascare otros fenómenos de gran influencia en el contexto actual.

Al final, el principal riesgo derivó de la dependencia de los datos de una entidad privada, sujetos a los cambios en la dirección y en las políticas corporativas.

La preocupación respecto a la dependencia de datos controlados por entidades privadas no es nueva. **Hurst y Mauron** (2008) alertaban de que los intereses comerciales pueden influir en la dirección de la investigación. Lo hemos visto.

En el contexto de la ciudadanía digital, en que las empresas privadas juegan un papel cada vez más importante como compiladoras de las ingentes cantidades de datos que deja nuestra huella digital, la investigación comunicativa adolece de una gran dependencia de datos corporativos. Las tensiones éticas y metodológicas de esta relación hacen preciso apostar por un espacio sostenible para estas interacciones entre las corporaciones y Academia (**Favaretto et al.** 2023).

En este sentido, **Campos-Domínguez y Díez-Garrido** (2023) señalan la preocupación investigadora por la transparencia de las grandes corporaciones digitales. No debemos obviar que las redes sociales son un producto sociotécnico, fruto de las relaciones sociales de quienes las producen y quienes las utilizan y, por tanto, sus datos no deberían equipararse a otra información corporativa.

Mientras tanto, hemos tenido que revisar enfoques y estrategias de investigación para hacer frente a la incertidumbre y desafíos de la política de acceso de *X*. Acostumbrada a interpretar el lenguaje de *Twitter*, la comunidad investigadora comienza a introducirse en plataformas alternativas y a buscar nuevas fórmulas para mapear esta esfera digital, pero deberíamos tener presentes las consecuencias de depender de una única plataforma.

Parafraseando a Job: *Twitter* me lo dio, *X* me lo quitó. Loado sea *Twitter*.

“En el contexto de la ciudadanía digital, en que las empresas privadas se han convertido en compiladoras de registros de nuestra huella digital, la investigación comunicativa adolece de una gran dependencia de datos corporativos”

4. Notas

1. La diferencia entre *Twitter* y *X* es consciente y buscada.
2. Una API, o Interfaz de Programación de Aplicaciones, es un conjunto de reglas y definiciones que permite que diferentes softwares o aplicaciones se comuniquen entre sí y compartan funcionalidades y datos.
3. *Agitate*. Asimetrías de género en la comunicación política digital. Prácticas, estructuras de poder y violencias en la tuitesfera española, financiado por el Instituto de las Mujeres.
<https://agitate.udc.es>

5. Referencias

Campos-Domínguez, Eva (2017). “*Twitter* y la comunicación política”. *El profesional de la información*, v. 26, n. 5, pp. 785-793.
<https://doi.org/10.3145/epi.2017.sep.01>

Campos-Domínguez, Eva; Díez-Garrido, María (2023). “Digital transparency and political communication”. *Profesional de la información*, v. 32, n. 1, e320104.
<https://doi.org/10.3145/epi.2023.ene.04>

Casero-Ripollés, Andreu (2018). "Research on political information and social media: Key points and challenges for the future". *El profesional de la información*, v. 27, n. 5, pp. 964-974.
<https://doi.org/10.3145/epi.2018.sep.01>

Congosto, Mariluz (2023). *t-hoarder_R*.
https://github.com/congosto/t-hoarder_R

Congosto, Mariluz; Basanta-Val, Pablo; Sánchez-Fernández, Luis (2017). "T-hoarder: A framework to process Twitter data streams". *Journal of network and computer applications*, v. 83, pp. 28-39.
<https://doi.org/10.1016/j.jnca.2017.01.029>

Favaretto, Madalena; De-Clercq, Eva; Caplan, Artur; Elger, Bernice-Simone (2023). "United in big data? Exploring scholars' opinions on academic-industry partnership and the use of corporate data in digital behavioral research". *PLoS one*, v. 18, n. 1, e0280542.
<https://doi.org/10.1371/journal.pone.0280542>

Fersini, Elisabetta; Nozza, Debora; Rosso, Paolo (2018). "Overview of the Evalita 2018 task on automatic misogyny identification (AMI)". In: *Proceedings of the sixth evaluation campaign of natural language processing and speech tools for Italian*. Final workshop (Evalita 2018).
<https://ceur-ws.org/Vol-2263/paper009.pdf>

Hurst, Samia; Mauron, Alex (2008). "A question of method. The ethics of managing conflicts of interest". *EMBO reports*, v. 9, pp. 119-123.
<https://doi.org/10.1038/sj.embor.2008.4>

Martínez-Rolán, Xabier; Piñeiro-Otero, Teresa; Castro-Souto, Laura; Padilla-Castillo, Graciela; Barrientos, Almudena; Sánchez-Vázquez, Clara (2021). *Mulleres pontevedresas na rede II. Análise do discurso de odio e outras violencias contra xornalistas e políticas*.
<https://micromachismos.webs.uvigo.es/proxecto2021/>

Martínez-Rolán, Xabier; Piñeiro-Otero, Teresa (2017). "Lazos invisibles de la comunicación política. Comunidades de partidos políticos en Twitter en unas elecciones municipales". *El profesional de la información*, v. 26, n. 5, pp. 859-870.
<https://doi.org/10.3145/epi.2017.sep.08>

Pfeffer, Jürgen; Matter, Daniel; Jaidka, Kokil; Varol, Onur; Mashhadi, Afra; Lasser, Jana; Assenmacher, Dennis; Wu, Siq; Yang, Diy; Branthner, Cornelia; Romero, Daniel; Otterbacher, Jahna; Schwemmer, Carsten; Joseph, Kenneth; García-García, David; Morstatter, Fred (2023). "Just another day on Twitter: A complete 24 hours of Twitter Data". arXiv:2301.11429v2 [cs.SI], 11 Apr. <https://doi.org/10.48550/arXiv.2301.11429>

Twitter (2019). "Twitter es la red donde la información política tiene mayor relevancia". *Twitter blog*, 9 abril.
<https://bit.ly/3RMCUhH>

Si te interesan los

INDICADORES EN CIENCIA Y TECNOLOGÍA,

y todos los temas relacionados con la medición de la ciencia, tales como:

Análisis de citas, Normalización de nombres e instituciones, Impacto de la ciencia en la sociedad, Indicadores, Sociología de la ciencia, Política científica, Comunicación de la ciencia, Revistas, Bases de datos, Índices de impacto, Políticas de open access, Análisis de la nueva economía, Mujer y ciencia, etc.

Entonces **INCYT** es tu lista. Suscríbete en:

<http://www.rediris.es/list/info/incyt.html>