

Del dato al grafo: rigor, ética y replicabilidad en el análisis de redes sociales en comunicación y ciencia política en América Latina (2009-2022)*

From Data to Graph: Rigor, Ethics, and Replicability in Social Network Analysis of Communication and Political Science in Latin America (2009-2022)

Dos dados ao gráfico: rigor, ética e replicabilidade na análise de redes sociais em comunicação e ciência política na América Latina (2009-2022)

Juan Federico Pino-Urbe^a
Facultad Latinoamericana de Ciencias Sociales
(FLACSO), Ecuador

jfpino@flacso.edu.ec

ORCID: <https://orcid.org/0000-0003-0799-3001>

DOI: <https://doi.org/10.11144/Javeriana.syp44.dgre>

Recibido: 30 julio 2025

Aceptado: 05 septiembre 2025

Publicado: 30 diciembre 2025

Andrés Lombana-Bermúdez
Pontificia Universidad Javeriana, Colombia
ORCID: <https://orcid.org/0000-0002-1246-5994>

Anaí Oñate-Bolaños
Facultad Latinoamericana de Ciencias Sociales
(FLACSO), Ecuador
ORCID: <https://orcid.org/0000-0002-3676-2301>

Manuel Elkin Castiblanco Briceño
Facultad Latinoamericana de Ciencias Sociales
(FLACSO), Ecuador
ORCID: <https://orcid.org/0009-0000-7940-9594>

Resumen:

Este estudio examina el uso del análisis de redes sociales en investigaciones sobre comunicación y ciencia política en plataformas digitales en América Latina entre 2009 y 2022. A partir de un diseño secuencial de métodos mixtos, se construyó un corpus de 68 artículos seleccionados en Google Scholar. La fase cualitativa incluyó la elaboración y validación de un libro de códigos para clasificar cada estudio según nueve indicadores de transparencia y replicabilidad. Posteriormente, la fase cuantitativa transformó esta información en un índice de claridad metodológica mediante la teoría de respuesta al ítem, lo que permitió comparar tendencias y evaluar fortalezas y debilidades de la producción académica. Los resultados muestran un crecimiento sostenido de publicaciones, con Twitter/X como fuente de datos predominante y un sesgo hacia enfoques cuantitativos (71 %). Sin embargo, se identifican carencias: la mayoría de los estudios no especifica el tipo de red analizada ni documenta procesos de limpieza, manejo ético de datos y publicación abierta de materiales. El artículo propone la ruta de claridad metodológica (RCM), una guía en cuatro etapas para estandarizar prácticas de recolección, análisis, presentación y validación, con el objetivo de fortalecer el rigor, la ética y la replicabilidad en la investigación latinoamericana.

Palabras clave: redes sociales, plataformas digitales, X, Twitter, Facebook, ARS.

Abstract:

This study examines the use of social network analysis in research on communication and political science focused on digital platforms in Latin America between 2009 and 2022. Using a sequential mixed-methods design, we built a corpus of 68 peer-reviewed articles identified on Google Scholar. The qualitative phase involved the development and validation of a coding book to classify each study according to nine indicators of transparency and replicability. Subsequently, the quantitative phase transformed this information into a methodological clarity index using item response theory, allowing for the comparison of trends and the assessment of strengths and weaknesses in the academic production. The results reveal a steady growth of publications, with Twitter/X emerging as the predominant data source (66 out of 72 articles) and a strong bias toward quantitative approaches (71 %). However, significant gaps were identified: most studies fail to specify the type of network analyzed and do not adequately

Notas de autor

^a Autor de correspondencia. Correo electrónico: jfpino@flacso.edu.ec

report processes of data cleaning, ethical handling of information, or open publication of materials. The article proposes the methodological clarity route (RCM), a four-stage guide aimed at standardizing practices of data collection, analysis, presentation, and validation, with the goal of strengthening rigor, ethics, and replicability in Latin American research.

Keywords: Social Networks, Digital Platforms, X, Twitter, Facebook, SNA.

Resumo:

Este estudo examina o uso da análise de redes sociais em pesquisas sobre comunicação e ciência política em plataformas digitais na América Latina entre 2009 e 2022. A partir de um desenho sequencial de métodos mistos, foi construído um corpus de 68 artigos selecionados no Google Scholar. A fase qualitativa incluiu a elaboração e validação de um livro de códigos para classificar cada estudo de acordo com nove indicadores de transparência e replicabilidade. Posteriormente, a fase quantitativa transformou essas informações num índice de clareza metodológica por meio da teoria da resposta ao item, permitindo comparar tendências e avaliar pontos fortes e fracos da produção acadêmica. Os resultados mostram um crescimento sustentado de publicações, com o Twitter/X como fonte de dados predominante e um viés para abordagens quantitativas (71 %). No entanto, foram identificadas lacunas: a maioria dos estudos não especifica o tipo de rede analisada nem documenta processos de limpeza, gestão ética de dados e publicação aberta de materiais. O artigo propõe a rota de clareza metodológica (RCM), um guia em quatro etapas para padronizar práticas de recolha, análise, apresentação e validação, com o objetivo de fortalecer o rigor, a ética e a replicabilidade na investigação latino-americana.

Palavras-chave: redes sociais, plataformas digitais, X, Twitter, Facebook, ARS.

Introducción

En la actualidad, las plataformas digitales (PD) se han convertido en un escenario vibrante para la vida política y comunicacional (Manfredi y González, 2019). Estas plataformas han dejado de ser espacios de interacción marginales para convertirse en complejos entornos, de producción y circulación de significados, altamente interconectados, que afectan la construcción de agendas públicas, la deliberación y las dinámicas de poder (Poell *et al.*, 2022). En el siglo XXI, plataformas como X (antes Twitter), Facebook, TikTok e Instagram alcanzaron una relevancia comparable a arenas de poder como los sistemas judiciales, electorales y mediáticos, de manera que transformaron las dinámicas de comunicación y movilización política. El impacto de estas transformaciones se ha evidenciado en hitos como la campaña presidencial de Barack Obama en 2008, pionera en la integración estratégica de internet y las plataformas digitales para elecciones (Ballard *et al.*, 2016) y en procesos posteriores como las victorias de Donald Trump en Estados Unidos y Jair Bolsonaro en Brasil, donde estas plataformas se consolidaron como herramientas para articular discursos polarizantes y movilizar a la ciudadanía (Matos *et al.*, 2020; Alexandre *et al.*, 2022). En Latinoamérica, las redes sociales (RR. SS.) han desempeñado un papel relevante en episodios de movilización, como las protestas en Ecuador (2019), Chile (2019), Perú (2020) y Colombia (2021), potenciando tanto agendas progresistas como conservadoras y evidencian su carácter dual como espacios de visibilización y confrontación política (Argüello, 2021; Lombana-Bermúdez y Rodríguez Gómez, 2023).

Este doble papel ha intensificado el interés por comprender su funcionamiento y su impacto en las sociedades. Investigaciones recientes han mostrado que las PD afectan dimensiones como el pluralismo informativo, la calidad del debate público y la propagación de desorden informativo (Levendusky y Stecula, 2021; Borge-Holthoefer y González-Bailón, 2017; González-Bailón, 2017; Lazer *et al.*, 2021). Esta situación plantea desafíos para la investigación, que debe lidiar con un volumen de datos sin precedentes y, al mismo tiempo, con dilemas éticos y metodológicos sobre cómo recolectar, analizar y presentar información. En Latinoamérica, estos desafíos son complejos, pues la región enfrenta condiciones como desigualdad, fragilidad institucional y polarización política (Cifuentes y Pino Uribe, 2018; Abadía *et al.*, 2024) que influyen en la producción y consumo de información. Sin embargo, gran parte de la literatura sobre ARS aplicada a estudios en comunicación y ciencia política proviene del norte global, lo que genera marcos analíticos.

Frente a este escenario, las preguntas de este artículo son ¿cómo se ha utilizado el ARS en investigaciones de comunicación y ciencia política en Latinoamérica? y ¿en qué medida estos estudios cumplen con estándares de transparencia, replicabilidad y ética en el manejo de datos? Responder estas preguntas resulta pertinente para comprender y evaluar las fortalezas y debilidades metodológicas de las investigaciones que utilizan ARS. Durante la última década, el número de artículos que emplean ARS ha crecido, pero este aumento cuantitativo no siempre se ha traducido en avances de calidad. Muchos estudios carecen de descripciones sobre cómo se definieron las poblaciones analizadas, cuáles fueron los criterios de recolección y limpieza de datos y cómo se interpretaron los resultados en relación con teorías. Además, la dependencia de herramientas propietarias y el acceso restringido a ciertas API (*application programming interface*, por su sigla en inglés) —como la de X/Twitter, cada vez más limitada tras la compra de la plataforma por Elon Musk— condicionan la posibilidad de obtener datos abiertos y replicables. Mientras otras plataformas, como Facebook, Instagram y YouTube, imponen barreras de acceso aún mayores (González-Bailón, 2017; Lazer *et al.*, 2021).

En este contexto, el ARS se ha consolidado como una herramienta para abordar estas dinámicas digitales. Basado en la teoría de grafos, permite modelar interacciones y visualizar estructuras relacionales, identificando patrones colectivos que no se observan mediante análisis centrados únicamente en actores individuales (Lozares, 2005; González, 2019). Este asume que los sistemas sociales pueden comprenderse como redes de nodos y vínculos, en los que propiedades estructurales como la densidad, la centralidad y la cohesión revelan dinámicas de poder y circulación de información (Carrington *et al.*, 2005; Ramos-Vidal y Ricaurte, 2015; Scott, 2013). El crecimiento exponencial de datos digitales y capacidad computacional ha revitalizado este método, posibilitando investigaciones a gran escala que hace apenas unas décadas eran inviables. No obstante, estas plataformas no son entornos neutrales: su diseño algorítmico y la presencia de actores no humanos, como *bots* y cuentas automatizadas, moldean la visibilidad y jerarquías de interacción e influyen en qué contenidos se vuelven centrales y cuáles permanecen en la periferia (Varol *et al.*, 2019). Ignorar estas mediaciones implica un riesgo: reducir fenómenos comunicacionales y políticos a simples patrones de conexión, sin captar los significados, estrategias y prácticas que los producen (Boyd y Ellison, 2007; Papacharissi y Yuan, 2011).

Este artículo busca ofrecer una respuesta a las preguntas planteadas a través de un análisis de la producción académica sobre ARS en la región entre 2009 y 2022. Para ello, se construyeron y analizaron dos corpus: uno de artículos en español y otro en inglés que estudian fenómenos latinoamericanos. Se privilegiaron publicaciones en revistas indexadas, excluyendo literatura gris como tesis, ponencias y reportes no revisados por pares, con el objetivo de garantizar la calidad y comparabilidad de los hallazgos. A partir de este trabajo, se propone una Ruta de Claridad Metodológica (RCM) que identifica cuatro etapas para el desarrollo de investigaciones rigurosas con ARS: recopilación de datos, análisis, presentación de resultados y validación. Cada etapa incluye indicadores para evaluar la transparencia y replicabilidad y la responsabilidad ética en el uso del ARS (Borge-Holthoefer y González-Bailón, 2017; Lazer *et al.*, 2021).

El artículo se organiza en cuatro secciones: la primera presenta la revisión de la literatura; la segunda describe la metodología empleada en la construcción y análisis de los corpus; la tercera desarrolla la propuesta conceptual y operativa de la RCM, y la cuarta discute los hallazgos del artículo y las recomendaciones para consolidar prácticas que permitan estudios de ARS transparentes y éticos.

Revisión de literatura

El ARS no debe entenderse como un recurso complementario en los estudios de medios y comunicación política, pues su relevancia emerge en el contexto de la evolución de la sociedad de la información y la consolidación de la sociedad en red. El desarrollo de la ciencia de redes en otras disciplinas y la posibilidad de acceder a datos estructurados y semiestructurados —como el mapeo de hipervínculos o la cartografía de Internet— han sido relevantes para estudiar interacciones sociales en plataformas digitales. Así, el ARS se ha

consolidado como una metodología que permite analizar fenómenos comunicativos, sociales y políticos en plataformas digitales, capturando dinámicas relacionales como la circulación de información, la influencia y la movilización con fidelidad en múltiples niveles (Nie *et al.*, 2023). Las revisiones coinciden en que, frente a la volatilidad de los comportamientos en plataformas y la insuficiencia de enfoques centrados únicamente en variables individuales, el ARS ofrece un marco estructural y un puente entre niveles micro, meso y macro. Esta literatura enfatiza, además, que su potencial depende de explicitar decisiones técnicas —delimitación de la red, criterios de captura y limpieza, elección de métricas y *layouts*¹— y de vincular los resultados con marcos teóricos, para alcanzar rigor, acumulación de conocimiento y replicabilidad (Moya Padilla y García Sánchez, 2023).

En términos metodológicos, varias investigaciones señalan que el ARS funciona como un cambio de paradigma en estudios de medios sociales al desplazar el foco desde la búsqueda incesante de factores hacia el análisis de resultados relacionales (nodos, vínculos, estructuras) (Aguirre, 2011), lo que habilita explicaciones de conductas situadas (Nie *et al.*, 2023). Este giro se sostiene en el uso de modelos y estimadores propios de la teoría de grafos —por ejemplo, modelos exponenciales aleatorios de grafos para estimar probabilidades estructurales y técnicas de máxima verosimilitud por Monte Carlo— que articulan microcomportamientos con patrones colectivos (densidad, triadas, centralizaciones) (Nie *et al.*, 2023).

Las revisiones centradas en usos predictivos de datos de plataformas digitales muestran cómo el ARS se integra con *scrapping web* y aprendizaje automático para construir grafos de interacción y semánticos, calcular centralidades y detectar comunidades (Cano-Marín *et al.*, 2023). En este frente, la literatura sistematiza procedimientos que van desde la construcción de redes de retuits y de menciones hasta la segmentación comunitaria con heurísticas modulares y reclama avanzar hacia *pipelines* reproducibles que integren extracción, preprocesamiento y modelado de redes de coocurrencias o dependencia semántica (Cano-Marín *et al.*, 2023). No obstante, también advierte zonas frágiles: la documentación incompleta de parámetros (por ejemplo, filtros, ventanas temporales, *layouts*), la opacidad de algunos pasos de transformación “texto→grafo” y la escasez de materiales abiertos (datos/*scripts*) que permitan replicar hallazgos.

Desde la literatura en español, los balances recientes confirman la centralidad del ARS con datos de algunas plataformas digitales y visibilizan desplazamientos metodológicos. Rivadeneira y Loor (2025) muestran que, pese a la desaceleración asociada a restricciones de acceso de la API de X/Twitter, se expandió con fuerza la integración ARS-PLN en el estudio de conversaciones públicas y recomiendan fortalecer estándares éticos y de transparencia en el reporte metodológico. Complementariamente, Jordán *et al.* (2024) evidencian efectos conductuales significativos y, a la vez, la necesidad de estudios regionales con mayor especificación metodológica, en los que el ARS sirva para diferenciar estructuras de difusión, comunidades de contenido y posiciones de autoridad en ecosistemas latinoamericanos.

Algunas investigaciones indican, también, una asimetría entre popularidad de plataformas y atención académica. En su análisis bibliométrico, Maciuk *et al.* (2025) señalan que Facebook y Twitter concentran la mayor parte de estudios, mientras Instagram y TikTok —aun con bases de usuarios masivas— permanecen relativamente subestudiadas (Rejeb *et al.*, 2024). Para el ARS, esto implica un reto de actualización: migrar técnicas consolidadas hacia plataformas con lógicas algorítmicas y formatos multimodales distintos, documentando decisiones de muestreo, construcción de aristas y homologación de entidades entre medios.

De la comparación sistemática se desprenden vacíos de investigación. Primero, persiste una brecha de claridad y replicabilidad: con frecuencia no se delimitan con precisión los límites de la red (sociocéntrica vs. egocéntrica), los periodos observados, los filtros de inclusión/exclusión ni los pasos de limpieza; tampoco se reportan abiertamente *scripts* o conjuntos de datos, lo que dificulta la verificación independiente. Esta carencia contrasta con lo que la literatura prescribe como núcleo del giro metodológico del ARS —explicitar unidades, estructuras y supuestos de modelado— para capturar contextos dinámicos (Nie *et al.*, 2023).

Segundo, hay un sesgo de accesibilidad que ha sobrerrepresentado a X/Twitter en los corpus empíricos y ha dejado rezagadas plataformas visuales o de video corto, cuya arquitectura de datos y lógicas algorítmicas plantean desafíos distintos para la construcción de grafos y la detección de comunidades. La literatura en español lo corrobora al documentar la expansión ARS-PLN en X/Twitter y, a la vez, la necesidad de trasladar esa sofisticación hacia TikTok y otros entornos con datos menos accesibles y más multimodales (Rivadeneira y Loor, 2025; Jordán *et al.*, 2024). Tercero, faltan modelos dinámicos, multinivel y multicapas que integren tiempo, capas semánticas y relaciones heterogéneas; aunque el ARS dispone de técnicas para estimar probabilidades estructurales y explorar micro-macro de manera conjunta, su despliegue empírico en la región es aún acotado (Nie *et al.*, 2023).

Cuarto, la dimensión ética aparece insuficientemente tratada: varias revisiones llaman a incorporar protocolos de anonimización, discusión de riesgos y publicación responsable de datos, especialmente cuando el análisis versa sobre comportamientos políticos y actores vulnerables (García Ramírez y Lombana-Bermúdez, 2024). En sintonía con esto, artículos de revisión, como el de Olteanu *et al.* (2019) y Cano-Marín *et al.* (2023), piden *pipelines* trazables de extracción-preprocesamiento-análisis que dejen constancia de decisiones y umbrales, algo que es a la vez requisito ético y condición de replicabilidad (Cano-Marín *et al.*, 2023).

A la luz de este estado del arte, el manuscrito responde a los vacíos identificados al proponer una RCM que exige definir de forma verificable la delimitación de la red y su temporalidad; transparentar la captura (API, palabras clave, paginación), la limpieza (criterios *antispam*, detección de automatización), la transformación “dato→grafo” (tipo de arista, dirección, ponderación) y el análisis (métricas reportadas, algoritmos de detección comunitaria y *layouts* con parámetros), además de documentar materiales reproducibles y protocolos éticos. Esta ruta busca crear las condiciones para trasladar al terreno regional las prácticas que las revisiones internacionales asocian con rigor, ética y replicabilidad en el ARS (Nie *et al.*, 2023; Cano-Marín *et al.*, 2023).

Metodología

El artículo emplea un diseño secuencial de métodos mixtos, en el cual la fase cualitativa se utiliza para construir una sistematización de los estudios, mientras que la fase cuantitativa permitió medir, comparar y sintetizar la información de forma replicable. Esta elección metodológica responde a la necesidad de analizar el uso del ARS en investigaciones de comunicación y ciencia política sobre plataformas digitales en América Latina, considerando tanto la diversidad conceptual y metodológica de los estudios como la necesidad de evaluar su transparencia y rigor científico (Adu *et al.*, 2022; Timans *et al.*, 2019). En este diseño, la fase cualitativa precede a la cuantitativa y le da sentido, definiendo las categorías que luego serán operacionalizadas en los análisis estadísticos. La fase cuantitativa, por su parte, válida los hallazgos cualitativos y busca lograr una integración entre interpretación y medición.

La primera fase se concentró en la identificación, selección y depuración del corpus de artículos. Para asegurar la calidad y relevancia de los estudios incluidos, se definieron criterios de inclusión y exclusión. Solo se consideraron artículos revisados por pares bajo la modalidad doble ciego y publicados en revistas indexadas en bases reconocidas, como Web of Science, Scopus, Dialnet, Latindex y SJR. Se excluyó deliberadamente la literatura gris —como tesis, ponencias y reportes técnicos—, debido a que, aunque pueden contener información útil, carecen de procesos estandarizados de evaluación. La búsqueda se realizó en Google Scholar. En español, se adoptó una estrategia minimalista que utilizó términos específicos como “ARS” o “análisis de redes sociales” en combinación con los nombres de plataformas digitales (por ejemplo: Twitter, Facebook, YouTube, TikTok) y países de América Latina. Este enfoque evitó la inclusión de artículos ajenos a la temática, como aquellos que emplean análisis de redes en campos como la biología o la ingeniería. En inglés, el reto

consistió en integrar tres dimensiones: metodología (*networkanalysis*), temática (*communicationy political science*) y contexto geográfico (*Latin America* y sus diferentes países). Para abordar esta dificultad, se empleó ChatGPT (versión GPT-4, 2023) como herramienta de apoyo para la generación de sintaxis. El *prompt* utilizado fue documentado para asegurar transparencia y replicabilidad:²

Elabora una *sintaxis de búsqueda en inglés para Google Scholar que identifique artículos académicos revisados por pares que utilicen network analysis aplicado a redes sociales digitales, dentro de los campos de la comunicación y la ciencia política, y que se enfoquen en países de América Latina.*³

Sin embargo, la revisión y depuración final de los resultados se realizó manualmente, para garantizar que la decisión final sobre inclusión respondiera a criterios de la investigación y no a la curaduría algorítmica de la plataforma (Timans *et al.*, 2019).

Reconociendo la influencia de la curaduría algorítmica de Google Scholar, que ordena los resultados según factores internos como el idioma, el número de citas y patrones de búsqueda previos entre otros (Rovira *et al.*, 2021), se estableció un control metodológico específico para evitar sesgos. La búsqueda se realizó bajo condiciones controladas, eliminando historiales y ajustando filtros de idioma y región. Además, no se limitaron las revisiones a las primeras páginas de resultados, sino que se exploraron ventanas más amplias y se emplearon ordenamientos alternativos. Esto garantizó que el corpus final no estuviera totalmente determinado por la lógica del algoritmo.

Una vez conformado el corpus inicial, compuesto por 68 artículos, se inició la clasificación cualitativa. Para guiar este proceso, se elaboró un libro de códigos que contenía definiciones operativas y criterios de clasificación para registrar aspectos como las plataformas digitales estudiadas, los niveles de análisis, las estrategias metodológicas empleadas y la forma en que cada artículo conceptualizaba el ARS. El equipo de trabajo estuvo conformado por cuatro investigadores.

Antes de proceder a la codificación completa, se realizó un pretest con 17 artículos, que representaban el 23,6 % del corpus total. Este *pretest* tuvo como objetivo evaluar la consistencia y confiabilidad de la interpretación cualitativa, así como verificar que el libro de códigos pudiera ser aplicado de forma homogénea por diferentes miembros del equipo. Para medir la fiabilidad intercodificador, se emplearon dos indicadores estadísticos utilizados en ciencias sociales: la alfa de Krippendorff y la kappa de Cohen (Rodríguez Pérez *et al.*, 2021). Cada artículo fue evaluado por dos codificadores de manera independiente y la fiabilidad del proceso fue confirmada mediante el cálculo de la alfa de Krippendorff y de la kappa de Cohen.

La alfa de Krippendorff ($\alpha = 0.92$) se utilizó para evaluar el nivel de acuerdo entre los codificadores en variables nominales y categóricas, considerando la probabilidad de que dicho acuerdo ocurriera por azar. Este coeficiente es útil en investigaciones con múltiples categorías, pues ofrece una medida ajustada del grado de consistencia interna en la codificación. Por su parte, la kappa de Cohen ($\kappa = 0.90$) permitió medir la concordancia entre dos codificadores en variables dicotómicas o categóricas, corrigiendo también el efecto del azar. El uso conjunto de ambos indicadores proporcionó una validación técnica de la estabilidad del proceso de codificación, lo cual aseguró que los resultados no dependieran de interpretaciones individuales, sino que reflejaran una aplicación sistemática de los criterios. Según Rodríguez Pérez *et al.* (2021), este tipo de validación permite confirmar que los juicios cualitativos alcanzan niveles adecuados de replicabilidad y que la información recopilada puede ser utilizada como base confiable para análisis posteriores. Una vez validado el instrumento, se procedió a la codificación completa. Esto generó una matriz estructurada en Excel que recogía la presencia o ausencia de indicadores vinculados a la RCM.

Con estos datos, se pasó a la fase cuantitativa, en la que la información cualitativa fue transformada en indicadores para la construcción de un índice de la RCM. Este índice se desarrolló mediante la teoría de la respuesta al ítem (TRI), lo que permitió asignar pesos diferenciados a cada indicador según su capacidad para discriminar niveles de claridad metodológica. El índice sintetizó la calidad metodológica de los estudios revisados en una sola métrica.

La dimensión evaluativa, por su parte, se centró en la aplicación de la RCM para diagnosticar la claridad metodológica de los artículos. Este diagnóstico no cuestionó los aportes de los estudios, sino que identificó los déficits metodológicos. Además de estas recomendaciones, se incluyó la elaboración de una guía⁴ en R orientada a mostrar un proceso técnico y transparente que utilicen ARS.

El ARS en práctica: una propuesta de Ruta de Claridad Metodológica (RCM)

La investigación en ARS, especialmente aquella basada en datos provenientes de PD, enfrenta el reto de definir con precisión qué significa pensar en términos de red y cómo traducir esa idea en una estrategia metodológica. Tradicionalmente, el ARS se ha apoyado en la construcción de grafos que permiten visualizar y medir relaciones entre actores. Sin embargo, como han advertido distintos autores, el grafo es solo una representación parcial de procesos más complejos. Venturini (2010) propone distinguir entre la red como concepto teórico, la red como dato digital y la red como resultado analítico para mostrar que no son equivalentes ni intercambiables. En esta línea, Latour (2011) señala que las redes no son cosas que existen allá afuera, sino una manera de redistribuir la agencia y cuestionar jerarquías establecidas entre estructuras y actores. De forma similar, Venturini, Munk y Jacomy (2019) advierten que mezclar sin reflexión estos distintos sentidos de red genera confusión y limita la interpretación de los hallazgos. De ahí que pensar en redes vaya mucho más allá del grafo: implica reconocer que los nodos y aristas que aparecen en una visualización no son la totalidad de la vida social, sino el resultado de un proceso de selección, traducción y codificación que debe explicitarse.

Responder a esta necesidad teórica y empírica motiva el diseño de la RCM, que ofrece una guía para comprender, sistematizar y replicar cada decisión en la investigación. La RCM organiza el proceso en cuatro etapas interconectadas: 1) recopilación de datos; 2) análisis de datos; 3) presentación y contribución, y 4) validación. Su propósito es describir procedimientos técnicos e integrar reflexiones conceptuales sobre cómo se definen los actores, las relaciones y los límites de las redes estudiadas. Cada etapa incluye indicadores evaluados mediante una escala continua de 0 a 1, donde 1 representa un cumplimiento total, 0.5 refleja un cumplimiento ambiguo y 0 indica ausencia de información suficiente. Este enfoque permite diferenciar estudios documentados de aquellos con vacíos metodológicos y proporciona una estructura para reflexionar sobre las implicaciones epistemológicas de las decisiones investigativas.

La primera etapa, recopilación de datos, constituye el cimiento sobre el cual se construirá el grafo y, con él, la interpretación de las redes. Este proceso incluye la captura, organización y preparación de la información obtenida, comúnmente denominada minería de datos o minería *web* (Rodríguez Cano, 2022). En contextos del norte global, algunas plataformas han desarrollado programas de colaboración con la academia que facilitan el acceso a datos, como ocurrió con la alianza Twitter - MIT Media Lab, que durante cinco años permitió a investigadores trabajar con información directa del *firehose* de Twitter antes de concluir en 2019. No obstante, mientras algunas iniciativas de este tipo han finalizado, otras han emergido recientemente, como el programa de Meta y el Center for Open Science, que busca proveer datos de Instagram para investigaciones sobre bienestar juvenil bajo protocolos de privacidad y transparencia. En América Latina, en cambio, la recolección suele depender casi exclusivamente del trabajo directo de los equipos de investigación, quienes enfrentan mayores restricciones de acceso y deben diseñar estrategias para obtener la información. En este contexto, se abren dos modalidades: *técnicas manuales*, que implican observación directa y registro detallado de interacciones, y *técnicas automatizadas*, que emplean API, *software* y/o herramientas de código abierto para descargar datos masivos en formatos estructurados como hojas de cálculo o archivos XML.

El primer componente evaluado en esta etapa es la delimitación de la población o red, entendida como el grado de precisión con el que se establecen los límites analíticos. Esto incluye definir el tipo de actores que integran la red, el periodo de observación, los criterios de inclusión y exclusión y la aproximación

metodológica adoptada. Puede tratarse de un enfoque sociocéntrico que busca abarcar la red completa, como en el caso de todos los participantes en un grupo de Facebook, o de un enfoque egocéntrico, centrado en un actor específico y sus conexiones, como menciones, retuits o respuestas en X (antes Twitter). Se asignó un puntaje de 1 cuando el artículo describe explícitamente todos estos elementos, como en el ejemplo: “Se analizaron los usuarios que participaron en el hashtag #Elecciones2023 durante los 15 días previos a la votación, excluyendo cuentas inactivas y bots identificados mediante criterios de actividad y metadatos”.⁵ Se calificó con 0.5 cuando la delimitación era parcial y 0 cuando no se ofrecía ninguna información. Este paso refleja una tensión: los bordes de la red no existen de forma natural, estos son contruidos a partir de decisiones teóricas y prácticas que determinan qué vínculos serán visibles en el grafo (Latour, 2011; Venturini *et al.*, 2019).

El segundo componente es la captura de datos. En la captura manual, la observación directa y el registro sistemático permiten un control detallado, pero limitan el volumen de datos recolectado. En la captura automatizada, las API y herramientas de *scraping web* permiten obtener grandes volúmenes de información, aunque introducen dependencias técnicas y restricciones impuestas por las plataformas. Se otorgó una calificación de 1 cuando el artículo describe con precisión las herramientas, parámetros y justificaciones técnicas utilizadas, como en el ejemplo: “Los datos fueron recolectados mediante la API de Twitter utilizando el paquete X de Python, con palabras clave . y ., entre el 1 y el 15 de enero de 2023”. Se asignó 0.5 cuando la descripción era incompleta y 0 cuando estaba ausente.

Una vez recolectados, los datos deben depurarse y organizarse antes de ser analizados. Este proceso se evaluó en la categoría de limpieza y manejo de datos, que considera la calidad de los procedimientos empleados para eliminar duplicados, cuentas falsas, usuarios inactivos y cualquier información irrelevante. Este aspecto resulta relevante, dado que los datos capturados suelen contener elementos externos a la población objetivo, como *bots* o cuentas inactivas. Se asignó una calificación de 1 cuando el artículo describe de forma detallada los pasos seguidos para depurar la información, como “se eliminaron cuentas con más de 50 publicaciones idénticas mediante un algoritmo de detección de *spam*, así como aquellas creadas en la semana previa al evento”. Se otorgó 0.5 cuando la limpieza se mencionaba de manera genérica, por ejemplo, “los datos fueron depurados antes del análisis”, y 0 cuando este proceso no se reportaba.

El cierre de esta primera etapa implica la transformación de los datos en matrices relacionales, requisito para el ARS. Los datos recolectados suelen presentarse inicialmente en tablas que describen atributos individuales de los actores, pero, para que puedan ser analizados, deben reorganizarse en matrices que representan las relaciones entre nodos y aristas. Por ejemplo, una mención en X/Twitter, una amistad en Facebook o una suscripción en YouTube se convierten en una conexión dentro de la red. Este paso permite visualizar y analizar las interacciones de manera estructurada. Además, en esta fase se enfatiza la necesidad de respetar las normas éticas y de privacidad, para asegurar que la información se maneje conforme a las políticas de las plataformas y a los principios de protección de datos personales.

La segunda etapa, análisis de datos, busca modelar la red, describir su estructura y comprender sus dinámicas. El ARS se apoya en la teoría de grafos, que proporciona métricas y algoritmos para representar relaciones, pero limitarse a este nivel sería insuficiente. Venturini (2010) propone la *cartografía de controversias* como un enfoque que combina lo cuantitativo y lo cualitativo para capturar tensiones, mediaciones y contextos que no se reflejan en el grafo. En esta etapa se distinguen tres niveles de análisis (González-Bailón, 2017; Rodríguez Cano, 2022). En el nivel macro, se estudia la red completa, calculando métricas globales como densidad, cohesión y reciprocidad. Se asignó 1 cuando estas métricas fueron definidas y justificadas, 0.5 cuando se mencionaron sin explicación y 0 cuando no se usaron. En el nivel meso, se identifican comunidades mediante algoritmos, útiles para analizar fenómenos como campañas electorales o movimientos sociales. Aquí también se evaluó la claridad en la descripción del procedimiento. En el nivel micro, el análisis se centra en actores individuales y en su posición dentro de la red, utilizando métricas como centralidad de grado, intermediación y vector propio (Himmelboim, 2017). La *visualización* es un componente transversal en esta

segunda etapa y se definió como la representación gráfica de la red mediante parámetros reproducibles. Para obtener la calificación máxima de 1, la visualización debía incluir información sobre el algoritmo de *layout*, el número de iteraciones y las visualizaciones utilizadas. Por ejemplo: “Se empleó el *layout* ForceAtlas2 con 500 iteraciones, en el que los colores indican comunidades y el tamaño de los nodos refleja su centralidad”. Un puntaje de 0.5 se asignó cuando la visualización aparecía sin explicación técnica y 0 cuando no se presentaban visualizaciones, aun cuando se declaraba el uso de ARS.

La tercera etapa, presentación y contribución, evalúa la forma en que los hallazgos son comunicados y cómo se vinculan con teorías y/o conceptos. La categoría central es el *alcance analítico*, entendido como la profundidad con la que los resultados se articulan con interpretaciones teóricas. Se establecieron tres niveles. El nivel básico, calificado con 0, corresponde a casos en los que el artículo se limita a reportar datos sin ningún esfuerzo de análisis, mostrando tablas, métricas o visualizaciones sin explicaciones. El nivel descriptivo, calificado con 0.5, se refiere a estudios que presentan hallazgos con cierta organización narrativa y resumen de patrones, pero sin establecer vínculos con teorías o hipótesis. Finalmente, el nivel explicativo, calificado con 1, se asigna a investigaciones que interpretan los resultados a la luz de marcos conceptuales sólidos o que contrastan redes de distintos casos o periodos utilizando criterios explícitos. Por ejemplo: “La estructura polarizada observada confirma la hipótesis de segmentación ideológica propuesta por Sunstein (1999)”. Esta graduación asegura coherencia, diferenciando entre la mera presentación de datos, la descripción estructurada de patrones y la interpretación teórica.

La cuarta etapa, validación, busca garantizar la transparencia y la replicabilidad de la investigación. Incluye dos dimensiones: ética y documentación abierta. La ética es un principio transversal que debe estar presente en las fases de la investigación, desde el diseño metodológico hasta la comunicación de resultados. Sin embargo, se ubica en esta etapa porque, al evaluar el documento en su conjunto, es posible apreciar cómo se aplicaron estas consideraciones a lo largo del proceso. Se calificó con 1 cuando el artículo discutía riesgos, detallaba procesos de anonimización y explicaba cómo se resguardaron los datos; con 0.5 cuando la referencia era genérica, como “los datos son públicos”, y con 0 cuando no se abordaba el tema. La documentación abierta evalúa la disponibilidad de materiales que permitan reproducir el estudio, asignando 1 cuando se proporcionaban bases y *scripts* completos, 0.5 cuando eran parciales y 0 cuando no existían materiales accesibles.

Durante la codificación se identificaron problemas recurrentes: el uso ambiguo del término *análisis de redes* para describir coocurrencias sin grafo social, visualizaciones sin detalle técnico, enlaces inactivos o falta de fechas de captura. Estas deficiencias dificultan la replicación y confunden el análisis. Por ello, la RCM no mide la validez sustantiva de los hallazgos, sino la claridad metodológica y la reflexión conceptual reportada. Un artículo puede estar bien documentado y aun así presentar sesgos teóricos, problemas de diseño o vacíos éticos. En consecuencia, la RCM debe verse como un punto de partida para mejorar la transparencia y la credibilidad del ARS integrando medición y teoría y situando cada estudio en un diálogo sobre qué significa hacer redes (Figura 1).



FIGURA 1.
Ruta de la claridad metodológica para el ARS

Fuente: elaboración propia.

En las cuatro etapas se establecen nueve indicadores, cada uno con un valor de 1, que permiten construir un índice continuo para clasificar los artículos. Los artículos que cumplen con todos los indicadores obtienen un valor de 9, lo que indica que presentan un ARS transparente, sistemático, ético y replicable. En contraste, aquellos con valores más cercanos a 0 carecen de claridad metodológica.

Resultados

Dimensión descriptiva

El uso ARS en el estudio de PD es una práctica relativamente reciente. En el ámbito de la academia latinoamericana, se identifica que el primer trabajo sobre este tema se remonta al año 2009. No obstante, dicho estudio presentó una RCM limitada. Con el paso del tiempo, se ha observado una evolución en la aplicación del ARS, de manera que logra un nivel de sofisticación cada vez mayor. Actualmente, los estudios que emplean esta metodología incluyen descripciones sobre la captura de datos, el uso de *softwares* especializados y los procesos de manejo y limpieza de la información. Además, los alcances de las investigaciones han mostrado una transformación. Inicialmente, predominan los trabajos de enfoque descriptivo. Sin embargo, se ha evidenciado una transición hacia investigaciones centradas en la explicación de fenómenos sociopolíticos. Este progreso se ha acompañado de un aumento en la producción científica que utiliza ARS para el estudio de PD. Este crecimiento alcanzó su punto más alto en 2012, con la publicación de 17 artículos en total (Figura 2).

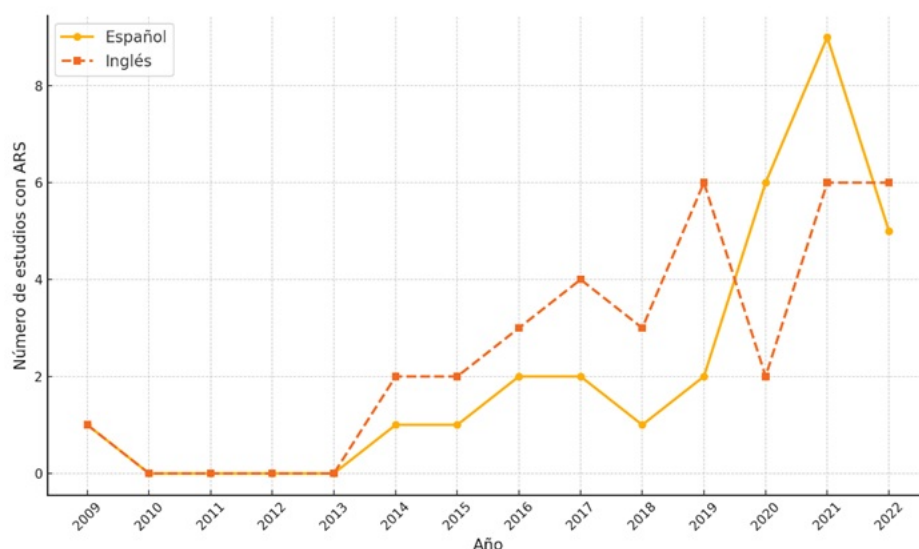


FIGURA 2.
Estudios con ARS a lo largo del tiempo
Fuente: elaboración propia.

Otro aspecto relevante en este contexto es el idioma en el que se publican estas investigaciones. En el ámbito académico, es común asumir que las innovaciones metodológicas se difunden con mayor rapidez en publicaciones anglosajonas. Bajo esta premisa, cabría esperar que la mayoría de los estudios estuvieran escritos en inglés. Sin embargo, se ha encontrado una paridad en la cantidad de investigaciones sobre ARS en plataformas digitales publicadas en inglés (33) y en español (35). Este fenómeno podría explicarse por el hecho de que muchas de estas investigaciones se centran en países latinoamericanos, lo que hace razonable que sean publicadas en español.

En cuanto a los países objeto de estudio en el contexto del ARS, se observa una marcada concentración en aquellos con mayor peso demográfico y territorial dentro de la región. México encabeza la lista con 22 menciones, seguido por Colombia con 11, Chile con 9 y Brasil con 8, lo que refleja una relación entre la magnitud de estos países y la atención académica que reciben. Sin embargo, también se identifican patrones que rompen con esta tendencia. Argentina, pese a su alto nivel de producción científica, aparece en 4 estudios, mientras que Ecuador alcanza una presencia destacable con 6 trabajos, una cifra relevante considerando su tamaño poblacional. Otros países como Perú y Venezuela se mencionan en 3 casos cada uno, mientras que Honduras, El Salvador y Cuba figuran en 2 ocasiones cada uno. Finalmente, se registran menciones puntuales a Bolivia, Guatemala, Puerto Rico y a categorías regionales como Latinoamérica/Sudamérica, que en conjunto suman 7 casos. En total, se contabilizan 82 menciones de países o regiones. Esta cifra supera los 68 artículos analizados porque varios trabajos abordan más de un país simultáneamente, ya sea mediante comparaciones o análisis de fenómenos transnacionales (Figura 3).

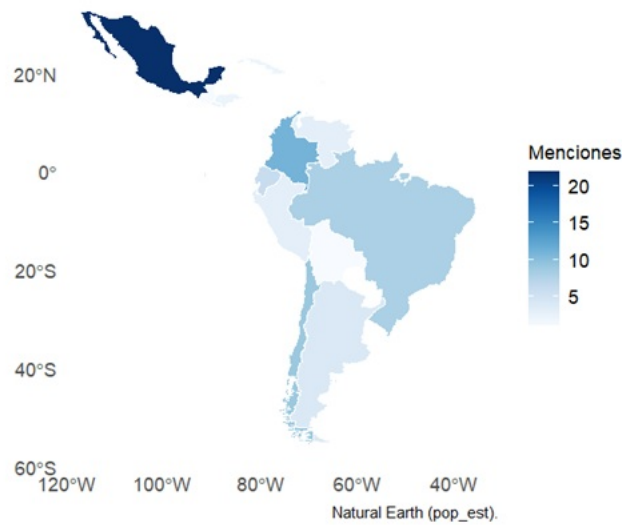


FIGURA 3.

Menciones por país en artículos sobre ARS en Latinoamérica (2009-2022)

Fuente: elaboración propia.

También se identifica que la mayoría de los artículos usa X/Twitter como fuente de datos para los análisis de ARS. Esta tendencia se explica por la facilidad de acceso a los datos en esta plataforma, en comparación con otras como Facebook, Instagram o YouTube, que imponen mayores restricciones antes de su conversión a X. De los 68 artículos analizados, 66 recurren a Twitter como la principal herramienta para la recopilación de datos, lo cual la consolida como la plataforma más utilizada en este campo de investigación.⁶

Las figuras 4 y 5 se construyeron con un mismo *pipeline* para asegurar comparabilidad entre idiomas y transparencia metodológica. El texto de los artículos se extrajo desde los PDF, se normalizó a minúsculas y sin diacríticos, se limpiaron URL, dígitos y signos y luego se lematizó con *spaCy*. Además de las *stopwords* estándar, aplicamos una lista de términos genéricos (por ejemplo, nivel, parte, forma) con el fin de reducir ruido sin eliminar vocabulario sustantivo del campo (plataformas, procesos político-institucionales, dinámicas de comunicación y emociones/polarización). El tratamiento por separado de español e inglés evita que diferencias morfológicas y de uso distorsionen conteos, TF-IDF y coocurrencias.

Los *tokens* resultantes se mapearon a cuatro categorías analíticas —comunicación, ciencia política, plataformas y emociones/polarización—. Sobre ese universo calculamos tres métricas complementarias: frecuencia (saliencia), TF-IDF⁷ con suavizado logarítmico (relevancia temática, penalizando términos ubicuos) y centralidad de coocurrencias en una red semántica construida con ventanas deslizantes de 6 *tokens*⁸ (grado ponderado). Para el mapa de burbujas, normalizamos centralidad y TF-IDF con una escala log-min-max,⁹ el tamaño representa la frecuencia y el color la categoría.

La elección visual obedece a una lógica de doble lectura. Las barras 2×2 muestran los diez términos más frecuentes por categoría, lo cual ofrece una comparación de magnitudes y permite detectar sesgos de agenda —por ejemplo, el peso de Twitter/X entre las plataformas—. El mapa de burbujas integra las tres métricas en un solo plano, lo que facilita pasar de la descripción (qué aparece y cuánto) a un diagnóstico (qué términos son, además, distintivos y articuladores del discurso). Para evitar que uno o dos términos extremadamente frecuentes aplasten la escala gráfica se atenuó su dominancia (sin alterar los *rankings*).

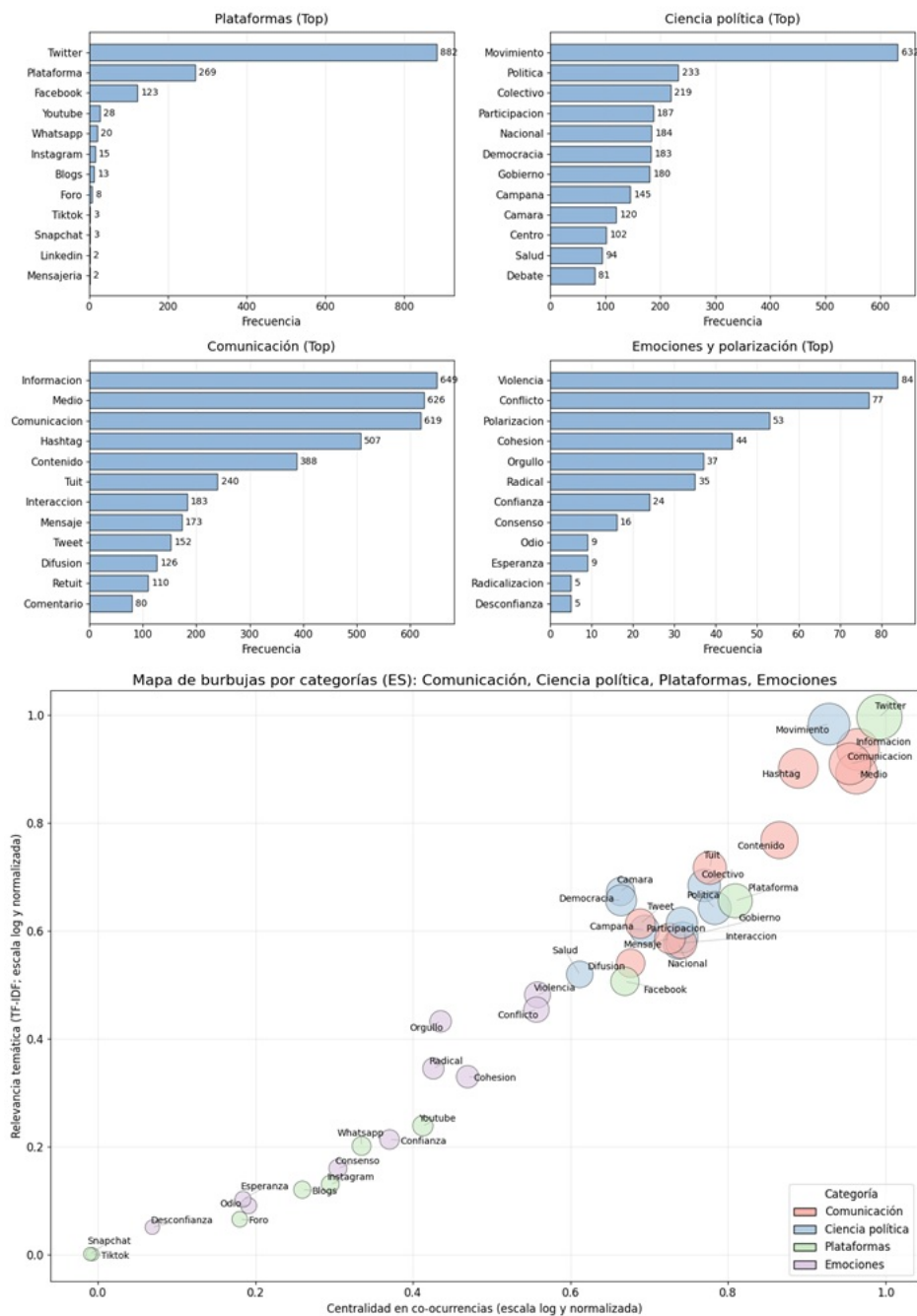


FIGURA 4.
Panorama temático del corpus en español: barras por categorías y mapa de burbujas

Fuente: elaboración propia.

La figura 4 expone el vocabulario en los artículos académicos en español que aplican ARS a PD. En los paneles de barras, Twitter/X concentra la referencia como principal PD, muy por encima de otras plataformas; en comunicación prevalecen información, medio, comunicación y *hashtag*, lo que refleja que buena parte de los estudios describe y explora flujos de contenido y mecanismos de visibilidad propios del entorno sociotécnico. En ciencia política, se destacan movimiento, política, colectivo, participación, Gobierno y democracia. Esto evidencia que las investigaciones se orientan a movilización social, acción colectiva y configuración de la opinión pública más que a instituciones formales. En emociones y polarización,

predominan conflicto y violencia por encima de confianza o consenso, lo que sugiere que la literatura analiza con mayor frecuencia episodios contenciosos.

El mapa de burbujas profundiza esa lectura: los términos más relevantes y conectores del campo combinan dispositivos de plataforma (X/twitter, *hashtag*) con procesos político-sociales (movimiento, participación) y tópicos comunicacionales (información, comunicación). Su posición en el cuadrante de alta centralidad y alta relevancia indica que aparecen con frecuencia en los textos y estructuran el discurso académico sobre ARS en español.

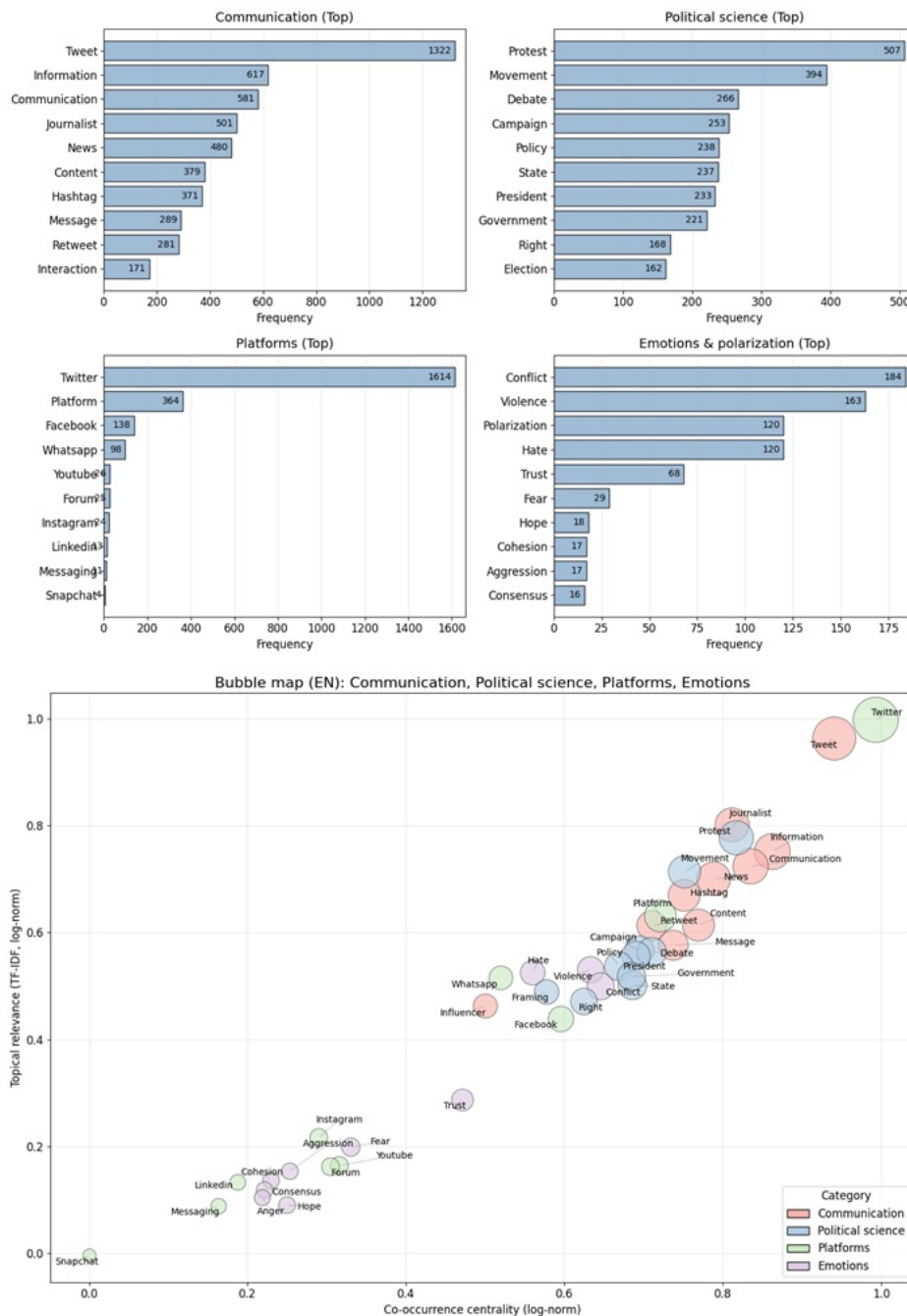


FIGURA 5.

Panorama temático del corpus en inglés: barras por categorías y mapa de burbujas

Fuente: elaboración propia.

La figura 5 correspondiente al corpus en inglés, muestra un predominio del léxico de comunicación: *tweet, information, communication, journalist, news* y *content* concentran buena parte de las menciones y orientan la lectura hacia la circulación informativa y las rutinas mediáticas. En plataformas, Twitter/X se impone sobre Facebook, Instagram o YouTube y se consolida como la plataforma digital más utilizada en los análisis de ARS. En el eje de ciencia política, sobresalen *protest, movement, debate, campaign, policy, state, president* y *government*, lo que sugiere una articulación frecuente entre redes sociales, repertorios de protesta y competencia política. En emociones y polarización, destacan *conflict, violence* y *polarization*, por encima de *trust* o *consensus*, lo que indica un foco hacia episodios contenciosos.

El mapa de burbujas de la misma figura confirma estos patrones: *twitter* y *tweet* se sitúan en el cuadrante de alta centralidad y alta relevancia temática y a su alrededor orbitan términos que conectan comunicación y política —*news, journalist, hashtag, content, campaign, protest, movement*—. Las emociones funcionan como moduladores del análisis más que como objeto autónomo: *conflict* y *violence* se enlazan con los núcleos comunicacionales y políticos, mientras *trust* y *consensus* ocupan posiciones periféricas. Así, la figura perfila una agenda, en la que la observación de flujos informativos y las dinámicas de movilización y disputa electoral aparecen entrelazadas, con Twitter/X como PD principal.

En contraste con la figura 4 (corpus en español), se mantienen semejanzas: Twitter/X ocupa un lugar central en ambos conjuntos y las dinámicas de movilización contenciosa estructuran buena parte de las investigaciones, con un registro afectivo en el que conflicto, violencia y polarización superan a confianza o consenso. La diferencia principal es de énfasis: el corpus en inglés se estructura más en el léxico mediático-informacional y en la competencia política (*journalist, news, debate, campaign*), mientras que el corpus en español otorga mayor peso a la movilización social y la participación (movimiento, participación, colectivo) y a la configuración de la opinión pública (información, comunicación, medio, *hashtag*).

El análisis metodológico de los artículos revisados muestra una predominancia clara del enfoque cuantitativo, que representa el 71 % de los estudios, como se aprecia en la figura 6. Este hallazgo subraya el carácter técnico y sistemático del ARS en el ámbito académico, en el que las herramientas cuantitativas permiten modelar y explorar datos a gran escala. Por su lado, el 22 % de los estudios adoptan un enfoque mixto (Figura 6).

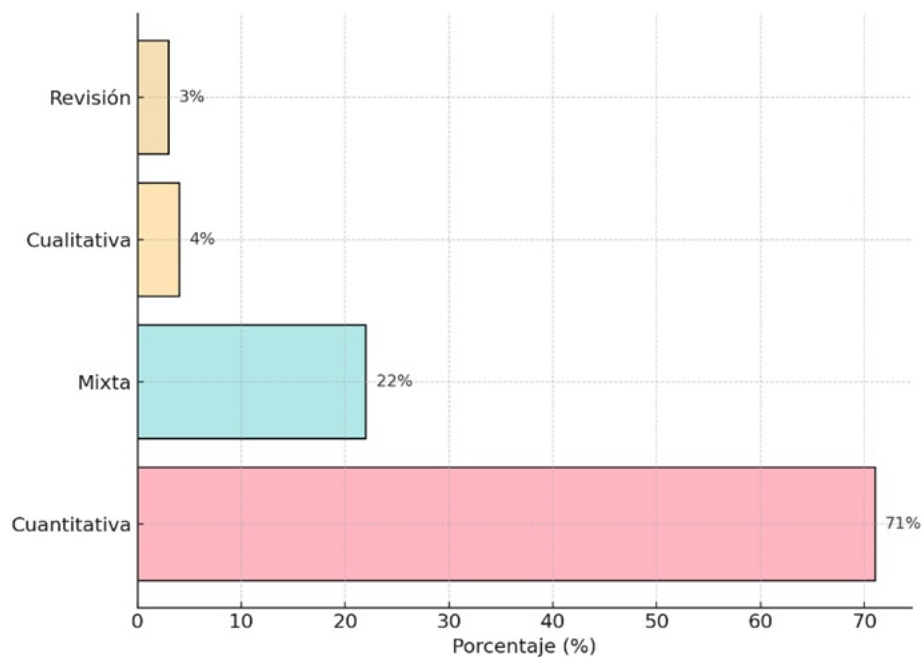


FIGURA 6.
Artículos de acuerdo con la metodología

Fuente: elaboración propia.

En contraste, los estudios cualitativos son minoritarios, representando apenas el 4 %, mientras que las revisiones metodológicas constituyen un 3 %. Esta tendencia sugiere que el ARS, en su mayoría, se enfoca en mediciones y modelaciones estructurales más que en análisis interpretativos o reflexivos. La limitada presencia de enfoques cualitativos puede indicar desafíos en la aplicación de estos métodos en investigaciones de PD al volumen de los datos generados en PD.

En cuanto a los procesos y herramientas empleados para la captura y análisis de datos, se observa que, en general, esta información es documentada en los artículos revisados, formando parte del diseño metodológico. El método más común de captura de datos consiste en el uso de API de Twitter/X combinado con *software* gratuito para la descarga de información. Cabe destacar que solo cuatro estudios recurren a *data brokers* o *software* pagos para este propósito. Aunque inicialmente se consideró la posibilidad de que los datos fueran obtenidos mediante asociaciones con PD, no se encontraron ejemplos de este tipo de colaboración en los artículos analizados. En lo referente a los programas utilizados, NodeXL¹⁰ destaca como la herramienta más empleada, presente en 19 trabajos, aunque se evidencia una diversidad significativa de *softwares* complementarios en el corpus estudiado.

Adicionalmente, se examinó la mención de las herramientas empleadas para el análisis y la visualización de redes en los estudios de ARS. En comparación con los procesos de captura, este aspecto es menos documentado. Como se observa en la figura 7, solo 45 de los artículos especifican el *software* utilizado para analizar y graficar las redes, Gephi¹¹ destaca como la herramienta más empleada. No obstante, un hallazgo preocupante en términos metodológicos es que 11 estudios mencionan el uso de ARS como método, pero no presentan las redes generadas, y, de estos, 8 tampoco especifican el *software* utilizado, lo que dificulta la replicabilidad y transparencia de sus resultados.

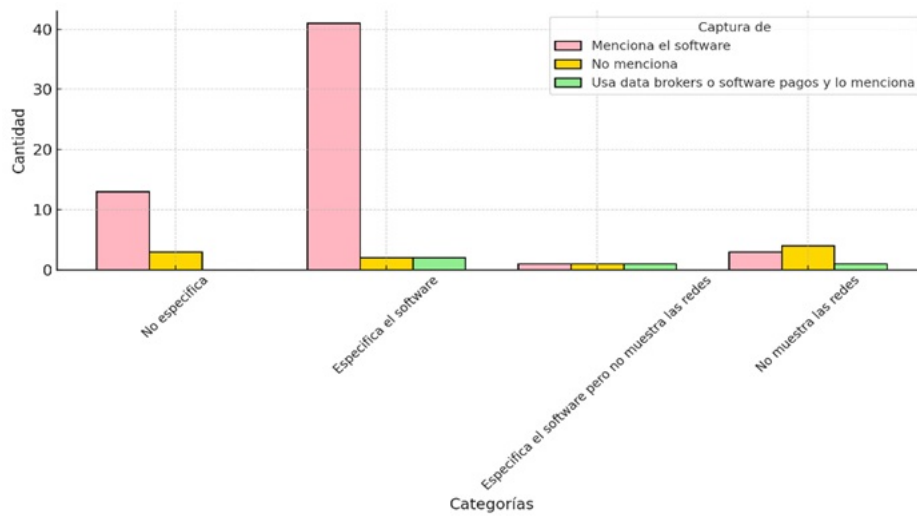


FIGURA 7.
Publicitación de las herramientas usadas para el manejo de los datos
Fuente: elaboración propia

En la dimensión relacionada con el manejo de los datos, se identificó que 48 artículos detallan los procesos empleados para organizar y limpiar la información, mientras que 20 no lo mencionan.

En la figura 8, se observa que la colaboración interdisciplinaria en investigaciones que emplean ARS sigue siendo muy limitada y concentrada en la producción anglosajona. De los 68 artículos analizados, únicamente 12 presentan algún grado de interdisciplinariedad y, de estos, 11 están escritos en inglés, lo que confirma que este tipo de cooperación se desarrolla casi exclusivamente en este ámbito lingüístico. Al examinar si la interdisciplinariedad se da dentro de las ciencias sociales, los resultados muestran que esta práctica también es poco frecuente: solo 9 artículos incluyen coautorías entre investigadores de distintas disciplinas de las ciencias sociales

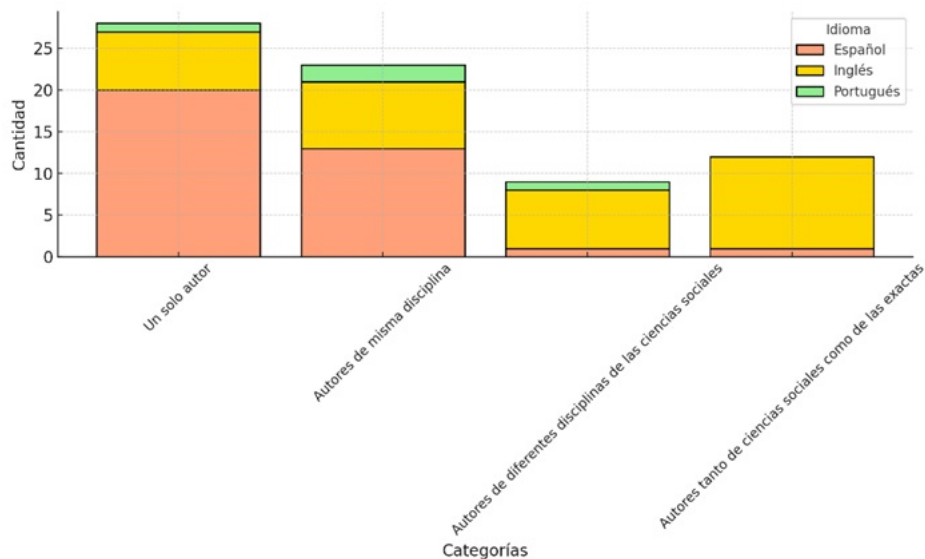


FIGURA 8.
Interdisciplinariedad en función del idioma del artículo
Fuente: elaboración propia.

Dimensión de diagnóstico

Una vez identificadas las características de los artículos que utilizan ARS se van a exponer los principales resultados derivados de la evaluación de los artículos con base en los nueve indicadores de la RCM (Figura 9).

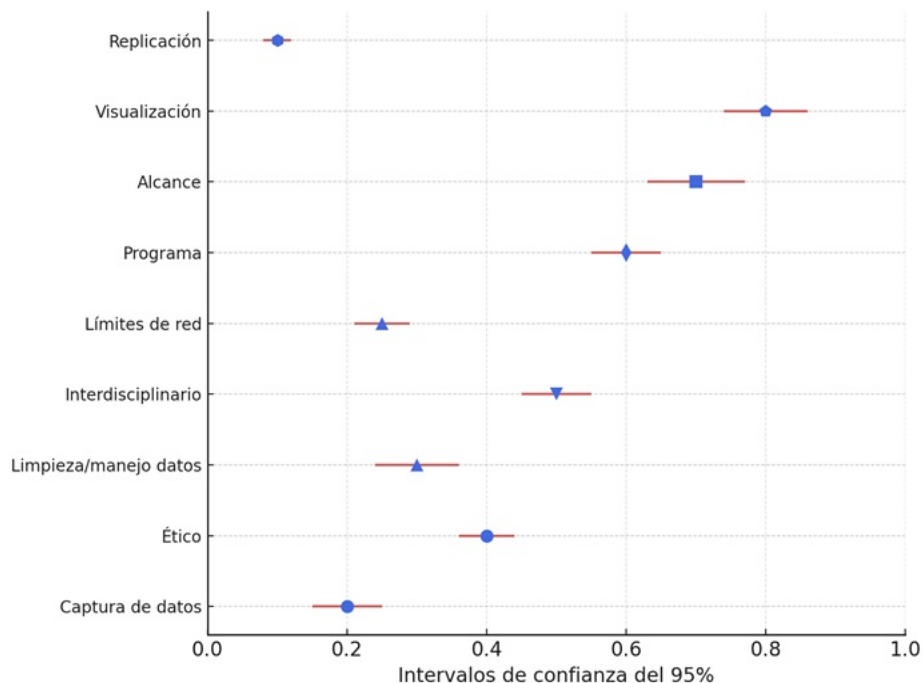


FIGURA 9.

Fuente: elaboración propia.

La heterogeneidad observada en los indicadores de la RCM (Figura 8) evidencia tanto avances como carencias en la implementación del ARS en el contexto de Latinoamérica. Mientras que una mayoría de estudios incluye descripciones sobre los programas utilizados para la captura de datos y las visualizaciones de las redes, otros aspectos, como los procesos de limpieza y manejo de datos, siguen siendo insuficientemente documentados. Esta disparidad sugiere una necesidad de estandarización y cualificación metodológica, especialmente considerando que el ARS requiere procesos transparentes y replicables para garantizar la validez de sus hallazgos.

Un punto relevante es la ausencia de discusión ética en una parte de los estudios. La falta de reflexión sobre el uso de datos personales y la publicación abierta de las bases de datos limita no solo la replicabilidad de los análisis, sino también la posibilidad de que otros investigadores construyan sobre estos trabajos. Esta carencia afecta la credibilidad de la investigación y también plantea riesgos éticos, especialmente cuando se trabajan datos sensibles o en contextos sociopolíticos conflictivos. Adicionalmente, se observa que la mayoría de los artículos no especifican el tipo de red analizada (sociocéntrica o egocéntrica), un aspecto relevante para el diseño y la interpretación de los estudios. La omisión de esta información dificulta la replicación y puede conducir a conclusiones teóricas imprecisas o mal fundamentadas.

Otro elemento destacado es el alcance. Esto refleja la tendencia de los estudios a centrarse en temas descriptivos, sin una integración en discusiones teóricas más amplias. Esto podría estar relacionado con la falta de colaboración interdisciplinaria, identificada como una debilidad generalizada en el corpus analizado. La incorporación de perspectivas de diferentes disciplinas podría complementar la capacidad de estos estudios para abordar fenómenos sociopolíticos y contextualizarlos dentro de marcos teóricos y metodológicos interdisciplinarios.

Por último, es importante mencionar que la dependencia de plataformas como X/Twitter o Facebook introduce sesgos algorítmicos en los datos recolectados que pueden distorsionar los hallazgos y limitar la representatividad de los estudios.

La figura 10 muestra la evolución temporal de la RCM por medio de las medias de los artículos. Estas muestran un incremento entre 2008 y 2015, seguido de fluctuaciones y una estabilización en los años posteriores. Aunque el progreso inicial es evidente, los valores promedio se mantienen entre 3 y 4 en una escala de 9, lo que indica que la mayoría de los estudios aún no logran alcanzar niveles altos de RCM.

El crecimiento observado entre 2010 y 2015 puede atribuirse a una mayor adopción de herramientas específicas para el análisis de redes y al fortalecimiento de las capacidades de los investigadores en el uso del ARS. Sin embargo, la estabilización y el ligero descenso observados a partir de 2018 plantean preocupaciones sobre un posible estancamiento en la adopción de mejores prácticas, como las señaladas en este artículo.

Otro punto relevante es el descenso observado en 2021, que podría estar relacionado con un aumento en la producción de estudios impulsados por eventos coyunturales, como la pandemia de la COVID-19 o movimientos sociales recientes. Estos estudios, al ser más rápidos en su producción, podrían haber priorizado el análisis de datos por encima de la documentación y la replicabilidad metodológica. Este patrón sugiere que, aunque la producción académica ha crecido, no siempre ha estado acompañada de un rigor metodológico acorde.

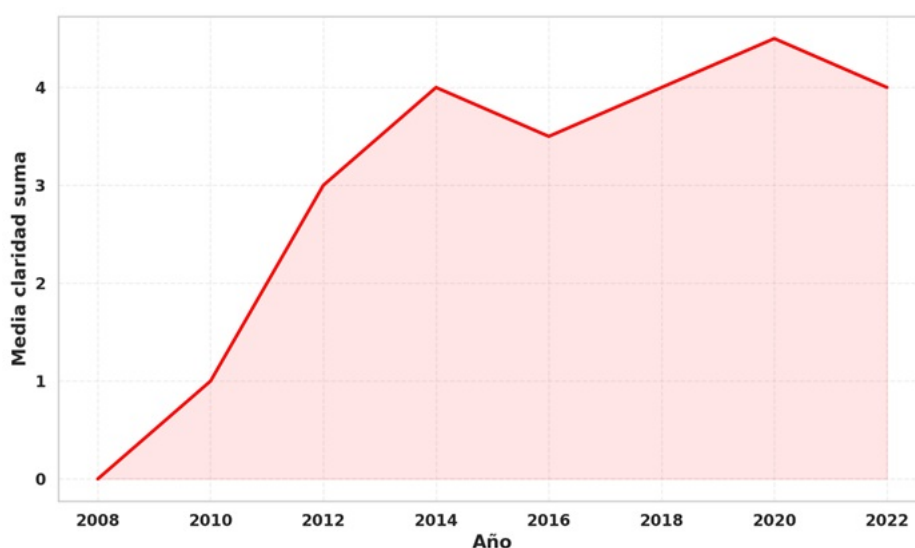


FIGURA 10.
Claridad metodológica en el tiempo

Fuente: elaboración propia.

La figura 11 muestra diferencias en la claridad metodológica de los artículos que emplean ARS según el idioma de publicación. Los artículos en inglés tienen una mediana más alta (4,68) frente a los publicados en español (3,34), lo que sugiere estándares más rigurosos en documentación y aplicación metodológica en la literatura anglófona. El rango intercuartílico es mayor en inglés, lo cual evidencia una mayor variabilidad: algunos artículos alcanzan altos niveles de claridad, mientras que otros se ubican en valores bajos, incluso comparables o inferiores a los de la literatura en español. En contraste, los artículos en español muestran una distribución más homogénea, pero con menor alcance, con pocos trabajos que superan la media del corpus en inglés. Este patrón podría estar influido por factores como acceso a recursos tecnológicos, formación metodológica avanzada o la orientación editorial hacia estándares internacionales. Finalmente, se destaca un valor atípico en español, cercano a cero, que refleja graves omisiones en procesos clave como limpieza de datos, delimitación de redes o publicación de bases, lo que muestra carencias extremas en algunos estudios.

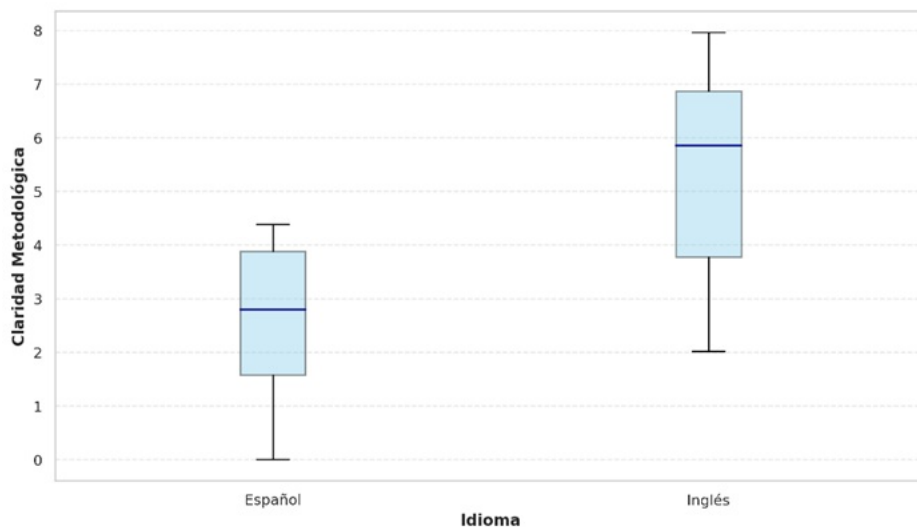


FIGURA 11.
Claridad metodológica por idioma

Fuente: elaboración propia.

La figura 12 muestra cómo distintos indicadores metodológicos se relacionan con los niveles de claridad metodológica en los estudios de ARS. En primer lugar, los trabajos con un enfoque explicativo presentan valores promedio más altos que los descriptivos, lo que refleja que las investigaciones orientadas a explicar fenómenos tienden a ser más rigurosas. En cuanto a la captura de datos, los estudios que especifican sus métodos alcanzan niveles superiores de RCM, lo cual subraya la importancia de documentar este proceso para garantizar replicabilidad. De manera similar, los artículos que describen explícitamente los procedimientos de limpieza y manejo de datos también logran mejores resultados en términos de RCM y destaca este aspecto como central para la calidad del análisis. Finalmente, la identificación del tipo de red analizada (sociocéntrica o egocéntrica) también se asocia con mayores niveles de CM. La ausencia de esta especificación en algunos estudios puede limitar la interpretación de los resultados y su validez.

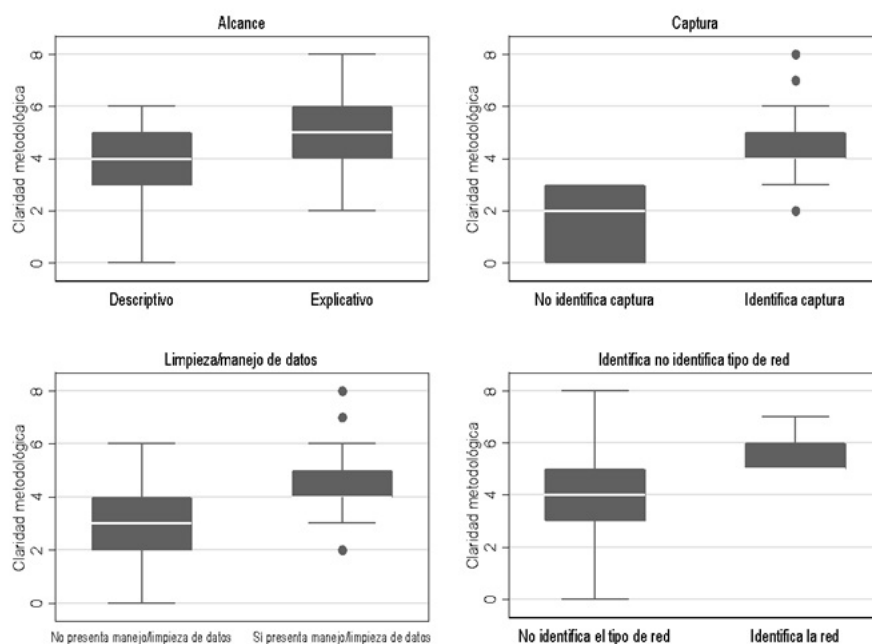


FIGURA 12

Niveles de claridad metodológica por alcance, captura, limpieza e identificación de red

Fuente: elaboración propia.

Conclusiones

Este estudio examinó el uso del ARS en artículos sobre comunicación y ciencia política entre 2009 y 2022 en Latinoamérica, con el objetivo de evaluar su desarrollo en términos de transparencia, replicabilidad y ética en el manejo de datos. A partir de un corpus de 68 artículos, se estableció un diagnóstico sobre el estado actual de este campo, para identificar fortalezas, vacíos y desafíos.

El primer hallazgo muestra una expansión sostenida en la aplicación del ARS, tanto en el crecimiento del número de publicaciones como en la diversificación temática de los estudios. Este desarrollo refleja la centralidad de las RR. SS. como espacios en los que se producen, circulan y disputan significados políticos (Nie *et al.*, 2023; Recuero *et al.*, 2019). Sin embargo, gran parte de esta literatura se ha centrado en métricas y visualizaciones, dejando de lado la comprensión de la red como entramado sociotécnico y político. Reducir la complejidad de las interacciones a grafos y números invisibiliza la agencia de actores no humanos, como algoritmos y plataformas, que configuran las formas de interacción y los patrones observados (Venturini *et al.*, 2019). Este vacío conceptual limita la capacidad del ARS para capturar los procesos híbridos que emergen en la intersección entre tecnología y política.

En el plano metodológico, se evidencian avances importantes en la captura y visualización de datos, así como en el uso de herramientas especializadas que han fortalecido la capacidad analítica de la región. Sin embargo, persisten vacíos en la limpieza y clasificación de datos, la definición explícita de los tipos de redes analizadas y la integración sistemática de principios éticos a lo largo de todo el proceso de investigación. Estas deficiencias afectan la transparencia y la replicabilidad, lo cual obstaculiza la construcción de conocimiento acumulativo (Brass, 2022; Nie *et al.*, 2023). La RCM propuesta busca responder a estas limitaciones, ofreciendo un marco estructurado en cuatro etapas —recopilación, análisis, presentación y validación— que proporciona indicadores precisos para evaluar la calidad de los estudios y elevar los estándares de investigación en el campo.

El tercer hallazgo identifica limitaciones que condicionan el desarrollo de la investigación en la región. La dependencia casi exclusiva de Twitter/X como fuente de datos, presente en la mayoría de los artículos analizados, ha generado sesgos importantes que restringen la diversidad de fenómenos observados y dificultan la generalización de los hallazgos. Esta concentración invisibiliza dinámicas propias de otras plataformas, como TikTok, Instagram o YouTube, que operan bajo arquitecturas algorítmicas distintas y producen formas de interacción que no pueden ser capturadas por un solo tipo de dato. Este problema se ha intensificado con la reducción progresiva en el acceso a información tras cambios recientes en las políticas de la plataforma, incrementando el riesgo de *algorithmic confounding*, en el que las señales algorítmicas distorsionan la interpretación de los patrones de comportamiento humano (Lazer *et al.*, 2021). A esto se suma la falta de reflexión ética sobre la protección de datos y la seguridad de los usuarios, especialmente en contextos de represión estatal y violencia política, donde la exposición de actores vulnerables puede tener consecuencias (Recuero *et al.*, 2019). Estos desafíos evidencian la necesidad de establecer protocolos éticos que protejan a las comunidades estudiadas.

Este artículo presenta limitaciones que deben ser consideradas al interpretar sus hallazgos. El corpus analizado, compuesto por 68 artículos publicados entre 2009 y 2022, ofrece una visión amplia, pero acotada a un periodo específico, lo que implica que transformaciones recientes en los ecosistemas digitales, como la irrupción de nuevas plataformas y las modificaciones en las políticas de acceso a datos, no se reflejan en el análisis. Además, el estudio se concentró exclusivamente en publicaciones académicas indexadas, dejando por fuera informes técnicos, bases de datos institucionales y trabajos elaborados por organizaciones de la sociedad civil que podrían aportar evidencia complementaria y perspectivas metodológicas alternativas. Por su parte, aunque la RCM fue aplicada y testeada en este artículo para el campo de la comunicación y ciencia política, su diseño se basa en las características de este corpus particular. Será necesario evaluar su desempeño en otros campos de estudio y con diferentes tipos de redes y datos.

El tránsito del dato al grafo resume el desafío identificado en este estudio. No se trata solo de recolectar grandes volúmenes de datos y transformarlos en visualizaciones sofisticadas, sino de construir investigaciones que combinen precisión técnica con responsabilidad ética y rigor interpretativo. La RCM se inscribe en este esfuerzo, ofreciendo una guía práctica para elevar los estándares de calidad y fomentar la transparencia y la apertura en el RSA. Este enfoque reconoce que los datos no son neutrales y que las representaciones gráficas, aunque poderosas, deben ser interpretadas a la luz de marcos teóricos críticos que den cuenta de las dimensiones políticas y tecnológicas que las atraviesan.

Referencias

- Abadía, A. A., Manfredi, L. C. y Rodríguez, J. L. (2024). Political Engagement and Aggressive Use of Social Networks: Presidential Campaigns in a Highly Polarized Electoral Scenario. En *Legal Challenges and Political Strategies in the Post-Truth Era* (pp. 67-90). Universidade da Beira Interior. <https://doi.org/10.18046/EUI/ohst.v1>
- Adu, J., Owusu, M. F., Martin-Yeboah, E., Pino Gavidia, L. A. y Gyamfi, S. (2022). A Discussion of Some Controversies in Mixed Methods Research for Emerging Researchers. *Methodological Innovations*, 15(3), 321-330. <https://doi.org/10.1177/20597991221123398>
- Aguirre, J. L. (2011). *Introducción al análisis de redes sociales* (Documentos de Trabajo, n° 82). Centro Interdisciplinario para el Estudio de Políticas Públicas.
- Alexandre, I., Jai-sung Yoo, J. y Murthy, D. (2022). Make Tweets Great Again: Who Are Opinion Leaders, and What Did They Tweet About Donald Trump? *Social Science Computer Review*, 40(6), 1456-1477. <https://doi.org/10.1177/08944393211008859>
- Argüello Pazmiño, S. (2021). Cruzadas *ciborg* en Ecuador: la disputa por el matrimonio igualitario en Twitter. *Cahiers des Amériques Latines*, 98, 67-102. <https://doi.org/10.4000/cal.13945>

- Ballard, A. O., Hillygus, D. S. y Konitzer, T. (2016). Campaigning Online: Web Display Ads in the 2012 Presidential Campaign. *PS: Political Science & Politics*, 49(3), 414-419. <https://doi.org/10.1017/S1049096516000780>
- Brass, D. (2022). New Developments in Social Network Analysis. *Annual Review Organizational Psychology and Organizational Behavior*, 9, 225-246. <https://doi.org/10.1146/annurev-orgpsych-012420-090628>
- Borge-Holthoefer, J. y González-Bailón, S. (2017). Scale, Time, and Activity Patterns: Advanced Methods for the Analysis of Online Networks. En N. Fielding, R. Lee y G. Blank (eds.), *The SAGE Handbook of Online Research Methods* (pp. 259-276). SAGE Publications Ltd. <https://doi.org/10.4135/9781473957992.n15>
- Boyd, D. y Ellison, N. B. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13, 210-230. <https://doi.org/10.1111/j.1083-6101.2007.00393.x>
- Cano-Marín, E., Mora-Cantalops, M. y Sánchez-Alonso, S. (2023). Twitter as a Predictive Tool: A Systematic Review of its Uses and Implications. *Journal of Business Research*, 159, 113688. <https://doi.org/10.1016/j.jbusres.2022.113561>
- Carrington, P. J., Scott, J. y Wasserman, S. (Eds.). (2005). *Models and Methods in Social Network Analysis* (vol. 28). Cambridge University Press.
- Cifuentes, C. F. y Pino Uribe, J. F. (2018). Conmigo o contra mí: análisis de la concordancia y estrategias temáticas del Centro Democrático en Twitter. *Palabra Clave - Revista de Comunicación*, 21(3), 885-916. <https://doi.org/10.5294/pacla.2018.21.3.10>
- García Ramírez, D., y Lombana-Bermúdez, A. (2024). Presentación. Desórdenes informativos en ecosistemas mediáticos latinoamericanos. *Dixit*, 38, e4354. <https://doi.org/10.22235/d.v38.4354>
- Gonzales-Bailón, S. (2017). *Decoding the Social World*. MIT Press.
- González, J. (2019). Análisis de Redes Sociales (ARS): estado del arte del caso mexicano. *Espacio Abierto*, 28(3), 5-24.
- Himmelboim, I. (2017). Social Network Analysis (Social Media). En J. Matthes, C. S. Davis y R. F. Potter (eds.), *The International Encyclopedia of Communication Research Methods* (pp. 1-15). Wiley-Blackwell.
- Jordán, D., Izaguirre, J. A. y López, A. L. (2024). TikTok y cambio de comportamiento: una revisión bibliométrica de la literatura. *Intersecciones en Comunicación*, 10(2), 78-96.
- Latour, B. (2011). Networks, Societies, Spheres: Reflections of an Actor-Network Theorist. *International Journal of Communication*, 5, 796-810.
- Lazer, D., Hargittai, E., Freelon, D., et al. (2021). Meaningful Measures of Human Society in the Twenty-First Century. *Nature*, 595, 189-196. <https://doi.org/10.1038/s41586-021-03660-7>
- Levendusky, M. y Stecula, D. (2021). *We Need to Talk: How Cross-Party Dialogue Reduces Affective Polarization*. Cambridge University Press.
- Lombana-Bermúdez, A. y Rodríguez Gómez, S. (2023). Desbordando *hashtags* de Twitter: la protesta digital k-pop en el Paro Nacional de 2021 en Colombia. *Anuario Electrónico de Estudios en Comunicación Social "Disertaciones"*, 16(2), 1-25.
- Lozares Colina, C. (2005). Bases socio-metodológicas para el Análisis de Redes Sociales (ARS). *Empiria. Revista de Metodología de Ciencias Sociales*, 0(10), 9. <https://doi.org/10.5944/empiria.10.2005.1042>
- Maciuk, K., Szczepańska, A. y Wilk, J. (2025). Social Media Platforms in Scientific Research: A Bibliometric and Critical Review. *Journal of Media and Communication Studies*, 37(2), 221-239. <https://doi.org/10.1016/j.ijinfomgt.2022.102510>
- Manfredi, L. y González, J. M. (2019). Comunicación y competencia en Twitter: un análisis en las elecciones presidenciales Colombia 2018. *Revista Estudios Institucionales*, 6(11), 133-130. <https://doi.org/10.5944/ceii.vol.6.n.11.2019.25086>
- Matos, F. F., Magalhães, L. H. de y Souza, R. R. (2020). Recuperação e classificação de sentimentos de usuários do Twitter em período eleitoral. *Informação & Informação*, 25(1), 92. <https://doi.org/10.5433/1981-8920.2020v25n1p92>
- Olteanu, A., Castillo, C., Diaz, F. y Kiciman, E. (2019). Social data: Biases, Methodological Pitfalls, and Ethical Boundaries. *Frontiers in Big Data*, 2, 13. <https://doi.org/10.3389/fdata.2019.00013>

- Moya Padilla, N. E. y García Sánchez, S. (2023). El análisis de redes sociales en las ciencias sociales. *Universidad y Sociedad*, 16(5), 206-212.
- Nie, Z., Waheed, M., Kasimon, D. y Wan Abas, W. A. B. (2023). The Role of Social Network Analysis in Social Media Research. *Applied Sciences*, 13(17), 9486. <https://doi.org/10.3390/app13179486>
- Papacharissi, Z. y Yuan, E. (2011). What if the Internet did not Speak English? New and old Language for Studying Newer Media Technologies. En N. Jankowski, S. Jones y D. Park (eds.), *The Long History of New Media* (pp. 89-108). Peter Lang.
- Poell, T., Nieborg, D. y Van Dijck, J. (2022). Plataformización. *Revista Latinoamericana de Economía y Sociedad Digital*, 1-17. <https://doi.org/10.53857/tsfe1722>
- Ramos-Vidal, I. y Ricaurte Quijano, P. (2015). Niveles de análisis y estrategias metodológicas en la ciencia de las redes. *Virtualis*, 6(11), 139-163. <https://doi.org/10.2123/virtualis.v6i11.116>
- Recuero, R., Zago, G. y Soares, F. (2019). Using Social Network Analysis and Social Capital to Identify User Roles on Polarized Political Conversations on Twitter. *Social Media + Society*, 5(2). <https://doi.org/10.1177/2056305119848745>
- Rejeb, A., Rejeb, K., Appolloni, A., Treiblmaier, H. y Iranmanesh, M. (2024). Mapping the TikTok Phenomenon: A Bibliometric Review of the Literature. *Digital Business*, 5(1), 100042. <https://doi.org/10.1016/j.digbus.2024.100075>
- Rivadeneira, L. y Loo, I. (2025). X Data-Based Research: A Review of Trends and Challenges. *Revista Latina de Comunicación Social*, 83, 95-113.
- Rodríguez Cano, C. A. (2022). *Hipermétodos: repertorios de la investigación social en entornos digitales*. Universidad Autónoma Metropolitana.
- Rodríguez-Pérez, C., Ortiz-Calderón, L. S. y Esquivel-Coronado, J. P. (2021). Desinformación en contextos de polarización social: el paro nacional en Colombia del 21N. *Anagramas: Rumbos y Sentidos de la Comunicación*, 19(38), 129-156. <https://doi.org/10.22395/angr.v19n38a7>
- Rovira, C., Codina, L. y Lopezosa, C. (2021). Language Bias in the Google Scholar Ranking Algorithm. *Future Internet*, 13(2), 31. <https://doi.org/10.3390/fi13020031>
- Scott, J. (2013). *Social Network Analysis* (3.ª ed.). Sage Publications.
- Timans, R., Wouters, P. y Heilbron, J. (2019). Mixed Methods Research: What it is and what it Could Be. *Theory and Society*, 48, 193-216. <https://doi.org/10.1007/s11186-019-09345-5>
- Varol, O., Ferrara, E., Davis, C., Menczer, F. y Flammini, A. (2019). Bot Stamina: Examining the Influence and Staying Power of Bots in Online Social Networks. *Applied Network Science*, 4(55). <https://doi.org/10.1007/s41109-019-0164-x>
- Venturini, T. (2010). Diving in Magma: How to Explore Controversies with Actor-Network Theory. *Public Understanding of Science*, 19(3), 258-273. <https://doi.org/10.1177/0963662509102694>
- Venturini, T., Munk, A. K. y Jacomy, M. (2019). Actor-Network vs Network Analysis vs Digital Networks: Are We Talking about the Same Networks? En J. Vertesi y D. Ribes (eds.), *DigitalSTS: A Handbook and Fieldguide* (pp. 510-523). Princeton University Press. <https://hal.science/hal-01672289v1>

Notas

* Artículo de investigación

1 Los algoritmos de espacialización, conocidos como *layouts*, permiten organizar y visualizar redes complejas de forma legible. Herramientas como Gephi han popularizado *layouts* como Force Atlas2, especialmente útiles para representar redes con gran número de nodos y aristas, y que facilitan la interpretación de macrodatos.

2 Sintaxis 1: “network analysis” AND “twitter” AND “communication” AND “political science” AND “Latin America” AND “Argentina” OR “Brazil” OR “Chile” OR “Colombia” OR “Mexico” OR “Peru” OR “Ecuador” OR “Uruguay” OR “Venezuela” OR “Paraguay” OR “Panama” Sintaxis 2: (“network analysis” OR NSA) AND “twitter” AND “communication”

AND “political science” AND “Latin America” AND (Argentina OR Brazil OR Chile OR Colombia OR Mexico OR Peru OR Ecuador OR Uruguay OR Venezuela OR Paraguay OR Panama), Syntaxis 3: (“network analysis” OR NSA) AND “twitter” AND “communication” AND “political science” AND (“Latin America” OR “Central America” OR “Caribbean”) AND (Argentina OR Brazil OR Chile OR Colombia OR Mexico OR Peru OR Ecuador OR Uruguay OR Venezuela OR Paraguay OR Panama OR Costa Rica OR Cuba OR Dominican Republic OR Guatemala OR Haiti OR Honduras OR Jamaica OR Nicaragua OR Puerto Rico OR El Salvador).

3 Elaborado por Juan Federico Pino-Uribe en 2023.

4 Esta guía, desarrollada en lenguaje ., pero publicada en Google Colab para facilitar su acceso y uso, contiene un *script* diseñado para ilustrar paso a paso el flujo completo de un análisis de redes sociales (ARS). El recurso permite a los investigadores reproducir de forma transparente procesos como la captura de datos, la limpieza y depuración, la construcción de matrices relacionales, la generación de grafos y la visualización de redes. Además, incluye ejemplos comentados que estandarizan la documentación y promueven la replicabilidad de los estudios. La guía está disponible en el siguiente enlace: <https://colab.research.google.com/drive/1Vu61bB-rEeWHSuWU0xsQs4TtVJ5EuwOV?usp=sharing>

5 Estas citas son ejemplos didácticos para ilustrar los enunciados de cada uno de los apartados de la RCM.

6 Sin embargo, asumir que los investigadores eligen Twitter/X únicamente por la facilidad de acceso resulta problemático. Tras la compra de la empresa por Elon Musk, la plataforma ha experimentado cambios que han restringido progresivamente el acceso a los datos mediante su API, dificultando la investigación académica. Además, Twitter/X posee prácticas dinámicas de apropiación política que lo hacen especialmente atractivo para el análisis de fenómenos de comunicación política, lo que complejiza la relación entre la elección de la plataforma y la naturaleza de los estudios analizados.

7 F-IDF (*Term Frequency-Inverse Document Frequency*) es una métrica que pondera la relevancia de una palabra considerando su frecuencia en un documento y su rareza en el conjunto de textos, lo que permite identificar términos clave y filtrar palabras comunes.

8 En procesamiento de texto, un *token* es la unidad mínima en la que se segmenta un texto —por lo general, palabras o signos— que sirve como base para el análisis computacional.

9 Una escala *log-min-max* combina la transformación logarítmica con la normalización *min-max*, reduciendo la influencia de valores extremos y ajustando los datos a un rango común, típicamente entre 0 y 1.

10 NodeXL es un complemento para Microsoft Excel que facilita el análisis y visualización de redes sociales, el cual permite explorar relaciones, calcular métricas clave y generar gráficos interactivos. Está disponible en versiones gratuitas y Pro y puede descargarse en <https://www.smrfoundation.org/nodexl/>. Requiere Windows y Excel.

11 Gephi es una herramienta de *software* libre para analizar y visualizar redes complejas, como redes sociales, gráficas de relaciones o estructuras de datos interconectados. Permite explorar grandes conjuntos de datos, calcular métricas clave y generar visualizaciones dinámicas e interactivas. Es compatible con sistemas Windows, macOS y Linux y puede descargarse gratuitamente desde su página oficial: <https://gephi.org/>

Licencia Creative Commons CC BY 4.0

Cómo citar: Pino-Uribe, J. F., Lombana-Bermúdez, A., Oñate-Bolaños, A. y Castiblanco Briceño, M. E. (2025). Del dato al grafo: rigor, ética y replicabilidad en el análisis de redes sociales en comunicación y ciencia política en América Latina (2009-2022). *Signo y Pensamiento*, 44. <https://doi.org/10.11144/Javeriana.syp44.dgre>