Chinese Journal of Communication

ISSN: 1754-4750 (Print) 1754-4769 (Online) Journal homepage: www.tandfonline.com/journals/rcjc20

Unraveling China's digital traces: evaluating communication scholarship through a sociotechnical lens

Kaiping Chen, Yingdan Lu & Yiming Wang

To cite this article: Kaiping Chen, Yingdan Lu & Yiming Wang (2024) Unraveling China's digital traces: evaluating communication scholarship through a sociotechnical lens, Chinese Journal of Communication, 17:2, 127-150, DOI: 10.1080/17544750.2023.2264406

To link to this article: https://doi.org/10.1080/17544750.2023.2264406

Published online: 11 Oct 2023.

CHINESE JOURNAL OF COMMUNICATION
2024, VOL. 17, NO. 2, 127–150
https://doi.org/10.1080/17544750.2023.2264406

Routledge
Taylor & Francis Group

Unraveling China's digital traces: evaluating communication scholarship through a sociotechnical lens

Kaiping Chen[a], Yingdan Lu[b] and Yiming Wang[c]

[a]Department of Life Sciences Communication, University of Wisconsin-Madison, WI, USA; [b]Department of Communication Studies, Northwestern University, Evanston, IL, USA; [c]School of Journalism and Mass Communication, University of Wisconsin, Madison, WI, USA

ABSTRACT

In the growing trend of research using digital trace data to study human activities and opinions across different contexts, networked China has emerged as a prominent area of interest. However, research that critically examines the use, strengths, and weaknesses of existing digital trace methods, and the extent to which they can reveal the true landscape of digital China remains limited. To address these gaps, this study proposes a framework for examining and evaluating the knowledge production of digital trace research within a sociotechnical system comprising state actors, platform governance, digital civil society, and international forces. We then provide the first empirical examination of the knowledge claims and epistemic approaches used in digital trace communication scholarship that has studied China across different phases in the past 30 years. Grounded in the resulting empirical evidence, we discuss two common practices in existing digital trace research on China, how these approaches and perspectives could affect the validity and reliability of offering diverse viewpoints for studying and understanding digital China, and directions for improving these practices.

ARTICLE HISTORY
Received 12 January 2023
Accepted 28 August 2023

KEYWORDS
Digital trace; knowledge production; evidence-driven approach; China; sociotechnical system; computational social science

Introduction

In China, the internet's infiltration into everyday life has significantly changed how people acquire information and communicate. As a result, many human activities can be tracked on the internet through digital trace data—evidence of human activity recorded digitally (Howison et al., 2011)—such as messages, images, mobile logs, and hyperlinks. Digital trace data are not only indicators of human activities; they also signify more general underlying social phenomena, such as public opinion and state–society interactions. An increasing stream of research uses these data to examine the

CONTACT Yingdan Lu yingdan@northwestern.edu
All three authors contributed equally to this work.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/17544750.2023.2264406.

sociotechnical-cultural landscape of digital China (Chen & Reese, 2015; Qiu et al., 2022), such as how online affordances facilitate economic reform (Meng & Li, 2002), cultural prosperity (Yang, 2003), and state influence (King et al., 2013).

With this growing trend of using digital trace data to understand communication phenomena in China, there is a need for a critical and systematic reflection on how digital trace data are used and the knowledge claims made from these data. However, there is little discussion about the strengths and weaknesses of the epistemic approaches that scholars have adopted in using digital trace data to explain communication phenomena in China, which constrains our ability to assess the validity and reliability of such claims. This limited reflection further restricts opportunities to develop creative and quality data sources for studying the Chinese digital public sphere.

To address these gaps, we offer a sociotechnical and evidence-driven approach to understanding digital trace communication scholarship that studies China, with special attention to three areas: (a) the need for an empirical investigation to understand the knowledge claims and epistemic approaches utilized in existing digital trace research on China in different phases; (b) the need for an analytical framework that assesses the validity and reliability of existing digital trace research on China in a dynamic system of state, platform, and civil society interactions and international influences; and (c) the need for a more China-specific reflection to identify the unique challenges of using digital trace data in the context of digital China.

The paper is organized as follows. Section 1 introduces the framework for assessing the validity and reliability of digital trace research in a dynamic sociotechnical system that consists of state, platform, and civil society interactions and international influences. Section 2 presents the findings from our empirical analysis of digital trace research that studies China in the past 30 years in communication journals. By presenting the results in different temporal phases, we reveal and compare how researchers collected and analyzed digital trace data, the topics they studied and the positionalities of authors, and how these shifts mirrored evolving domestic sociotechnical system and international influences. Section 3 applies the evaluation framework in Section 1 to discuss two common practices in existing digital trace research on China and how these approaches and perspectives could affect the validity and reliability of offering diverse viewpoints for studying and understanding digital China. In the same section, we also provide suggestions for pathways forward.

Evaluating digital trace research in a dynamic system of state–platform–civil society interactions and international influences

Despite recent efforts to build integrated frameworks that guide digital trace studies globally (Freelon, 2014; Howison et al., 2011) and despite the increased interest in digital trace studies on China, little reflection has been done on digital trace data in non-Western contexts. China has a unique media environment with a highly localized media market (Montag et al., 2018) and ubiquitous and sustained government control of both domestic and transnational information flows (King et al., 2013; Lu et al., 2022). As such, assessing the strengths and weaknesses of using digital trace data to study digital phenomena in China without frameworks that take its political and social contexts into account is difficult.

When it comes to reflecting on academic scholarship, sociologists have proposed the concept of epistemic reflexivity. Originating from Bourdieu's reflexive sociology, epistemic reflexivity has motivated researchers to scrutinize the presuppositions they use to understand the world (Bourdieu, 1990, p. 382)—how they build concepts, construct instruments (e.g. statistical techniques), and operate their research (e.g. coding, data cleaning, and rules of thumb) (Wacquant, 1992, p. 40). This suggests the importance of examining who the scholars are, the subjects they have studied, the knowledge claims they have made, and the epistemic approaches they have used (Maton, 2003).

Indeed, previous works have offered a valuable framework for dissecting existing scholarship using a science-of-science approach that primarily concentrates on the knowledge production process, involving epistemic reflection on how researchers develop concepts, apply them for analysis, and address their own biases. However, the framework lacks depth in explaining how epistemic reflexivity can extend to encompass the wider social, cultural, and political systems. As recent scholarship has highlighted, in the context of digital trace research on China, it is crucial to examine its knowledge production against the backdrop of a broader sociotechnical system that can shape the measurement regimes (e.g. how platforms manage their internal audience analytics to administer their user ecosystems while also structuring their advertising markets) in digital trace research; such a system consists of domestic political and technological factors, such as state influence and platform governance, as well as international factors, such as geopolitical influences (Y. Chen et al., 2023; Qiu et al., 2022; Wu & Taneja, 2021). Joining this recent scholarship that emphasizes the importance of evaluating the knowledge production of digital trace research in a sociotechnical system, we explicate the components of such a sociotechnical system (domestic and international), their changing relationships, and how these components might challenge the validity and reliability of digital trace research in order to provide evidential values for understanding digital China (Figure 1, Table 1).

Examining the **validity** of these knowledge claims involves asking whether the digital trace data used in a publication can comprehensively measure the communication concepts that the authors intend to measure. The validity challenge facing digital trace studies highlights the nature of digital data as found data (Howison et al., 2011). Because the data are found unobtrusively, studies using these data might not be able to answer research questions that need strong interventions (e.g. fully controlled experiments to

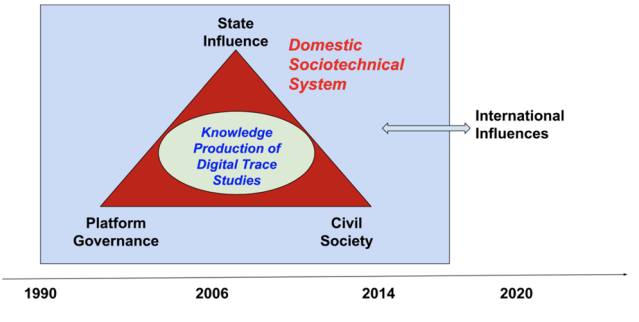

Figure 1. A sociotechnical lens for evaluating digital trace research.

Table 1. Assessing the validity and reliability of digital trace research through a sociotechnical lens: an analytical framework.

State influence	Operations that state actors deploy (either manually or by affecting private agents) to change the ecosystem on which digital traces rely	Challenge the internal validity (e.g. content and discriminant validity) of using digital trace data to measure the communication concepts that researchers intend to study Challenge the reliability of research conclusions because of inconsistency in how data are generated and collected over time. This further lowers the reliability of open science that stresses reproducibility.
Platform governance	Operations that platforms deploy for their administrative purposes or for compliance with state instructions	Challenge the validity of using digital trace data to measure communication concepts, as researchers often overlook or fail to investigate the commercial and political dimensions of platforms that can influence the actors, content, and interactions on digital platforms Challenge the reliability of using digital trace data, especially when content or users are silently removed. This makes it difficult to ensure sample consistency when researchers attempt to replicate previous findings.
Civil society	The changing media diets across generations and the constantly evolving coded languages used by netizens to contest state actors	Challenge the validity of using digital trace data, as researchers need to be aware of the evolving coded languages and changing media diets that affect which platforms users post content on. Without this awareness, researchers may fail to develop a comprehensive list of keywords and platforms from which to collect the discourse related to their research topics. Challenge the reliability of using digital trace data, as researchers need to constantly update the keywords they use to sample discourse about an issue, as well as to update the list of users included in their samples
International influence	The international geopolitical actors who influence how Chinese state actors interact with platforms and digital civil society	Challenge the validity of using data trace data, as researchers need to investigate how the relationships between China and other geopolitical actors might influence cyber policies in China and the content and interactions on the digital public sphere Challenge the reliability of using digital trace data, as international geopolitical tensions may prompt platforms in China to revise their data access policies, potentially making data collection inconsistent

study mechanisms) and sampling designs (e.g. targeting specific populations to answer a research question). Examining the **reliability** of these knowledge claims involves asking whether the digital trace data used in a study can *consistently* measure the concepts of interest.

Domestic sociotechnical systems and their impact on digital trace research

In the context of China, we highlight three major interrelated domestic sociotechnical forces that have shaped how academics collect, analyze, and make sense of digital trace data—*state influence*, *platform governance*, and the dynamics of *digital civil society*.

State influence and digital trace research

State influence refers to how various levels of state actors use information control strategies to shape the content and interactions on digital platforms, including but not limited to censorship (King et al., 2013), clickbait (Lu & Pan, 2021), and ideotainment (Zou, 2023). As Yang (2022) noted, there is a long history of "the political and cultural shaping of the internet in China" (p. 37). Party-sponsored platformization includes both traditional methods of content moderation and new strategies, such as the use of positive energy discourse or direct participation by government official accounts to occupy the digital public sphere (Yang, 2022). Recognizing that state actors are not monolithic is important. As Qiu et al. (2022) stressed, there are different segments of officialdom that extend beyond the central government, such as business and society, some members of which oppose Beijing's policies. Assessing the validity of research on the resulting data requires us to ask the extent to which state intervention might affect the processes of collecting and analyzing digital trace data, which may differ across the types of research questions and topics studied and the epistemic approaches used by researchers.

The influence of various levels of state actors also affects the reliability of digital trace work. For researchers who collect and sample data from social media platforms, some actors who used to be on the platforms and were thus in the dataset might be removed at a certain point by state actors, or they might have closed their accounts, making it difficult to reproduce the findings from prior studies. This also makes it difficult to ensure sampling consistency over time, as certain actors in a sampling quota might disappear.

Platform governance and digital trace research

While state actors can influence platforms, platforms themselves also shape digital trace data from their own perspectives. For instance, platform companies are motivated by their commercial and administrative purposes to design logging, analysis, and ad placements that help sustain their business models (Y. Chen et al., 2023; Wu & Taneja, 2021). Besides the commercial logic, platforms in China need to follow the political logic and comply with state instructions in order to secure their operations, such as to ensure a clean and civilized internet and thus meet the state's cyberspace guidelines (Horsley, 2022). Taking Weibo Trending Topics as an example, some evidence suggests that trending topics do not simply aggregate the most popular search terms to rank topics; instead, the platform has a hidden agenda to comply with state actors by banning certain search terms from reaching the top rank (Romano, 2022).

The commercial and political logics of platform governance can affect the validity and reliability of digital trace research. Platform datafication, in which "platform data are generated as platforms' records of their own behavioral experimentation … [and] a platform sets up its measurement infrastructure to create data that improve design features, fuel advertising analytics, and please investors" (Y. Chen et al., 2023, p. 260), can bias research results. Wu and Taneja (2021) emphasized that a persistent problem facing quantitative social science is the following: "Studying platform user behavior while maintaining the positivist commitments requires the researcher to position herself as an outsider to this behavior, and to achieve this entails taking hold not

only of platform log data but also methodical data about platform architectures and internal administrations. The absence of the latter, usually the case with academics, effectively undercuts the epistemic integrity of quantitative social science" (p. 2656). While platforms sometimes reconstruct their technology features to comply with state instructions, these practices are not known to researchers (Freiling et al., 2021).

Platform governance also affects the reliability of digital trace work. Take application programming interfaces (APIs) as an example; APIs modify data in ways that are not transparent to researchers, as companies change them without warning for reasons including but not limited to user privacy (Y. Chen et al., 2023). These undetected changes in APIs or data gatherers' sampling strategies make it impossible to study the same phenomenon consistently over time (Howison et al., 2011).

Digital civil society and digital trace research

While state actors and platforms have been actively influencing digital discourse, the dynamics of and interactions within digital civil society, consisting of individual users and profit and non-profit organizations, also influence the creation and content of digital trace. For instance, netizens have used platforms to criticize or praise the policies of different governments (A. Chen et al., 2023; Han, 2018; King et al., 2013; Pan & Chen, 2018). Besides these persistent interactions between the state and the public, Guan (2019) noted that many China-focused political communication studies run the risk of falling along the authoritarian determinism and reductionism approach, in which the central focus is repression and resistance. However, recognizing other relationships and networks on the internet beyond the state–society conflict paradigm is crucial. For example, the rise of online celebrities and influencers has motivated more research to use digital trace data for examining fan communities' participatory cultures, such as how fan communities harness platform functions to build communities (Yin & Xie, 2021) or engage with state propaganda on social media for more visibility (Lu, 2023). Recognizing heterogeneity within digital civil society is also essential. For instance, research shows that while some users on WeChat are activists or watchdogs when they encounter alternative political news, others are lurkers (Guo, 2017).

The evolving and heterogeneous nature of the digital public sphere can challenge the validity and reliability of digital trace research. This evolving nature includes generational changes that result in corresponding changes in users' social media diet and habits (Fietkiewicz et al., 2016; Q. Hu & Cheong, 2021) and the coded languages developed by users to build civil communities or to counter state influence. For instance, in the early 2000s, netizens often used Bulletin Board System (BBS) forums to share opinions (Yang, 2012), while in recent years, Douyin and Xiaohongshu have become new forums in which people discuss political and non-political issues. Without investigating the changing social media diet of users over time, researchers can miss important discussion threads and platforms in which they need to sample data in order to examine a communication phenomenon. Besides generational changes, the coded languages that netizens use and update to counter state influence pose difficulties in measurement and data collection. For instance, during the COVID-19 pandemic, netizens in China used various words from games and anticensorship relay campaigns (Sun, 2023) to counter censorship about local misconduct. Without a deep understanding of how netizens chose and updated their

languages to bypass state intervention over time, the dataset that researchers collect will not be complete, and the measurements for developing content analysis codebooks to capture these civil resistance complexities will also be deficient.

The evolving nature of the digital public sphere also challenges the reliability of conducting digital trace work. With content being removed or user accounts being censored in a shorter time span, researchers find it difficult to reproduce findings in previous scholarship and to measure other open science practices, such as data sharing.

The changing domestic sociotechnical system and digital trace research

It is important to note that the influence of these three domestic sociotechnical forces on digital trace research is dynamic, as each force and its interactions are changing over time. Looking at the history and development of the internet in China, we can see that the late 1990s to 2000s witnessed a rapid growth of online civil society with the development of information and communications technologies. Online bulletin board systems, chat rooms, and personal home pages have started to grow (Yang, 2009). People from different socioeconomic statuses use these spaces not only to share non-political topics but also to form cultural expressions and networks to protest against local governance (Qiu, 2009; Yang, 2009). Social media companies in the 2000s also utilized this golden period to build and increase their user bases by developing technology functions (e.g. initiating topics, voting, and posting content in a multimodal format) to attract users' attention, boost clicks, and build and sustain user communities.

The year 2014 was a critical moment for the Chinese internet when a new national security commission was established to reconstruct cybersecurity regulations. Three years later, a new cybersecurity law was promulgated, requiring any purchase of foreign software and hardware to be reviewed by the state cybersecurity office. These new rules allowed state actors to exert more influence in managing the content and actors in the domestic digital public sphere. Despite rising regulation and stricter control on the internet, civil society also reacts with more coded languages and metaphors among netizens to navigate through state censorship and manipulation (Lei, 2018; Link & Qiang, 2013; Yang, 2022).

These interesting and dynamic relationships between the state, platform companies, and digital civil society highlight the necessity for digital trace research not only to attend to the content and actors on digital platforms but also to take a step further to investigate how content creation and interaction among netizens in digital China mirror or are shaped by the changing sociotechnical system. As Guan (2019) stressed, it is crucial for communication research about China to break from the event-based approach and instead adopt a temporal perspective (i.e. wider periods) or longitudinal designs to capture "the differentiation and segmentation of views within the Chinese media and related public discussions" (p. 742). Otherwise, researchers will run the risk of producing an outdated understanding of digital discourses.

International forces and digital trace research

Besides assessing digital trace research by understanding how it is influenced by the domestic sociotechnical system, scholars stress the necessity of understanding

the development of the Chinese internet through a techno-geopolitics lens (Qiu et al., 2022). The Chinese internet not only operates at the national level but is also influenced by events and actors in other parts of the world. This requires scholars to adopt a multilayered approach in digital trace research that analyzes the Chinese internet while attending to how it interacts with these international influences (Qiu et al., 2022). Some examples include how transnational firms, such as Zoom (Chen, 2022), Apple, and IBM (M. Tang, 2022), negotiate solutions with Chinese authorities in order to continue business in China, and how Chinese companies, such as Alibaba (Shen & He, 2022), need to adjust their business models when they go global.

The geopolitics between China and the West and between China and Africa and the Middle East also shape the interaction between the state and digital civil society. As Hong and Harwit (2020) noted, foreign policy debates around the world and the ongoing US–China trade dispute have stimulated technology companies, such as TikTok and Huawei, into geopolitical actions and countermeasures. In response to these international influences, the authors called for researchers to examine the internet as an analytic construct that consists of "supranational entities … [and a] networked public within and across which ideas, interests, and arrangements emanating from China mingle, align, or delink with transnational forces" (p. 2).

These international forces can influence the validity and reliability of digital trace research. In terms of validity, without understanding the influence of international forces, researchers might blindly examine the Chinese internet through an authoritarian gaze (i.e. CCP's control approach). For instance, most studies that examine information control in China tend to focus on state actors in order to study their information control strategies. However, as Qiu et al. (2022) flagged, it is important to approach the Chinese internet using a multilayered strategy for investigating how regional and international factors might shape state actions and users' content. Without putting international factors into the equation of understanding content creation on the Chinese internet, researchers can miss the crucial players that shape the Chinese internet. Geopolitical conditions can also influence the stability of data collection. For instance, Sina Weibo used to provide API access to collect data, which made it easier for researchers to study large-scale Chinese online discourses. However, around 2019, with the heightened tensions between the United States and China, such as cybersecurity issues, Sina Weibo updated its user agreement and required login credentials for its API access; researchers found that this made it more difficult to collect data on the same scale as before (D. Hu et al., 2020). These constantly changing data access policies from platforms could challenge researchers in terms of reproducing findings from prior work and ensuring consistency in data collection for longitudinal studies that track digital opinions.

Research questions

This study proposes the following research questions to examine how the existing body of scholarship uses digital trace data to study China and how the shift in this

scholarship in terms of data, methodology, and area of focus reflects the broader sociotechnical system in China and international influences.

RQ1: What knowledge claims were made by digital trace research about China in our field's journals between 1990 and 2020, and how were these claims made?

RQ2: How do the knowledge claims and methods in digital trace research reflect the broader sociotechnical system in China and international influences?

RQ3: In what ways are common practices in digital trace research valid and reliable for studying communication phenomena in China, and what are future pathways to enhance these practices?

Empirical investigation: uses of digital trace data to study China from 1990 to 2020

This study presents the first empirical investigation into how existing scholarship uses digital trace data to study Chinese media and communication (Figure 2). We collected all the abstracts and related metadata ($N = 85{,}671$) published in the 95 communication journals indexed in the Web of Science from 1990 to 2020, covering the 30 years since the invention of the World Wide Web. Our dataset was retrieved in February 2022. We then developed a lexicon of digital trace-related words drawn from Stier et al. (2020) and Freelon (2014) to identify abstracts that focused on Chinese media and communication and potentially used digital trace data. The keywords included: (1) trace sources, such as hyperlinks and smartphone logs; (2) collection methods, such as APIs; and (3) Chinese online platforms, such as Weibo and Douyin (see Supplementary Material I). This keyword-based computational approach identified 980 abstracts that likely employed digital trace data to study China. Three researchers then reviewed all the abstracts for the use of digital trace data on the basis of the typology we developed from Freelon (2014; see Supplementary Material I). After excluding the abstracts that relied solely on non-digital trace data and those with data that were not clearly explained, we had 121 papers for in-depth content analysis (Fleiss' kappa = 0.841). The results presented below focus on these 121 papers.

Given the changing dynamics of the sociotechnical system laid out in Section 1, our empirical analyses offer a temporal examination of how digital trace research was conducted over three major phases: (1) 1990–2009, which captures the overall enthusiasm of a new public sphere across scholarship; (2) 2010–2014, which features scholarships focusing on topics related to state and capital power; and (3) 2015–2020, when more diverse topics and epistemic routes were approached (Figure 3). For digital trace research in each phase, we examined the authors' social positionalities, the objects of the studies, and the epistemic approaches that the authors used to make knowledge claims from their full papers (see codebook in Supplementary Material II). By combining a longitudinal lens and a sociotechnical approach, we presented how academic practices of digital trace research shifted (Table 2) and how these shifts interacted with state politics, platform governance, civil society, and geopolitical contexts in different phases.

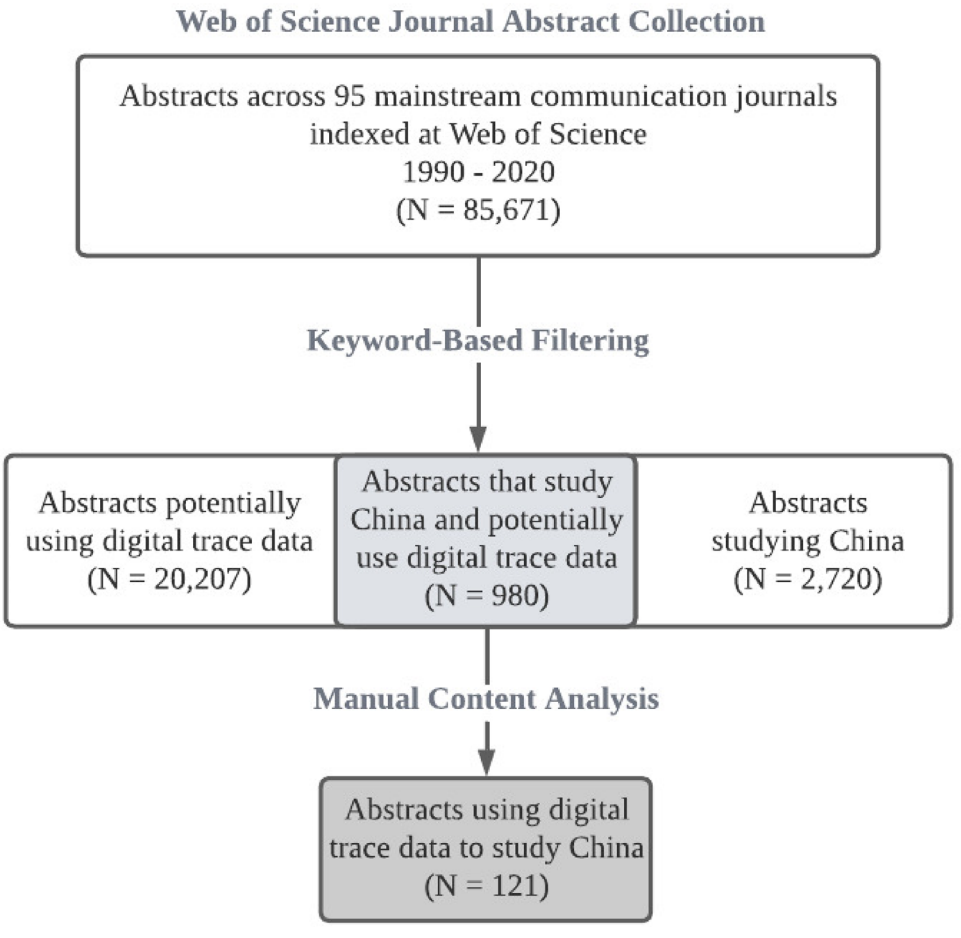

Figure 2. Data collection flowchart.

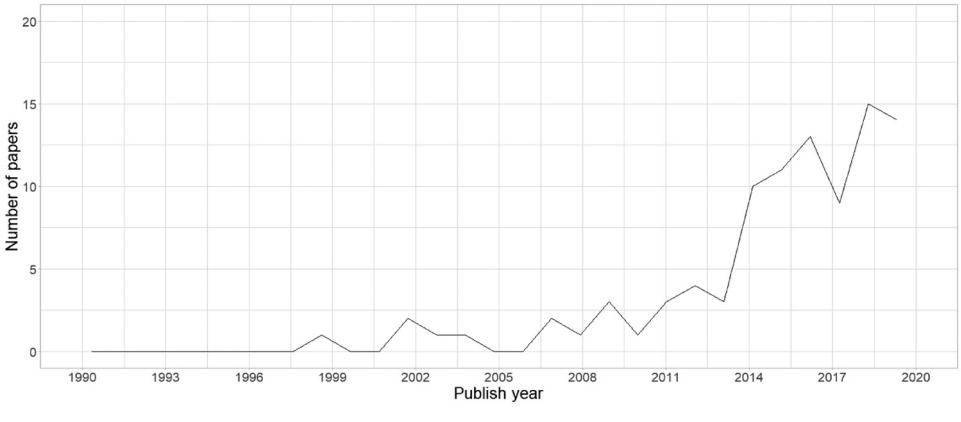

Figure 3. Number of papers using digital trace data to study China (1990–2020).

Phase I (1990–2009): enthusiasm of a new public sphere

The adoption of the World Wide Web in China motivated researchers to start using digital trace data to understand China in the late 1990s. From the emergence of digital trace research to the surge in social media use in China (1990–2009), a small number of journal articles ($n = 11$, 9.1% of 121 papers) employed digital trace data as a new means of understanding China. The authors' backgrounds, their positionalities, and the subjects being studied during this period reflected the domestic sociotechnical system and international conditions at the turn of the century. Professors in academic institutions were the first to adopt digital trace data in China studies ($n = 11$, 100%). The regions of the authors' institutions were based outside mainland China ($n = 9$, 81.8%), with 54.5% hailing from Hong Kong, Taiwan, and other regions in Asia, and 27.3% from North America. More studies during this period focused on phenomena happening in Hong Kong (63.6%) rather than in mainland China (54.5%). These findings suggest that the field of Chinese digital trace research initially emerged from scholars located outside mainland China and that they focused on communication phenomena outside mainland China.

Studies during this period featured authors' enthusiasm surrounding the advent of new technology for political change and cultural prosperity, with 36.4% of papers investigating the internet as a public sphere for political and cultural communication within China and beyond. Identity (18.2%) was a popular topic among these early China digital trace studies, which was explored against the backdrop of geopolitics and social changes in the 1990s. For instance, the local identity of Hong Kong residents was studied following Hong Kong's handover in 1997 (Fung, 2002). Nakamura (2009) highlighted the racialization of Chinese player workers in the context of virtual gaming, shedding light on the internet's potential for facilitating racial disparities. Studies published during this period also examined the cross-border flow of information (18.2%) and the online strategic communication of international brands (18.2%), underscoring a broader interest in the World Wide Web's connectivity and global impact at the dawn of the twenty first century.

During the pre-social media era, digital data were limited in terms of availability and accessibility. As such, a significant proportion of studies have focused on websites and BBS forums as the primary sites and data sources for examining institutional behaviors and public communication in China (90.9%). For example, one study during this period used news articles from Sina.com, a news portal site, and discussion data from the Qiangguo forum (强国论坛) to explore the impact of online discussions of local events on the nationalization of these events.

The studies published during this early period demonstrated a diverse range of methodologies for analyzing digital trace data. Digital ethnography (9.1%), qualitative content and discourse analysis (81.8%), and case studies (27.3%) were used. Computational methods, including network analysis and computer-assisted content analysis, were employed in two of the papers (18.2%). Digital trace data were triangulated with other methods, including surveys and in-depth interviews, in 36.4% of the publications. Notably, the very first digital trace study on China adopted a more flexible and open-minded approach in ethnographic and content analyses. Overall,

Table 2. Evolving features of China's digital trace studies (1990–2020).

	1990–2009	2010–2014	2015–2020	1990–2020
Number of papers	11 (9.1%)	21 (17.4%)	89 (73.6%)	121
Author's job				
Professor	11 (100%)	15 (71.4%)	47 (52.8%)	73 (60.3%)
Student	0 (0%)	4 (19.0%)	36 (40.4%)	40 (33.1%)
Industry researcher	0 (0%)	1 (4.8%)	1 (1.1%)	2 (1.7%)
Others	0 (0%)	1 (4.8%)	5 (5.6%)	6 (5.0%)
Author's region				
Mainland China	2 (18.2%)	5 (23.8%)	19 (21.3%)	26 (21.5%)
Other Asian regions/ countries	6 (54.5%)	1 (4.8%)	22 (24.7%)	29 (24.0%)
North America	3 (27.3%)	8 (38.1%)	30 (33.7%)	41 (33.9%)
Europe	0 (0%)	6 (28.6%)	12 (13.5%)	18 (14.9%)
Australia	0 (0%)	1 (4.8%)	5 (5.6%)	6 (5.0%)
Africa	0 (0%)	0 (0%)	1 (1.1%)	1 (0.8%)
Region of study (not mutually exclusive)				
Within mainland China	6 (54.5%)	19 (90.5%)	67 (75.3%)	92 (76.0%)
In Hong Kong/Macau/ Taiwan	7 (63.6%)	1 (4.8%)	18 (20.2%)	26 (21.5%)
In other Asian regions/ countries	4 (36.4%)	0 (0%)	6 (6.7%)	10 (8.3%)
In other non-Asian regions/countries	5 (45.5%)	3 (14.3%)	13 (14.6%)	21 (17.4%)
Data type (not mutually exclusive)				
Websites/software	10 (90.9%)	12 (57.1%)	39 (43.8%)	61 (50.4%)
Social media / API	0 (0%)	5 (23.8%)	57 (64.0%)	62 (51.2%)
PC/smartphone	0 (0%)	0 (0%)	1 (1.1%)	1 (0.8%)
Sensors	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Others	1 (9.1%)	4 (19.0%)	4 (4.5%)	9 (7.4%)
Specific method used				
Computational method	2 (18.2%)	5 (23.8%)	28 (31.5%)	35 (28.9%)
Digital ethnography	1 (9.1%)	3 (14.3%)	9 (10.1%)	13 (10.7%)
Data triangulation	4 (36.4%)	6 (28.6%)	28 (31.5%)	38 (31.4%)
Domain				
Political communication	5 (45.5%)	8 (38.1%)	35 (39.3%)	48 (39.7%)
Culture, race, and gender	3 (27.3%)	6 (28.6%)	30 (33.7%)	38 (31.4%)
PR, ads, or strategic communication	2 (18.2%)	5 (23.8%)	4 (4.5%)	11 (9.1%)
Interpersonal communication	0 (0%)	1 (4.8%)	5 (5.6%)	6 (5.0%)
Organizational communication	0 (0%)	0 (0%)	6 (6.7%)	6 (5.0%)
Health communication	0 (0%)	0 (0%)	6 (6.7%)	6 (5.0%)
Media psychology	0 (0%)	0 (0%)	3 (3.4%)	3 (2.5%)
Journalism studies	0 (0%)	1 (4.8%)	2 (2.2%)	3 (2.5%)
Methodology	1 (9.1%)	0 (0%)	0 (0%)	1 (0.8%)
Science communication	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Unclear/hard to determine	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Topic				
Identity	2 (18.2%)	0 (0%)	16 (18.0%)	18 (14.9%)
Online culture	0 (0%)	5 (23.8%)	12 (13.5%)	17 (14.0%)
User behavior	0 (0%)	0 (0%)	13 (14.6%)	13 (10.7%)
Activism	0 (0%)	3 (14.3%)	9 (10.1%)	12 (9.9%)
Public sphere	4 (36.4%)	1 (4.8%)	6 (6.7%)	11 (9.1%)
Propaganda and censorship	0 (0%)	2 (9.5%)	7 (7.9%)	9 (7.4%)
Information flow	2 (18.2%)	1 (4.8%)	6 (6.7%)	9 (7.4%)
Brand strategy	2 (18.2%)	4 (19.0%)	1 (1.1%)	7 (5.8%)
NGO communication	0 (0%)	0 (0%)	4 (4.5%)	4 (3.3%)
International communication	0 (0%)	0 (0%)	3 (3.4%)	3 (2.5%)
Privacy	0 (0%)	1 (4.8%)	2 (2.2%)	3 (2.5%)

Note: The topic table only presents the topics with a total frequency > 2.

these studies emphasized the importance of researchers entering the sites with a mind free of preconceptions (e.g. Yang, 2003).

Phase II (2010–2014): the influence of the state and capital power

The years from 2010 to 2014 witnessed a remarkable increase in the number of digital trace studies conducted on China ($n = 21$, accounting for 17.4% of 121 papers). This coincided with the rise of social media platforms and the involvement of multiple stakeholders, such as the state tightening its online regulation and control, platforms coping with the regulation while also maximizing profit, and international organizations promoting internet freedom, in challenging the power dynamics surrounding the internet in China. During this period, while professors remained the dominant force in employing digital trace data to comprehend China (71.4%), student researchers also emerged as a new force (19.0%). This period also saw greater domestic interest in digital trace research, with a larger proportion of papers published by scholars based in mainland China (23.8%), and the majority of these studies also focused on communication phenomena within mainland China (90.5%).

The domains and topics of these studies reveal the dynamic and ever-evolving nature of China's online landscape and the complex interplay between technology, politics, and society. Political communication remained the most significant area of focus during this time (38.1%). Among political communication studies, those that examine online activism emerged during the five-year period (25.7%), highlighting how Chinese netizens use online platforms for collective action to challenge government legitimacy. Another topic of interest that arose within political communication was censorship and internet control (25%). These studies did not simply view censorship as a state program but explored delegated censorship systems involving the central government, local institutions, service providers, and webmasters (Dong, 2012; Wright, 2014). Additionally, cultural studies received a fair amount of attention, with 23.8% of papers focusing on online cultures, including but not limited to new cultural phenomena, such as linguistic patterns in social media communication, translation groups, and cultural symbols on the internet. These cultural studies illustrate the role of social media in facilitating diverse cultures and subcultures within China's online sphere.

During this period, websites and BBS data remained the major sources of data, accounting for 57.1% of the data used. However, the types of data used expanded significantly. The availability of social media data through APIs provided by platforms, such as Weibo Open Platform, has made social media a significant source of data for studying social and political phenomena in China, comprising 23.8% of the data used. Other types of data used included Virtual Private Network (VPN) and Domain Name System (DNS) data, which accounted for 19.0% of the total data used. Similar to the first phase of the China digital trace study, 28.6% of the papers in this second phase used more than one type of data for triangulation purposes.

In terms of methodology, there was a significant increase in computational tools for analyzing digital trace data (23.8%). Researchers used network analysis, computational content labeling, and geolocation matching to analyze larger amounts of data. Still, many digital trace studies during these five years sought an in-depth

understanding of digital sites in the Chinese context, particularly those examining digital cultures. Inductive, qualitative data assessments and digital ethnography remained the dominant methods for understanding the power dynamics among online players, accounting for 59.1% of all China digital trace studies published during this period. The use of digital ethnography (14.3%) underscored the significance of "exploratory studies entering the space without preconceptions" (Dong, 2012, p.407), allowing researchers to explore Chinese-specific narratives rather than adopting a Western-centric perspective.

Phase III (2015–2020): a more diverse landscape of research topics and methodologies

As the internet has entered a new age of use with the rapid advancement of broadband cellular network technologies, research on digital trace in China has seen a boost since 2015, accounting for 73.6% ($n = 89$) of all China digital trace studies within the past 30 years. During this period, student authors became a strong force, publishing 40.4% of China digital trace papers. Notably, more scholars from non-Asian regions, such as North America (33.7%) and Europe (13.5%), used digital trace data to study China, except for authors based in mainland China (21.3%) and other Asian regions (24.7%).

During this period, most papers focused on communication phenomena in mainland China (75.3%), with a significant number also examining those in Hong Kong and Taiwan (20.2%). While political communication remained the most studied domain (39.3%), there was a substantial increase in the focus on cultural communication (33.7%). Identity studies spanned both the political and cultural communication domains, accounting for 18.0% of the studies during this period. Among identity studies, a significant focus was placed on the political and cultural aspects of group identities related to nationality and territories (43.8%), with Chinese nationalism studies emerging during this period (12.5% of all identity studies). These studies reflected the interplay between state nationalistic propaganda, civil society's cultural creativity, and the rising Sino–US and other geopolitical tensions.

Compared with the studies in the two previous phases, studies published after 2015 showed less emphasis on top-down political processes, such as propaganda and state censorship (7.9%). Instead, they focused more on the interaction between digital platforms and civil society. For instance, continuing with the second phase, activism studies (10.1%) during this period explored the bottom-up resistance of civil society in the online sphere. Studies on online cultures (13.5%) also delved into how platforms facilitate diverse online cultures, such as coded language, humor, and other tactics, to evade censorship and express dissent. Moreover, significant attention has been given to the factors and implications of users' choices in using different digital tools, posting and sharing information, and forming social networks online (14.6%).

In terms of the types of digital trace data, social media data—used by 64.0% of the papers—have become more prevalent than data from other internet-based portals. This reflects both the growing interest in social media activities and the increasing accessibility of this data type through the open APIs offered by platforms and third-party data providers. Nevertheless, website data remained significant, with 43.8%

of studies published from 2015 to 2020 using them. This indicates the consistent importance of websites in digital trace studies on China. Scholars have also begun using smartphone data for communication inquiries (1.1%). Data triangulation remained an important focus, with 31.5% of the papers using more than one type of data.

The methodologies used to analyze digital trace data have also benefited from the rapid development of computational analytical tools and packages. During this period, 31.5% of the published papers used computational methods, such as textual analysis, network analysis, and machine learning approaches. Compared with the earlier computational works in Phases I and II, which mainly used computers as calculation tools and usually adopted codebook and algorithmic rules built by the researchers themselves, 57.1% of the computational studies during this period overly relied on existing packages and algorithms. Notably, the popularity of computational methods did not take away the importance of digital ethnography (10.1%) and the qualitative investigation of digital trace data, such as critical discourse analysis (31.5%).

Epistemic reflection on digital trace research on China

Overreliance on social media data

Our empirical investigation shows an increasing reliance on social media data as the major type of digital trace data. While we acknowledge the strength of social media data in understanding individuals' online activities, we found that most social media studies, regardless of methodology, did not reflect on their data generation processes. Specifically, they did not consider the extent to which the tools used for data collection, such as APIs, crawlers, and other third-party data collection tools, provided accountable data for use. Data validation was rare among these studies. This overreliance on social media data raises critical validity and reliability concerns, especially when considering the platform, the state, civil society, and international factors.

First, as elaborated in our theory, the commercial and political logics of platform governance could affect data generation and analysis, introducing both validity and reliability concerns. This is evident in the overreliance on Weibo data by previous digital trace studies. In particular, Weibo's recommendation algorithm may undermine the validity of certain knowledge claims in these works. For example, the results of studies measuring the participation of online nationalist movements through the dissemination of a nationalist discourse could be biased because platform companies can recommend users with more state-favored content to either comply with state directives or to compete for state resources. The blackboxing view of algorithms in data generation and collection can also undermine the evidential value of understanding communication phenomena in China. For instance, several of the studies we reviewed used IssueCrawler, a mainstream web network mapping software program, for hyperlink analysis to portray social networks. Such studies did not consider the fact that at the time, IssueCrawler was not able to capture newer websites written in JavaScript, which could bias knowledge claims, such as the reach of non-government organizations' websites and the scale of globalized information flow. Similar concerns also occur in those studies that apply digital ethnography because the context and practices with which researchers interact can be algorithmically crafted. Meanwhile,

the frequent structural changes of platform APIs may bring a sudden crackdown to data collection, which endangers the reliability of the digital trace data collected if the researcher is not aware of the data instability, especially for computational analyses.

While researchers studying Chinese communication generally assume that state intervention could influence digital trace research on China, our empirical analysis of existing publications found only two studies reflecting on the potential influence of the state on the data collection and analysis process. In fact, state actors could affect the validity and reliability of digital trace research in different ways. For instance, most studies in political communication that rely on social media have used compiled keyword lists or user lists for data collection. However, state actors frequently filter posts with political keywords and block posts by political activists, strongly affecting the validity of data collection in these studies, as removed content is invisible to researchers. At the same time, state actors mobilize different state or social actors, including local government actors and influencers, to flood the social media environment in a more subtle way. Without accounting for this potential state influence, conclusions such as the increasing pro-government public discourse or the rising attention to political topics may lose their internal validity.

While the pervasive presence of the state on social media is facilitated by local agents at different levels (Lu & Pan, 2022), the opening or closure of local accounts could affect the reliability of longitudinal analyses of state behaviors if only social media data are used. Compared to studies using computational analysis, social media studies that apply digital ethnography are less vulnerable to state influence, as the researchers often stay on the digital site doing participatory observation and are thus more likely to collect and capture real-time content and interaction before state actors take action. However, in many computational analyses, what the researchers want to collect mostly relies on retrospective data, meaning that at the time when the researchers want to retrieve historical data about an issue, some historical content might have already been removed by state actors or filtered through platform governance. Thus, researchers will miss many real-time interactions between the state and digital civil society.

In addition to platform and state factors, the changing digital civil society and international environment can affect the validity of studies that rely solely on social media data. Specifically, we found that some social media studies using content analysis employed pre-existing keywords or coding schemes to classify content, which may result in missing new coded languages or variants of vocabularies that are no longer used. Moreover, the shifting relationships between China and the rest of the world present both opportunities and challenges to the understanding of social media activities. For instance, ordinary users can bring information flows from the world to China (Lu et al., 2022), raising questions about the extent to which the collected social media data represent domestic narratives or China-originated public opinions.

Given the aforementioned validity and reliability concerns, two potential pathways could be applied in future research that uses social media data to study China. First, we suggest a more reflexive use of social media data by examining the data generation process and the factors that may affect the validity of using digital trace data to explain communication phenomena in China. Our theoretical framework lays out how domestic and international sociotechnical systems might influence the validity

and reliability of the data generation process. When researchers collect social media data, they can actively investigate how the data are obtained. For example, in the use of a third-party tool or platform, researchers should evaluate how the tool or platform collects data and the representativeness of the data collected, the power dynamics related to the tool or platform and the sociotechnical system in which it is located, and the potential limitations of third-party data. When reporting methods, researchers can also provide more concrete rationales for how each step (e.g. keyword development and sampling method) is made and why it is used to analyze social media data.

Second, we suggest using "multiple methods and sources of data in the execution of a study" (Mathison, 1988, p.13)—the triangulation process. One salient development that we found from our empirics is a rising trend in data triangulation, both for quantitative and qualitative studies. In our dataset, the combination of interviews and surveys was the main path for data triangulation. Data triangulation can take the form of unifying different sources of data, such as triangulating metadata from other sources to provide more information for the digital trace data used (Taneja & Wu, 2014). We argue that another way to overcome overreliance on social media data is to triangulate the latter with non-digital trace data to understand media activities. For instance, when studying the digital trace of social media influencers, researchers can go into the field to observe how influencers created the data and whether any potential biases or challenges were encountered in this process. In this way, researchers will not only focus on reading the digital trace or numbers but also more reflexively study the factors that influence how people conduct media practices, which can improve the validity of their research.

Potential concerns behind online activism studies

Our empirical evidence shows that the topics of digital trace research on China reflect the changing political landscape of the country. Among the wide range of domestic and international political issues portrayed over time, online activism has accounted for a substantial proportion of digital trace data studies in political communication on China-related issues. While these studies have expanded our knowledge of how the state and civil society interact, we highlight some potential validity and reliability concerns.

Among the activism studies captured in our dataset, the majority used content analysis of activists' discursive strategies and public discourse. The discourse dataset that the authors collected to study certain mobilizations may not represent the true landscape of related online public opinion, as it can be moderated and filtered by state actors. As researchers have found, state actors have taken an active moderating role on activism topics, such as #MeToo discussions (Zeng, 2020), collective action expressions (King et al., 2013), and lockdown diaries during the recent pandemic (Yang, 2022). The removed contents are invisible to researchers, unless they are collected with real-time data collection infrastructure. Thus, when researchers conduct content analyses to explore online mobilization strategies or their effects on activism

engagement, state influence can always interfere, but it is extremely difficult to measure.

In addition, separating platform governance from its commercial logic may lead to the inaccurate measurement of state influence. For example, platforms may prioritize pro-government activism messages by manipulating their search engines or recommendation algorithms in exchange for state resources. Thus, the potential moderation and mediation effects of platforms could substantially affect content and engagement in online activism. Without investigating the sociotechnical system of platform governance and state intervention, researchers who use digital trace data to study online activism may miss crucial confounding factors that influence their conclusions about online mobilization strategies. Moreover, as highlighted in the theory section, digital civil societies are evolving. Researchers have shown that netizens use and update coded languages to build fan communities or to counter state influence (Sun, 2023; Yin & Xie, 2021). However, most activism studies have tended to use prior keyword lists to collect data, raising substantial validity issues, as these lists may not capture the changing languages that netizens use in activism discourse.

Furthermore, when examining the theoretical frameworks used across digital trace studies, we found that most activism-related studies tend to adopt a repression–resistance framework. While the Chinese government is adopting various measures of digital repression like other authoritarian governments (Earl et al., 2022), civil society in China is constantly evolving and affecting the state–public relationship. For example, activism leaders may be compliant with the state, such as fan girls in China mobilized by the state to protest against pro-Hong Kong independence groups (Zhuang et al., 2023). The increasing competition on digital media also motivates state actors to innovate their strategies in coping with online and offline activism. For example, the presence of the 50 Cent Party—the government-sponsored online commenters—may result in the misclassification of activism languages and dilute the activism-related online discourse collected by researchers.

Given the validity limitations, we call for a multifaceted perspective in using digital trace data to understand activism in China. This perspective should include how the state co-opts, how a platform regulates, how the public reacts, and how all of these interact in shaping the political processes of forming activism. While political communication studies dominate and tend to conceptualize activism from the repression–resistance lens, few studies have investigated the cultural perspectives behind online collectivism. For instance, L. Tang and Yang (2011) perceived the role of the internet and its *symbolic power*, as it generates follow-up discourse once a powerful symbol has appeared. This work highlights how symbols, which are important elements in many collective actions, are made powerful on the internet. Gong (2020) examined activism in the sports industry by studying how football fans build intimacy and community through emotional expression and *virtual collectivity*. These works, although rare in our dataset, illustrate that future studies can adopt a diverse perspective to examine activism not only from a political perspective but also through cultural and social lenses.

Methodologically, we suggest data triangulation for studies on online activism. For example, studies on activism discourse on social media can triangulate and compare both pre-censorship (data collected in nearly real-time fashion) and post-censorship

datasets to identify removed content (Lu et al., 2021), which implies the state's intention to manage protest content. Researchers can triangulate different types of digital trace data, such as screen tracking data (Ohme et al., 2023) and social media data, or triangulate digital trace data with non-digital trace data, such as surveys and diaries. Through data triangulation, researchers can explore in greater depth the intentions and mechanisms of how online activism is organized.

Conclusion

Digital trace research shows promise as an approach for studying China. This study serves as a starting point for discussing the state and challenges of this stream of research. We propose an evaluation framework to assess the validity and reliability of digital trace research on China through a sociotechnical view that includes the state, the platform, civil society, and international factors. By conducting an empirical examination of digital trace communication scholarship that studies China and by evaluating two major practices in these works and the potential validity and reliability concerns, we hope to provide future researchers with a critical thought map that includes both temporal layers and evaluation components for designing and describing their digital trace research to understand China.

Our research joins a growing effort to reflect on the use of digital trace data. For example, the latest work by Ohme et al. (2023) compared the advantages and limitations of three methods of digital trace data collection—APIs, data donation, and screen tracking—to understand social media effects. Many other studies have reflected on the validity and reliability of using a single type of digital trace data or method, or digital trace data in general (Choi, 2020; Howison et al., 2011). In comparison, to our knowledge, no study has provided empirical evidence and reflected on digital trace research with a regional focus. Our study takes this first step and elaborates on how domestic sociotechnical factors and international factors shape scientific work, which should not be ignored in method reflection.

We would like to acknowledge several limitations regarding our framework for examining and evaluating digital trace research. First, the validity and reliability criteria of our sociotechnical framework might be inadequate for assessing qualitative research that utilizes digital trace data. Other valuable criteria should be considered. For example, Tracy (2010) included triangulation and thick description as credibility criteria in their *eight big-tent criteria for excellent qualitative research*. Reflexivity is emphasized by ethnographers and critical scholars as an important criterion when evaluating the data collection process (Cohen & Crabtree, 2008; Yadav, 2022). Moreover, the scope of validity and reliability in evaluating qualitative research can be broader than what has been described in our framework. Mays and Pope (2000), for example, included both triangulation and reflexivity in their definition of qualitative research validity. Finally, our study did not include other factors that could bias digital trace data collection and analysis. For example, we did not consider the potential bias introduced by researchers' personal beliefs, preconceptions, training backgrounds, and vested interests, as well as the influence of research funding agencies and sponsorship.

Many components in our evaluation framework can be applied to the knowledge claims and epistemic approaches of digital trace research undertaken to understand other societies. For instance, state intervention has been shown to influence digital discourses in other contexts, such as in Russia (Golovchenko, 2022), North Africa (Weber, 2011), and the Middle East (Rand, 2013). Thus, researchers who use digital trace data from these regions need to estimate how state intervention can influence the validity of conducting digital trace work in their contexts. As the nature of information communication varies and evolves, a critical examination of digital trace research in diverse contexts will provide valuable comparative insights for understanding digital trace work through sociotechnical systems across borders.

Acknowledgments

We would like to thank Steve Meyer, the data strategist from the University of Wisconsin–Madison Libraries, for helping us retrieve data from the Web of Science database. We are also grateful to Wenhong Chen, Zhongdang Pan, Stephen D. Reese, the anonymized reviewers, and the participants from the National Communication Association 107th Annual Convention for providing feedback at different stages of this project.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by University of Wisconsin Madison, the Wisconsin Alumni Research Foundation.

Notes on contributors

Kaiping Chen (PhD, Stanford University) is an assistant professor in computational communication at the Department of Life Sciences Communication, University of Wisconsin–Madison. Her research interests are public deliberation, science communication, and computational social science, and she has published in journals such as the *Journal of Communication*, *Journal of Computer-Mediated Communication*, *New Media & Society*, *The American Political Science Review*, *Public Opinion Quarterly*, and *PNAS*.

Yingdan Lu (PhD, Stanford University) is an assistant professor in the Department of Communication Studies at Northwestern University. Her research focuses on digital technology, political communication, and information manipulation in authoritarian and democratic contexts using computational and qualitative methods. Her work has appeared in peer-reviewed journals, such as *Political Communication*, *New Media & Society*, *Human–Computer Interaction*, and *Computational Communication Research*.

Yiming Wang is a PhD candidate in mass communications at the School of Journalism and Mass Communication, University of Wisconsin–Madison. Her research focuses on the interplay between information ecology, identity politics, and other political attitudes and behaviors.

ORCID

Kaiping Chen http://orcid.org/0000-0002-5864-5333
Yingdan Lu http://orcid.org/0000-0002-9955-6070
Yiming Wang http://orcid.org/0000-0001-6088-9233

References

Bourdieu, P. (1990). The scholastic point of view. *Cultural Anthropology*, *5*(4), 380–391. https://doi.org/10.1525/can.1990.5.4.02a00030

Chen, A., Chen, K., Zhang, J., Meng, J., & Shen, C. (2023). When national identity meets conspiracies: The contagion of national identity language in public engagement and discourse about COVID-19 conspiracy theories. *Journal of Computer-Mediated Communication*, *28*(1), zmac034. https://doi.org/10.1093/jcmc/zmac034

Chen, W. (2022). Zoom in and zoom out the glocalized network: When transnationalism meets geopolitics and technopolitics. *Information, Communication & Society*, *25*(16), 2381–2396. https://doi.org/10.1080/1369118X.2022.2118545

Chen, W., & Reese, S. D. (Eds.). (2015). *Networked China: Global dynamics of digital media and civic engagement: New agendas in communication*. Routledge.

Chen, Y., Lu, A. J., & Wu, A. X. (2023). 'China' as a 'black box?' Rethinking methods through a sociotechnical perspective. *Information, Communication & Society*, *26*(2), 253–269. https://doi.org/10.1080/1369118X.2022.2159488

Choi, S. (2020). When digital trace data meet traditional communication theory: Theoretical/methodological directions. *Social Science Computer Review*, *38*(1), 91–107. https://doi.org/10.1177/0894439318788618

Cohen, D. J., & Crabtree, B. F. (2008). Evaluative criteria for qualitative research in health care: Controversies and recommendations. *Annals of Family Medicine*, *6*(4), 331–339. https://doi.org/10.1370/afm.818

Dong, F. (2012). Controlling the internet in China: The real story. *Convergence: The International Journal of Research into New Media Technologies*, *18*(4), 403–425. https://doi.org/10.1177/1354856512439500

Earl, J., Maher, T. V., & Pan, J. (2022). The digital repression of social movements, protest, and activism: A synthetic review. *Science Advances*, *8*(10), eabl8198. https://doi.org/10.1126/sciadv.abl8198

Fietkiewicz, K. J., Lins, E., Baran, K. S., & Stock, W. G. (2016, January). Inter-generational comparison of social media use: Investigating the online behavior of different generational cohorts. In *The 2016 49th Hawaii International Conference on System Sciences (HICSS)* (pp. 3829–3838). IEEE. https://doi.org/10.1109/HICSS.2016.477

Freelon, D. (2014). On the interpretation of digital trace data in communication and social computing research. *Journal of Broadcasting & Electronic Media*, *58*(1), 59–75. https://doi.org/10.1080/08838151.2013.875018

Freiling, I., Krause, N. M., Scheufele, D. A., & Chen, K. (2021). The science of open (communication) science: Toward an evidence-driven understanding of quality criteria in communication research. *Journal of Communication*, *71*(5), 686–714. https://doi.org/10.1093/joc/jqab032

Fung, A. Y. (2002). Identity politics, resistance and new media technologies: A Foucauldian approach to the study of the HKnet. *New Media & Society*, *4*(2), 185–204. https://doi.org/10.1177/14614440222226334

Golovchenko, Y. (2022). Fighting propaganda with censorship: A study of the Ukrainian ban on Russian social media. *The Journal of Politics*, *84*(2), 639–654. https://doi.org/10.1086/716949

Gong, Y. (2020). Virtual collectivity through second screen: Chinese fans' WeChat use in televised spectatorship of European football. *Television & New Media*, *21*(8), 807–824. https://doi.org/10.1177/1527476419857199

Guan, T. (2019). The 'authoritarian determinism' and reductionisms in China-focused political communication studies. *Media, Culture & Society*, *41*(5), 738–750. https://doi.org/10.1177/0163443719831184

Guo, L. (2017). WeChat as a semipublic alternative sphere: Exploring the use of WeChat among Chinese older adults. *International Journal of Communication*, *11*, 21.

Han, R. (2018). *Contesting cyberspace in China: Online expression and authoritarian resilience*. Columbia University Press.

Hong, Y., & Harwit, E. (2020). China's globalizing internet: History, power, and governance. *Chinese Journal of Communication*, *13*(1), 1–7. https://doi.org/10.1080/17544750.2020.1722903

Horsley, J. (2022, August 8). *Behind the façade of China's cyber super-regulator: What we think we know—and what we don't—about the Cyberspace Administration of China*. Stanford University.

Howison, J., Wiggins, A., & Crowston, K. (2011). Validity issues in the use of social network analysis with digital trace data. *Journal of the Association for Information Systems*, *12*(12), 767–797. https://doi.org/10.17705/1jais.00282

Hu, D., Martin, C., Dredze, M., & Broniatowski, D. A. (2020). Chinese social media suggest decreased vaccine acceptance in China: An observational study on Weibo following the 2018 Changchun Changsheng vaccine incident. *Vaccine*, *38*(13), 2764–2770. https://doi.org/10.1016/j.vaccine.2020.02.027

Hu, Q., & Cheong, P. H. (2021). Understanding digital generations: Social media habitus, memetic engagements, and digital social inequalities in China. *International Journal of Communication*, *15*, 22.

King, G., Pan, J., & Roberts, M. E. (2013). How censorship in China allows government criticism but silences collective expression. *American Political Science Review*, *107*(2), 326–343. https://doi.org/10.1017/S0003055413000014

Lei, Y. W. (2018). *The contentious public sphere: Law, media, and authoritarian rule in China*. Princeton University Press.

Link, P., & Qiang, X. (2013). China at the tipping point? From "fart people" to citizens. *Journal of Democracy*, *24*(1), 79–85. https://doi.org/10.1353/jod.2013.0014

Lu, Y. (2023). [*Performative propaganda engagement: How celebrity fandom engages with state propaganda in China*] [Dissertation]. Stanford University.

Lu, Y., & Pan, J. (2021). Capturing clicks: How the Chinese government uses clickbait to compete for visibility. *Political Communication*, *38*(1-2), 23–54. https://doi.org/10.1080/10584609.2020.1765914

Lu, Y., & Pan, J. (2022). The pervasive presence of Chinese government content on Douyin trending videos. *Computational Communication Research*, *4*(1), 68–97. https://doi.org/10.5117/CCR2022.2.002.LU

Lu, Y., Pan, J., & Xu, Y. (2021). Public sentiment on Chinese social media during the emergence of COVID19. *Journal of Quantitative Description: Digital Media*, *1*, 1–47. https://doi.org/10.51685/jqd.2021.013

Lu, Y., Schaefer, J., Park, K., Joo, J., & Pan, J. (2022). How information flows from the world to China. *The International Journal of Press/Politics*, *0*(0), 194016122211174. https://doi.org/10.1177/19401612221117470

Mathison, S. (1988). Why triangulate? *Educational Researcher*, *17*(2), 13–17. https://doi.org/10.3102/0013189X017002013

Maton, K. (2003). Reflexivity, relationism, & research: Pierre Bourdieu and the epistemic conditions of social scientific knowledge. *Space and Culture*, *6*(1), 52–65. https://doi.org/10.1177/1206331202238962

Mays, N., & Pope, C. (2000). Assessing quality in qualitative research. *BMJ (Clinical Research ed.)*, *320*(7226), 50–52. https://doi.org/10.1136/bmj.320.7226.50

Meng, Q., & Li, M. (2002). New economy and ICT development in China. *Information Economics and Policy*, *14*(2), 275–295. https://doi.org/10.1016/S0167-6245(01)00070-1

Montag, C., Becker, B., & Gan, C. (2018). The multipurpose application WeChat: A review on recent research. *Frontiers in Psychology*, *9*, 2247. https://doi.org/10.3389/fpsyg.2018.02247

Nakamura, L. (2009). Don't hate the player, hate the game: The racialization of labor in World of Warcraft. *Critical Studies in Media Communication*, *26*(2), 128–144. https://doi.org/10.1080/15295030902860252

Ohme, J., Araujo, T., Boeschoten, L., Freelon, D., Ram, N., Reeves, B. B., & Robinson, T. N. (2023). Digital trace data collection for social media effects research: APIs, data donation, and (screen) tracking. *Communication Methods and Measures*, 1–18. https://doi.org/10.1080/19312458.2023.2181319

Pan, J., & Chen, K. (2018). Concealing corruption: How Chinese officials distort upward reporting of online grievances. *American Political Science Review*, *112*(3), 602–620.

Qiu, J. L. (2009). *Working-class network society: Communication technology and the information have-less in urban China*. MIT Press.

Qiu, J. L., Yu, P. K., & Oreglia, E. (2022). A new approach to the geopolitics of Chinese internets. *Information, Communication & Society*, *25*(16), 2335–2341. https://doi.org/10.1080/1369118X.2022.2136856

Rand, D. H. (2013). *Roots of the Arab Spring: Contested authority and political change in the Middle East*. University of Pennsylvania Press.

Romano, A. (2022, October 17). The Chinese government's unlikeliest standoff is with … fandom: Xi Jinping versus the stans. *Vox*. https://www.vox.com/culture/23404571/china-vs-fandom-danmei-censorship-qinglang-social-media

Shen, H., & He, Y. (2022). The geopolitics of infrastructuralized platforms: The case of Alibaba. *Information, Communication & Society*, *25*(16), 2363–2380. https://doi.org/10.1080/1369118X.2022.2128599

Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2020). Integrating survey data and digital trace data: Key issues in developing an emerging field. *Social Science Computer Review*, *38*(5), 503–516.

Sun, Z. (2023). Let Dr Ai's story flee: A minor practice against China's internet censorship during the COVID-19 pandemic. *Visual Studies*, *38*(3-4), 681–686. https://doi.org/10.1080/1472586X.2022.2070535

Taneja, H., & Wu, A. X. (2014). Does the Great Firewall really isolate the Chinese? Integrating access blockage with cultural factors to explain Web user behavior. *The Information Society*, *30*(5), 297–309. https://doi.org/10.1080/01972243.2014.944728

Tang, L., & Yang, P. (2011). Symbolic power and the internet: The power of a 'horse'. *Media, Culture & Society*, *33*(5), 675–691.

Tang, M. (2022). The challenge of the cloud: Between transnational capitalism and data sovereignty. *Information, Communication & Society*, *25*(16), 2397–2411. https://doi.org/10.1080/1369118X.2022.2128598

Tracy, S. J. (2010). Qualitative quality: Eight "big-tent" criteria for excellent qualitative research. *Qualitative Inquiry*, *16*(10), 837–851. https://doi.org/10.1177/1077800410383121

Wacquant, L. J. D. (1992). Toward a social praxeology: The structure and logic of Bourdieu's sociology. In P. Bourdieu & L. J. D. Wacquant (Eds.), *An invitation to reflexive sociology* (pp.1–60). Polity Press.

Weber, R. H. (2011). The Arab Spring| Politics through social networks and politics by government blocking: Do we need new rules? *International Journal of Communication*, *5*, 9.

Wright, J. (2014). Regional variation in Chinese internet filtering. *Information, Communication & Society*, *17*(1), 121–141. https://doi.org/10.1080/1369118X.2013.853818

Wu, A. X., & Taneja, H. (2021). Platform enclosure of human behavior and its measurement: Using behavioral trace data against platform episteme. *New Media & Society*, *23*(9), 2650–2667. https://doi.org/10.1177/1461444820933547

Yadav, D. (2022). Criteria for good qualitative research: A comprehensive review. *The Asia-Pacific Education Researcher*, *31*(6), 679–689. https://doi.org/10.1007/s40299-021-00619-0

Yang, G. (2003). The Internet and the rise of a transnational Chinese cultural sphere. *Media, Culture & Society*, *25*(4), 469–490. https://doi.org/10.1177/01634437030254003

Yang, G. (2009). *The power of the internet in China: Citizen activism online*. Columbia University Press.

Yang, G. (2012). A Chinese internet? History, practice, and globalization. *Chinese Journal of Communication*, *5*(1), 49–54. https://doi.org/10.1080/17544750.2011.647744

Yang, G. (2022). *The Wuhan lockdown*. Columbia University Press.

Yin, Y., & Xie, Z. (2021). Playing platformized language games: Social media logic and the mutation of participatory cultures in Chinese online fandom. *New Media & Society*, *0*(0), 146144482110594. https://doi.org/10.1177/14614448211059489

Zeng, J. (2020). # MeToo as connective action: A study of the anti-sexual violence and anti-sexual harassment campaign on Chinese social media in 2018. *Journalism Practice*, *14*(2), 171–190. https://doi.org/10.1080/17512786.2019.1706622

Zhuang, Y., Huang, S., & Chen, C. (2023). Idolizing the nation: Chinese fandom nationalism through the Fangirl Expedition. *Chinese Journal of Communication*, *16*(1), 53–72. https://doi.org/10.1080/17544750.2022.2108861

Zou, S. (2023). Restyling propaganda: Popularized party press and the making of soft propaganda in China. *Information, Communication & Society*, *26*(1), 201–217. https://doi.org/10.1080/1369118X.2021.1942954