MedieKultur | Journal of media and communication research | ISSN 1901-9726

Article – Open section

Using digital traces in user research

A mixed-method literature review of the (dis)connections between media research and computational methods

David Mathieu[1], Elisabetta Petrucci[2] & Jakob Bæk Kristensen[3]

1. Department of Communication and Arts, Roskilde University, mathieu@ruc.dk
2. Department of Communication and Arts, Roskilde University, elpe@ruc.dk
3. Department of Communication and Arts, Roskilde University, jakobbk@ruc.dk

Abstract

This article presents a mixed-method literature review exploring the (dis)connections in user research between media and communication and computational methods. Drawing on a large dataset from Scopus, the study combines network and keyword analyses with qualitative inquiries to assess interdisciplinary engagement regarding literature that relies on digital traces to produce user studies. Findings reveal limited interaction between the media and communication and computer science literatures, each field maintaining distinct research agendas, vocabularies, and epistemologies. While computational work emphasizes methodological innovation and data analysis, media and communication research critically contextualizes digital traces within cultural and theoretical frameworks. The review identifies emerging interdisciplinary journals and calls for greater integration through shared language, collaborative research, and methodological reflection.

Keywords

User research, digital traces, literature review, mixed-method design, computational methods

Introduction

Audiences and media reception are key interests of media studies (Schrøder & Mathieu, 2021). As a qualitative approach, reception analysis was developed for the study of television audiences but has been applied to digital and social media (Mathieu, 2015, 2016). With the computational turn in social sciences and humanities, however, new methods emerged (Berry, 2011) that could make use of users' digital footprints (clicks, links, comments) to conduct something akin to reception analysis with the "assumption that digital traces left on social media can provide insights into the nature of human interaction" (Neumayer et al., 2021, p. 1).

In his presentation of the computational turn, Berry (2011) argues that research methodologies have changed substantially with the rise of digitalisation and computation. These transformations have also impacted user research. Computational methods, making use of large scale digital data, can provide new answers and new perspectives in user studies. At the same time, these methods have allowed user research to focus on new topics, such as disinformation and polarisation on social media (e.g., Kristensen, 2020; Benkler et al., 2018), and to provide new ways to understand media uses via visualisation of networks (Venturini et al., 2018).

Given that digital media account for a large part of the media diet of contemporary users, digital data have become unavoidable in media research. Yet, questions remain whether digital traces provide sufficient indications of how users make sense, use, and relate to digital media. Our study is motivated by this premise, which we have discussed amongst ourselves and with colleagues at seminars, that computational methods in user research might be used independently of the knowledge that has been developed in audience research, creating tensions and reviving old divides between different ways of doing research. Can user-generated content be taken as an indication of the sense-making that users experience with their use of media?

The opening of a new research trend that is interdisciplinary in nature never proceeds without frictions. Media and communication research is full of these frictions, given that it shares borders with a multiplicity of disciplines. The advent of big data research brought a shift in paradigmatic thinking that challenged the core of humanistic and interpretative research (Kitchin, 2014). The move has been so drastic that the then editor-in-chief of *Wired* called it "the end of theory" (Anderson, 2000).

Our interest is in the nature and extent of this interdisciplinary dialogue and its capacity to generate a new and fertile cross-disciplinary field that can bring new light to the study of media users. Given the potential that computational methods have for the study of users of digital and social media, we wish to better understand the nature of the interchange that is taking place between user studies that rely on computational methods within and outside the borders of media research. Are there any interactions and synergies between media scholars and those who apply computational methods to the study of media users?

Our inquiry is shaped by our positionality as media and communication researchers. We are all media scholars in a media and communication department where audience research has had a strong presence. We are all interested in these new developments and committed to bringing the highest standards of research to the study of media users. Therefore, this review of literature is also an attempt to discuss, validate, and critically reflect on the use of digital data for the purpose of user research, while contributing to qualifying the use of computational methods for media and audience research.

This article presents a literature review on the (dis)connections between media and audience research and user studies conducted with computational methods. Specifically, we asked the following research questions:

- To what extent do user research employing computational methods and media research inform one another, given a common object of study?
- Where and how do these two scholarships connect and disconnect in terms of journal citations, keywords, and agendas?
- What conceptual, methodological, theoretical, and epistemological discussions exist on the use of digital traces for user research?

First, we present the discussions that arose while preparing this literature review and introduce readers to past research in the area. Then, we present the mixed-method approach used to develop this review of literature. Starting from a quantitative network analysis, we then present three subsequent analyses: a keyword analysis, an analysis of interdisciplinary literature in computer science, and an analysis of the use of computational methods in media research.

State of the art

In preparing this work, we found that terms are ascribed different meanings and are not consistently used across disciplines, which challenged our comparative literature review. There seems to be a lack of common language between the different traditions of research that rely on digital data, which made it difficult to identify a common pool of relevant literature and look for parallel trends.

The terms "big data" and "social media data" are used mostly in the social sciences and computer science literature, while humanistic media research prefers to talk about "digital footprints" or "digital traces", defined as "any inscription produced by a digital medium in its mediation of collective actions – for instance, a post published on a blog and a hyperlink connecting two websites or the log of an e-commerce transaction" (Venturini et al., 2018, p. 4). The term "footprint" is perhaps an attempt to index the agentic role of humans in the production of these digital traces. However, Hepp et al. (2018) remind us that these footprints are not always deliberately or consciously produced (for instance,

while scrolling, viewing, or searching), making their inclusion in an analysis of agency problematic.

Additionally, Koed Madsen (2023, p. 3) remarks that digital traces are produced under constantly changing conditions: "Each new project needs to reassess how digital traces make sense as inscriptions of the specific empirical phenomenon one is trying to produce knowledge about". As these considerations indicate, the term "digital traces" possesses a more complex, historicised, and contextual definition than terms such as "big data", suggesting that the import of digital data in media research is a task that involves more than the automated processing of large quantities of data.

We also note that "digital methods" is an ambiguous concept. While it has gained popularity in circles related to media research, it is not a term used in computer science, where the term "computational methods" (or simply "methods") is preferred. In analyzing the material for this literature review, we could identify three different uses of the term "digital methods", referring to: 1) investigations that rely on computational methods to study media (see Vicari, 2017); 2) the study of non-media phenomena with the aid of repurposed digital media data, following Roger's (2015) groundwork; and 3) the study of digital media through a variety of different methods, including, for example, digital ethnography such as the walkthrough method (Duguay & Gold-Apel, 2023; Light et al., 2018).

In audience research, there is a number of publications that reflect on the use of digital traces for the purpose of audience and reception research (Schrøder, 2019; Mathieu, 2015, 2016; Livingstone, 2019; Zeller, 2015). These reflections are mostly critical and theoretical in nature, and do not engage in empirical research using computational methods. They are often written from the perspective of media research, with audience at its center and computational methods in the periphery. For example, Livingstone (2019, p. 176) reminds us that the "enhanced visibility" provided by digital data "obscures more than it reveals", while Zeller (2015, p. 271) argues that it is the raison d'être of audience research that is abandoned if one blindly follows the traces of big data.

Extensive literature reviews about the role of social media data have been conducted. Esfahani et al. (2019) looked at the growth of big data research in the study of social media, while Fan (2022) looked at the divides between "data-rich and data-poor" research. Özkula et al. (2023) showed that there is a preponderance of using easily accessible data for activism research. Overall, several literature reviews have been conducted about social media in connection to a variety of topics and fields, mostly tourism (e.g., Leung et al., 2013), health (e.g., Wang et al., 2014), education (e.g., Tess, 2013), and marketing (e.g., King et al., 2014). To our knowledge, there has not been any literature review looking specifically at the connections between media research and computational methods.

Methodology

The literature review employs a mixed-method strategy best described as a combination of computational methods and qualitative inquiry. The advantage of using computational methods to review literature is that a large dataset can be explored and connections between individual works can be mapped in terms of keywords, citations, journals, and other bibliographical information.[1] This allows the possibility to gather a systematic overview of the main trends in the literature, which we complemented with a qualitative analysis of subsets of literature that pays attention to details, argumentation, and context.

Besides the journal *Participation*, which focuses on cultural studies, there are no journals dedicated to audience research which could encapsulate relevant discussions on the use of digital traces for the purpose of audience research. Rather, these discussions are spread amongst a variety of journals dedicated to both media and communication studies. Similarly, to our knowledge, there was no publication dedicated to the application of computational methods to media research. Rather, research that made use of these methods to study phenomena that fall within media audiences, such as media use and reception, are spread among different journals. Thus, in order to find a common ground for our comparative literature review, we decided to first search for literature using the keyword "social media". This search strategy was a compromise given that the research covering our interest behind this literature review – the use of digital traces in user research – did not share a common vocabulary. The term big data would have biased the search towards computer science, while digital traces or footprints would have resulted in few hits in the same literature. Therefore, the more general term "social media" was chosen as it provided a common entry point in this vast literature.[2]

The study draws on data from Scopus, which covers most established peer-reviewed journals within computer science and engineering as well as the social sciences and humanities. Scopus does not carry all journals, but given its size, we assumed that trends identified within its database would be representative of a wider pool of literature. The reliance on computational methods for this literature review involved the implementation of specialized algorithms for cleaning and standardizing references and citations logged by Scopus. For this reason, the benefits of focusing solely on Scopus outweighed the costs of adding additional databases, which would have created a relatively large overlap with the searches we conducted. The review also limits itself to journal articles published in English between 2000 and 2023, as computational methods became widely used after the turn of the century, following the development of the Internet and the introduction of platforms such as Facebook and Twitter.

An issue with using computational methods for the purpose of reviewing literature is that we do not know what criteria Scopus uses to categorize literature by subject, and whether this reflects already established divides present in research (for example, through journal publication or institutional affiliation). We found its classification to be unreliable for the purpose of our study, as some journal articles relying on qualitative methodologies

were classified as belonging to computer sciences, while some literature clearly relying on computational methods was not included in this classification. This confirmed our choice to investigate the literature qualitatively, and not solely rely on computational methods, which are dependent on the affordances and design choices of the supporting systems (Rogers, 2015).

This literature review cannot be considered systematic in the traditional sense of the term. Rather, it relied on a mixed-method approach, involving an explanatory sequential design (Creswell, 2013) often used in studies that involve computational methods (e.g., Munk et al., 2022), in which quantitative exploration precedes qualitative explanation. Given the challenge of analysing an emerging literature that is not precisely anchored in a set of journals, or even fields of study, the approach known as systematic literature review felt out of grasp. Hence, we consider the process of selection and deselection of relevant literature to be a matter of operationalization which follows the logics of either computational methods or qualitative analysis. On the one hand, computational methods take advantage but are also limited by the native digital environment from which they extract data. On the other hand, qualitative sampling often proceeds from theoretical assumptions where the sample is retained, not for its exhaustivity, but for its capacity to represent something of concern for research. These logics do not correspond to what is traditionally dictated by systematic literature review. With that said, we have aimed for providing transparency in our choices and assumptions.

The article consists of four different analyses organised sequentially: the computational analysis of large bibliographical data provides general trends, which are then looked at in detail through qualitative analysis. The first analysis, the network analysis, mapped the relations between research on social media. The second analysis was based on the most frequent keywords in two relevant clusters identified in the network analysis. We subsequently used these keywords to identify subsets of interdisciplinary literature that do the most for bridging and connecting the two traditions of research. We performed two qualitative analyses of the most interdisciplinary literature from two clusters identified in the previous analysis: We looked at a subset of literature in computer science that we found connects heavily with media research, and we looked at a subset of the media research literature that connects heavily with computational methods. Beyond these general considerations, we provide further specific methodological information on each analysis in their respective sections below.

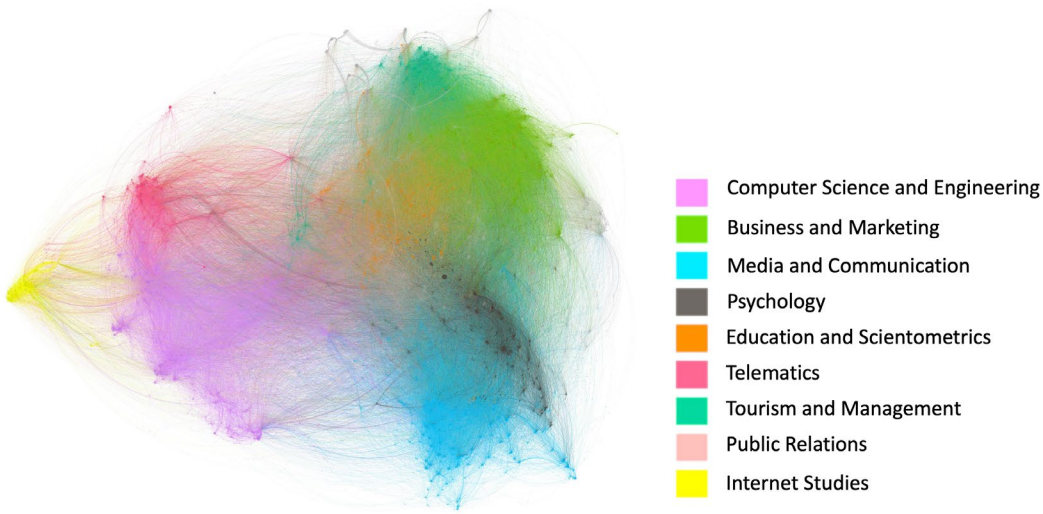
Analysis 1: Network analysis

A network analysis was first conducted on a large set of literature (N = 68,898), yielded by a single search for the term "social media" in Scopus, in order to investigate connections between research from different fields. For the purpose of the network analysis, journals are treated as nodes and citations as edges, such that two journals will be connected if

any of their articles have cited the same reference. If this occurs multiple times, that edge will have a correspondingly stronger weight. Thus, the more references articles in two journals have in common, the more tightly they are connected.

Visualizing the network (see figure 1), we identified two distinct clusters relevant to our analysis – media and communication (MAC) and computer science and engineering (COMPSCI) – that have relatively little overlap. Normally, fields like computer science and engineering would appear more distinct; however, since only studies on social media are included, almost all fields pertaining to natural science cluster together. Noteworthy for the purpose of this overview is the presence of a third major cluster, identified as literature pertaining to business and marketing studies. As can be seen in figure 1, these three clusters form the bulk of the literature on social media research. The network representation shows distinct clusters with lots of interaction within, and less between the clusters. Thus, the network analysis reveals the presence of separate traditions of research with some, but limited crossovers.

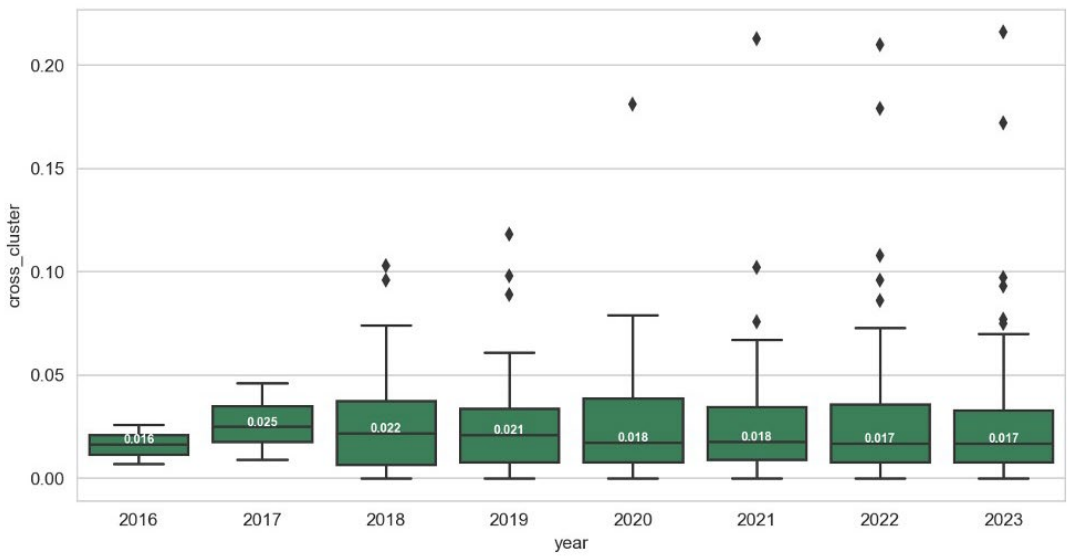
Figure 1
Clusters of relations in the social media literature



The two traditions of MAC and COMPSCI rarely cite the same research; authors belonging to one tradition tend to focus on research in a restricted pool of recognised journals, rarely extending a hand to other traditions of research. Thus, the operational interpretation of interconnection between clusters is the degree to which they orient themselves towards the same types of research, though some of this research may not belong to a clear-cut computer science or media studies journal per se, but could be from another

distinct field such as marketing. We can quantify the interconnection between these two clusters by counting the proportion of edges that lie between them. This comes to 0.019, which means that a little less than 2% of MAC literature is cited by papers published in COMPSCI journals, and vice versa. Furthermore, the findings clearly indicate that research on social media in the two traditions are not moving towards increased interdisciplinarity (see figure 2, showing the degree of interconnectedness between MAC and COMPSCI over time). However, from 2018 onwards, we do start to see a few significant outlier journals starting to establish themselves as more interdisciplinary compared to most journals in their field.

Figure 2
Proportion of edges running between MAC and COMPSCI clusters by year[3]



Zooming in on the two clusters of interest, it is possible to visually detect some of the interdisciplinary nodal points (journals), such as *PLoS One*, *Big Data and Society*, *Computational Communication Research*, and *Journal of Computational Social Science*. While the overall trend does not point towards increasing interdisciplinarity between MAC and COMPSCI, we do see the emergence of some interdisciplinary frontrunners in research on social media.

Analysis 2: Keywords analysis

Given the few interconnections between the two disciplinary clusters on research in social media, we hypothesised that each cluster followed an autonomous research agenda. The network analysis was, therefore, complemented by an analysis of keyword

frequency in the MAC and COMPSCI clusters. The goal here was to obtain an overview of the main topics covered in each cluster in order to better understand the differences between these two sets of literature. We selected the top keywords for each cluster by sorting them on their min-max normalized TF-IDF weighted counts. This entails first calculating the number of occurrences of a given keyword in one cluster and then converting it to its TF-IDF weighted frequency (Řehůřek & Sojka, 2010). The TF-IDF score is then normalized by min-max scaling in order to turn the score to a number between 0 and 1. Using TF-IDF weighting allowed us to consider both keywords that have been purposely attached to the papers by the authors (the keywords often featured on the front page of the paper) and keywords found within the abstract by filtering out common words. Furthermore, we calculated the proportion of a given keyword by considering how many papers within a cluster feature the said keyword. We consider both uni-gram and bi-gram keywords, since many important terms consist of two words (i.e., social media), but tri-grams are relatively rare and were thus excluded to reduce noise. Finally, we checked for the most frequent keywords in each cluster relative to their infrequency in the other clusters. This gave us an indication of which keywords are considerably more important in each cluster compared to the other.

In the tables, we retained keywords that were the least ambiguous in terms of the context to which they likely belong, and accordingly we left out keywords that were too ambiguous regarding their intended meaning in specific contexts (such as online, social, role, new, etc.). Despite this limitation, the findings show interesting trends. We divided the list of keywords obtained into four categories: 1) keywords on aspects of computational methods (in green), 2) keywords on media (in blue), 3) keywords expressing research interests common in MAC (in yellow), and 4) other keywords we deemed relevant for the interest of the current study, which we explain below (in purple).

The keyword analysis shows further gaps between the two clusters of literature. Looking at the most prominent keywords in the COMPSCI literature (see table 1), we note an abundance of keywords related to aspects of computational methods. A telling figure is the discrepancy in the keyword "data", mentioned by 46.3% of publications in COMPSCI compared to 14.5% in MAC. This may indicate that questions about methods take precedence in the COMPSCI literature. We can also see that there is overabundance of research on Twitter, compared to other social media, confirming earlier findings that Twitter has been heavily researched due to the ease of accessibility of its data (Burgess & Brun, 2015; Özkula et al., 2023). The COMPSCI literature presents an overrepresentation of research on fake news; however, there is not a significant difference regarding research on "news" between the two clusters (based on the norm score). Thus, news research has the potential to have a bridging effect on the two clusters. Finally, we can remark that the generic term "user(s)" is overly used in the COMPSCI literature compared to MAC.

Table 1

Prominent keywords in COMPSCI relative to MAC

Keyword	Proportion COMPSCI	Proportion MAC	Norm score COMPSCI	Norm score MAC
data	46,3%	14,5%	0,851295587	0,441030711
model	24,8%	4,4%	0,725504395	0,225746038
users	30,3%	13,0%	0,709990159	0,37167262
detection	12,0%	0,4%	0,669144931	0,030164276
sentiment_analysis	14,4%	0,8%	0,668771308	0,063404325
user	16,1%	3,1%	0,651531651	0,138357177
twitter	23,8%	16,2%	0,639390871	0,454355516
analysis	23,5%	13,3%	0,632032602	0,463137128
machine_learning	12,3%	0,7%	0,618340691	0,054836293
models	12,4%	1,7%	0,606384246	0,110382975
classification	10,3%	0,4%	0,587605748	0,032466148
methods	13,3%	2,6%	0,582366159	0,152454573
accuracy	9,9%	0,7%	0,581381064	0,045218594
network	13,5%	3,5%	0,580492217	0,184942833
tweets	14,5%	4,5%	0,57382023	0,237146004
dataset	9,9%	0,6%	0,572929212	0,035277801
method	12,6%	1,9%	0,561692634	0,120626196
text	12,8%	2,0%	0,56080428	0,119187442
deep_learning	8,0%	0,2%	0,557379606	0,01916463
sentiment	8,9%	0,9%	0,53640408	0,064681802
fake_news	7,8%	2,2%	0,507409956	0,161692364
datasets	7,3%	0,4%	0,50515511	0,023996823
techniques	8,6%	1,3%	0,471402822	0,078928012
processing	7,2%	0,9%	0,460316464	0,06989875
framework	8,5%	3,0%	0,446814566	0,179463961
learning	9,4%	2,6%	0,445715615	0,156483734
news	9,7%	11,2%	0,443724057	0,355980434
social_networks	9,1%	2,4%	0,443667788	0,146740918
platforms	10,8%	11,2%	0,439770681	0,397766851
networks	8,1%	3,7%	0,420359534	0,201033964
natural_language	5,4%	0,3%	0,419946873	0,023427747
social_network	7,4%	2,9%	0,413188761	0,171385269
web	7,9%	3,7%	0,41075011	0,211817782

Legend: green = COMPSCI-related, blue = media-related, yellow = MAC-related, purple = to be specified

Table 2, showing prominent keywords in MAC relative to COMPSCI, presents an abundance of keywords that show common research interests within MAC studies. These interests appear to be equally distributed, perhaps with the exception of "political" and "communication".

There is little vocabulary relating to computational methods in the MAC literature, with only "analysis" and "data" making the top of the list (while being quite general and potentially relating to various methods, not exclusively to computational methods), and

much less frequently than in the COMPSCI literature. In contrast, the keyword "interview" is mentioned more frequently in the MAC literature. We note that Facebook and Instagram seem to be more studied by MAC scholars, while both communities pay similar attention to "platforms", the "internet", and "technologies".

The keyword "users" appears to be frequent also in the MAC literature (13%), but not as frequent as in the COMPSCI literature (30.3%). However, well represented in both clusters is the keyword "use". Yet, we notice that the MAC literature employs a more varied conceptual vocabulary to refer to media users. The keywords "participants", "citizens", and "audiences" are used more frequently, while "practice(s)" is also mentioned.

Table 2

Prominent keywords in MAC relative to COMPSCI

Keyword	Proportion COMPSCI	Proportion MAC	Norm score COMPSCI	Norm score MAC
use	18,8%	22,3%	0,536714369	0,585335767
political	3,2%	12,4%	0,190209154	0,507096744
communication	6,6%	13,3%	0,293323374	0,489214909
practices	1,1%	7,9%	0,072313134	0,479941949
facebook	6,6%	13,4%	0,285004166	0,468549353
analysis	23,5%	13,3%	0,632032602	0,463137128
twitter	23,8%	16,2%	0,639390871	0,454355516
data	46,3%	14,5%	0,851295587	0,441030711
platforms	10,8%	11,2%	0,439770681	0,397766851
users	30,3%	13,0%	0,709990159	0,37167262
identity	1,1%	4,9%	0,115069294	0,370782764
discourse	1,0%	4,7%	0,098586645	0,364184151
internet	7,1%	7,7%	0,314982082	0,35881475
youth	0,5%	3,6%	0,057685332	0,357274339
news	9,7%	11,2%	0,443724057	0,355980434
engagement	2,2%	5,5%	0,147842265	0,340171124
interviews	0,8%	4,4%	0,064257277	0,334992723
participants	2,8%	5,7%	0,143992708	0,331166198
politics	0,9%	4,3%	0,082041392	0,325492173
instagram	2,6%	5,3%	0,217905411	0,315609364
participation	1,6%	4,1%	0,108460864	0,315037639
gender	1,4%	4,3%	0,126966858	0,314875191
context	6,2%	6,1%	0,304141364	0,314613728
citizens	1,7%	4,1%	0,126682275	0,313857474
community	5,4%	6,2%	0,297019623	0,307101864
technology	4,2%	4,8%	0,236132332	0,297151789
practice	1,3%	3,5%	0,094936389	0,28081545
activism	0,1%	2,8%	0,018865025	0,280766274
audiences	0,5%	2,8%	0,054610671	0,253720762

Legend: green = COMPSCI-related, blue = media-related, yellow = MAC-related, purple = to be specified

From computational analysis to qualitative analysis

The two analyses above provided an overview of research interested in social media, which suggested that the literature in computer science and the literature in media and communication studies do not connect much to one another. The keyword analysis substantiated this claim by showing an agenda focused on questions of method in the COMPSCI literature, while the MAC literature revealed a broader palette of interests. These findings, however, are limited by the primary selector we chose for our review of literature, namely the term "social media", as we do not know with certainty whether the literature concerns the use of digital traces in media research – our main interest behind this review – or other aspects of social media. Therefore, it was important to dive into the material in a qualitative way, in order to secure relevant material for our analysis. We also wanted to investigate further the claims made above in a qualitative way, paying attention to details in ways that help us understand the nature of the divide. The focus here was not on comparison, but on understanding. On the one hand, we were curious to investigate the extent to which research employing computational methods for the study of users included or avoided references to media scholars. On the other hand, knowing that research in media studies was progressively integrating computational methods into their arsenal, we wanted to better understand the nature of the discussions that such research had on the use of digital traces. We therefore devised two qualitative inquiries that allowed us to delve into these questions.

A main challenge was to move from a large sample of references to a corpus that could be analysed qualitatively. We identified corpuses suitable for qualitative analysis by attending to the two main clusters (COMPSCI and MAC) identified in the network. We then identified interdisciplinary journals in each cluster based on the proportion of edges that go towards both the COMPSCI and MAC clusters. We then looked for keywords within those journals that reflected interdisciplinary potential. For the first qualitative analysis, we selected works within COMPSCI journals that engaged with the keyword "news", as we found this keyword to be frequent in both clusters, thus asserting a meaningful connection. We assumed that news is an important topic in MAC research, and its treatment in the COMPSCI literature should provide occasions to draw on the insights developed in MAC. We further reduced the size of this corpus by selecting the keyword "Twitter_data" on the assumption that Twitter (called X since July 2023) data is of interest in both clusters. For the second qualitative analysis, we selected references within MAC that engage with computational methods. In order to gather this corpus, we selected the keywords "digital methods" and "media" based on the assumption that works identified under this label revolve around similar historical and epistemological interest as MAC, while integrating developments in computational methods.[4]

Analysis 3: Qualitative analysis of interdisciplinarity

The first qualitative dive into the network aimed to investigate the extent to which user research employing computational methods draws on the scholarship of media and communication research to inform their work. The corpus retained for analysis contains 25 articles using the keywords "news" and "Twitter_data"[5] from the COMPSCI journals that appeared to be most interdisciplinary in the network analysis: *IEEE Access*, *Information Processing and Management*, *Plos One*, *Scientific Reports*, and *Social Network Analysis and Mining*. The analysis followed these steps: First, we thematically categorized the articles based on title and abstract; then we went through the reference list of each article to check whether it contained media and communication sources. We identified media and communication sources based on title, author, and journal of publication.

The COMPSCI literature can be divided into two groups: articles conducting user/sentiment analysis (n = 10) and methodological articles introducing or testing/assessing the performance of data and methods (n = 15). Out of these 25 articles, we found a total of 6 referenced media and communication sources – 3 from each group. For publications that, in the network analysis, appear to bridge with MAC research, we find this figure to be low. Even though the articles had different purposes, they made use of the MAC literature in similar ways. The MAC literature referenced was mostly empirical studies about social media use and news sharing on Twitter and was used to provide brief contextual information in the introduction. Other times, studies were referenced in the state of the art, but mostly without being individually addressed or discussed.

One article, for instance, refers to MAC studies by Harlow and Harp (2012) and Bouilianne (2015) in the introduction, mentioning that "studies have looked at the use of social media to facilitate activism and civic engagement [14-16]" (Boeck et al., 2021, p. 16). Another article references boyd et al. (2010) to describe in the introduction that retweets "contribute to the propagation of ideas and opinions across the network" (Gaumont et al., 2018, p. 9).

All in all, the analysis shows that COMPSCI seems to reference MAC empirical studies to provide context at the start of their own work, rather than to engage with MAC concepts and ideas in the development of their work. Yet, we find that they concern themselves with similar issues as MAC scholars, such as polarization and the spread of fake news, which they approach from a methodological angle, focusing on the assessment of the appropriateness of methods and data sources. While discussions in the MAC literature tend to address epistemological issues – such as whether digital traces can be considered indicative of sense-making – the COMPSCI scholars concern themselves with empirically testing the usefulness of specific data and datasets to measure phenomena or make prediction. As such, the difference between the two kinds of literature reproduces known distinctions, such as between critical and administrative research (Lazarsfeld, 1941) or between humanistic-interpretative and social sciences research.

Given that news research is a well-developed field in media and communication research, covering a broad range of perspectives ranging from political-economy, audience research, content analysis, etc., and given that quite influential work in the broad scholarship of MAC has emerged from news research, we are surprised to see little mention of this research in the COMPSCI literature. This could be explained by several factors. The format of article writing may encourage drawing on news research for establishing the state of the art, with little incentive to integrate this literature in other parts of the article. It could be that these articles are written for a community of scholars who are mostly interested in aspects of methods, and that the literature from MAC is seen to provide little help to the development of computational methods. However, if the results of computational research need to impact MAC news research, it is important that these results are discussed with reference to the rich pool of insights that news research has to offer.

The sample considered in this analysis is small and it would be interesting to validate these findings in different corpuses reflecting other common interests between COMPSCI and MAC. However, it is a highly relevant sample for our questions, and we cannot see any bias in our sampling method that could account for such a low number of MAC citations found. We simply assumed there would be more interaction, given that news is a common interest between MAC and COMPSCI.

Analysis 4: Qualitative analysis of media research using digital methods

In the second part of the qualitative analysis, we built a corpus of literature answering to the keywords "digital methods" and "media" ($N = 426$) from MAC journals which contained at least two occurrences, so as to obtain a size appropriate for qualitative analysis ($N = 67$, with 12 journals represented). This search was conducted to gather literature specifically concerned with the use of computational methods for the purpose of media studies. As this literature appears in media studies journals, the retained references are interesting because they are deemed relevant to MAC research by their authors, reviewers, and journal editors. There are therefore expectations that these works will engage with epistemological positions, theoretical bases, and methodological considerations constitutive of MAC research when discussing the role of computational methods or digital data in their research. The thematic analysis involved a close reading of the abstracts and aimed to understand and analyze how digital/computational methods and digital data are being adapted to fit the interests, practices, and assumptions of MAC studies.

It is noteworthy to mention that *Social Media + Society* is the journal with the most publications on digital methods (17). Thus, we believe this journal to be a frontrunner in facilitating interdisciplinary dialogue between media and computer science research. This journal is followed by *Information Communication and Society* (11), *New Media & Society*

(8), *International Journal of Communication* (6), *Big Data & Society* (5), as well as *Convergence* (4).

We identified four main trends in the MAC literature on digital methods: 1) it problematizes aspects of digital methods, inserting these within larger methodologies and theoretical frameworks, often in combination with qualitative methods, 2) it relies on mixed-methods approaches, often combining computational methods with qualitative approaches, 3) it anchors digital methods within cultural analysis, seeing digital data as spaces of user agency or practices that emphasise its performative role, and 4) it situates digital data within contexts of production, especially affordances and cross-platform realities.

The first trend is an emphasis on *questioning or problematizing aspects of digital methods*.[6] In the MAC literature, the use of digital methods participates in a mode of inquiry that is critical and inquisitive, often with explicit research questions guiding the study. These papers have in common that they criticize the status quo of big data and digital data research established in the early days of its expansion and root their research in the problematization of the universalistic status of data in research. Some articles even make this the main topic of intervention (e.g., Leckner & Severson, 2019; Madsen & Munk, 2019; Munk, 2019; Gerlitz & Rieder, 2018).

Four articles in particular address the challenges that are at the heart of this review of literature. Hepp et al. (2021)[7] argue that the digital has become the new norm in MAC research and that mixed-method designs provide the best opportunities for contemporary media research. van Es et al. (2021) argue for the need of "tool criticism" given that digital tools are opaque, but active mediators of epistemological assumptions that may harbor tensions with humanistic research. Neumayer et al. (2021) bring our attention to all that is around data, but which is less visible and explicit, inviting more reflective and ethical study of social media data. Finally, Fuchs (2017) suggests the adoption of a critical approach to digital methods as an antidote to the administrative and positivistic stances in research brought by big data analytics.

A second trend concerns the *reliance on mixed methods*.[8] In this literature, digital methods are often set in relation to qualitative methods, such as interview (e.g., Hoyng, 2023) or action research (e.g., Kennedy et al., 2015), as digital data are understood to require contextualization. Hepp et al. (2018) argue that three kinds of context are relevant for properly understanding digital traces: the scientific discourse in which these are inserted, the methods used to collect and analyse them, and finally the context of use in which they are produced and acquire meaning.

Although there has been a desire to transcend the long-established divide between qualitative and quantitative methods (see Lindgren, 2022), some works, like Vicari and Kirby (2022, p. 1738), recognize that in practice,

> digital methods designs are actually better suited to progress from quantitative to qualitative methodological steps: after exploiting the potential of quantitative (i.e., automated) techniques for data access, collection, and handling, they can turn to qualitative practices to produce thick data analyses. (see also Munk, 2019, for other kinds of combination)

Others, like Leckner and Severson (2019, p. 79), suggest the idea of "digital methods triangulation" to understand the complexity of online and offline media practices through the intersection of "conventional and digital methods".

The third trend is that this literature is *anchored in cultural analysis*.[9] Vicari and Kirby (2023, p. 1737) speak of an "emerging scholarship that situates the digital methods paradigm within cultural research (e.g., Caliandro & Gandini, 2016)". They argue that early conceptions of platforms emphasized their technical characteristics, to the detriment of their cultural context, and advocate for considering users' lived experiences as part of platform inquiries.

Anchoring digital methods in cultural analysis seems to be tied to heavy reliance on theory and to a complex conception of the user as fan, public, community, or audience. Furthermore, these studies conceive of digital use as a cultural, discursive, or digital practice that relates to other offline practices (Madsen & Munk, 2019). They stress the performativity of digital spaces (Murru & Vicari, 2021) – and by extension the traces collected in these spaces – and bring attention to user agency.

Digital method projects often start with the identification of relevant digital objects – specific keywords or hashtags – that are used to collect the data. In order to increase the emic perspective of such studies, Vicari and Kirby (2022) recommend to pre-research digital objects amongst relevant communities rather than simply following the researcher's intuition or platform logics that may introduce biases of popularity or virality.

The fourth trend of the MAC literature is that it *situates digital data within contexts of media production and reception*.[10] Research investigates the affordances that make communication possible on social media and situates digital data in its "infrastructures". Data is considered to be the expression of a socio-technical environment and is often analyzed as part of the larger ecology of cross-platform uses. For example, Gerlitz and Rieder (2018, p. 528) challenge the universal status of Twitter metrics and argue that digital data need be considered as traces produced through "the situated accomplishment of users, platforms, interfaces and developers".

In that sense, digital traces are more than simply the expression or reflection of user cognitions, interests, or practices; for instance, phenomena such as the "visibility boost" (Matamoros-Fernández et al., 2021), in which platform algorithms guide user engagement towards specific content, cannot be accounted for solely with the use of metrics that supposedly reflect user interest. The literature criticizes single-platform research of social media (see, e.g., Pearce et al., 2020) and argues that digital data is most fruitful if analyzed as part of the larger ecology of cross-media/platform uses.[11] This signals an overall understanding of data in the literature as constructed and contextual.

Finally, we present an interpretation of our results in a table that articulates a series of oppositions between COMPSCI and MAC studies. This table is not a definitive narrative about the divides between these two fields, but we hope that such a table can be used to engage into a discussion about what makes these two scholarships different and how to resolve these differences in concrete analysis.

Table 3
Contrasts between COMPSCI and MAC studies in their consideration of digital data

COMPSCI	MAC
Method	Methodology
Cognitive referent	Cultural referent
Representation	Performance
Flat social world	Variety of practices
Found data	Constructed data
Revealing aggregated behaviors	Hiding underlying meanings

Problematizing aspects of digital methods in the literature brings up discussions about the distinction between method and methodology: While in media studies, there is recognition of the importance of an underlying methodology in the employment of a specific method, in the computer science field, the relevance is given to the technical and computational workings of a specific method. The situatedness of trace data in cultural analysis comes in contrast with the conceptualization of computational approaches, which sees digital data as a direct referent for individual or collective cognitions and emotions. As the MAC literature understands digital traces as part of their contexts of production and reception, underlining the agency of actors and the performative role that metrics such as clicks and likes play in those contexts, it avoids situating data in a representational view of media and communication (see, for example, Matzner, 2016). The COMPSCI literature relies heavily on the generic term "user" (as seen in table 1), a move that contributes to "flatten" the social world (Couldry, 2020), essentially replacing the need for theory with method. By contrast, literature in MAC emphasises the variety of practices and positions involved in the production of digital traces, as users are situated as *activists* (Caldeira, 2023), *fans* (Hagen & Stauff, 2022), or *publics* (Duguay, 2023).

While digital spaces are often seen as spaces of "found data" (Jensen, 2012) amenable for capture, research anchored in traditional humanistic approaches tend to see digital spaces as constructed spaces with "blurred boundaries" (Broer & Schmidt, 2022), requiring cultural and contextual analysis. Computational methods tend to approach big data

as a space that reveals aggregated and behavioral realities that otherwise would remain hidden, while humanistic research requires a "hermeneutic of suspicion" (Gadamer, 1985) to be able to interpret the surface layer of digital spaces and reveal its deeper, hidden meanings (Livingstone, 2019).

Conclusion

This review of literature provides different snapshots on the points of connection and disconnection between user research on social media from the computational method literature and media and communication research. The lack of common language between COMPSCI and MAC turned out to be an unforeseen limitation of our study. Furthermore, our lack of familiarity with the COMPSCI literature potentially created a blind spot in our review. Yet, we wish to underline that the mixed-methods approach adopted in this review allowed us to look at relevant literature from different perspectives, which, put together, provide a rather coherent picture. Overall, there is little connection between COMPSCI user research on social media and media research on the same topic, but there are a few outliers, in terms of both journals and researchers, who are pushing interdisciplinary connection forward.

It should come as no surprise for anyone following these discussions that the connections between the two scholarships are weak. However, we see the value of this literature review to be elsewhere. By addressing questions of dialogue and interdisciplinarity, we are interested in shedding light on the development of an interdisciplinary research agenda and in providing a hub of considerations for those who are interested in developing computational methods for the purpose of user studies.

There seems to be a lack of common epistemological ground between the two different disciplines – an issue that demands more scholarly discussion from both sides of the divide. Additionally, these oppositions seem to be driven by the conventions in place in each field, not least in the world of publishing. The lack of interaction in terms of citations could be addressed if, on the one hand, scholars in MAC acquired the literacy necessary to integrate findings produced by computational methods, and on the other hand, computer scientists made their findings more robust in relation to the scholarship in MAC. This could also be achieved by encouraging multidisciplinary teams in research projects or by promoting a generation of scholars who are comfortable in both fields. Journals – some already identified in this review – can also play a role in bridging the two fields by encouraging the publication of special issues that are thematically driven and that provide an adequate format for research publishing that favors interdisciplinary connection.

The development of a common language can facilitate this process. Often, critical scholars wish to take their distance towards the normativity or status quo of data analysis and its embedded assumptions, and they develop their own analytical vocabulary to do so. The drawback is that this language might only be appreciated by likeminded scholars

and fail to enter other communities of research. Terms like traces, footprints, or digital methods are cases in point.

In pushing for an integrated agenda of research between computational methods and media research, we may be drawing an ideal for research that may require sustained effort to be put into practice, an effort that is not always well-supported by publishing, institutional boundaries, and conference networks. The good news is that this development is already underway, and this review of literature provides examples of this as well as inspiration for developing further this kind of work. We also underline that the use of computational methods in media research needs to be normalized rather than marginalized. Yet, we also recognize that this is done in the context of a media landscape that is in constant development, both technically, socially, and legally, bringing ongoing challenges and insecurities to scholars interested in furthering connections.

Funding statement

The authorship of this article has been partially funded by the Carlsberg Foundation, funding code CF20-0247.

Notes

1 Digital methods have the capacity to revolutionize working with literature reviews and have begun to be used for such purposes. See, e.g., Flensburg and Lomborg (2021), who use digital methods as a way to explore bibliographical information on a larger scale and with more complexity than what was possible to achieve manually in traditional review of literature.

2 A limitation of this choice is that not all research on social media concerns the use of digital traces. However, it remains interesting to see whether research on social media in the computational method literature does draw on media studies in their endeavor, while equally interesting is to see the extent to which computational methods have penetrated research on social media.

3 Yearly development is illustrated as boxplots where the middle line inside the box denotes the median value for the year. In 2017, the median proportion of cross citations among journals in the MAC and COMPSCI clusters was 0.025 (2.5%). The two middle parts of the box are the 2nd and 3rd quartile groups indicating that 50% of all data points lie in that area. The points above the top whisker are significant outliers indicating journals that have a significantly higher proportion of cross citations compared to the median.

4 The search term "digital methods" was used to dive into the MAC literature on social media research based on our knowledge that the term was gaining popularity in media studies to refer to social media studies using computational methods. While the search term "media" may appear redundant to the first selector "social media", it allowed us to identify the subset of literature on digital methods that was most relevant to media studies, and provided us with 426 relevant results.

5 The term "news" was chosen given it was identified as a bridging research interest in analysis 2. However, this search resulted in too many references to manage qualitatively. Therefore, a decision was made to use the keyword "Twitter data" to narrow the search further, on the assumption that the results would be relevant for the purpose of this analysis.

6 Within the corpus of 67 references, these are the works that engage with the first trend: Adler-Nissen et al., 2021; Barnes, 2020; Bogers et al., 2020; Broer & Schmidt, 2022; Caldeira, 2023; d'Andrea & Mintz, 2019; Duguay, 2023; Duguay & Gold-Apel, 2023; van Es et al., 2021; Fuchs, 2017; Geboers & Van De Wiele, 2020; Gerlitz & Rieder, 2018; Gray et al., 2020; Hepp et al., 2021; Karsgaard & MacDonald, 2020; Koed Madsen, 2023; Koenen, 2021; Leckner & Severson, 2019; Luka & Millette, 2018; Madsen & Munk, 2019; McKelvey et al., 2022; Munk, 2019; Neumayer et al., 2021; Özkula et al., 2023; Pearce et al., 2020; Vicari et al., 2020; Vicari & Kirby, 2020, 2023; Koed Madsen, 2023; Koenen, 2021; Leckner & Severson, 2019; Luka & Millette, 2018; Madsen & Munk, 2019; McKelvey et al., 2022; Munk, 2019; Neumayer et al., 2021; Özkula et al., 2023; Pearce et al., 2020; Vicari et al., 2020; Vicari & Kirby, 2023.

7 This reference is the introduction to a German-language special issue entitled *Beyond the computational turn: Methodological development and research software in communication and media studies*, published in the journal *Medien & Kommunikationswissenschaft*. The reference appeared amongst the results of our quantitative analysis because an English translation of the title and abstract were provided in *Scopus*. Given its centrality to the main thematic of this literature review, we decided to report on it in spite of being written in German.

8 Within this trend, the literature found in the corpus covers: Adler-Nissen et al., 2021; Ben-David & Fernández, 2016; Birchall, 2020; Bonini et al., 2016; Brooker et al., 2018; Caldeira, 2023; Cossu, 2018; Hagen & Stauff, 2022; Hoyng, 2023; Kennedy et al., 2015; Kumar, 2018; Leckner & Severson, 2019; Matamoros-Fernández et al., 2021; Munk, 2019; Ryan Bengtsson & Edlom, 2021; Schneider, 2016; Vicari et al., 2020; Vicari & Kirby, 2023.

9 Amongst the works that follow that trend in our corpus are: Adler-Nissen et al., 2021; Barnes, 2020; Ben-David & Fernández, 2016; Brooker et al., 2018; Davies, 2018; Elmer et al., 2021; van Es et al., 2021; Geboers & Van De Wiele, 2020; Gerlitz & Rieder, 2018; Hepp et al., 2018; Hinzo & Clark, 2019; Koenen, 2021; Light et al., 2018; Lindell, 2017; Luka & Millette, 2018; Madsen & Munk, 2019; Marres & Moats, 2015; Matamoros-Fernández, 2017; Middha, 2018; Vicari, 2017; Vicari et al., 2020; Vicari & Kirby, 2023.

10 Within that trend, the literature identified cover: Ben-David & Fernández, 2016; Bogers et al., 2020; Burgess et al., 2019; d'Andrea & Mintz, 2019; Davies, 2018; Elmer et al., 2021; Geboers & Van De Wiele, 2020; Gerlitz & Rieder, 2018; Gray et al., 2020; Hagen & Stauff, 2022; Karsgaard & MacDonald, 2020; Light et al., 2018; Madsen & Munk, 2019; Matamoros-Fernández, 2017; Matamoros-Fernández et al., 2021; McKelvey et al., 2022; Omena et al., 2020; Pearce et al., 2020; Sánchez-Querubín & Rogers, 2018; Schneider, 2016; Vicari, 2017; Vicari & Murru, 2020.

11 See Hagen & Stauff, 2022; McKelvey et al., 2022; Elmer et al., 2021; Bogers et al., 2020; Vicari & Murru, 2020; Karsgaard & MacDonald, 2020; Pearce et al., 2020; McKelvey et al., 2022.

Litterature

Adler-Nissen, R., Eggeling, K. A., & Wangen, P. (2021). Machine anthropology: A view from international relations. *Big Data & Society*, *8*(2). https://doi.org/10.1177/20539517211063690

Anderson, C. (2000, June 23). The end of theory: The data deluge makes the scientific method obsolete. *Wired*. https://www.wired.com/2008/06/pb-theory/

Barnes, N. (2020). Trace publics as a qualitative critical network tool: Exploring the dark matter in the #MeToo movement. *New Media & Society*, *22*(7), 1305-1319. https://doi.org/10.1177/1461444820912532

Bechmann, A., & Nielbo, K. L. (2018). Are we exposed to the same "news" in the news feed? An empirical analysis of filter bubbles as information similarity for Danish Facebook users. *Digital Journalism*, *6*(8), 990-1002. https://doi.org/10.1080/21670811.2018.1510741

Ben-David, A., & Fernández, A. M. (2016). Hate speech and covert discrimination on social media: Monitoring the Facebook pages of extreme-right political parties in Spain. *International Journal of Communication, 10*, 1167-1193. https://ijoc.org/index.php/ijoc/article/view/3697

Benkler, Y., Faris, R., & Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press. https://doi.org/10.1093/oso/9780190923624.001.0001

Berry, D. (2011). The computational turn: Thinking about the digital humanities. *Culture Machine*, *12*, 1-22.

Birchall, C. (2020). Trying not to fall out: The importance of non-political social ties in online political conversation. *Information, Communication & Society, 23*(7), 963-979. https://doi.org/10.1080/1369118X.2018.1539758

Boeck, M. A., Juillard, C. J., Dicker, R. A., Joseph, B.A., & Sakran, J.V. (2021). Turning value into action: Healthcare workers using digital media advocacy to drive change. *PLoS ONE*, *16*(4). https://doi.org/10.1371/journal.pone.0250875

Bogers, L., Niederer, S., Bardelli, F., & De Gaetano, C. (2020). Confronting bias in the online representation of pregnancy. Convergence: *The International Journal of Research into New Media Technologies, 26*(5-6), 1037-1059. https://doi.org/10.1177/1354856520938606

Bonini, T., Caliandro, A., & Massarelli, A. (2016). Understanding the value of networked publics in radio: Employing digital methods and social network analysis to understand the Twitter publics of two Italian national radio stations. *Information, Communication & Society, 19*(1), 40-58. https://doi.org/10.1080/1369118X.2015.1093532

Bonini, T., & Gandini, A. (2021). The field as a black box: Ethnographic research in the age of platforms. *Social Media + Society*, *6*(4). https://doi.org/10.1177/2056305120984477

Boulianne, S. (2015). Social media use and participation: A meta-analysis of current research. *Information, Communication & Society, 18*(5), 524-538. https://doi.org/10.1080/1369118X.2015.1008542

boyd, d., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society, 15*(5), 662-679. https://doi.org/10.1080/1369118X.2012.678878

boyd, d., Golder S., & Lotan, G. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. *43rd Hawaii International Conference on System Sciences*, Honolulu, HI, USA, pp. 1-10. https://www.danah.org/papers/TweetTweetRetweet.pdf

Broer, I., & Schmidt, J.-H. (2022). Dasein: Die Integration kulturanthropologischer Konzepte in die Journalismusforschung am Beispiel einer hybriden Ethnografie beim SMC Germany [Being there: Integrating concepts from cultural anthropology into journalism research using the example of a hybrid ethnography at SMC Germany]. *Medien & Kommunikationswissenschaft, 70*(1-2), 79-96. https://doi.org/10.5771/1615-634X-2022-1-2-79

Brooker, P., Barnett, J., Vines, J., Lawson, S., Feltwell, T., Long, K., & Wood, G. (2018). Researching with Twitter timeline data: A demonstration via "everyday" socio-political talk around welfare provision. *Big Data & Society*, *5*(1). https://doi.org/10.1177/2053951718766624

Bruns, A., Moon, B., Münch, F., & Sadkowsky, T. (2017). The Australian Twittersphere in 2016: Mapping the follower/followee network. *Social Media + Society*, *3*(4). https://doi.org/10.1177/2056305117748162

Burgess, J., & Brun, A. (2015). Easy data, hard data: The politics and pragmatics of Twitter research after the computational turn. In G. Langlois, J. Redden, & G. Elmer (Eds.), *Compromised data: From social media to big data* (pp. 93-111). Bloomsbury Academic.

Burgess, J., Bruns, A., & Osman, K. (2019). Analysing scholarly contributions to public debate with social media issue mapping: A case study of the Australia Day controversy. *Communication Research and Practice, 5*(4), 393-407. https://doi.org/10.1080/22041451.2019.1689688

Caldeira, S. P. (2023). The pluralization of feminist hashtag landscapes: An exploratory mapping of feminist hashtags on Portuguese Instagram. *Social Media + Society*, *9*(2). https://doi.org/10.1177/20563051231171638

Cossu, A. (2018). Beyond social media determinism? How artists reshape the organization of social movements. *Social Media + Society*, *4*(1). https://doi.org/10.1177/2056305117750717

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage.

d'Andrea, C., & Mintz, A. (2019). Studying the live cross-platform circulation of images with computer vision API: An experiment based on a sports media event. *International Journal of Communication, 13*, 1825-1845. https://ijoc.org/index.php/ijoc/article/view/10423

Davies, H. C. (2018). Learning to Google: Understanding classed and gendered practices when young people use the Internet for research. *New Media & Society*, *20*(8), 2764-2780. https://doi.org/10.1177/1461444817732326

Duguay, S. (2023). TikTok's queer potential: Identity, methods, movements. *Social Media + Society*, *9*(1). https://doi.org/10.1177/20563051231157594

Duguay, S., & Gold-Apel, H. (2023). Stumbling blocks and alternative paths: Reconsidering the walkthrough method for analyzing apps. *Social Media + Society*, *9*(1). https://doi.org/10.1177/20563051231158822

Elmer, G., Neville, S. J., Burton, A., & Ward-Kimola, S. (2021). Zoombombing during a global pandemic. *Social Media + Society*, *7*(3). https://doi.org/10.1177/20563051211035356

Esfahani, H. J., Tavasoli, K., & Jabbarzadeh, A. (2019). Big data and social media: A scientometrics analysis. *International Journal of Data and Network Science, 3*, 145-164. https://doi.org/10.5267/j.ijdns.2019.2.007

Fan, Y. (2022). *Research in the era of social media: Communities, methodologies & author trajectories*. [Doctoral dissertation, University of Copenhagen, Denmark].

Flensburg, S., & Lomborg, S. (2021). Datafication research: Mapping the field for a future agenda. *New Media & Society*, *25*(6), 1451-1469. https://doi.org/10.1177/14614448211046616

Fuchs, C. (2017). From digital positivism and administrative big data analytics towards critical digital and social media research! *European Journal of Communication*, *32*(1), 37-49. https://doi.org/10.1177/0267323116682804

Gadamer, H.-G. (1985). The hermeneutics of suspicion. In J. N. Mohanty (Ed.), *Phenomenology and the Human Sciences* (pp. 73-83). Springer. http://link.springer.com/chapter/10.1007/978-94-009-5081-8_6

Gaumont, N., Panahi, M., & Chavalarias, D. (2018). Reconstruction of the socio-semantic dynamics of political activist Twitter networks—Method and application to the 2017 French presidential election. *PLoS ONE*, *13*(9). https://doi.org/10.1371/journal.pone.0201879

Geboers, M. A., & Van De Wiele, C. T. (2020). Machine vision and social media images: Why hashtags matter. *Social Media + Society*, *6*(2). https://doi.org/10.1177/2056305120928485

Gerlitz, C., & Rieder, B. (2018). Digital traces in context: tweets are not created equal: A platform perspective on social media metrics. *International Journal of Communication, 12*, 528-547. https://ijoc.org/index.php/ijoc/article/view/5974

Gray, J., Bounegru, L., & Venturini, T. (2020). 'Fake news' as infrastructural uncanny. *New Media & Society, 22*(2), 317-341. https://doi.org/10.1177/1461444819856912

Hagen, S. H., & Stauff, M. (2022). Enhancing #TdF2017: Cross-media controversies and forensic fandom during live sports events. Convergence: *The International Journal of Research into New Media Technologies*, *28*(1), 192-213. https://doi.org/10.1177/13548565211010481

Harlow, S., & Harp, D. (2012). Collective action on the web: A cross-cultural study of social networking sites and online and offline activism in the United States and Latin America. *Information, Communication & Society, 15*(2), 196-216. https://doi.org/10.1080/1369118X.2011.591411

Hepp, A., Breiter, A., & Friemel, T. N. (2018). Digital traces in context: digital traces in context—An introduction. *International Journal of Communication, 12*, 439-449. https://ijoc.org/index.php/ijoc/article/view/8650

Hepp, A., Loosen, W., & Hasebrink, U. (2021). Jenseits des computational turn: Methodenentwicklung und Forschungssoftware in der Kommunikations- und Medienwissenschaft [Beyond the computational turn: Methodological development and research software in communication and media studies]. *Medien & Kommunikationswissenschaft*, *69*(1), 4-24. https://doi.org/10.5771/1615-634X-2021-1-3-1

Hinzo, A. M., & Clark, L. S. (2019). Digital survivance and Trickster humor: Exploring visual and digital Indigenous epistemologies in the #NoDAPL movement. *Information, Communication & Society, 22*(6), 791-807. https://doi.org/10.1080/1369118X.2019.1573911

Hoyng, R. (2023). From Bitcoin to Farm Bank: An idiotic inquiry into blockchain speculation. *Convergence: The International Journal of Research into New Media Technologies*, *29*(4), 1015-1032. https://doi.org/10.1177/13548565231154104

Jensen, K. B. (2012). Lost, found, and made: Qualitative data in the study of three-step flows of communication. In I. Volkmer (Ed.), *The handbook of global media research* (1st ed.) (pp. 435-450). Wiley.

Karsgaard, C., & MacDonald, M. (2020). Picturing the pipeline: Mapping settler colonialism on Instagram. *New Media & Society, 22*(7), 1206-1226. https://doi.org/10.1177/1461444820912541

Kennedy, H., Moss, G., Birchall, C., & Moshonas, S. (2015). Balancing the potential and problems of digital methods through action research: Methodological reflections. *Information, Communication & Society, 18*(2), 172-186. https://doi.org/10.1080/1369118X.2014.946434

Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society, 1*(1). https://doi.org/10.1177/2053951714528481

King, R. A., Racherla, P., & Bush, V. D. (2014). What we know and don't know about online word-of-mouth: A review and synthesis of the literature. *Journal of Interactive Marketing*, *28*(3), 167-183. https://doi.org/10.1016/j.intmar.2014.02.001

Koed Madsen, A. (2023). Digital methods as 'experimental a priori' – How to navigate vague empirical situations as an operationalist pragmatist. *Convergence: The International Journal of Research into New Media Technologies*, *30*(1), 94-115. https://doi.org/10.1177/13548565221144260

Koenen, E. (2021). Forschungssoftware für die Kommunikations- und Mediengeschichte. Epistemologische Herausforderungen und Perspektiven [Research software for communication and media history: Epistemological challenges and perspectives]. *Medien & Kommunikationswissenschaft*, *69*(1), 117-135. https://doi.org/10.5771/1615-634X-2021-1-136

Kristensen, J. B. (2020). *Public opinion formation on social media in a big data perspective: measuring political homophily and cross-cutting agreement.* University of Canterbury Library. http://dx.doi.org/10.26021/10858

Kumar, P. (2018). Rerouting the narrative: Mapping the online identity politics of the Tamil and Palestinian diaspora. *Social Media + Society, 4*(1). https://doi.org/10.1177/2056305118764429

Lazarsfeld, P. F. (1941). Remarks on administrative and critical communications research. *Studies in Philosophy and Social Science*, *91*, 2-16.

Leckner, S., & Severson, P. (2019). Exploring the meaning problem of big and small data through digital method triangulation. *Nordicom Review, 40*(s1), 79-94. https://doi.org/10.2478/nor-2019-0015

Leung, D., Law, R., Van Hoof, H., & Buhalis, D. (2013). Social media in tourism and hospitality: A literature review. *Journal of Travel & Tourism Marketing*, *30*(1-2), 3-22. https://doi.org/10.1080/10548408.2013.750919

Light, B., Burgess, J., & Duguay, S. (2018). The walkthrough method: An approach to the study of apps. *New Media & Society*, *20*(3), 881-900. https://doi.org/10.1177/1461444816675438

Lindell, J. (2017). Bringing field theory to social media, and vice-versa: Network-crawling an economy of recognition on Facebook. *Social Media + Society*, *3*(4). https://doi.org/10.1177/2056305117735752

Lindgren, S. (2022). *Digital media & society* (2nd ed.). Sage.

Livingstone, S. (2019). Audiences in an age of datafication: Critical questions for media research. *Television & New Media*, *20*(2), 170-183. https://doi.org/10.1177/1527476418811118

Luka, M. E., & Millette, M. (2018). (Re)framing big data: Activating situated knowledges and a feminist ethics of care in social media research. *Social Media + Society*, *4*(2). https://doi.org/10.1177/2056305118768297

Madsen, A. K., & Munk, A. K. (2019). Experiments with a data-public: Moving digital methods into critical proximity with political practice. *Big Data & Society*, *6*(1). https://doi.org/10.1177/2053951718825357

Marres, N., & Moats, D. (2015). Mapping controversies with social media: The case for symmetry. *Social Media + Society*, *1*(2). https://doi.org/10.1177/2056305115604176

Matamoros-Fernández, A. (2017). Platformed racism: The mediation and circulation of an Australian race-based controversy on Twitter, Facebook and YouTube. *Information, Communication & Society*, *20*(6), 930-946. https://doi.org/10.1080/1369118X.2017.1293130

Matamoros-Fernández, A., Gray, J. E., Bartolo, L., Burgess, J., & Suzor, N. (2021). What's "up next"? Investigating algorithmic recommendations on YouTube across issues and over time. *Media and Communication*, *9*(4), 234-249. https://doi.org/10.17645/mac.v9i4.4184

Mathieu, D. (2015). The continued relevance of reception analysis in the age of social media. *Trípodos*, (36), 13-34. https://doi.org/10.51698/tripodos.2015.36.13-34

Mathieu, D. (2016). Users' encounter with normative discourses on Facebook: A three-pronged analysis of user agency as power structure, nexus, and reception. *Social Media + Society*, *2*(4). https://doi.org/10.1177/2056305116683206

Matzner, T. (2016). Beyond data as representation: The performativity of big data in surveillance. *Surveillance & Society*, *14*(2), 197-210. https://doi.org/10.24908/ss.v14i2.5831

McKelvey, F., DeJong, S., Kowalchuck, S., & Donovan, E. (2022). Is the alt-right popular in Canada? Image sharing, popular culture, and social media. *Canadian Journal of Communication*, *47*(4), 702-729. https://doi.org/10.3138/cjc.2022-0021

Middha, B. (2018). Everyday digital engagements: Using food selfies on Facebook to explore eating practices. *Communication Research and Practice*, *4*(3), 291-306. https://doi.org/10.1080/22041451.2018.1476796

Munk, A. K., Olesen, A. G., & Jacomy, M. (2022). The thick machine: Anthropological AI between explanation and explication. *Big Data & Society*, *9*(1). https://doi.org/10.1177/20539517211069891

Munk, A. K. (2019). Four styles of quali-quantitative analysis: Making sense of the new Nordic food movement on the web. *Nordicom Review*, *40*(s1), 159-176. https://doi.org/10.2478/nor-2019-0020

Murru, M. F., & Vicari, S. (2021). Memetising the pandemic: Memes, covid-19 mundanity and political cultures. *Information, Communication & Society*, *24*(16), 2422-2441. https://doi.org/10.1080/1369118X.2021.1974518

Neumayer, C., Rossi, L., & Struthers, D. M. (2021). Invisible data: A framework for understanding visibility processes in social media data. *Social Media + Society*, *7*(1). https://doi.org/10.1177/2056305120984472

Omena, J. J., Rabello, E. T., & Mintz, A. G. (2020). Digital methods for hashtag engagement research. *Social Media + Society*, *6*(3). https://doi.org/10.1177/2056305120940697

Özkula, S. M., Reilly, P. J., & Hayes, J. (2023). Easy data, same old platforms? A systematic review of digital activism methodologies. *Information, Communication & Society, 26*(7), 1470-1489. https://doi.org/10.1080/1369118X.2021.2013918

Pearce, W., Özkula, S. M., Greene, A. K., Teeling, L., Bansard, J. S., Omena, J. J., & Rabello, E. T. (2020). Visual cross-platform analysis: Digital methods to research social media images. *Information, Communication & Society, 23*(2), 161-180. https://doi.org/10.1080/1369118X.2018.1486871

Peeters, S., & Hagen, S. (2022). The 4CAT capture and analysis toolkit: A modular tool for transparent and traceable social media research. *Computational Communication Research*, *4*(2), 571-589. https://dx.doi.org/10.2139/ssrn.3914892

Pentland, A. (2014). *Social physics: How good ideas spread-the lessons from a new science*. Penguin.

Řehůřek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. *Proceedings of the LREC 2010 Workshop on New Challenges for NLP*. https://doi.org/10.13140/2.1.2393.1847

Rogers, R. (2015). *Digital methods*. The MIT Press.

Ryan Bengtsson, L., & Edlom, J. (2021). Mapping transmedia marketing in the music industry: A methodology. *Media and Communication*, *9*(3), 164-174. https://doi.org/10.17645/mac.v9i3.4064

Sánchez-Querubín, N., & Rogers, R. (2018). Connected routes: Migration studies with digital devices and platforms. *Social Media + Society*, *4*(1). https://doi.org/10.1177/2056305118764427

Schneider, F. (2016). China's 'info-web': How Beijing governs online political communication about Japan. *New Media & Society, 18*(11), 2664-2684. https://doi.org/10.1177/1461444815600379

Schrøder, K. C. (2019). Audience reception research in a post-broadcasting digital age. *Television & New Media*, *20*(2), 155-169. https://doi.org/10.1177/1527476418811114

Schrøder, K. C., & Mathieu, D. (2021). Receptionsteori. In M. Fulg Eskjær, & M. Mortensen (Eds.), *Klassisk og moderne medieteori* (pp. 299-325). Hans Reitzels Forlag.

Tess, P. A. (2013). The role of social media in higher education classes (real and virtual) – A literature review. *Computers in Human Behavior*, *29*(5), A60-A68. https://doi.org/10.1016/j.chb.2012.12.032

Tufekci, Z. (2014). Big questions for social media big data: Representativeness, validity and other methodological pitfalls. *Proceedings of the international AAAI conference on web and social media*, *8*(1), 505-514. https://doi.org/10.1609/icwsm.v8i1.14517

van Dijck, J. (2014). Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society, 12*(2), 197-208. https://doi.org/10.24908/ss.v12i2.4776

van Es, K., Schäfer, M. T., & Wieringa, M. (2021). Tool criticism and the computational turn: A "methodological moment" in media and communication studies. *Medien & Kommunikationswissenschaft*, *69*(1), 46-64. https://doi.org/10.5771/1615-634X-2021-1-46

Venturini, T., Jacomy, M., & Jensen, P. (2019). What do we see when we look at networks. *arXiv:1905.02202*. https://doi.org/10.48550/arXiv.1905.02202

Venturini, T., Bounegru, L., Gray, J., & Rogers, R. (2018). A reality check(list) for digital methods. *New Media & Society, 20*(11), 4195-4217. https://doi.org/10.1177/1461444818769236

Vicari, S. (2017). Twitter and non-elites: Interpreting power dynamics in the life story of the (#)BRCA Twitter stream. *Social Media + Society, 3*(3). https://doi.org/10.1177/2056305117733224

Vicari, S., Iannelli, L., & Zurovac, E. (2020). Political hashtag publics and counter-visuality: A case study of #fertilityday in Italy. *Information, Communication & Society, 23*(9), 1235-1254. https://doi.org/10.1080/1369118X.2018.1555271

Vicari, S., & Kirby, D. (2023). Digital platforms as socio-cultural artifacts: Developing digital methods for cultural research. *Information, Communication & Society, 26*(9), 1733-1755. https://doi.org/10.1080/1369118X.2022.2027498

Vicari, S., & Murru, M. F. (2020). One platform, a thousand worlds: On Twitter irony in the early response to the COVID-19 pandemic in Italy. *Social Media + Society*, *6*(3). https://doi.org/10.1177/2056305120948254

Wang, Y., McKee, M., Torbica, A., & Stuckler, D. (2019). Systematic literature review on the spread of health-related misinformation on social media. *Social Science & Medicine*, *240*, 112552. https://doi.org/10.1016/j.socscimed.2019.112552

Zeller, F. (2015). Big data in audience research: a critical perspective. In F. Zeller, C. Ponte, & B. O'Neill (Eds.), *Revitalising audience research: Innovations in European audience research* (pp. 261-278). Routledge. https://doi.org/10.4324/9781315762821