Easy Data, Usual Suspects, Same Old Places? A Systematic Review of Methodological Approaches in Digital Activism Research

Suay M. Özkula, Paul J. Reilly & Jenny Hayes

Abstract

Burgess and Bruns (2015) have linked the computational turn in social media research to an increase in the number of studies focussing exclusively on 'easy data', such as the 'low hanging fruit' provided by Twitter hashtags. This paper explores whether there is a preponderance of such easy data in digital activism research through a systematic review of relevant journal articles published between 2011 and 2018 (N = 315). Specifically, it examines whether computational digital methods have become increasingly prominent in digital activism research during this period. A key focus of the paper is the extent to which digital activism research focused on easily accessible Twitter data, and whether these were obtained via standard API services. Results indicate that (1) traditional research methodologies were more commonly deployed in these articles than digital methods, but (2) Twitter was the most researched platform in the corpus, and (3) single-platform hashtag studies were an archetype of digital activism research alongside single-platform Facebook studies and holistic approaches (hybrid, multi-method & multi-sited, e.g., ethnography). The paper concludes by advocating for greater diversity in terms of the methodological approaches adopted in digital activism research.

Keywords: easy data; digital activism; digital bias; digital methods; APIs.

Introduction

The computational turn in social media has resulted in an ever-increasing number of studies based on 'easy data'. such as the 'low-hanging fruit' of Twitter hashtags (Burgess & Bruns, 2015; Hargittai, 2020). These are relatively accessible data that can be collected and analysed by researchers using software-based methods. This paper sets out to explore this apparent hegemony of easy data by reviewing methodological approaches in digital activism research in the era of social media.

While there have been several quantitative meta-analyses examining digital media and political engagement (e.g., Boulianne et al., 2020; Matthes et al., 2019), few have explored digital

activism, and none (to our knowledge) have focused on the methodological approaches in this body of research. To address this gap, we conducted a systematic review of thematically relevant, peer-reviewed journal articles. A key aim was to empirically explore the type of data and methodological approaches adopted in digital activism research during this period.

This paper is organised as follows. First, we contextualise this study with reference to existing reviews in the field. We then introduce the easy data hypothesis and explain why digital activism scholarship might conform to the methodological and empirical characteristics of such research in cognate areas. We use Bruns and Burgess' work as a proxy for other studies expressing similar concerns about the representativeness of freely available Twitter data (e.g., Hargittai, 2020). The next section outlines the research design, including the search strategy used to generate the corpus and how articles were coded. Finally, the results of the review are presented and the implications for future digital activism research elucidated.

Reviews of digital activism scholarship

The rapid penetration of information and communication technologies (ICT) has created a range of new opportunities for individuals to engage in activism across the globe, especially within the Western hemisphere.. While there are various definitions of the term, this study adopts the comprehensive definition of digital activism provided by Karatzogianni (2015, p. 1):

> 'Political participation, activities and protests organized in digital networks beyond representational politics. It refers to political conduct aiming for reform or revolution by non-state actors and new socio-political formations such as social movements, protest organizations and individuals and groups from civil society, that is by social actors outside government and corporate influence'.

Digital activism scholarship can be traced back as far as the Zapatista netwar in January 1994. The Ejército Zapatista de Liberación Nacional (EZLN) mobilisation against the Mexican army in January 1994 was differentiated from previous Maoist insurgencies courtesy of the internet-enabled 'swarm networks' which mobilised in their favour (Arquilla & Ronfeldt, 2001; Cleaver, 1997). Much of this research focussed on the use of newsgroups to share information about the Chiapas uprising, thus enabling EZLN supporters to circumvent the censorship of the Mexican government. The following 25 years saw a proliferation in both the case studies and

methods employed to research digital activism.

However, there have been no systematic reviews exploring how this field of research has evolved methodologically to date. Meta-analyses published prior to 2020 have tended to focus more on the political implications of using ICTs, or the link between online and offline political participation rather than activism per se. These have generally found a correlation between the use of online platforms for political expression and offline forms of participation, albeit it remains unclear whether there is a causal link between these forms of engagement and participating in elections (Boulianne, 2015; Boulianne, 2019; Boulianne & Theocharis, 2020). Other reviews and meta-analyses have explored the relationship between internet use and political participation (Boulianne, 2009; 2015; Boulianne et al., 2020; Boulianne & Theocharis, 2020; Chae et al., 2019) and the effects of disagreement or difference in opinion on political engagement (Matthes et al., 2019).

Probably the most comprehensive systematic review of digital activism scholarship to date was provided by Neumayer and Rossi (2016). Their review of papers (N=150) published between 2000 and 2014 identified three components in the academic discourse, namely the development of technologies, the methods and techniques used by researchers, and the social phenomena being investigated. This paper will build on this work by exploring the extent to which the computational turn in social sciences and humanities has shaped digital activism scholarship, including how the foci of this research have evolved during this period.

Although systematic reviews often include methodological reflections in their assessment (e.g. Zanaboni et al., 2018; Zhu et al., 2018), purely methodologically oriented reviews remain rare. One significant obstacle lies in the lack of detail provided in the methodology sections of published articles. There are widely recognised disciplinary differences in terms of academic writing conventions (Nagano, 2015), as well as issues around the lack of transparency in methodological assessments of the analysed studies (e.g. Hammersley, 2020). Thus, although methodological considerations are integral to evaluations of empirical validity in systematic reviews, they are rarely their primary focus.

A review of methodological approaches is particularly valuable when one considers how the prevalence of certain approaches might affect the type of empirical and theoretical knowledge produced. This paper was motivated by such a case. Burgess and Bruns (2015) suggested that the computational turn in social media research was leading to an increase in the number of

studies involving the collection and analysis of data that can be easily accessed by researchers. Corroborative evidence for this thesis was found in a number of systematic reviews which found a preponderance of Twitter studies in internet scholarship (Puschmann & Pentzold, 2020) and discussions of climate change on social media (Pearce et al., 2019).

Although the growth of Twitter research arguably reflects the platform's popularity among users, it could be viewed as a manifestation of the 'digital bias' towards specific online environments (Marres, 2017). This 'partiality' limits data not only to the most popular social media platforms, but also to the demographics who populate these sites. Marres identifies at least three areas of digital bias: **(1)** in the selection of data and content (from specific digital spaces or solely from digital space, and without a contextualised critique of what or whom these data represent), **(2)** in the individual research instruments (e.g. machine bias, algorithmic influences, API restrictions, query limitations), and **(3)** wider methodological biases. Hence, the decisions about which platforms to study not only establish a tradition of easy data research, but also arguably produce unrepresentative data. This paper will empirically explore these biases in the context of digital activism scholarship.

The Easy data hypothesis

Burgess and Bruns (2015) suggest that the computational turn in social media research has led to an increase in studies based on easy data, and in particular the 'low-hanging fruit' provided by Twitter hashtags (this is problematised by other researchers including Blank, 2017; Driscoll & Walker, 2014; Hargittai, 2020; Hino & Fahey, 2019; Lorentzen & Nolin, 2017; Tromble, 2021). This paper explores whether this hegemony of easy data was evident in digital activism research. Burgess and Bruns' argument was therefore operationalised in a way that would allow for such hypothesis testing. Three propositions are central to their hypothesis:

The ***first proposition*** is that methodologies have been substantially changed by the 'computational turn' in social sciences and humanities (Berry, 2011). The production of both digitised data (information converted into digital formats) and natively digital data (data created for and in digital space) has led to *epistemic changes* such the increasing *prevalence of digital, digitised, and hybrid (i.e. mixed) methodologies.* In response to the rapid growth of natively digital data, contemporary research has been drawing on software-based automated internet research methods, often labelled 'digital methods' (see Rogers, 2013; 2017; 2019).

Burgess and Bruns' ***second proposition*** is that the increase in the number of easy data studies can be partly explained by the growth of digital methods research. They argue that these changes have led to the preponderance of more easily accessible data, i.e., 'easy data', as these media environments are easier to use, adapt, or manipulate than conventional media spaces. Most notably, access to platform Application Programming Interfaces (APIs) has allowed researchers to search and filter publicly available data without incurring any significant financial cost (Hino & Fahey, 2019). Twitter is perhaps the best known of these 'easy access' platforms, with there being significantly more Google Scholar searches for 'Twitter API' than those of other platforms (Burgess & Bruns, 2015). Other scholars have provided anecdotal evidence indicating the popularity of Twitter as a research space (Hargittai, 2020; Tufekci, 2014).

Methodological choices for the study of online platforms are shaped to a certain extent by the data they make available to researchers. Data scandals, such as the Cambridge Analytica revelations in March 2018, have resulted in changes to social media APIs which have left Twitter as the 'easiest' platform to study. Bruns (2019) is among those scholars to express concerns over the implications of this 'API-calypse', which has effectively shut down access for third-party app developers and researchers. This has clear implications for those wishing to study digital activism, who may not be able to access relevant data from certain social media platforms.

Their ***final proposition*** is that the aforementioned methodological choices and platform policies have resulted in a glut of studies of hashtags and @reply networks. These are typically studies based on 'modestly sized sets of tweets and certain associated, pre-determined metadata matching a keyword search over a short, recent period of time' (Burgess & Bruns, 2015, n.p.). In contrast, hard data is the 'more comprehensive, longitudinal data sets and/or any of the "missing" metadata' (ibid, n.p.). This has traditionally been the preserve of well-resourced research labs within elite universities, who have the resources to purchase access to these data and the software packages to analyse it at scale. Consequently, they argue that a large proportion of digital research relies on the 'low-hanging fruit of Twitter data', suggesting that easier access leads to a preponderance of reductionist Twitter studies that provide smaller snapshots of the wider field.

While acknowledging the contentiousness of labelling hashtag studies as 'reductionist', this final proposition captures a number of arguments about the validity and representativeness of

digital social research. Such potentially skewed representations have been described as 'digital bias' (Marres, 2017) or sampling bias (Hargittai, 2020). Scholars have questioned the representativeness of Twitter research on numerous grounds, including: (1) incompleteness of Twitter data where API access is subject to several limitations, e.g., samples limited to 1% of the full data or incomplete datasets where deleted tweets are no longer available for analysis; (2) biases resulting through non-random samples the API access provides; (3) missing data when some participants choose to participate publicly via hashtags or @mentions under investigation and others purposefully choose not to do so; (4) the relatively low level of representativeness of Twitter users as a privileged, high-income, educated, highly-skilled, and Western-centric (above all US) demographic; and (5) limitations arising from single-platform Twitter studies when movements take place across different platforms and Twitter may not be the most significant or popular space of engagement (Blank, 2017; Gerlitz & Rieder, 2013; Hino & Fahey, 2019; Lorentzen & Nolin, 2017; Tromble, 2021; Venturini et al., 2018). We concur that Twitter constitutes a significant research space for exploring public expression, and, in the case of digital activism, new arenas of contention and collective action. Nevertheless, these concerns highlight the epistemological and empirical fallacies that arise from a dominant focus on easy data platforms.

In response to these propositions, we formulated two hypotheses and a question for our review of digital activism scholarship:

> **H1:** Following the computational turn, research methodologies draw predominantly on digital, digitised, and hybrid methods in their data collection.
>
> **H2:** Empirical digital activism research predominantly focuses on single-platform Twitter data (H2a) that is relatively easy to access and monitor through APIs such as hashtags and reply networks (H2b).

Bruns and Burgess' propositions include a third hypothesis: Twitter studies of activism are biased and unrepresentative due to their reliance on easy data platforms and access. However, the precise method(s) used by researchers to access platforms may be difficult to establish in a systematic review, given that this information is not always included in methodology sections of journal articles. We therefore converted the third proposition into a research question:

To what extent are Twitter studies of activism prone to digital bias or issues of

representativeness?

This was used to explore how far regimes of access shaped digital activism research, and whether scholars' efforts to mitigate these produced easy data research.

Materials and methods

Systematic reviews

In order to test the hypotheses, a quantitatively oriented systematic review of relevant journal articles was conducted between February 2019 and August 2020. The corpus was created by running queries spanning 21 relevant keywords describing digitally enabled activism (e.g. digital activism, online activism, cf. Appendix 1) on the Scopus database. While the review was largely quantitatively oriented in order to test the above hypotheses, it was not a meta-analysis of similar to those conducted by Shelley Boulianne and colleagues (e.g. Boulianne, 2009; Boulianne, 2015; Boulianne & Theocharis, 2020). The aim here was to evaluate the methodological approaches in digital activism research, rather than examine the empirical results of these articles (see Liberati et al., 2009).

The study implemented principles from high quality systematic reviews in order to create a reliable and valid corpus. These included focused research questions on common methodological approaches taken in digital activism literature, a set of hypotheses for testing, a defined methodology with both inclusion and exclusion criteria, and the application of a search strategy and an extensive review protocol. These criteria were derived from systematic reviews conducted in health sciences and social policy research, where the method has been most commonly used to date (Cochrane Library, 2019; The Campbell Collaboration, 2019; PRISMA, 2015). There were a number of adaptations such as the pragmatic decision to cap the initial number of articles at 2000 (after a ranking by relevance by the Scopus database[1]), and the manual coding of abstracts and the methods sections of each article by the researchers. In addition, some of the common foci and related wording in systematic reviews were omitted due to them not being relevant in the field of activism. For example, interventions, (randomised control) trials, evidence-based research, informed decision-making, effects and evaluations were not applicable here. A PRISMA flow chart explaining the search protocols was created

[1] This ranking is subject to the site's algorithms, which are not fully transparent. To address this issue, we reviewed the titles and abstracts in order to ensure articles met the study's inclusion criteria.

for the review (see Appendix 2).

Article collection & search strategy

Searches were conducted on Scopus (www.scopus.com), one of the largest searchable databases of peer-reviewed research. It was chosen due to its functionality, which enables researchers to filter by abstract, title, year, publication type, and discipline. It also allowed several search terms to be searched simultaneously, a pragmatic benefit for multi-query searches. Articles were identified by search queries using 21 keyword combinations relevant to digital activism. For a judgement of what constituted a paper on digital activism, Karatzogianni's definition was applied (cf. section 2). Queries were initially based on synonyms drawn from existing scholarly literature. These included terms such as cyberactivism (see Illia, 2002; McCaughey & Ayers, 2003), online activism (Earl et al., 2010; Yang, 2009), internet activism (Earl & Kimport, 2011; Kahn & Kellner, 2004; Tatarchevskiy, 2011), web activism (Dartnell, 2006), net activism (Meikle, 2010), e-activism (Carty, 2010), mobile activism (Cullum, 2010), social media activism (Miller, 2017), hashtag activism (Ofori-Parku & Moscato, 2018; Jackson et al., 2020), as well as keyword combinations of synonyms for both "digital" and "activism". The final set of queries was a result of a reiterative process comprising one pilot test run and three consecutive runs using different sets of queries with 44, 13, and 21 queries respectively. The aim was to produce a dataset that was exhaustive but not too exclusive, which resulted in 21 queries, producing an initial list of 4492 articles. This corpus was part of a larger exploration into digital activism and was later filtered to produce a dataset more reflective of this paper's methodological focus.

Filtering

Queries were limited to searches of article titles, abstracts, and keywords between 1995 and 2018. Searches conducted within the 'Social Sciences' category produced an initial sample of 2668 articles, from which the 2000 most relevant entries were selected. The focus here was therefore on articles published in academic journals, which had been subject to peer review prior to acceptance and publication. Additionally, only articles written in English were analysed due to the language orientation of the queries and the researchers' language skills. This resulted in an initial corpus of 1447 abstracts.

Data analysis

Abstracts were coded manually in Excel according to their methodological attributes such as

methods of data collection and analysis, digital data sources (i.e., studied platforms), and case studies. When they provided insufficient detail, the introduction and methodology sections of articles were reviewed for the coding. Articles were initially included if they incorporated empirical research on digital activism and published in academic journals between 01.01.1995 and 01.01.2019 (for the larger scoping study). Following the review of abstracts, they were excluded on the following grounds: (1) digital activism constituted a relatively minor part of the empirical or theoretical work, (2) methodological information was not included (e.g., worked examples of the 'Arab Spring'), (3) the article focused on states, political parties or election campaigns and were therefore judged not to be a case of activism (as per Karatzogianni's definition), (4) they did not include empirical data (e.g., essays, theoretical contributions, opinion pieces, and reviews), and (5) digital activism or cognate disciplines/fields were not the primary focus – for example, articles from education journals were excluded when digital activism or protest was mentioned but was not the primary focus of the study. Based on these exclusion criteria, the final corpus consisted of 315 articles, covering the timespan 2011–2018.

Easy data hypothesis testing

These articles were split into (a) the traditional dataset: 117 articles that used exclusively traditional research methods[2] (e.g., interviews and surveys, ≃37.1%, published in 74 journals), and (b) the digital dataset: 198 articles that included some form of digital data as their source (≃62.9%, published in 91 journals), though not necessarily their only source. Both datasets were initially coded on broader methodological attributes (e.g., methods of data collection and analysis, combinations of digital and traditional methods). The digital dataset was then coded separately to test H2. A distinction was made between data collected manually and via automated software.[3] Further coding categories were added to assess the 'easiness' of the data, including the choice of platform, data access (e.g., type of API access), the use of computational methods, as well as the wider methodological approaches (e.g., hashtag studies, single- or multi-platform studies, triangulation).

[2] We did not distinguish between traditional research methods (e.g., offline/in-person interviews or surveys), digitised methods (e.g., online interviews, surveys), and software-based digital methods (e.g., data mining, web crawling) due to frequently missing information on digitisation in conventional methods. When traditional research methods were conducted using software and video conferencing they were therefore coded as traditional.

[3] These distinctions were also applied to the coding of data analysis methods. We focussed primarily on methods of data collection, however, as information on how data were analysed was often missing from the articles.

The final categories were based not only on the notion of easiness in the work of Burgess and Bruns (2015), but also on wider discussions about the fallacies of predominantly Twitter-focused research (e.g., Blank, 2017; Gerlitz & Rieder, 2013; Hargittai, 2020; Tufekci, 2014). This was refined through a reiterative coding process, in which certain studies threw up new questions around what constituted easy data. Categories added after such discussions included sample size, data collection period, and concentration on hashtag studies.

A qualitative dimension was added based on the researchers' observations, which helped identify themes in the methodological approaches evident in the corpus. We thematically examined the dataset for archetypal approaches and whether these could be judged reductionist in themselves. Themes on archetypes were identified first individually by the researchers and then discussed in group sessions.

Methodological adaptations

There were some limitations that need to be acknowledged. Some of these revolve around ambiguities in defining digital activism (Karatzogianni's definition was used for consistency), different perceptions about what constituted empirical research (e.g. in the case of worked examples), and the blurring of the boundaries between traditional and online research methods. In all these cases, intercoder reliability checks were conducted to address inconsistencies and differences in interpretation.

Other limitations resulted from inconsistencies in methodological sections in these articles. There were differences in **(a)** journal expectations and writing conventions, as well as in **(b)** units of measure in data collection (e.g. tweets or posts). In the case of the former, differing conventions meant that not all papers indicated whether social media data was collected through an automated process or what tools were used for this. In almost all cases, the researchers were able to judge whether automated approaches were used based on the language used to describe the data size and methodological approach. For example, an article describing the results of a qualitative study of eight images was deemed to be non-automated when no information was provided on APIs or automated tools used for data collection. Conversely, a statistical analysis of half a million tweets was judged to be automated if no information was provided on how data were collected and analysed. One issue that proved problematic was determining whether researchers had high-level or standard API access. Therefore, broad methodological approaches (e.g., single- versus multi-platform approaches), sample sizes,

periods in which data were created, and themes from the qualitative coding were interpreted as more reliable measures.

Intercoder reliability checks were conducted in order to address any coding discrepancies. 25% of the corpus was coded by the three researchers, who then discussed results and any disagreements. The level of agreement for each coding category was over 90%, well above the recommended reliability level of 0.67% (Krippendorff, 2004). Where disagreements were identified, categories were further refined to prevent any further inconsistencies in the coding process.

Results

The review partially rejected H1, but supported Bruns and Burgess' second and third propositions (H2). While research methodologies using digital data did indeed favour Twitter data above all other platforms, traditional methods were still significantly represented in digital activism articles published during this period.

#H1: Research methodologies following the computational turn draw significantly, if not predominantly, on digital and hybrid methods.

The composition of the initial corpus suggested that internet research methods were indeed prevalent, with 63% of articles being categorised in the digital data corpus, compared to 37% using traditional methods. However, a closer inspection revealed that a large number of studies combined traditional methods with social media data (e.g., interviews and social media data mining[4]; 56≃17.8%) or other digital data sources such as websites, donation data, or conventional blogs (11 ≃3.5%). Thus, a similar number of studies deployed purely traditional methods (55–58.4% depending on the inclusion of 'unconventional/other data sources') as did digital data along with non-digital sources (≃62%).

#H1 was judged only partially accurate. Although computational methods were prevalent, they were not the most frequently used method in the corpus. Rather, the most commonly deployed method was the semi-structured interview (120 ≃38.1%), followed by social media data mining (112 ≃35.6%; see Table 1 and Figure 1; includes both single and multi-method studies).

[4] Social media data mining is defined here as a software-based method of data collection for social media data (sometimes termed 'scraping'), as opposed to manually collected/copied data.

However, digital methods did not constitute a majority across the entire corpus. While they have become increasingly popular in digital activism research during this period, the results indicate that traditional methods were deployed in a majority of articles in this field.

Rank	Articles using traditional methods	Count (*N* = 117)	Rank	Articles using digital data sources	Count (*N* = 198)	Rank	Combined corpus	Count (*N* = 315)
1	Interview	67	**1**	Social media data mining	112	**1**	Interview	120
2	Survey	47	**2**	Social media data (manual)	68	**2**	Social media data mining	112
3	Observation / Ethnography	11 (each)	**3**	Interview	53	**3**	Social media data (manual)	68
4	Observation / Ethnography	11 (each)	**4**	Ethnography	29	**4**	Survey	52
5	Focus group / Experiment	5 (each)	**5**	Participant observation	23	**5**	Ethnography	40
6	Focus group / Experiment	5 (each)	**6**	Survey	5	**6**	Participant observation	34

Table 1. Comparison of popular methods by corpus

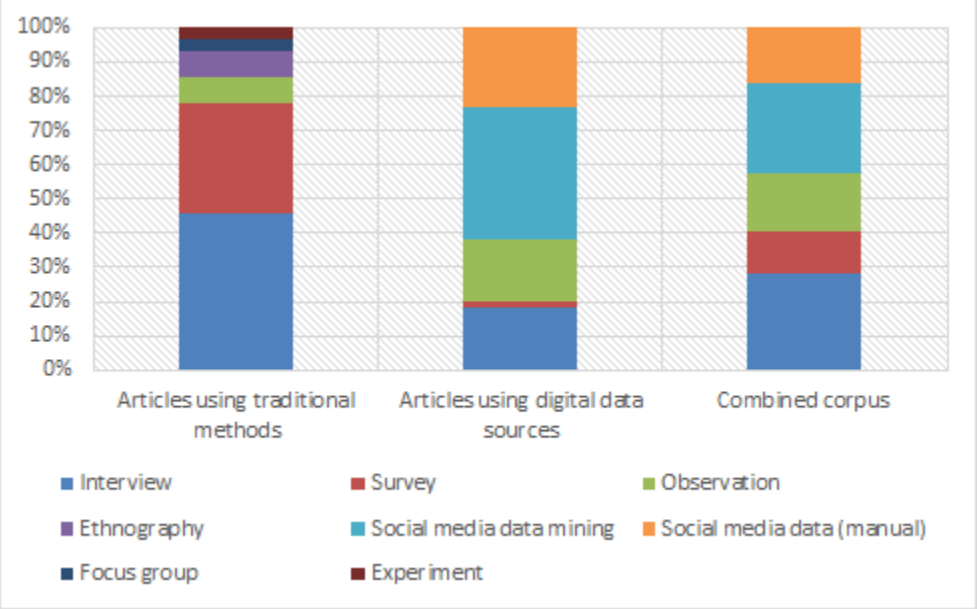

Figure 1. Comparison of popular methods by corpus in a distribution out of 100%

While articles reporting the findings from empirical digital activism research increased in frequency between 2011 and 2018, those using purely traditional methods (not including hybrid/mixed methods) decreased between 2017–2018 (see Figure 2). As such, it remains to be seen whether digital social research and particularly software- based digital methods will

become more prevalent in the field in the future.

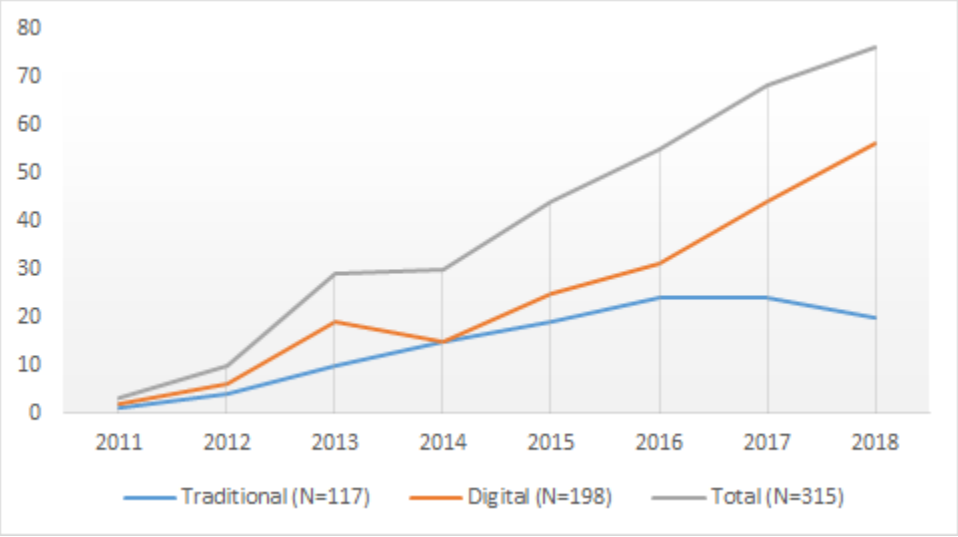

Figure 2. Timeline of articles using traditional methods (only) and studies using digital data sources (with or without traditional methods).

#H2: Digital activism research based on digital data predominantly focuses on Twitter data, and hashtags & reply networks in particular (= 'low-hanging fruit').

Overall, there were an equal number of articles based on Twitter and interview data in the whole corpus (both 120 out of 315 ≔38.1%). However, Twitter was a particularly prominent data source within the subset of studies drawing on digital data (120 out of 198 ≔60.6%; includes both single- and multi-platform studies). The microblogging site was the most researched platform in the corpus, followed by Facebook (23.5%), websites (7.7%), YouTube and blogs (6% each) (see Figure 3, and Tables 2 and 3).

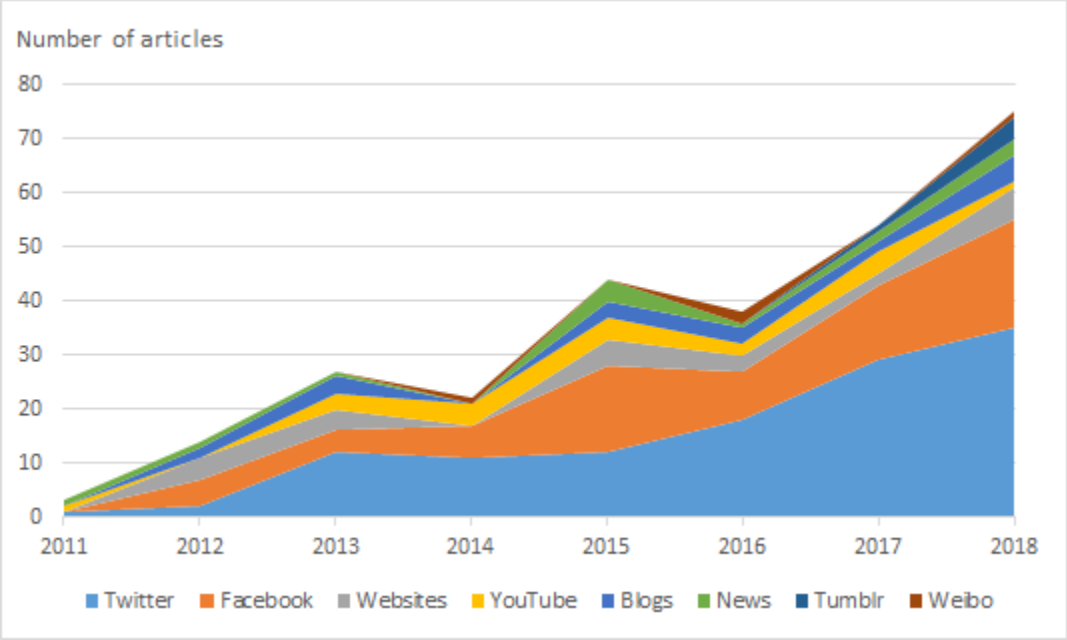

Figure 3. Digital data sources used in articles, 2011-2018 (N=198).

Digital data source /platform	Count in digital dataset ($N = 198$)	% in digital dataset ($N = 198$)	Count in % overall ($N = 315$)
Twitter	120	60.61%	38.10%
Facebook	74	37.37%	23.49%
Websites	24	12.12%	7.70%
YouTube	19	9.60%	6.03%
Blogs	19	9.60%	6.03%
Various, unspecified, or unclear	16	8.08%	5.08%
News reports (digital or non-digital)	13	6.57%	4.13%
Tumblr	5	2.53%	1.59%
Sina Weibo	4	2.02%	1.27%

Table 2. Common Digital Sources

It should be noted that the number of articles drawing on websites and blogs remained consistent throughout the timeframe of the review, while articles drawing on Twitter and Facebook data increased significantly (see Figure 3).

Single-platform studies were considerably more prevalent than those drawing on data collected from multiple platforms. Of 198 articles drawing on digital data, 131 (≔66.2%) drew on data collected from a single platform, 47 (≔23.7%) on data from multiple platforms, and 17 (≔8.6%) combined several digital but not social media platforms (e.g., generic blogs). Twitter was the most commonly researched site in articles reporting the results of both single- and multi-platform research, although more so in single-platform studies. Twitter constituted 58% of all

articles drawing on single-platform research, followed by Facebook with 28.2% (see Table 3). Single-platform Twitter studies were additionally highly likely to use automated data-mining techniques, whether it was the only form of data collection or one of several methods used (combined: 61 ≔80.3%, N = 76).

Digital data source /platform	Count ($N = 131$)	% ($N = 131$)
Twitter	76	58.02%
Facebook	37	28.24%
YouTube	6	4.58%
Sina Weibo	4	3.05%

Table 3. Common Digital Sources in single-platform Studies

Articles relying purely on Facebook and Twitter data were relatively rare (12.8% of those reporting results of multi- platform studies). However, more than half of the articles using a multi-platform approach (63.8%) triangulated data from Facebook, Twitter and other sources. In particular, Twitter constituted a highly popular choice for the vast majority of articles drawing on multi-platform approaches (≔93.6%). One interpretation of this finding was that it showed how researchers were increasingly investigating events and issues synonymous with these platforms, such as the 'Arab Spring', Occupy Wall Street, and Black Lives Matter. Nevertheless, the review found that the use of Facebook data in digital activism research remained relatively static from 2015 onwards, compared to a sharp increase in the use of Twitter data between 2015 and 2018 (see Figure 4). While both Facebook and Twitter data were used extensively for contextualising the findings of studies based on other forms of data (e.g., traditional ethnography), the latter was dominant in both single- and multi-platform studies in the corpus.

A substantial difference was also found in the use of single-platform (the majority) versus multi-platform approaches. Out of 198 articles drawing on digital data (the digital dataset), 131 (N=198; ≔66.16%) utilised single-platform approaches, 47 (≔23.74%) multi-platform approaches, and 17 (≔8.59%) combined several digital but not social media platforms (e.g. news platforms, generic blogs). Twitter was the most commonly researched platform in both articles that reported the results of single- and multi-platform research, although more so in single-platform approaches. Twitter constituted 58.02% of all articles drawing on single-platform research, followed by Facebook with 28.24% (see Table 3). Within these articles

reporting results from single-platform Twitter studies, approaches utilising software-based social media data mining as either their only form of data collection or as part of their enquiry were both prevalent (combined: 61 ≔80.26%, N=76).

Articles relying purely on Facebook and Twitter data were relatively rare (12.8% of those reporting results of multi- platform studies). However, more than half of the articles using a multi-platform approach (63.8%) triangulated data from Facebook, Twitter and other sources. In particular, Twitter constituted a highly popular choice for the vast majority of articles drawing on multi-platform approaches (≔93.6%). One interpretation of this finding was that it showed how researchers were increasingly investigating events and issues synonymous with these platforms, such as the 'Arab Spring', Occupy Wall Street, and Black Lives Matter. Nevertheless, the review found that the use of Facebook data in digital activism research remained relatively static from 2015 onwards, compared to a sharp increase in the use of Twitter data between 2015 and 2018 (see Figure 4). While both Facebook and Twitter data were used extensively for contextualising the findings of studies based on other forms of data (e.g., traditional ethnography), the latter was dominant in both single- and multi-platform studies in the corpus.

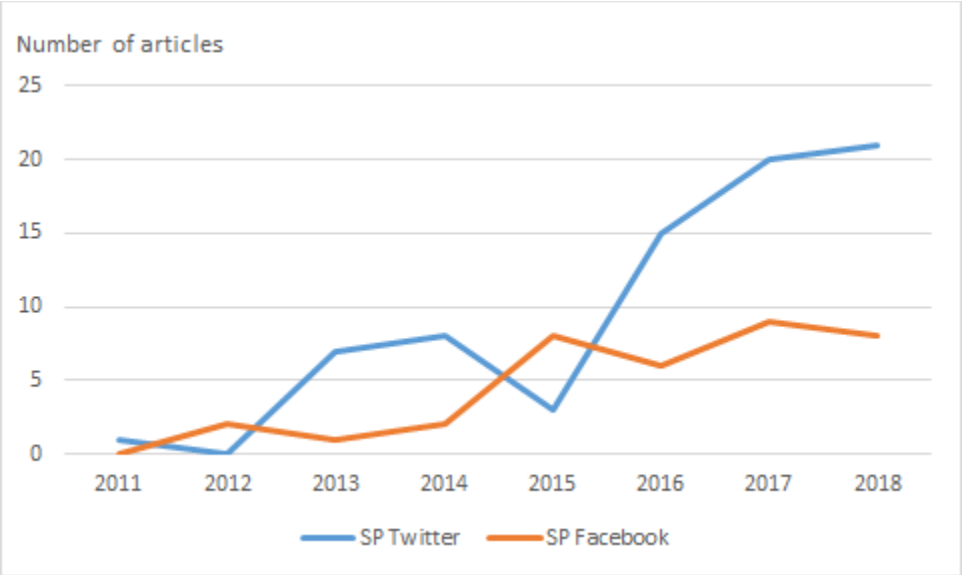

Figure 4. Number of articles featuring single-platform Facebook and Twitter studies, 2011-2018.

Hashtag studies featured prominently but were not the only sampling method used in articles

focussing on Twitter. Of 120 articles, 51 (42.5%) focused entirely on hashtags and 32 (26.7%; combined: 69.2%) included them in their analysis, often amongst mention or reply networks. Given Twitter's overall dominance in the dataset, #H2 was therefore judged to be predominantly true as Twitter data (a) constituted a substantial portion of the overall corpus, (b) constituted the dominant field of enquiry in both single- and multi-platform studies, and (c) was typically collected through software-based approaches with a dominant focus on hashtag publics. However, the time series also showed that these are very recent trends. Although articles spanned the 2011–2018 timespan, single-platform Twitter studies only really became a significant portion of the corpus post-2014.

While empirical digital activism research appeared to rely heavily on easy data, the scant information provided on sample sizes and API access made it difficult to say for certain whether these studies were reductionist or unrepresentative. The time periods of data collection provided starting points for understanding scope in easy data research. These ranged from tweets collected during a 24-hour period to ethnographic studies lasting over a year. In relation to the latter, most of the sample sizes of articles in the corpus were below 5000 units of measurement (= 68 out of 198 articles using digital data ≃34.3%). However, one of the limitations relates to the coding of time frames in which data were collected. While our interest was in the timeframe in which the social media data was created or monitored in real-time, many articles instead listed the time spent on collecting data through online ethnographies or participant observation. This was in part a manifestation of the differences between the methodologies applied in digital activism research. Whereas many traditional methods such as interviews typically create data during the timeline of data collection, digital research often consists of 'live' data, archival, and reappearing data (e.g., the re-sharing or updating of older content). As such, timelines remain difficult to judge in digital social research due to the complex ways in which data is produced and collected; shorter timeframes may not necessarily be a reliable indicator of the 'easiness' of the data used in these articles. This raises critical questions around what indicates reductionism: limitations through API access, short-term research, single-method studies, single-platform approaches, lack of triangulation, or a combination of these elements.

Archetypes in digital activism research

The coding revealed three methodological archetypes of digital activism research (see Table 4). Congruent with previous social media research (Hargittai, 2020), single-platform research, above all on Twitter and Facebook, constituted the most popular methodological approach, accounting for 46% of articles drawing on digital data sources. While these were not necessarily considered reductionist or unrepresentative by the researchers (there were, for example, several single-platform ethnographies), they did revolve around the easy data identified by Burgess and Bruns (2015). This type of Twitter research (Type #1) typically focussed on hashtag data mined from the platform, before being analysed using social network analysis or qualitative coding methods like thematic content analysis. As such, their results were prone to key limitations of Twitter hashtag studies, such as the omission of replies and non-hashtagged tweets, and the skewed demographics of those contributing to these public conversations.

	Description
#1: Twitter	• Single-platform computational studies focused on hashtags and using data-mining for short timeframes • Analysis through social network analysis or qualitative coding
#2: Facebook	• Single-platform studies focused on Facebook pages or profiles and using data-mining or longer-term observation • Analysis through content or discourse analysis
#3: Holistic approaches	• Multi-method and multi-sited approaches combining traditional and digital data in rich ways (often ethnography)

Table 4. Methodological archetypes.

Facebook approaches (Type #2) differed in that mined and manually collected data were equally represented, for example through observational methods such as ethnography or approaches including elements of participant observation. The focus here was also typically on account pages and profiles (e.g., civil society organisations or leading figures in protest movements), rather than issue-based/ hashtag publics, that were observed over a longer period of time and often coded using content or discourse analysis. While it does not necessarily follow that studies drawing on manually collected data are methodologically more rigorous than those using data mining, they were not subject to the same limitations in data collection caused by the use of APIs. These distinct approaches (Type #1 on Twitter and Type #2 on

Facebook) further illustrate how platform affordances not only shape dynamics of digital activism, but also methodological approaches in the field. As such, extant digital activism research is, to a meaningful degree, a reflection of platform cultures, or at the very least protest vernaculars within the norms of specific popular platforms.

Multi-method ethnographies (Type #3) were the third archetype of digital activism research in the corpus. These typically involved a combination of traditional and internet research methods, along with observational methods of some kind (online or offline, and often but not always labelled ethnography). Due to differing methodological conventions in ethnographic research, the precise methods of data collection and analysis were often unclear. For instance, several articles referred to 'social media data' without providing any specific information on the specific platforms under observation. In itself, the use of multi-method approaches suggested a higher degree of triangulation, as well as the coverage of wider movement dynamics beyond a single platform. However, the dearth of information on sample size, timelines, and types of data, often made it difficult to assess the rigour that was applied in these studies. A complicating factor was that social media data was often used to contextualise data gathered via other methods, which perhaps explained why so little space was devoted to explaining how it was collected and analysed. Overall, while single-platform Twitter studies relying on API access constituted a prominent type of research, there was also some diversity in methodological approaches in digital activism research.

Discussion & conclusion

The review provided evidence that Burgess and Bruns' easy data hypothesis holds partially true for digital activism literature. One significant finding was that traditional methods accounted for a majority in the article corpus. While less than half of the studies drew exclusively on traditional data sources, a large array of studies using social media data also drew on traditional sources such as interviews and surveys. Consequently, more studies drew on traditional methods than did on social media data.

Even so, Twitter, closely followed by Facebook, was the most researched platform in those articles drawing on social media data (whether through social media data mining, digital ethnography, or the manual collection of tweets). The microblogging site was increasingly dominant in both single and multi-platform studies, as well as a distinct methodological

archetype in digital activism research that relied heavily (in line with Burgess and Bruns' propositions) on content or social network analysis of hashtag publics. While the review highlighted some diversity in digital activism research, Twitter data was analysed in a substantial proportion of the overall corpus, and single-platform computational Twitter studies were the most common methodological archetype. As such, we found Burgess and Bruns' hypothesis around representativeness to be true to an extent.

Representativeness in digital activism methodologies

The findings of the review provide some evidence to support the notion that there is representation bias in digital activism research (Hargittai, 2020; Marres, 2017). Empirical research in the field is largely based on data gathered from a small number of online platforms, Twitter in particular, and focuses on forms of participation synonymous with these sites, such as hashtag publics. That is not to say that research on ad hoc issue publics and hashtag movements is reductionist; clearly these are sites of great importance to both activists and those researching their practices. Even so, the micro-blogging platform cannot be deployed as a proxy for other online platforms due to its specific platform affordances and the representativeness (or lack thereof) of its technologically skilled, educated, high earning user base (see Blank, 2017; Hargittai, 2020). Hence, digital activism research disproportionately produces knowledge of particular social groups as well as very specific dynamics of activism., i.e., ad-hoc issue publics aggregating around hashtags and through short messages on sites like Twitter. While these dynamics often attract high visibility and media attention, this form of activism is often ephemeral and based on weak ties. Even where such dynamics may be more complex and sustained, computational Twitter research (as described in the archetypes above) in and of itself cannot capture backstage activism. We acknowledge that research is necessarily shaped by socio-economic conditions (see Bruns, 2019; Burgess & Bruns, 2015; Rogers, 2019). Unequal access to software and non-random samples may continue to be problematic, allowing wealthier institutions to work with larger datasets and leaving impoverished ones languishing behind in areas such as computer science and data science. However, perhaps in response to concerns over the implications of the 'API-calypse' (Bruns, 2019), Twitter's decision in January 2021 to open up its API for academic research may make it easier for scholars to access these data in the future. Whether these researchers have the necessary skills and expertise to analyse such datasets remains to be seen.

Reconsidering easy data

While the review appeared to provide some support for Bruns and Burgess's propositions, it also highlights the need to reconsider the notion of reductionism. For example, in many articles hashtagged tweets were often triangulated with data gathered from other sources via non-computational methods. While Twitter might be the most well-known easy data platform, it would be misleading to categorise all studies focussing on the microblogging site as reductionist or unrepresentative. The variety of methodological approaches in the corpus suggested that digital activism researchers frequently reached beyond the 'low-hanging fruit' of hashtags; many contextualised their findings through the use of other data sources. Such approaches better contextualise high-visibility hashtag data within wider activism dynamics taking place on these platforms, regardless of whether these attract media attention.

Decisions about methodological approaches are, of course, based primarily on the specific questions that researchers are investigating. It may be appropriate to rely on exclusively easy data to investigate specific research questions, rather than triangulate with data gathered via other methods. Small-scale qualitative research is no less rigorous than large quantitative studies of hard data, assuming that best practices in these methods have been applied by the researcher(s). Rather, the research questions and theoretical framework deployed by researchers may ultimately determine the type of data collected, whether it is easy or hard.

Future directions

In this paper we have argued that future digital activism scholarship would benefit from greater diversity in methodological approaches. Whether multi-platform or triangulated with other data sources, there is a need for scholars in the field to move away from the easy data of hashtags and explore the less visible dynamics of digital activism which are often under-researched.We Also join other scholars in calling for researchers to (a) consider and discuss the social and behavioural mechanisms produced by digital platforms (Flanagin, 2020), (b) limit research questions to social media (rather than movements as a whole) and the specific dynamics on these platforms (Hargittai, 2020), (c) reflexively contextualise the dynamics of activism revealed in this research (Driscoll & Walker, 2014; Lorentzen & Nolin, 2017; Venturini et al., 2018), and (d) consider methodologies that produce better quality data, such as triangulation through multi-platform analyses, qualitative analysis of subsamples of large datasets, and not using digital methods for studies requiring granular data (Tromble, 2021; Tufekci, 2014). This

would produce digital activism research that contextualises specific cases of activism within wider movement dynamics as well as the platform vernaculars under scrutiny.

Beyond research design, digital activism researchers should provide more information in the methodology sections of their articles to demonstrate the rigour of their work. A limitation of this systematic review was that there were articles in the corpus that provided little information pertaining to their methods of data collection and analysis. Although journal expectations vary in terms of these disclosures, methods should be described in sufficient detail in order for others to assess their scope, representativeness, and validity (Driscoll & Walker, 2014). Such transparency would illuminate how issues of representativeness are reflected in the choices of methods, platforms and case studies in digital activism research. This includes explicit mentions of sample size in the collected social media data - regardless of methodological approach, the timeframe of data creation and collection, the use of computational methods, the chosen software tools including the API access that these provide, as well as a brief commentary on limitations set by these specific choices. This would aid not only reproducibility in quantitative work, but also add validity and rigour to digital activism research.

References

Arquilla, J., & Ronfeldt, D. (2001). *Networks and netwars: The future of terror, crime, and militancy*. RAND.

Berry, D. (2011). The computational turn: Thinking about the digital humanities. *Culture Machine*,12,1–22. https://culturemachine.net/wp-content/uploads/2019/01/10-Computational-Turn-440-893-1-PB.pdf

Blank, G. (2017). The digital divide among Twitter users and its implications for social research. *Social Science Computer Review*, 35(6), 679–697. https://doi.org/10.1177/0894439316671698

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

Boulianne, S. (2019). Revolution in the making? Social media effects across the globe. *Information,Communication & Society*, 22(1), 39–54. https://doi.org/10.1080/1369118X.2017.1353641

Boulianne, S., Koc-Michalska, K., & Bimber, B. (2020). Mobilizing media: Comparing TV and social media effects on protest mobilization. *Information, Communication & Society*, 23(5),1–23. https://doi.org/10.1080/1369118X.2020.1713847

Boulianne, S., & Theocharis, Y. (2020). Young people, digital media, and engagement: A metaanalysis of research. *Social Science Computer Review,* 38(2), 111–127. https://doi.org/10.1177/0894439318814190

Bruns, A. (2019). After the 'APIcalypse': Social media platforms and their fight against critical scholarly research. *Information, Communication & Society,* 22(11), 1544–1566. https://doi.org/10.1080/1369118X.2019.1637447

Burgess, J., & Bruns, 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 Publishing.

Campbell Collaboration. (2019). Campbell systematic reviews: Policies and guidelines. *Campbell Policies and Guidelines Series No. 1*. https://doi.org/10.4073/cpg.2016.1

Chae, Y., Lee, S., & Kim, Y. (2019). Meta-analysis of the relationship between internet use and political participation: Examining main and moderating effects. *Asian Journal of Communication,* 29(1), 35–54. https://doi.org/10.1080/01292986.2018.1499121

Cochrane Library. (2019). *Cochrane Database of Systematic Reviews*. Chrane / John Wiley & Sons [Online]. Available at: https://www.cochranelibrary.com/cdsr/about-cdsr

Driscoll, K., & Walker, S. (2014). Big data, big questions: Working within a black box: Transparency in the collection and production of big Twitter data. *International Journal of Communication*, 8(20), 1745–1764. https://ijoc.org/index.php/ijoc/article/view/2171/1159

Flanagin, A. J. (2020). The conduct and consequence of research on digital communication. *Journal of Computer-Mediated Communication,* 25(1), 23–31. https://doi.org/10.1093/jcmc/zmz019

Gerlitz, C., & Rieder, B. (2013). Mining One percent of twitter: Collections, baselines, sampling. *M/C Journal*, 16(2), https://doi.org/10.5204/mcj.620

Hammersley, M. (2020). Reflections on the methodological approach of systematic reviews. In O.Zawacki-Richter, M. Kerres, S. Bedenlier, M. Bond, & K. Buntins (Eds.), *Systematic reviews in educational research* (pp. 23–39). Springer V. https://doi.org/10.1007/978-3-658-27602-7_2

Hargittai, E. (2020). Potential biases in Big data: Omitted voices on social media. *Social Science Computer Review*, 38(1), 10–24. https://doi.org/10.1177/0894439318788322

Hino, A., & Fahey, R. A. (2019). Representing the Twittersphere: Archiving a representative sample of Twitter data under resource constraints. *International Journal of Information Managemen*t, 48, 175–184. https://doi.org/10.1016/j.ijinfomgt.2019.01.019

Jackson, S. J., Bailey, M., & Foucault Welles, B. (2020). *#Hashtag activism: Networks of race and gender justice*. MIT Press.

Karatzogianni, A. (2015). *Firebrand waves of digital activism 1994-2014: The rise and spread of hacktivism and cyberconflic*t. Springer.

Krippendorff, K. (2004). Measuring the reliability of qualitative text analysis data. *Quality and Quantity*, 38(6), 787–800. https://doi.org/10.1007/s11135-004-8107-7

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions:Explanation and elaboration. *PLOS Medicine,* 6(7), e1–e34. https://doi.org/10.1371/journal.pmed.1000100

Lorentzen, D. G., & Nolin, J. (2017). Approaching completeness: Capturing a hashtagged Twitter conversation and its follow-on conversation. *Social Science Computer Review*, 35(2), 277–286. https://doi.org/10.1177/0894439315607018

Marres, N. (2017). Digital sociology: The reinvention of social research. Polity Press.

Matthes, J., Knoll, J., Valenzuela, S., Hopmann, D. N., & Von Sikorski, C. (2019). A meta-analysis of the effects of cross- cutting exposure on political participation. *Political Communicatio*n, 36(4), 523–542. https://doi.org/10.1080/10584609.2019.1619638

McCaughey, M., & Ayers, M. D. (2003). *Cyberactivism: Online activism in theory and practic*e. Routledge.

Meikle, G. (2010). Intercreativity: Mapping online activism. In J. Hunsinger, L. Klastrup, & M.Allen (Eds.), *International handbook of internet research* (pp. 363–377). Springer.

Miller, V. (2017). Phatic culture and the status quo: Reconsidering the purpose of social mediaactivism. *Convergence*, 23(3), 251–269. https://doi.org/10.1177/1354856515592512

Nagano, R. L. (2015). Research article titles and disciplinary conventions: A corpus study of eight disciplines. *Journal of Academic Writing*, 5(1), 133–144. https://doi.org/10.18552/joaw.v5i1.168

Neumayer, C., & Rossi, L. (2016). 15 years of protest and media technologies scholarship: A sociotechnical timeline. *Social Media + Society*, 2(3), 1–13. https://doi.org/10.1177/2056305116662180

Özkula, S. M. (2021). What is digital activism anyway? Social constructions of the "digital" in contemporaryactivism. *Journal of Digital Social Research*, 3(3), 60–84. https://doi.org/10.33621/jdsr.v3i3.44

Pearce, W., Niederer, S., Özkula, S. M., & Sánchez Querubín, N. (2018). The social media life of climate change: Platforms, publics, and future imaginaries. *Wiley Interdisciplinary Reviews:Climate Chang*e, 10(2), e569. https://doi.org/10.1002/wcc.569

PRISMA. (2015). PRISMA Checklist. PRISMA Transparent Reporting of Systematic Reviews and Meta-Analyses [Online]. Available at: http://prisma-statement.org/PRISMAStatement/Checklist.aspx

Puschmann, C., & Pentzold, C. (2020). A field comes of age: Tracking research on the internet within communication studies, 1994 to 2018. *Internet Histories*, 5(2), 1–19. https://doi.org/10.1080/24701475.2020.1749805

Rogers, R. A. (2017). Digital methods for cross-platform analysis. In J. Burgess, A. Marwick, & T. Poell (Eds.), *SAGE Handbook of social media* (pp. 91–110). Sage.

Rogers, R. A. (2019). *Doing digital method*s. Sage.

Tatarchevskiy, T. (2011). The 'popular' culture of internet activism. *New Media & Society*, 13(2), 297–313. https://doi.org/10.1177/1461444810372785

Tromble, R. (2021). Where have All the data gone? A critical reflection on academic digital research in the post-API Age. *Social Media + Society*, 7(1), 1–8. https://doi.org/10.1177/2056305121988929

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://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewPaper/8062

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

Yang, G. (2009). Online activism. *Journal of Democracy*, 20(3), 33–36. https://doi.org/10.1353/jod.0.0094

Zanaboni, P., Ngangue, P., Mbemba, G. I. C., Schopf, T. R., Bergmo, T. S., & Gagnon, M. P. (2018). Methods to evaluate the effects of internet-based digital health interventions for

citizens: Systematic review of reviews. *Journal of Medical Internet Researc*h, 20(6), e10202. https://doi.org/10.2196/10202

Zhu, M., Sari, A., & Lee, M. M. (2018). A systematic review of research methods and topics of the empirical MOOC literature (2014–2016). *The Internet and Higher Education*, 37, 31–39. https://doi.org/10.1016/j.iheduc.2018.01.002

Appendices

Appendix 1: Search term list for query

	Queries
Set 1: Synonyms	“digital activism” “cyberactivism” “cyber-activism” “online activism” “virtual activism” “social media activism” “internet activism” “informational activism” “web activism” “e-activism” “hashtag activism” “net activism” “mobile activism”
Set 2: keyword combinations	"digital media" AND "activism" "social media" AND "activism" "digital media" AND "protest" "social media" AND "protest" "digital media" AND "social movement" "social media" AND "social movement" "digital media" AND "campaign" "social media" AND "campaign"

Appendix 2: PRISMA flowchart

Appendix 2. PRISMA flowchart