



Selected Papers of #AoIR2020:
The 21st Annual Conference of the
Association of Internet Researchers
Virtual Event / 27-31 October 2020

EASY DATA, USUAL SUSPECTS, SAME OLD PLACES? A SYSTEMATIC REVIEW OF METHODOLOGICAL APPROACHES IN DIGITAL ACTIVISM RESEARCH BETWEEN 1995-2019

Suay M. Özkula
University of Sheffield

Paul J. Reilly
University of Sheffield

Jennifer Hayes
University of Sheffield

Background

Burgess and Bruns (2015) are amongst those researchers to link the computational turn in social media research to significant transformations in digital research, such as a rise in studies that focus exclusively on “easy” data”, such as the ‘low hanging fruit’ provided by Twitter hashtags. This paper sets out to explore whether this “hegemony” of easy data and studies is evident in digital activism research between 1995 and 2018; this was a period of increasing media fragmentation and hybridization, as well as innovation in research methodologies. Therefore, a systematic review of methodological approaches towards the study of digital activism in relevant peer-reviewed journal articles was conducted.

Objective

The aim of the study was to provide an overview of the methodological explorations in the field to date, with a view to test Burgess and Bruns’ (2015) “easy data” hypothesis. Three arguments are central to their hypothesis. The *first argument* is that research methodologies have substantially changed through the computational turn, in which the introduction of computational methods and the production of both digitized (data that has been converted into digital formats) and natively digital data (data created for and in digital spaces) has led to epistemic changes. Their *second argument* is 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 than to adapt or manipulate. Their *third*

Suggested Citation (APA): Özkula, S., Reilly, P., Hayes, J. (2020, October) *Easy Data, Usual Suspects, Same Old Places? A Systematic Review of Methodological Approaches in Digital Activism Research Between 1995-2019*. Paper presented at AoIR 2020: The 21th Annual Conference of the Association of Internet Researchers. Virtual Event: AoIR. Retrieved from <http://spir.aoir.org>.

argument relates to the results of these preferences in light of platforms' politics and regimes of access. They say that, due to restrictions set by platforms, these (Twitter) studies often provide more easily or readily accessible data within those platforms such as reductionist hashtag studies and @reply network. Consequently, a considerable amount of research relies on what they label "easy data", data available through standard Twitter API services and therefore consist of 'some 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).

In response to these arguments, we formulated the following hypotheses to be tested in our systematic review of (empirical) digital activism scholarship:

H1: Following the computational turn, research methodologies draw predominantly on digital, digitized, and hybrid methods.

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

H3: Studies of activism on Twitter primarily rely on "easy data" obtained via standard API services.

Methods

In order to test these hypotheses, a mixed-methods 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) on the Scopus database. Post-filtering, a CSV file containing the meta-data and abstracts for 1447 articles was downloaded from Scopus. The abstracts were coded manually in Excel on methodological attributes such as chosen methods of data collection and data analysis, digital data sources (i.e. studied platforms), and case studies. The final set of articles coded on methodological attributes consisted of 364 articles, of which 202 articles were judged to analyze social media data in some form among its methods, and separated. The aim was to explore the applicability of Bruns and Burgess' easy data hypothesis by including additional coding categories, which were framed around how "easy" the data was to collect (e.g. hashtag studies, single- or multi-platform studies, triangulation).

Results

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

In terms of methodological approaches, traditional research methods and digital methods studies were prevalent. The study found that Twitter was the most researched platform in the corpus, closely followed by Facebook. However, (software-based) digital methods were not as commonly deployed in these articles as traditional methods, such as interviews and surveys. #H1 was therefore judged predominantly accurate in terms of a strong representation of computerized methods, particularly in pos-2010 articles, at least as a part of the wider methodological approach. However, these approaches did not constitute a majority. Thus, while computational methods have had a significant influence in the field, they do not dominate them.

#H2: Research methodologies using digital data predominantly draw on Twitter data, particularly hashtag and keyword searches (= “low-hanging fruit”).

While Twitter studies did not dominate in the article collection overall, they did so within the subset of studies drawing on social media data mining as well as in studies using manual collection of social media data, meaning that amongst computational methods Twitter was indeed prevalent. As such, Twitter dominated studies drawing on computational methods, but not the dataset overall. While both Twitter and Facebook were used extensively as additional or contextual data (particularly in observational studies), Twitter dominated the field both where multiple sources were used (in 74 out of 178 studies, compared to 31 Facebook studies) and in single-platform studies. Overall, #H2 was judged to be partially true as the “low-hanging fruit of Twitter data” constituted (a) a substantial but not majority portion of the overall corpus, and (b) a dominant approach in single-platform and software-based studies.

#H3: Twitter research in the field primarily draws on “easy data”, meaning modestly sized short-term data available through standard Twitter API services.

Since API access level remained difficult to judge based on the contents of the article methodologies, this hypothesis was primarily tested through what could be considered “reductionist approaches” beyond the platform and automation, for example sample size and timeframes. Within the limitations of some missing data (due to differing methodologies), #H3 was judged true in that the data mining articles primarily drew on “easy data” interpreted as modestly sized short-term data (typically available through standard Twitter API services). We acknowledge that the judgment of sample size and timeframe as high or substantial strongly relies on the type of activism and the size and spread of the specific movement overall, but considered these (in line with Bruns and Burgess’ line of argumentation) to be significant for an evaluation of easy data.

As part of the inductive element of the study, we further considered what methodological profiles constitute typical approaches to the study of digital activism. We found that single-platform Twitter studies drawing on either a form of content analysis or social network analysis (of hashtags, mention/reply, or link networks) were the single most consistent approach, with non-computerized single-platform Facebook page studies also constituting a common approach to the study of digital activism.

Future Work

The findings of the review suggest that future scholarship would benefit from including other methodological approaches such as hard data studies through high-level (and therefore more extensive) access, as well as non-Twitter single- or multi-platform studies, and more contextualized approaches (e.g. multi-platform, triangulated, and holistic approaches) towards producing knowledge that is somewhat more representative of the complex and wide digital landscapes. We do, however, acknowledge that future studies remain limited by the socio-economic conditions that produce these relations (as described by Bruns and Burgess) and these will consequently affect the type of research that is produced.

References

Burgess, J. & Bruns, A. (2015). Easy data, hard data: the politics and pragmatics of Twitter research after the computational turn. In Langlois, G., Redden, J., & Elmer, G. (Eds.), *Compromised Data: From Social Media to Big Data* (pp. 93-111). Bloomsbury Publishing: London.