



UvA-DARE (Digital Academic Repository)

Social Media Research after the Fake News Debacle

Rogers, R.

DOI

[10.1285/i20356609v11i2p557](https://doi.org/10.1285/i20356609v11i2p557)

Publication date

2018

Document Version

Final published version

Published in

Partecipazione e Conflitto

License

CC BY-NC-ND

[Link to publication](#)

Citation for published version (APA):

Rogers, R. (2018). Social Media Research after the Fake News Debacle. *Partecipazione e Conflitto*, 11(2), 557-570. <https://doi.org/10.1285/i20356609v11i2p557>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.



Partecipazione e Conflitto
*** The Open Journal of Sociopolitical Studies**
<http://siba-ese.unisalento.it/index.php/paco>
ISSN: 1972-7623 (print version)
ISSN: 2035-6609 (electronic version)
PACO, Issue 11(2) 2018: 557-570
DOI: 10.1285/i20356609v11i2p557

Published in July 15, 2018

Work licensed under a Creative Commons Attribution-Non commercial-Share alike 3.0 Italian License

SYMPOSIUM/7

SOCIAL MEDIA RESEARCH AFTER THE FAKE NEWS DEBACLE

Richard Rogers
University of Amsterdam

1. Introduction: The coming crisis in social media research

The purpose of the following is to reintroduce contemporary critiques of social media research, as they are gathering steam following the 'fake news debacle', which I come to. These are not social media or platform critiques per se, such as platformization which refers to how the web is becoming enclosed and overwritten by social media (Helmond 2015). Embedded in the research critique is some discussion of Facebook policy as well as Twitter rules (for example), but that is not the main effort here. Rather, the point is a larger academic one that discusses issues related to social media research, both concerning the use of the platforms for research generally as well as the data they collect. What are the implications for doing political and social research these days when employing social media platforms and their data? When one is studying (political) engagement online, for example, how to conceptualise platform effects?

Behind these questions is a digital methods approach to studying social media that revolves around the notion of 'repurposing' (Rogers 2013). Digital methods as an idea is built on the notion of using existing online data left behind, or collected for other purposes, and then repurposing it for research such as "tracing the spread of arguments, rumors, or positions about political and other issues" (Watts 2007; Lazer *et al.*

2009, 722). The data could be described as ‘traces’, as in that which was left behind like footprints in the snow. Social media data analysis thereby becomes akin to unobtrusive measures (Webb *et al.*, 1966). Or, the data could be ‘interactions’ expressly collected by the platforms. An early term that encapsulates platforms’ collecting user interactions is registrational interactivity (Jensen, 1998). As the user ‘likes’ or otherwise interacts with posts, her activities are registered. They then are ‘industrialized’ by the platforms, or made productive use of, for commercial as well as socio-epistemological purposes, in a manner similar to how hyperlinks are construed as valuation practices, and their measure may be transformed into commercial product (Brin and Page 1998; Turow 2008; Helmond 2015).

Recently, repurposing has been questioned, largely because of the current emphasis placed on how platforms capture user data, and how they encourage greater exposure of the self. Whether discussed as an ensnaring or an extractive practice, the platforms’ models of interaction and user experience also enable it to offer fine-grained ‘audience segmentation’ to those who wish to purchase ads, such as on Facebook. In the infamous case of the U.S. presidential elections in 2016 (but likely in many other cases, too), the ad systems were used to spread so-called hyperpartisan fake news, disinformation and other transgressive or malevolent content (Chen 2015; Commons Select Committee 2018a; 2018b). The use of traces and interactions for spreading fake news, especially to those with particular personality profiles, has led to a crisis in social media research, including calls for unplugging as well as developing alternative scientific instrumentarium for data collection. The question now reads, how could political and social researchers continue to use Facebook data to study engagement, when these systems are both normatively dubious in their data collection practices, and are being deployed for partisan, political ends?

In the following the discussion of social media research critique has five entry points: good data, human subjects, proprietary effects, repurposing and alternatives. The first concerns how social media have oftentimes been criticized for not being ‘good data’ at least in the sense that the fields in the databases are unstable overtime, and that the introduction of new ones leads to interactive complexity. For example, on Facebook the ‘reactions’ that were introduced in 2016 interfered with the stability of ‘likes’, given the new choices in how to react to a post. The critique extends beyond the data fields. Even the metrics used by the corporations evolve, such as the definition of reach on Facebook’s CrowdTangle, as a researcher found after publishing findings on Russian disinformation Pages on Facebook (Timberg 2017).

The second issue – social media users as ‘human subjects’ – has been raised in a well-known work by Metcalf and Crawford and is part of an ethics turn in social media

research and the coming ‘crisis’ in computer science and online research more generally (2016). Regarding the crisis, it has been argued that unlike other disciplines computer science has not had the ‘reckoning’ that chemistry had after dynamite and poison gas, physics after the nuclear bomb, human biology after eugenics, civil engineering after bridge, dam and building collapses, and so forth (Zunger 2018). The point is of course that the Cambridge Analytica affair could become such a reckoning. In the affair, a psychometrics researcher at the University of Cambridge delivered 80 million profiles to a political marketing firm intent on undertaking a ‘psyops-style’ political influence campaign on Facebook users, delivering ‘dark posts’ of hyperpartisan ‘fake news’ to those whose personality profile had been determined to have a high degree of ‘openness’ and ‘neuroticism’ (Commons Select Committee 2018a). In the terms of service of the app that collected the personality profile data, the researcher did not indicate that individuals’ answers would be deployed in such a manner, which captures the ‘ethics divide’ or “discontinuities between the research practices of data science and established tools of research ethics regulation” (Commons Select Committee 2018b; Metcalf and Crawford 2016, 1)

The third issue – proprietary effects – has been present in the background of research based on social media data for some time. Social media platforms as proprietary platforms have different goals from science, though such a distinction may be blurred given that there are behavioural and data scientists working and publishing academically at these companies. It could be said that data are being collected for dual purposes, advertising foremost, and research secondarily. Nevertheless, one of the main differences between two data collection means and ends is the reflexivity involved. In digital sociology (and sociology more generally), the effects of collecting and analysing data anticipate societal impact, rather than experiment with it (Marres 2018).

The fourth point concerns the repurposing issue touched upon above. Given that social media data have been gathered primarily for the purposes of selling ads, and that system interactivity and user experience are aimed primarily at furthering social media consumption by the users and granting more exposure of oneself in order to provide still more data, repurposing faces the issue of medium or platform effects. One may not straightforwardly separate activity on social media with activity in the wild. Liking may be overdetermined by the platform rather than an expression of feeling or preference.

The fifth discussion point concerns whether researchers should be studying and also pursuing alternatives to the current, dominant social media platforms for data collection for one’s own research as well as one’s own publicity practices. Research about Facebook is far vaster than that on alternative social media platforms that embed dif-

ferent design choices and values, such as the open-source, decentralised ones, Diaspora and Mastodon. Researcher use of social media for academics (researchgate.net and academia.edu but also ssrn.com) is far more widespread than on alternatives such as scholarlycommons.org and scholarhub.org (Matthews 2016). Recently normative questions have arisen, such as whether scholars ought to contribute to academia.edu, which is a “for-profit venture capital backed company” rather than the educational institution that the use of the .edu top-level domain would imply (Bond 2017; Tennant 2017).

2. Good data?

A starting point in the critique of social media research is that social media platforms are not instruments set up for the purposes of doing research, e.g., for tracking social discourse or ‘social listening,’ a term often used in this regard, imported from the business and marketing literature (Balduini *et al.* 2013; Cole-Lewis *et al.* 2015). The platforms are not the equivalent of specially crafted sensors for collecting carbon dioxide levels in the air, for example, as the Mauna Loa Observatory in Hawaii has undertaken since the 1950s. The data the platforms do collect (whether traces or registered interactions) are also not to be considered good data in the sense of data that is collected at the beginning of a phenomenon, is complete and remains stable over time (Borgman 2009). Rather, certain fields disappear, and other ones appear. When they do, there is what could be called ‘interactive complexity’ in the data (to borrow a term from technological systems theory), as certain data from the fields that were collected previously (e.g., ‘likes’) are then affected by new data fields that are introduced (e.g., ‘reactions’) (Perrow 1984). If one examines ‘likes’ over time (as a proxy for feeling or preference), dips may be platform-dependent rather than an indicator of a change of heart.

Not only are the data fields unstable, but so are the data themselves as well as the inbuilt metrics. The journalism researcher, Jonathan Albright, brought to light in October 2016 how Facebook deleted Russian disinformation pages from CrowdTangle, Facebook’s social media monitoring tool. Albright had captured the top posts from 6 Russian disinformation pages (Blacktivists, Heart of Texas, United Muslims of America, Being Patriotic, Secured Borders and LGBT United), and published his findings as a data visualisation concerning their engagement and reach (Albright 2016). The engagement and reach numbers Albright published were much larger than Facebook had originally indicated in Congressional testimony that focused on the ads purchased by the influence campaigners rather than the issue-oriented Facebook Pages that stirred discon-

tent. After Albright published his findings, Facebook “wiped” the Page data from CrowdTangle, arguing that the Pages should not have been available any longer because they are ‘inactive’, the term for suspended, or accounts that broke Facebook rules (Timberg and Dwoskin 2017). To Albright and others, the “public interest data” was removed for public relations reasons, and researchers have no recourse (Timberg, 2017). After all, Facebook owns its data as well as its de facto CrowdTangle archive that once held the content of interest. Albright also found that Facebook changed the in-built metrics that could also be considered data, or indicators to be repurposed. The second of CrowdTangle’s two metrics (“total engagement” and “total people shared to”) was renamed to “total followers”. To Albright, that change implies that “the thousands of propaganda posts (with tens of millions of shares) were not shared to ‘people,’ but rather to ‘accounts,’ which lowers the perceived impact” (2018).

Holes in the data may be created for a variety of reasons, the most common of which are set country restrictions, but they also occur when data are shared. Twitter is a case in point. For example, the German authorities may ask Twitter to ‘withhold’ far right extremist tweets to users, and Twitter likely would comply for the location Germany, as has been the case on numerous occasions including tweets not only by German extremists but British ones, too (Kulish 2012; Cox 2017). The tweets may be unavailable in Germany, but they are still available in the Netherlands (and elsewhere for that matter). Routine data collection of German extremist tweets may thus be better performed outside the country, in order to plug the holes.

Another occasion where data sets are depleted occurs through sharing data. One may not share a tweet collection proper, but rather only a collection’s tweet IDs. These tweet IDs may be recompiled as a collection by querying for them via one of Twitter’s APIs, but those tweets that have been withheld or deleted would be cleansed from the data set by Twitter. Twitter also asks tweet collectors to obey Twitter’s Rules and be a ‘good partner’ by routinely removing from one’s tweet collections those that have been withheld or deleted. It becomes a debatable norm when Twitter purges accounts that a researcher feels are worthy of study, such as Russian disinformation trolls or alt right figures, to name two examples. The third category of data hole that is created arises from privacy settings. Facebook Pages, for example, can have country and age restrictions set, and depending on where one collects the data or who collects it, some may be missing without the researcher having any knowledge of it.

3. Human Subjects

Are researchers 'covered' by the fact that users have signed on to platforms' terms of service, which indicate clearly (and, in the case of Twitter, repeatedly) that their data may be used not only for the improvement of the software but also for marketing research and other research purposes, including academic endeavours? If one buys historical data from Twitter, for example, is one able to use it for research purposes as one sees fit? The particular idea that researchers may use as cover platform terms of service or purchased data has come under scrutiny, not only in the debates that ensued from the 'outing' by Michael Zimmer of the (weakly) anonymised Facebook data set used in the taste and ties research at Harvard (2010), which is one marker in the ethics turn in social media research. The idea that 'the data are already public' (and users have agreed to share it) are points of departure in the debate surrounding notions of contextual privacy and contextual integrity, which puts forward the contrary position (Nissenbaum 2010). Respect for 'contextual privacy' implies an understanding that a user posting data online does not expect that same data to be used in a different context, e.g., for commercial activities or research purposes not knowingly consented to or reasonably expected, even if the terms of service, agreed to by the user, appear to grant a wide range of data uses, including to third parties who have acquired that data through purchase or the proper use of the API.

Data ethics in the context of internet-related research (as espoused by the Association of Internet Researchers' guidelines and elsewhere) would have as its point of departure that care be taken with 'data subjects' who are not 'objects' in a database but rather human subjects (Markham and Buchanan 2012). An ethics of care approach, which would consider establishing and maintaining a relationship with the data subjects, however, could be seen as incompatible with big data research, for its impracticality given the sheer number of subjects involved. When consent is not explicitly sought, one would consider publicizing one's research and inviting opt-out.

The third point concerns treating social media users as not only human subjects but also as authors. Is one using the subject's data, or is one citing and/or quoting them? The question of a tweet or a Facebook post as 'authored' work conventionally would consider if they are worthy forms of creative expression. An authored work is often considered as such owing to its originality or because it is the product of the sweat of one's brow (Beurskens 2014). These definitions are considered when imparting copyright and other author's rights. One case in point would be a particularly impactful tweet from an analytical point of view, such as one that was found (through emotion

analysis) to be the angriest tweet on the night of the U.S. presidential elections. That tweet could be considered a citable work by researchers.

4. 'Proprietary effects'

The question of the impact of proprietary data or operations on research normally would begin with the observation that social media data increasingly have been commodified, meaning that the media companies are in the advertising as well as in the data business (Puschmann and Burgess 2014). Such a state of affairs does not necessarily interfere with one repurposing the data for social research, if one can still acquire it. But the amount of free data (especially on Facebook), and the quality of free data (especially on Twitter), have gradually declined. Researchers have been coming to grips for some time with the consequences of relying on commodified APIs, starting with the disclosure that in-house data scientists (at Twitter) have higher quality data than those on the outside (boyd and Crawford 2012). There is a 'data divide' between those researchers without access to great pipelines, making due with narrow ones that are choked by rate limiting. Data has become expensive, such as a 'complete' "climate change" (hashtag and keyword) Twitter data set I estimated recently with the aid of Texifter at \$54,000. Accompanying the rise of the proprietary data is a price tag, lest the quality is reduced.

When the company holding the data is charging handsome sums for it, one could consider consulting the archives. Up until January of this year, there was the prospect that the U.S. Library of Congress would continue to hold all of Twitter's archive, and eventually make it available with query machines, but the December 2017 announcement put paid to associated research plans (Osterberg 2017). The Library related that it would cease collecting the entire Twitter archive, bravely reporting that it has its first twelve years (as text), which itself is a worthy collection. From thenceforward the Library would create special collections, and though it remains to be seen of which type, the plans would be to continue with its web collection policy, where there has been a preponderance of collections concerning disasters and elections (and transitions such as the papal or presidential) (Rogers 2018). When one is accustomed to querying a Twitter API for whichever keywords and hashtags and is now confronted with limited, curated data sets on special subject matters, research agendas are affected, certainly ones that explore wide-ranging contemporary social issues with approaches that seek competing hashtag publics, for example.

More to the point, the social media archives are now held solely by the companies, and these archives are 'updated' from time to time, given that the companies make accounts inactive, or suspended, as in the case of the Russian disinformation Pages on Facebook or the Alt Right accounts on Twitter. As Jonathan Albright has pointed out, this data has been "wiped", and there is no public archive that holds them for academic and other public research purposes. The 2018 Facebook initiative, 'Social Science One', that makes Facebook data available to social scientists for the study of disinformation could well have been cleaned of that which the researchers may be seeking (Gonzalez 2018; King and Persily 2018).

The last proprietary effect to be mentioned here concerns researcher treatment by social media companies. Social media APIs do not differentiate between academic researchers and marketing companies or potential data resellers. All are customers. If one strives to configure a system for more comprehensive data collection (using multiple accounts, funnelling all data collected into one repository), one is treated as a spammer or reseller, blocked and actively worked against. Researchers became spammy users, breaking terms of service, or not regarded as a 'good partner'.

Since the Cambridge Analytica scandal of 2018 (and the fake news debacle that accompanied it), researchers with tools sitting atop Twitter's APIs or running native apps on Facebook have been asked by the companies to reapply for accounts and permissions. The Facebook application form (with a 5-day deadline) is particularly worthy of study, since it actively seeks ethical lapses in one's prior data collection, unless one demonstrates otherwise.

5. 'Repurposing'

Recently 'repurposing' social media data for social research has been critiqued along normative and analytical lines (Marres 2018). As discussed above, platforms are not scientific instruments for collecting societal trend data, but rather are in the business of data extraction for the purposes of segmenting audiences and selling advertising. One queries keywords in the Facebook ad interface and an audience is returned. The company would like to increase the amount of data points per user so that the audience becomes ever more differentiated (segmented).

More critically, it is argued that social media companies, like natural resource firms before them, are the new extraction industries. They do not so much crowdsource as crowd-fleece (Scholz 2017). That researchers would rely on data that has been 'fleeced'

from the crowd is normatively problematic. At bottom, the companies also operate outside of the norms of science, whether Mertonian, Kuhnian or otherwise.

Moreover, data extraction requires interface and interaction engineering that invites users to expose themselves further and interact often with the system. When one is studying social media data, one could just as well be studying the success of engineered user interaction rather than 'genuine' behaviour (liking or endorsing), where for example measures of value, reputation or preference could be derived. On the contrary, so goes the argument, when one is studying social media data, one is primarily learning about social media consumption. In other words, the platform is so built to extract data from users in order for others to advertise to ever finer grained segmented audiences, rather than for other reasons such as to create community or enhance public debate.

6. Conclusion: Alternatives

By way of conclusion, I would like to discuss briefly the question of alternatives, both to API-driven research as well as to studying and using the dominant social media platforms. Scraping has been a method of online data collection that through the rise of the API became associated with breaking terms of service or 'partnership' guidelines (Marres and Weltevrede 2013). Rather than collect data through scraping, researchers complying with the terms have witnessed an array of changes to the APIs of the dominant platforms and have been asked on a number of occasions to re-apply for developer access as well as permission to deploy a research tool. On one specific occasion in 2016, applications made to Instagram (for the 'visual hashtag explorer') have failed (Rieder 2016). Others in 2018 have been highly time-sensitive – Facebook's multiple-page reapplication form due in 5 days, as recounted above. Still others have been only cumbersome, such as Twitter's demand in 2018 to re-apply for developer keys. Apart from calls to drop the API and return to digital ethnography, user studies and other small data research practices, reactions to such obstacles erected by social media companies more in line with digital methods include continuing technical fieldwork as well as API critique. How is research affected by the latest API version, and what kind of tool development could (still) result in valuable social research? Are we only increasingly studying the platforms' updates?

When discussing alternatives, one also may begin with the observation made by Tim Berners-Lee (the web's co-inventor) that the 'open web' is in decline, and one of the major reasons (apart from the rise of surveillance) he listed is the growth of the social

media platform, walling in users and content. As the Internet Archive has demonstrated, even public Facebook Pages are challenging to archive (and little has been retained); web ‘recording’ is one small-scale alternative. The overall result of platformization is a falling web. Whilst an empirical question, if one studies the health of the web, say sector by sector, it would appear that the governmental and commercial webs are still up and running, but the non-governmental web (for one) does not appear as vibrant as it once was.

There has been a series of proposals put forward to change the online landscape, including ones at an ownership level. Trebor Scholz’s call for “platform cooperativism” is a discussion about “cloning the technological heart” of sharing economy platforms, whilst basing the co-ops on principles of solidarity and innovation for all rather than the few (2016, 14).

The amount of scholarly output using Facebook and Twitter data is vast compared to that examining alternatives. But it is not just researcher interest in the dominant platforms over ‘secondary social media’; it is also researcher use of such sharing platforms as researchgate.net, academia.edu and ssrn.com that is of interest here. The Scholarly Commons is an alternative, implemented by universities as part of their domain such as scholarcommons.law.northwestern.edu or repository.upenn.edu. These systems tend to highlight a university or department’s output, rather than aggregate across universities. Another (at the demo phase) is ScholarlyHub, which (as Scholz calls for) emulates much of the functionality of academia.edu or researchgate.net but emphasises scholarly interests over ranking and score-keeping.

Acknowledgement

This project has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement no. 732942.

References

- Albright J. (2016), “Itemized Posts and Historical Engagement - 6 Now-Closed FB Pages,” Tableau Public, <https://public.tableau.com/profile/d1gi#!/vizhome/FB4/TotalReachbyPage>, last accessed 23 August 2018.
- Albright J. (2018), Personal conversation.

- Balduini M., E. Della Valle, D. Dell'Aglio, M. Tsytsarau, T. Palpanas, and C. Confalonieri (2013), "Social Listening of City Scale Events Using the Streaming Linked Data Framework", *ISWC 2013: The Semantic Web – ISWC 2013*, 1-16.
- Beurskens M. (2014), "Legal Questions of Twitter Research", in Weller, Katrin, Axel Bruns, Jean Burgess, Merja Mahrt and Cornelius Puschmann (eds.), *Twitter & Society*, New York, Peter Lang, 123-136.
- boyd d. and K. Crawford (2012), "Critical Questions for Big Data", *Information, Communication & Society*, 15(5): 662-679.
- Brin S. and L. Page (1998), "The Anatomy of a Large-Scale Hypertextual Web Search Engine," *Seventh International World-Wide Web Conference (WWW 1998)*, 14-18 April, Brisbane, Australia.
- Bond S. (2017), "Dear scholars, Delete your account at Academia.edu," *Forbes*, 23 June, <http://www.forbes.com/sites/drsarahbond/2017/01/23/dear-scholars-delete-your-account-at-academia-edu>, last accessed 23 August 2018.
- Borgman C. (2009), "The Digital Future is Now: A Call to Action for the Humanities," *Digital Humanities Quarterly*, 3(4).
- Chen A. (2015), "The Agency," *New York Times*, 2 June, <https://www.nytimes.com/2015/06/07/magazine/the-agency.html>, last accessed 23 August 2018.
- Cole-Lewis H., J. Pugatch, A. Sanders, A. Varghese, S. Posada, C. Yun, M. Schwarz, and E. Augustson (2015), "Social Listening: A Content Analysis of E-Cigarette Discussions on Twitter," *Journal of Medical Internet Research*, 17(10): e243.
- Commons Select Committee (2018a), Evidence from Christopher Wylie, Cambridge Analytica whistle-blower, *U.K. Parliament*, 28 April, <https://www.parliament.uk/business/committees/committees-a-z/commons-select/digital-culture-media-and-sport-committee/news/fake-news-evidence-wylie-correspondence-17-19/>, last accessed 23 August 2018.
- Commons Select Committee (2018b), Dr Aleksandr Kogan questioned by Committee, *U.K. Parliament*, 24 April, <https://www.parliament.uk/business/committees/committees-a-z/commons-select/digital-culture-media-and-sport-committee/news/fake-news-aleksandr-kogan-evidence-17-19/>, last accessed 23 August 2018.
- Cox J. (2017), "This Is How Twitter Blocks Far-Right Tweets in Germany," *Motherboard*, 13 June, https://motherboard.vice.com/en_us/article/3kz57j/this-is-how-twitter-blocks-far-right-tweets-in-germany, last accessed 23 August 2018.

- Gonzalez R. (2018), "Facebook is giving scientists its data to fight misinformation," *Wired.com*, 29 May, <https://www.wired.com/story/facebook-is-giving-scientists-its-data-to-fight-misinformation/>, last accessed 23 August 2018.
- Helmond A. (2015), "The Platformization of the Web: Making Web Data Platform Ready", *Social Media + Society*, July-December: 1–11.
- Jensen J. F. (1998), "Interactivity: Tracking a New Concept in Media and Communication Studies", *Nordicom Review*, 1: 185-204.
- King G. and N. Persily (2018), "A New Model for Industry-Academic Partnerships", *Working Paper*, Available at <http://j.mp/2q1IQpH>, last accessed 23 August 2018.
- Kulish N. (2012), "Twitter Blocks Germans' Access to Neo-Nazi Group", *New York Times*, 18 October, <https://www.nytimes.com/2012/10/19/world/europe/twitter-blocks-access-to-neo-nazi-group-in-germany.html>, last accessed 23 August 2018.
- Lazer D., Pentland, A. (S.), Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., Van Alstyne, M. (2009), "Life in the network: the coming age of computational social science", *Science*, 323(5915): 721–723.
- Markham A., Buchanan, E. (2012), *Ethical Decision-Making and Internet Research: Recommendations from the AoIR Ethics Working Committee (Version 2.0)*, Association of Internet Researchers.
- Marres N. (2018), *Digital Sociology*, Cambridge, Polity.
- Marres N. and E. Weltevrede (2013), "Scraping the social? Issues in live social research", *Journal of Cultural Economy*, 6(3): 313-335.
- Matthews D. (2016), "Do academic social networks share academics' interests?", *Times Higher Education*, 7 April.
- Metcalf J. and K. Crawford (2016), "Where are human subjects in Big Data research? The emerging ethics divide", *Big Data & Society*, January–June: 1–14.
- Nissenbaum H. (2010), *Privacy in Context: Technology, Policy and the Integrity of Social Life*, Palo Alto, Stanford University Press.
- Osterberg G. (2017), "Update on the Twitter Archive at the Library of Congress", *Library of Congress blog*, 26 December.
- Perrow C. (1984), *Normal Accidents: Living with High-risk Technologies*, New York, Basic Books.
- Puschmann C. and Burgess J. (2014), "The Politics of Twitter Data," in Weller K., Bruns A., Burgess J., Mahrt M., Puschmann C. (eds.), *Twitter & Society*. New York, Peter Lang, 43-54.

- Rieder B. (2016), "Closing APIs and the public scrutiny of very large online platforms," *Politics of Systems blog*, 27 May, <http://thepoliticsofsystems.net/2016/05/>, last accessed 23 August 2013.
- Rogers R. (2013), *Digital Methods*, Cambridge, MA: MIT Press.
- Rogers R. (2018), "Periodizing web archiving: Biographical, event-based, national and autobiographical traditions", in Brügger, N., Milligan, I. (eds.), *SAGE Handbook of Web History*, London, Sage.
- Scholz T. (2017), *Uberworked and Underpaid: How Workers Are Disrupting the Digital Economy*, Cambridge, Polity Press.
- Scholz T. (2016), *Platform Cooperativism: Challenging the Corporate Sharing Economy*, New York, Rosa Luxemburg Foundation.
- Tennant J. (2017), "Who Isn't Profiting Off the Backs of Researchers?", *Discover Magazine Blog*, 1 February, <http://blogs.discovermagazine.com/crux/2017/02/01/who-isnt-profiting-off-the-backs-of-researchers/#.W36XAS17EWo>, last accessed 23 August 2018.
- Timberg C. (2017), "Russian propaganda may have been shared hundreds of millions of times, new research says", *Washington Post*, 5 October.
- Timberg C. and E. Dvoskin (2016), "Facebook takes down data and thousands of posts, obscuring reach of Russian disinformation", *Washington Post*, 12 October.
- Turow J. (2008), "Introduction: On Not Taking the Hyperlink for Granted", in Turow J. and L. Tsui (eds.), *The Hyperlinked Society: Questioning Connections in the Digital Age*, Ann Arbor, MI, The University of Michigan Press, 1-18.
- Watts D. (2007), "A twenty-first century science", *Nature*, 445: 489.
- Webb E. J., D. T. Campbell, R. D. Schwartz, L. Sechrest (1966), *Unobtrusive Measures: Nonreactive Research in the Social Sciences*, Chicago, Rand McNally.
- Zimmer M. (2010), "'But the data is already public': On the ethics of research in Facebook", *Journal of Ethics and Information Technology*, 12(4): 313-325.
- Zunger J. (2018), "Computer science faces an ethics crisis. The Cambridge Analytica scandal proves it", *Boston Globe*, 22 March, <https://www.bostonglobe.com/ideas/2018/03/22/computer-science-faces-ethics-crisis-the-cambridge-analytica-scandal-proves/lzaXxl2BsYBtwM4nxezgcP/story.html>, last accessed 23 August 2018.

Author's information

Richard Rogers is Professor of New Media & Digital Culture at the Department of Media Studies, University of Amsterdam. Richard does research in Digital Methods, Digital Culture and Information Politics.