Book Review

Marta Cantijoch, Rachel Gibson and Stephen Ward, eds, 2014, Analysing Social Media Data and Web Networks, Palgrave Macmillan, London

The digital realm offers a multitude of opportunities for research. However, given the ever-changing nature of online environments, research focused on assessing such 'moving targets' [7] need to 'freeze the flow' [5] or make the data deluge available online suitable for scientific analysis in some other way. The volume at hand, *Analysing Social Media Data and Web Networks* [2] is edited by Marta Cantijoch, Rachel Gibson and Stephen Ward and offers a series of useful and often practical insights for those of us who take special interest in analysis of online media. Specifically, the book features ten chapters that all provide insights into (primarily) methodological issues, presented by some of the most well known authors in what could perhaps be described as the field of online political communication (and beyond). In this review of the book, I have arranged my comments around five main issues that permeate throughout the text. In so doing, I will provide examples from individual chapters featured in the title, as well as from other sources. I have chosen to label the five issues dealt with as follows: *The ever-changing nature of online services*; *Commercialization of data access*; *Socio-demographic perspectives*; *Ethical issues* and *Comparing with what is to come*.

First, digital methods are fickle. They need to be fashioned so as to be able to adapt to and catch the aforementioned online flows. Indeed, researchers have dealt with what I like to call the *ever-changing nature of online services* for some twenty years, painstakingly learning from previous mistakes and developing more efficient ways of data gathering from online sources. Often, the tools used for such endeavors are constructed and maintained by individual scholars and their respective research groups, making it somewhat difficult for the community at large to judge the merits of any particular tool in comparison to some other variety. For example, while I am certain that the services introduced by Thelwall and Hussain et al. in their chapters are of the utmost quality, the very fact that more and more purpose-built tools are launched could lead to difficulty in performing cumulative, comparative research as researchers select their tool of choice from an ever-increasing array of instruments. We should, of course, always strive to improve our tools, but the lament of the editors regarding the apparent lack of theoretical cohesion would appear to ring true also for these issues: 'the field has deviated from [systematic theoretical inquiry] in a rather chaotic fashion, which makes cross-country and longitudinal comparison extremely difficult' (pp. 16–17). A similar statement could arguably made with regards to the methodological development of online research, broadly defined.

Second, such tools for collecting are made subject to almost constantly updated rules of the social media platforms they allow us to study. Such changes often appear to be related to what is understood in this review as ongoing processes of *commercialization of data access*. We can, for example, point to relatively recent delimitations of free access to a variety of public application programming interfaces (APIs) as hosted by Twitter (e.g. [1]), or the delimitations of functionalities imposed by Facebook on the freely available Netvizz data extraction service [9]. Indeed, issues like these are touched upon in the chapter penned by Jungerr and Jürgens, but it would have been nice if the authors or editors had touched upon what could be labeled as critical interpretations of these developments. With such a view in mind, the chapter by Graham and Wright correctly suggests that 'people's online data is often commercially

valuable' (p. 204) – but what does such value entail for academic conduct? Arguably, the current developments are detrimental for scientists who, often with scarce funding, seek to perform research detailing services like these. As such, there is a clear risk that the increased commercialization of data access will contribute to a further widening of the already existing chasms between 'data-rich' and 'data-poor' scholars (e.g. [6]).

My third point considers *socio-demographic perspectives* of the users whose digital trace data often end up in our work sheets, research notebooks and eventually (or hopefully, perhaps) published works. Specifically, regardless of how data are collected, we must assess who the producers of these data are – at least in some overarching, structural sense. Here, many of the included authors do a good job at acknowledging the biases that societal divisions like these unequivocally place on the data we gather from online sources. Increased knowledge about such stratifications should help end the sometimes heard happy-go-lucky type argument that data, because it is so plentiful (or even 'big', if you will), would be representative of the public opinion. Certain groups of citizens will always be overrepresented for certain forms of media use – a difficult obstacle to overcome for scholars, but an obstacle to be acknowledged clearly, nonetheless (e.g. [4]). Of course, such over- or underrepresented groups could be expected to vary across countries and contexts – something that further underlines the necessity of and challenges with comparative research across the strata of your choice.

Fourth, *ethical issues* are, or at least should be, at the very heart of scholarship. Such choices and prioritizations seemingly become especially poignant in the online context, where data emanating from a variety of user profiles and interactions can be collected and systematized with relative ease. The openness of online platforms like Twitter or YouTube is sometimes discussed as providing a *carte blanche* for various forms of data collection. Therefore, it is refreshing to see such arguably simplistic approaches to methodology questioned in Thelwall's chapter, where it is suggested that '[d]espite this openness, there is of course a need for researchers to exercise discretion when personally identifying individuals in the course of their research' (p. 76). Related to such identification of individuals is the topic or theme dealt with in the tweets, Facebook posts or YouTube videos examined. While it might be technically true that 'the majority of this data is open for all to examine' [10, p. 296], special consideration should be taken when the content deals with what could be understood as sensitive topics, such as sexual preferences or political orientation [3,8]. A recent overview by Zimmer and Proferes suggests that at least for research into Twitter, reflection on ethical issues are seldom seen [11]. This reviewer would be surprised if the situation was different for scholarship detailing other, similar services.

One way to approach ethical issues has been to focus on content that has been actively put forward by users in such a way as to indicate their willingness to be seen in a specific thematic context. On Twitter, for example, this has been done by focusing on so-called hashtags – thematic keywords included by the users themselves to show thematic coherence. Such an approach is favored by a series of authors contributing to the volume at hand, like the previously mentioned chapter by Jungherr and Juĺrgens as well as the section penned by García-Albacete and Theocharis. Indeed, this way of approaching research could be seen as relatively unproblematic from an ethical point of view. However, the issue of what lies beyond the hashtag – in other words, what contents of relevance we are missing out on by delimiting our searches in this supposedly ethically sound way – remains unanswered.

My fifth point, *comparing with what is to come*, relates back to the first one. I mentioned at the beginning of this review that the methods discussed here could be seen as in constant flux, given the almost continuous changes taking place within the technical infrastructures we wish to study. For this final point, I'd like to stress the fact that not only do these infrastructures change – they will undoubtedly become out-of-date at some point, replaced by some new variety. Indeed, the services we study today

will most likely not be around tomorrow, and it would have been fruitful to see the authors and editors reflect to a higher degree on such issues of cross-platform comparability in the volume. For example, how do we secure longitudinal insights, comparing possible future online platforms with those in fashion today if we construct our data collection tools and phrase our research questions based on the affordances of those services currently available?

Finally, while studies assessing the use of various social media platforms are *all the rage*, it is good to see that *Analysing Social Media Data and Web Networks* also features a series of chapters dealing with analyses of web sites, particularly those provided by Rosalund Southern and Benjamin N. Lee. Indeed, while it might be tempting to study comparably new services like social media platforms, the important role of web pages within political campaigning should be acknowledged with a suitable amount of attention from researchers.

In conclusion, while the focus of the book is placed on issues primarily of concern to the broader field of political communication, such a thematic delimitation should not keep potential readers with mainly methodological interests at bay – the rich perspectives offered here are sure to be of use also to those coming to the study of online methods from some other disciplinary starting point.

Anders Olof Larsson, Westerdals – Oslo School of Arts, Communication and Technology,
Olso, Norway
E-mail: anders.larsson@westerdals.no

References

[1] Burgess J, Bruns A. Twitter Archives and the Challenges of 'Big Social Data' for Media and Communication Research. [big data;social media;internet studies]. M/C Journal. 2012; 15(5).
[2] Cantijoch M, Gibson R, Ward S. *Analysing Social Media Data and Web Networks*. London: Palgrave Macmillan, 2014.
[3] Ess C. *Digital Media Ethics* (Second ed.). Cambridge: Polity Press, 2013.
[4] Hargittai E, Litt E. Becoming a Tweep. Information, Communication & Society. 2012; 15(5): 680-702.
[5] Karlsson M, Strömbäck J. FREEZING THE FLOW OF ONLINE NEWS – Exploring approaches to the study of the liquidity of online news. Journalism Studies. 2010; 11(1): 2-19.
[6] Larsson AO. Studying Big Data – ethical and methodological considerations. In H. Fossheim & H. Ingierd (Eds.), *Internet research ethics*. Oslo: Cappelen Damm Akademisk, 2015.
[7] McMillan SJ. The Microscope and the Moving Target: The Challenge of Applying Content Analysis to the World Wide Web. Journalism & Mass Communication Quarterly. 2000; 77(1): 80-98.
[8] Moe H, Larsson AO. Methodological and Ethical Challenges Associated with Large-scale Analyses of Online Political Communication. Nordicom Review. 2012; 33(1): 117-124.
[9] Rieder B. the end of Netvizz (?) (2015). Retrieved from http://thepoliticsofsystems.net/2015/01/the-end-of-netvizz/.
[10] Vargo CJ, Guo L, McCombs M, Shaw DL. Network Issue Agendas on Twitter During the 2012 U.S. Presidential Election. Journal of Communication. 2014; 64(2): 296-316.
[11] Zimmer M, Proferes N. A Topology of Twitter Research: Disciplines, Methods, and Ethics. *Aslib Proceedings*, 2014.