



Caring as Critical Proximity: A Call for Toolmaking in Digital Migration Studies

Anders Munk

The field of digital methods has emerged over the past two decades at the interface between media studies, science and technology studies (STS), computer science, and information design (Marres 2017; Rogers 2013). The core questions and ambitions covered closely resemble the kind of digital migration studies presented in this book. First, a commitment to studying wider sociocultural phenomena on, through, and with digital traces from online media. Second, a desire to critically engage with the role of these media, their platforms, and algorithms as infrastructures of social life. Third, a curiosity about how to wrongfoot entrenched divides between offline and online, actual and virtual, or qualitative and quantitative. And finally, ongoing experimentation with computational techniques for large-scale data capture analysis in conjunction with ethnographic approaches to deep sensemaking and interpretation. Indeed, to begin

A. Munk (✉)

University of Aalborg in Copenhagen, Copenhagen SV, Denmark
e-mail: anderskm@hum.aau.dk

© The Author(s) 2022

M. Sandberg et al. (eds.), *Research Methodologies and Ethical Challenges in Digital Migration Studies*, Approaches to Social Inequality and Difference, https://doi.org/10.1007/978-3-030-81226-3_10

235

caring for big data, as argued by Sandberg and Rossi in the introductory chapter to this book, is not only a welcome call for digital migration research but also rings true in digital methods more broadly.

The fact is that digital migration studies were early and significant contributors to the evolution of digital methods, although this may be relatively unacknowledged in both fields today. When Dana Diminescu began developing her E-Diasporas project in 2006,¹ she also began collaborating with a young research engineer, Mathieu Jacomy, who would later become one of the key toolmakers in digital methods (the story is recounted in Jacomy's recent Ph.D. dissertation, Jacomy 2021). Together with other team members, and spurred by the ambition of the project to collect and map relations between websites from thirty different migrant communities online (Diminescu 2008), they devised the concepts and developed the prototypes for two iconic research tools, namely, Gephi² (Jacomy et al. 2014), which is today one of the world's most popular pieces of open-source software for visual network analysis, and the Navicrawler³ (Diminescu et al. 2011), which was a predecessor for the current state-of-the-art tool for web corpus curation in digital methods (Jacomy et al. 2016).

Reading through the timely and thought-provoking contributions in this book, it seems to me that the E-Diasporas collaboration has an important story to tell. What Jacomy and Diminescu practiced together in the borderland between digital methods and digital migration studies was not only a form of caring for big data, but a very particular form of caring that took place in what Bruno Latour would call "critical proximity" (Latour 2005) with the technical circumstances that surround the production and analysis of such data. Contrary to a critically distant position, which cares for the consequences of new technological practices (in this case data scientific practices) that encroach on a field (in this case digital migration studies), caring as critical proximity undertakes to intervene with such practices, reimagining what they could do and redesigning them accordingly. When Sandberg, Mollerup, and Rossi discuss what it would take to develop a "contrapuntal" analysis of digital connectedness (Chapter 3), or when Makrygianni, Kamal, Rossi, and Galis (Chapter 2) engage with how to map the relationality of digital migrant space, they do so in a form of critical proximity with the Facebook API and its shifting affordances. Like other forms of care, not least those described by scholars such as Annemarie Mol, Ingunn Moser, or Jeanette Pols (Mol 2008; Mol et al. 2015), Sandberg and Rossi draw on in their introduction; caring for big

data could thus be construed as a material practice in which data-intensive analysis is *done* differently. Crucially, this *doing* depends on the willingness and ability of researchers to engage practically with the affordances of digital tools and platforms. As the chapters in this book demonstrate, this is not to be mistaken for simply replacing digital migration studies with computational social science. Caring for big data in critical proximity must necessarily entail that data science methods are reconceived and experimentally reassembled from a position firmly inside the field in question.

CARING FOR DIGITAL TRACES AND RELATIONAL SPACES

In digital methods, caring as critical proximity has resulted in a wide variety of homegrown tools for harvesting and analysing digital traces from social media platforms such as Facebook (Rieder 2013) or Twitter (Borra and Rieder 2014). At the time of the E-Diasporas project, however, the main interest was still centred on websites. Several tools had already been developed for collecting websites, the so-called web corpora, when Jacomy and Diminescu began their collaboration, but their specific interest in online migrant communities made it necessary to do data curation and analysis differently.

Web pages at the time were (and to a large extent still are) written in hypertext markup language (html), which allows a browser to know which parts of a page to display as headings, body text, links, menus, images, etc. It also allows a piece of software known as a *scraper* to selectively harvest parts of that information, for example, all the links, all the images, or all the text found under certain headings. It does so by “repurposing” (Rogers 2013) the tags in the markup. A hyperlink (or hypertext reference) is marked up with the tag “href = ” followed by a web address. Thus, by identifying all the instances of a “href = ” in the html code and collecting the accompanying web addresses, a scraper piggybacks on the functionality and conventions of html, as well as the choices made by the authors of a web page, to build a data set. It follows that scrapers cannot simply harvest any kind of digital trace, at least not with equal ease, but must operate within an affordance space that has been designed elsewhere and by others, which is why Noortje Marres points out that such digital methods are “distributed” (Marres 2012), in the sense that they are conditioned by and constituted from a heterogeneous set of actors that are often extraneous to the immediate situation of the researcher.

Digital traces, then, are not simply given as data but actively taken as what Johanna Drucker (2011) calls “capta”—that which is taken as opposed to that which is given, in order to signal that there is nothing naturally occurring about it—with due regard to the specific sociotechnical circumstances of their construction. Caring for these sociotechnical circumstances is of critical importance to digital methods. The most obvious capta—those that are most straightforwardly capturable—are not necessarily the most interesting or amenable for a given research purpose. In the case of the E-Diasporas project, which was interested in mapping the online presence of specific diasporas through their websites and their linking practices, the generic hyperlink was not in itself a relevant digital trace to follow. Websites send links to other websites for a variety of reasons and clearly not exclusively to signal belonging to the same diaspora. The generic hyperlink, however, was the straightforwardly capturable digital trace. At the time, all the available web crawlers that made link scraping manageable and approachable for researchers did not distinguish between different types of hyperlinks. For good reasons: there were no tags or conventions in html that could be easily repurposed to make such a distinction, and in any case, the crawlers had been built for other purposes.

The predominant web crawler in digital methods, for instance, was called the Issue Crawler and had been developed with a particular media studies interest in mind (Marres and Rogers 2005). A crawler (or a spider) is a version of a scraper that follows hyperlinks from page to page according to a set of rules and thus automates part of the data collection process. The Issue Crawler took as input a list of seed pages dedicated to a specific issue (e.g., the GMO debate, a new vaccine, or a contested infrastructure project) and used them as starting points from which it followed all hyperlinks to more pages at a set crawl depth (number of link steps from the seed). This would produce a data set of more pages dedicated to the same issue, but also of pages about tangential issues, as well as news or social media sites, and internet infrastructure such as add servers, trackers, search engines, or content management systems. Rather than viewing this wider entanglement of web pages, which come into view when you generically follow hyperlinks from a seed, as noise, the Issue Crawler catered for a now long-standing tradition in digital methods for studying the infrastructures of online media as an integral part of the issue (Marres 2015).

When the knowledge interest is different, however, as was the case with E-Diasporas, the existence of ready-to-use tools like the Issue Crawler presents a challenge. Because they make certain digital traces, in this case the generic hyperlink, even more capturable than they already were, they require an extra level of care on behalf of the researcher, who must not only be capable of critically appreciating how capture could be different, but also intervene directly in the design of tools that make other forms of capta possible. This was exactly the approach taken by Jacomy and Diminescu in E-Diasporas when they realised that they would need to be able to curate multiple web corpora with a data collection approach that was simultaneously capable of leveraging the power of crawling to find all links from thousands of pages on the same website but also qualitatively select which of the discovered websites to include in a corpus. Rather than including all websites at a set distance from the seed, they needed a crawler that would continuously prompt the researcher to qualitatively decide if links should be followed or not. The solution was a purpose-built application for Firefox, the Navicrawler (Diminescu et al. 2011), which combined the qualitative element of browsing web pages with the quantitative power of crawling. The Navicrawler asks the user to navigate to a website from which to begin building the corpus. The crawler element visits all the pages on the site in question, scrapes all hyperlinks, and provides a list of discovered sites for the user, who then decides which to include in the corpus and which to leave out. When a site is included in the corpus, the crawler automatically reiterates the process on that site. In the case of E-Diasporas, the result was 30 different web corpora representing 30 different online migrant communities, each built from a known website of that community and curated through a selective crawling process whereby only sites relating to the same community were included.

The E-Diasporas web corpora, then, were made possible through caring as critical proximity. Indeed, this is true not only for the collection of the websites, which depended on the development of the Navicrawler, but also for the exploration of the relational spaces emerging from their linking practices, which depended on the development of the Gephi software (Jacomy et al. 2014). Visually exploring patterns in how websites were linked required network analysis and, in particular, force-directed network layouts. A force-directed layout produces a visualisation where nodes (in this case websites) are placed close to each other if they share many of the same connections (in this case hyperlinks). Force-directed

layouts are non-deterministic algorithms that begin from a random positioning of the nodes and introduce an energy model that pushes them apart if they are unconnected. The visualisation will be different each time you run the algorithm, and settings for the energy model can be changed, further changing the visualisation. The goal is to exploratively obtain visualisations that prompt curiosity and help the researcher generate questions about the relational space that can be pursued qualitatively. Or rather, that is *one* goal in the use of force-directed layouts and typically one that is associated with practices such as multi-sited ethnography (Munk and Ellern 2015) or actor-network theory (Venturini et al. 2019) in which the phenomenon in question is not presumed to exist in a pre-existing bounded space, and the ongoing construction of spatiality is a central object of study. This requires a tool that allows the user to adjust the energy settings for the force-directed layout and experiment with producing different network views. Gephi was born as an answer to that specific knowledge interest (Fig. 10.1).

IMPLICATIONS FOR DIGITAL MIGRATION STUDIES

It is perhaps unlikely that digital migration studies anno 2021 will have to care much about the curation of websites. As the chapters in this book demonstrate, interest has, as in most other fields, shifted to digital platforms, not least social media. The need for caring as critical proximity has never been greater, though. The dependence on platform APIs for harvesting data has, over the years, proven highly unstable, as endpoints have been changed or deprecated and access denied at short notice. Simply relying on tools that others have built to do digital methods research quickly lands you in situations where whole research projects become untenable from month to month as the infrastructure on which the tool is based changes or is entirely removed. We must therefore be agile enough to not only think about, but also to develop in practice, makeshift alternatives (Perriam et al. 2020; Venturini and Rogers 2019).

This comes on top of an even more central point about critical proximity, which has always been true regardless of API changes and deprecations, namely, that specific digital traces must be repurposed for specific knowledge interests. As the E-Diasporas example demonstrates, this repurposing is often a matter of not capturing the most straightforwardly capturable digital traces but working materially to make other

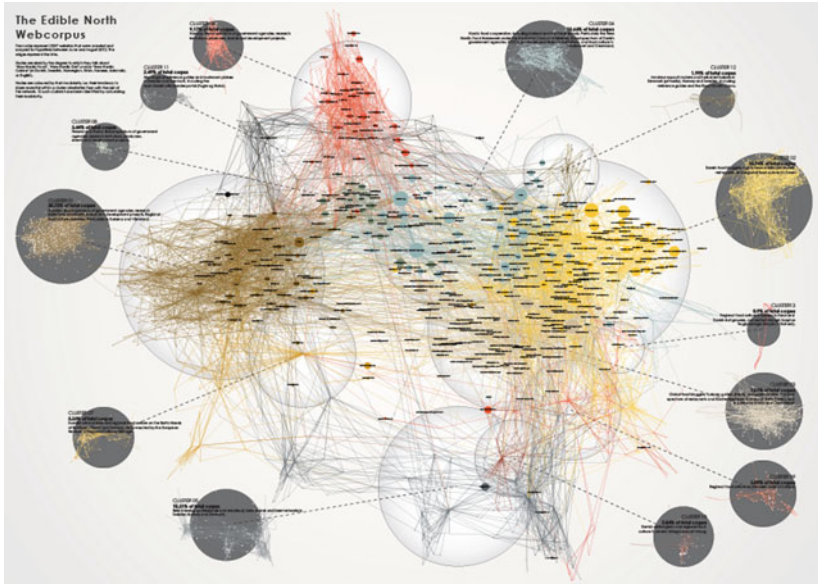


Fig. 10.1 An example of a web corpus collected with the Navicrawler and explored with a force-directed layout in Gephi (Munk 2019). By combining manual curation and automated crawling, I collected 2007 websites from food actors in Scandinavia. The relational space resulting from the linking practices of these actors becomes visible and explorable when the force-directed layout places clusters of interconnected websites in proximity with each other. The large yellow (right), red (top), and brown (left) clusters turn out to be national spheres of food-related websites (Danish, Norwegian, and Swedish), whereas some of the smaller clusters are communities of interest focussed on practices such as foraging, mushroom picking, beer brewing, or aquavit making. These relational spaces emerge from weblinks of the actors and become visible as such through the intervention of Gephi. This image is used with permission of the author [Rightsholder]

kinds of *capta* possible. When Sandberg, Mollerup, and Rossi convincingly argue for the merits of thinking about digital connectedness as “contrapuntal” (Chapter 3), then how should the Facebook API be repurposed to best support such an analysis? Is it even possible within the current restrictions on API-based research? If not, should we adopt a more activist stance (as suggested by Ben-David 2020) and begin scraping

Facebook instead of accepting how the platform makes itself available for scrutiny through the API, despite its ethical and legal complications? Like many related fields, digital migration research could conceivably very well find itself in a situation where the ethical and societal arguments in favour of documenting a phenomenon collide with the terms of service and/or API policies of a given platform. What if a platform that prohibits its users from scraping and offers no API access for research is also a key hub for human trafficking? What if a closed group with tens of thousands of members on Facebook, which is out of bounds for research except through scraping, is used for coordinating harassment or exploitation of migrants? Does digital migration research have an obligation to document such issues? If so, where is the toolbox that allows this?

Similar questions pertain to data analysis: Are the tools we have available for visualising patterns in Facebook data and making it available for quali-quantitative exploration also well suited for contrapuntal analysis? Is Tableau enabling us or preventing us from doing so? Similarly, when Makrygianni, Kamal, Rossi, and Galis want to “follow a relational approach (...) according to which space is not considered as a life container but as a derivative of social relations and interactions” (Chapter 2), are line graphs and custom error bar charts the visual techniques that really support that endeavour? And how would we go about collecting the relational user data today, now that the API no longer allows it?

These are questions for a research practice that cares for big data as a form of critical proximity. Care as a material practice obliges us to seriously consider how the sociotechnical circumstances of data capture and analysis can be engaged and transformed. This is as true in digital methods as it is in digital migration studies, and this book aptly demonstrates this point. It is not an easy task, however, as there will always be practical motivation to leave the data science to the data scientists and the question-posing to those in SSH with a digital research interest. It is simply convenient. But it is also dangerous and, I believe, the direct path to a version of caring for big data at a distance that will render our story of data science as either one of appropriation, where we accept remodelling our questions and research interests in order to become amenable to the methods of others, or one of antagonism, where we try to protect our qualitative and interpretive approaches by pointing out the biases, shortcomings, shallowness, and ethical problems of various kinds of big social data and computation. What this book promises to do instead—and

it deserves a lot of credit for it—is to begin the hard work of reassembling data-intensive computational methods from a position within digital migration studies, in dialogue with the existing methodological landscape of the field, and in response to the research interests that are already articulated by its researchers.

NOTES

1. <http://www.e-diasporas.fr/> [last accessed March 1, 2021].
2. <https://gephi.org/> [last accessed March 1, 2021].
3. <https://medialab.sciencespo.fr/en/tools/navicrawler/> [last accessed March 1, 2021].

BIBLIOGRAPHY

- Ben-David, Anat. 2020. “Counter-Archiving Facebook.” *European Journal of Communication*, 35 (3): 249–264. <https://doi.org/10.1177/0267323120922069>.
- Borra, Eris, and Bernhard Rieder. 2014. “Programmed Method: Developing a Toolset for Capturing and Analyzing Tweets.” *Aslib Journal of Information Management*, 66 (3): 262–278. <https://doi.org/10.1108/AJIM-09-2013-0094>.
- Diminescu, Dana. 2008. “The Connected Migrant: An Epistemological Manifesto.” *Social Science Information*, 47 (4): 565–579.
- Diminescu, Dana, Matthieu, Renault, Mehdi, Bourgeois and Jacomy, Mathieu. 2011. “Digital Diasporas Atlas Exploration and Cartography of Diasporas in Digital Networks.” *Proceedings of the International AAAI Conference on Web and Social Media*, 5 (1).
- Drucker, Johanna. 2011. “Humanities Approaches to Graphical Display.” *Digital Humanities Quarterly*, 5 (1): 1–21.
- Jacomy, Mathieu. 2021. *Situating Visual Network Analysis*. Aalborg University: Aalborg.
- Jacomy, Mathieu, Paul Girard, Benjamin Ooghe, and Tomaso Venturini. 2016. “Hyphe, a Curation-Oriented Approach to Web Crawling for the Social Sciences.” *International AAAI Conference on Web and Social Media*.
- Jacomy, Mathieu, Venturini, Tomaso, Heymann, S., and Bastian, M. 2014. “ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software.” *PLoS One*, 9 (6): e98679.
- Latour, Bruno. 2005. “Critical distance or critical proximity.” Unpublished Manuscript. Accessed March 31, 2014. <http://www.bruno-latour.fr/sites/default/files/p-113-haraway.pdf>.

- Marres, Nortje. 2012. "The Redistribution of Methods: On Intervention in Digital Social Research, Broadly Conceived." *The Sociological Review*, 60, 139–165.
- Marres, Nortje. 2015. "Why Map Issues? On Controversy Analysis as a Digital Method." *Science, Technology, & Human Values*, 40 (5): 655–686. <https://doi.org/10.1177/0162243915574602>.
- Marres, Nortje. 2017. *Digital Sociology: The Reinvention of Social Research*. John Polity & Sons: Cambridge.
- Marres, Northe and Richards Rogers. 2005. *Recipe for Tracing the Fate of Issues and their Publics on the Web*.
- Mol, Annemarie. 2008. *The Logic of Care: Health and the Problem of Patient Choice*. Abingdon: Routledge.
- Mol, Annemarie., Ingunn Moser and Jeanette Pols. 2015. *Care in Practice: On Tinkering in Clinics, Homes and Farms* (Vol. 8). Transcript Verlag: Bielefeld.
- Munk, Anders K. 2019. "Four Styles of Quali-Quantitative Analysis: Making Sense of the New Nordic Food Movement on the Web." *Nordicom Review*, 40 (s1): 159–176. <https://doi.org/10.2478/nor-2019-0020>.
- Munk, Anders K., and Ellern, A. B. 2015. "Mapping the New Nordic Issuescape: How to Navigate a Diffuse Controversy with Digital Methods." *Tourism Encounters and Controversies: Ontological Politics of Tourism Development*, eds. R. Duim et al, Ashgate.
- Perriam, Jessamy, Andreas Birkbak, and Andres Freeman. 2020. "Digital Methods in a Post-API Environment." *International Journal of Social Research Methodology*, 23 (3): 277–290.
- Rieder, B. 2013. "Studying Facebook Via Data Extraction: The Netvizz Application." *Proceedings of the 5th Annual ACM Web Science Conference*, 346–355. <https://doi.org/10.1145/2464464.2464475>.
- Rogers, Richard. 2013. *Digital Methods*. MIT Press.
- Venturini, Tomaso, Anders K. Munk, and Mathieu Jacomy. 2019. "Actor-Network Versus Network Analysis Versus Digital Networks: Are We Talking about the Same Networks?" In *Digital STS: A Field Guide for Science & Technology Studies*. Princeton University Press.
- Venturini, Tomaso and Richard Rogers. 2019. "'API-Based Research' or How can Digital Sociology and Journalism Studies Learn from the Facebook and Cambridge Analytical Data Breach." *Digital Journalism*, 7 (4): 532–540. <https://doi.org/10.1080/21670811.2019.1591927>.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

