



Unblackboxing Gephi: How a user culture shapes its scientific instrument

Mathieu Jacomy, Emilija Jokubauskaitė

► To cite this version:

Mathieu Jacomy, Emilija Jokubauskaitė. Unblackboxing Gephi: How a user culture shapes its scientific instrument. 2022. hal-03729114

HAL Id: hal-03729114

<https://hal.science/hal-03729114>

Preprint submitted on 21 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Unblackboxing Gephi: How a user culture shapes its scientific instrument

Authors:

- **Mathieu Jacomy** (corresponding author)
TANT Lab, Aalborg University, Copenhagen, Denmark
mathieu.jacomy@gmail.com
<https://orcid.org/0000-0002-6417-6895>
Twitter: @jacomyma
- **Emilija Jokubauskaitė**
PhD candidate at University of Amsterdam
e.jokubauskaite@uva.nl
Cultural Sociology / New Media Studies
Twitter: @emilijajok

Abstract. Network visualization tools are blamed for being black boxes, but depending on who criticizes, blackboxing has one of two meanings. Blackboxing by inscrutability blames opacity (inaccessibility of the method) while blackboxing by embodiment blames transparency (invisibility of the mediation), thus we propose to drop the confusing dichotomy of opacity/transparency. We focus on Gephi, a network visualization tool popular in the social sciences and humanities, addressing its criticism as a black box, in the hope of proposing useful interventions. By analyzing empirical material from participant observation, interviews, and a publication review, we find that Gephi's blackboxing problem is by embodiment. We observe that some researchers circulate Gephi images without a proper interpretative context, a situation we refer to as “storyletting”, or use methodological shortcuts in time-constrained situations, trading methodological awareness for efficiency. We acknowledge the existence of an epistemic subculture around Gephi, and we argue that its practices are only partially aligned with the instrument's design. We contend that Gephi's black box materializes in cultural artifacts such as online documentation and tutorials. We conclude that these artifacts are a valid place of intervention to try unblackboxing Gephi, and that researchers are natural actors of this critical technical practice.

Introduction

Technology is not neutral – to paraphrase Kranzberg's (1986) first law of technology – and accordingly, digital tools and data are said to have transformed the production of knowledge in the social sciences and humanities (Savage and Burrows, 2007; Berry, 2011; Manovich, 2013; Ruppert *et al.*, 2013; Meyer & Schroeder, 2015; Masson, 2017). However, digital tools and practices are too rarely given an active role in the narratives about big data, neither from enthusiasts (e.g. Lazer *et al.*, 2009) nor from critics (e.g. Boyd and Crawford, 2012; Ruppert *et al.*, 2013). The discussion tends to be about technology in general rather than tools in particular (or instruments, as we refer to them). Most authors agree on the material-semiotic nature of big data issues, yet the level where materiality can be directly observed, the instrument level, is not addressed often. This is surprising considering that the idea that tools shape practices (see e.g. Wertsch, 1991) is common in Science and Technology Studies (Callon, 2002) as well as in software development (Kelty, 2012). The premise of this paper is to focus on a given scientific instrument (Gephi, an open-source software), and address its criticism (being a black box) by studying the practices of researchers, to contribute to better understand how to improve it. In short, we wonder what makes Gephi a black box; but as we will see, that research question comes with complications: what does it mean to be a black box, and where does the problem lie, materially?

In this article we focus on a specific network analysis and visualization tool called Gephi (see figure 1). Some academic authors associate Gephi to a black box, as a point of criticism (Lemerrier and Zalc, 2019; Wieringa *et al.*, 2019) or not (Bruns, 2013; Paßmann and Boersma, 2017). Other authors make similar arguments but about a wider class of instruments, such as big data visualization (Ruppert and Scheel, 2019), computational instruments (Masson, 2017), or quantitative visualization (Drucker, 2012; Drucker, 2014). These authors cite each other, suggesting a relatively unified argument; but that assumption does not stand examination. Indeed, their arguments are on two distinct fronts, echoing different understandings of the notion of black box: either inability to access the method embedded in the instrument, that we call *blackboxing by inscrutability*, or invisibilization of the technology as a mediation, that we call *blackboxing by embodiment*. The coexistence of these two versions has caused confusion within the critical literature about digital instruments, and clarifying the situation is a necessary step towards addressing Gephi's criticism.

We study Gephi's blackboxing problem from different angles: as narrated by researchers in their publications (literature review), as explained by Gephi users in interviews, and through participant observation in teaching situations. From these complementary perspectives we account for different ways users engage with Gephi, we retrace the material-semiotic web of their interactions, and we explore how "different realities are being woven into being in different practices" (Law, 2019). We observe that a specific subculture has emerged around Gephi, and that Gephi's powers of embodiment and inscrutability do not only root in its software, but also in the cultural artifacts of its user community (e.g., teaching material, online discussions). This leads us to conclude that the tool design might not be the right place where to intervene, and

prompts us to reconsider who or what has enacted Gephi's blackboxing, and how to care for the Gephi users' community.

The case of Gephi

The instrument this research focuses on is Gephi (Bastian *et al.*, 2009), an open-source software tool for network analysis and visualization. It operates via a number of embedded algorithms (e.g. Blondel *et al.*, 2008; Jacomy *et al.*, 2009) and a graphical user interface (see figure 1). In practice, it allows users to visualize a relational data set (a network) and execute different analytical tasks, such as computing statistics or filtering data. It has been referenced in more than five thousand papers¹ across a variety of disciplines, inside and outside the social science and humanities. Its popularity within the SSH has led it to be discussed from that perspective (Grandjean and Jacomy, 2019; Wieringa *et al.*, 2019), in addition to a more general discussion, in computer science and information design, on the kind of visualizations it provides (Gibson *et al.*, 2012; Krzywinski *et al.*, 2012; Dunne *et al.*, 2015; Munzner, 2015).

[INSERT_FIGURE_1]

Gephi's most popular node placement algorithm, *Force Atlas 2*, has become an increasingly important part of contemporary network analysis. Algorithms of this type (i.e., "force-driven") work by simulating "a physical system in order to spatialize a network. Nodes repulse each other like charged particles, while edges attract their nodes, like springs. These forces create a movement that converges to a balanced state" (Jacomy *et al.*, 2009). We must emphasize, however, that network visualizations created this way are "not deterministic, and the coordinates of each point do not reflect any specific variable" (Jacomy *et al.*, 2009). While popular, force-driven placement algorithms have also been extensively criticized (for an overview, see Gibson *et al.*, 2012; Dunne *et al.*, 2015), notably for their "inherent unpredictability, inconsistency and lack of perceptual uniformity" (Krzywinski *et al.*, 2012) and "brittleness" (Munzner, 2015). Additionally, due to the many parameters and user interactions that need to be taken into consideration while using Gephi, it is notoriously complicated to document, communicate and reproduce how the network visualization came into being (Wieringa *et al.*, 2019).

Analytical perspectives

Our conceptual framework has been chosen to accommodate our two starting points: the criticism of data visualization tools in the SSH literature, and the empirical engagement with the Gephi user community. We aimed at remaining coherent with the voiced criticism while supporting an empirical engagement with Gephi as a device, its interactions with its users, and its community over internet. This led us to mobilize concepts such as epistemic culture (Knorr Cetina 1999) and technical mediation (Simondon, 1958). We will present this criticism first, and second our main concepts, as the latter have been picked to address the former.

The criticism of data visualization tools in the SSH literature

Within the social sciences and humanities (SSH) some authors criticize data visualization and its tools on the ground that they hide the conditions of their production and/or the layers of mediation they involve. As we briefly retrace in this subsection, this argument roots in science and technology studies (STS) through Haraway, Latour and Law (see figure 2).

[INSERT_FIGURE_2]

When Evelyn Ruppert and Stephan Scheel (2019) account for the politics of method involving big data, they argue that the visualization becomes “not only a **vehicle** for the data, but first and foremost **for its claimed self-evidence**” (emphasis added). Here they draw on Johanna Drucker (2011), Donna Haraway (1988) (both summarized below), and on John Law’s idea that “practices enact realities” (2012). Similarly, Eef Masson (2017) writes that “computational research tools **are often assigned** such values as reliability and transparency” (emphasis added), drawing on Drucker (2012, 2014) and Rieder and Röhle (2012, 2017).

Bernhard Rieder and Theo Röhle (2017) write: “[t]he problem, again, comes from the fact that tools such as Gephi have made network analysis accessible to broad audiences that happily produce network diagrams without having acquired robust understanding of the concepts and techniques the software mobilizes. This more often than not **leads to a lack of awareness of the layers of mediation** network analysis implies and thus to limited or **essentialist readings** of the produced outputs that miss its artificial, analytical character” (emphasis added). This argument is a refined version of a point from their 2012 paper.

Johanna Drucker writes that maps, charts and graphs “are a kind of **intellectual Trojan horse**, a **vehicle through which assumptions** about what constitutes information **swarm with potent force**. ... they pass as **unquestioned** representations of ‘what is.’ This is the hallmark of realist models of knowledge and needs to be subjected to a radical critique to return the humanistic tenets of constructedness and interpretation to the fore” (emphasis added). This passage from her book *Graphesis* (2014) is also in her 2011 paper. In 2012 she also explores the same idea: “[g]raphs and charts reify statistical information. They give it **a look of certainty**. Only a naive viewer, unskilled and untrained in matters of statistics or critical thought, would accept an information visualization at **face value**” (emphasis added).

Finally, Donna Haraway writes in her influential paper *Situated Knowledges* (1988): “all seems not just mythically about the god trick of seeing everything from nowhere, but to have put the myth into ordinary practice. ... There is **no unmediated** photograph or passive camera obscura in scientific accounts of bodies and machines; there are only highly specific visual possibilities, each with a wonderfully detailed, active, partial way of organizing worlds” (emphasis added). Through “situated knowledges,” Haraway reclaims the necessity for all knowledge to be critically re-examinable, to state explicitly the epistemic ground where it stands.

Main concepts

Epistemic (sub-)culture. We make use of Knorr Cetina's concept of epistemic culture.

"Epistemic cultures are cultures that create and warrant knowledge" (Knorr Cetina 1999, p. 26). In her own words, the notion of epistemic culture "disrupts the idea of the epistemic unity of the sciences and suggests that the sciences are in fact differentiated into cultures of knowledge that are characteristic of scientific fields or research areas, each reflecting a diverse array of practices and preferences coexisting under the blanket notion of science" (Knorr Cetina and Reichmann, 2015, p. 873). In the case of Gephi, as the scale is noticeably smaller, we argue an epistemic *subculture*.

Laboratory. Still following Knorr Cetina (1999), we situate the SSH use of Gephi in its own laboratory. For Knorr Cetina, and for us, the laboratory is not necessarily a physical place, but wherever takes place the action of working with the inscriptions produced by inscription devices (see below), as opposed to studying objects in nature. The laboratory is "the locus of mechanisms and processes which can be taken to account for the success of science" (Knorr Cetina 1992, p. 116). Gephi's laboratory consists of digital and physical spaces where users share their work and seek help and advice, notably Twitter, a dedicated Facebook groupⁱⁱ, and blogsⁱⁱⁱ. Some official teaching material does exist (website^{iv}, wiki^v) but most of it has been authored by the community, including three books.^{vi} Gephi is also part of a few regular cursus, for instance the Digital Methods Initiative's Winter and Summer schools^{vii}.

Inscription device. Latour and Woolgar (1986) introduce the notion of inscription device as "any item of apparatus or particular configuration of such items which can transform a material substance into a figure or diagram which is directly usable" (p. 51). We analyze Gephi as a scientific instrument but more precisely as an inscription device.

Black box. As an inscription device, Gephi is said to be a black box. That concept is problematic because it has at least two distinct meanings. We will return to that point in the next subsection, after we have introduced a few necessary concepts.

(Technical) mediation. In line with the conceptual framework mobilized to criticize Gephi and data visualization, we draw on Gilbert Simondon's thinking of the technical object (1958), and more recently on Madeleine Akrich (1993) and Bruno Latour (1994). For Latour (1994), the technical mediation refers to the fact that technical objects "act, displace goals, and contribute to their redefinition." An important feature of that framework is its flat ontology: objects act on people and symmetrically people act on objects. The mediation allows us inquiring into Gephi's material-semiotic ability to reorganize the goals of its users, and vice-versa.

Embodiment. To analyze the material-semiotic effects of Gephi we draw on Maurice Merleau-Ponty's (1962) notion of "embodiment," that he applies specifically to mediational objects (see also Heidegger, 1962). As he argues, one stops perceiving the mediation as an object, and instead perceives *through it* as if it were part of one's body. His classic example is the blind man's stick, but one could also think of eyeglasses or the computer mouse. This

phenomenological approach is widespread in the field of Human-Computer Interaction (HCI) with concepts such as “embodied interaction” (Dourish, 2004) and as a foundation of Activity Theory (Kaptelinin and Nardi, 2012).

Critical technical practice. As participants to the Gephi development team and community, we also draw on the tradition of critical technical practice (Agre, 1997; see also van Es and Schäfer, 2017), also popular in the field of HCI, “in which technology development is not simply an end in itself, but also becomes a means to reflectively explore the assumptions and attitudes that underpin ideas about technology and humanity” (Dourish *et al.*, 2004, p. 1727).

Storyletting (and not storytelling) is a notion we propose further in this text to refer to the act of sharing or circulating an image (a network visualization) without providing the context necessary to its interpretation. One is then *letting* the image speak for itself (hence the name), which suggests its “self-evidence” (Ruppert and Scheel, 2019) and fails to consist of “situated knowledges” (Haraway, 1988).

Two versions of blackboxing: by embodiment / by inscrutability

We find two distinct versions of the black box in the literature. Here we present their differences and give them distinct names for clarity. We call the first “by embodiment” because it grounds into phenomenology of perception, and the second “by inscrutability” because it points to a scrutiny problem.

Latour argues that blackboxing is “the way scientific and technical work is made invisible by its own success. When a machine runs efficiently, when a matter of fact is settled, one need focus only on its inputs and outputs and not on its internal complexity. Thus, paradoxically, the more science and technology succeed, the more opaque and obscure they become.” (1999, p. 304). It corresponds to what Paßmann and Boersma (2017) call *practical transparency*, when the user “does not have to be bothered with the underlying machinery or software” as Star *et al.* put it (2003, p. 242). We call this version *blackboxing by embodiment*, drawing on Maurice Merleau-Ponty’s (1962) notion of “embodiment” (see also Heidegger, 1962), typical of mediational objects. Embodiment is not a problem by itself; on the contrary, it is required for a mediation to function (Latour, 1994). Yet by invisibilizing the mediational object, it allows it to escape scrutiny. It acts against what Haraway (1988) calls “situated knowledges,” where the knower’s relation to the known must be re-examinable. The invisibilization of the mediation is the main problem pointed to by blackboxing by embodiment.

The alternative, that we call *blackboxing by inscrutability*, addresses the impossibility to know the method, as in the colloquial expression “opening the black box.” This understanding is typically present in algorithm accountability and explainable AI (overview in Wieringa, 2020) where black boxes are defined as “opaque decision systems” (Guidotti *et al.*, 2018). Latour attributes this version to cyberneticians, who use it “whenever a piece of machinery or a set of commands is too complex ... That is, no matter how controversial their history, how complex their inner workings, how large the commercial or academic networks that hold them in place,

only their input and output count.” (2000, p. 681) Such black boxes are inscrutable by design. *Blackboxing by inscrutability* is one of the wider-known points of tool criticism in academia, especially in the social sciences. Rieder and Röhle characterize it as the “practical [im]possibility to access the most obvious layer of functional specification, a tool’s source code” (2017, p. 113), hindering “our ability to understand the method, to see how it works, which assumptions it is built on, to reproduce and criticise it” (2017, p. 112). This blackboxing opposes what Paßmann and Boersma (2017) call *formalized transparency*, the ability to access methodological information.

The coexistence of these distinct understandings of blackboxing has brought confusion to the debate. First problem: when a critique mentions blackboxing, we do not always know what it refers to, notably at the intersection of STS (where *blackboxing by embodiment* is prioritized) and Computer Science (where *blackboxing by inscrutability* prevails), and in HCI. Second problem: the opacity/transparency dichotomy often used to clarify the notion of blackboxing is confusing on its own. In the metaphor, transparency opposes blackness; yet counter-intuitively, transparency is a defining feature of *blackboxing by embodiment*. Therefore, demarcation attempts that rely on the notion of transparency sound confusing. For instance, Paßmann and Boersma (2017) demarcate *practical transparency* from *formalized transparency*. The former *aligns* with blackboxing by embodiment, while the latter *opposes* blackboxing by inscrutability. Their demarcation is productive on a conceptual level, yet it is awkward to articulate with blackboxing because the metaphors clash on an intuitive level. Third problem: these blackboxings call for different countermeasures. One must avoid addressing the wrong kind of blackboxing to improve Gephi. The distinction matters to tool makers and users. Indeed, as we will see, Gephi is mainly blackboxed *by embodiment*: its methodological ground is relatively accessible, yet problems arise from its mediational properties.

[INSERT_TABLE_1]

Now equipped with these distinct concepts of blackboxing, we can articulate our general argument. We contend that Gephi’s is a black box *by embodiment*. The layers of mediation it involves are made invisible by its own success, but that success partially depends on practices invented in the laboratory and institutionalized by its subculture as cultural artifacts. We argue that the apparent self-evidence of network visualizations partially roots in Gephi’s ability, as a mediation, to displace the goals of its users: by facilitating the production of images that can easily circulate but not easily be understood, it incentivizes storytelling (images are let free to tell their own story). These knowledges therefore lack the situatedness required to reexamine the interpretative process.

Methodology

This paper is a collaboration around the material collected by one of the authors, Emilija Jokubauskaite, more extensively presented in her master’s thesis (2018). It contrasts the voice of users with Mathieu Jacomy’s designer and developer perspective on Gephi, to unpack the different dimensions of blackboxing.

We engaged in participant observation in a dozen of Gephi workshops over the last decade, often as trainers, and sometimes as trainees. We conducted and inductively analyzed semi-structured interviews with 6 Gephi users of varying levels of experience, who were chosen due to having conducted work that utilizes it for academic purposes within media and communication studies. The interviews consisted of open-ended questions aimed at understanding how researchers communicate their usage of the tool and how they represent the method. Parts of each interview followed an elicitation interviewing method (Johnson *et al.*, 2002): the interviewees were asked to use Gephi to briefly interact with a data sample (either familiar or new to them) prior to the interview and take notes on their use of the tool. During the interview, the tools' graphical user interface was open, and the interviewees narrated their methodological steps of using it, followed by a discussion of the decision-making process, questions about starting to use the tool and overall experience, as well as reflections on Gephi use in the academic field in general and other topics.

To triangulate the information gathered from the interviews with the methodologies presented in the documentation, we also analyzed 17 published academic papers. The papers were chosen through a systematic sampling, including recent papers from 2009 up to 2018, filtering 50-100 papers published each year for those of most relevance to the area of social media research in media and communication studies.

[INSERT_TABLE_2]

Gephi's blackboxing by embodiment

Here we present Gephi's blackboxing problem as painted in our empirical material. Our analysis diagnoses an issue of blackboxing by embodiment: researchers have issues accounting for what Gephi performs, not because they fail to scrutinize it, but because they do not understand, or are not aware of, the layers of mediation involved.

Scholars do not fully acknowledge Gephi's methodological dimension

One of the interviewed researchers reflected on the need to understand how the specific settings of Gephi function. In their eyes, the level of understanding should be at least "to the extent that they can explain what is happening ..., because it is also a part of your job as a researcher to actually communicate this stuff". They add that on the other hand, "if you know a lot about this stuff, you get into a potential problem that you communicate too much about the method ..., whereas what you want to say is actually something about the data."

Here we account for how academic publications frame the methodology where Gephi is involved. Our publication review shows that authors rarely make explicit which method they use to analyze networks. Some of these publications do not present the settings employed to apply the algorithms of Gephi [e.g., 8, 12, 13, 14, 15]; as the vignette above shows, the authors may

just consider it out of scope. In other papers, the explanation of methodological decisions draws on the settings of the tool [e.g., 16, 18] but their *effects* on the analysis are not discussed.

While some authors do elaborate on the method [e.g., 19, 20], their rationale does not directly relate to Gephi. An illustration of that situation can be found in the publication of Barnett and Benefield [19]. In their article, they expose and discuss the measures of in-degree centrality and betweenness centrality. Gephi is providing these statistical metrics but is just presented to illustrate the findings. Similarly, Bonini and Celli *et al.*'s articles [9, 15] present the method of network analysis in the methodology section, but do not mention Gephi until a visualization is presented later in the text, somewhat separating the tool from the method. Additionally, as Bruns (2013) writes, the referees for articles do not usually request such methodological information: "treatment of tools such as Gephi as black boxes whose interior operations can be ignored is not limited to scholarly authors alone", referring here to what we call blackboxing by embodiment.

To better understand the rationale behind the methodological decisions with Gephi, we turn to the interview data. In them, we observe, the users explain their methodological decisions from the viewpoint of the tool, for example stating that they used Force Atlas 2, a layout algorithm, "because it is actually optimized for Gephi" [6]; "because basically it was the one that worked best" [5] or because it "is the most convenient ... to use" [1]. These examples suggest that the Gephi users may be paying more attention to finding the settings that allow extracting the "correct" or "best" network visualization, rather than rationalizing their practice based on how the method operates. For many scholars, Gephi is put to use before they fully grasp its methodological effects, which also points to blackboxing by embodiment.

Gephi is rarely perceived as inscrutable by its users

The interviewees reflect on the level of scrutiny offered by the instrument. For example, some of them state that "it gives you a very direct feedback" [1] or that they don't necessarily know what it does, but it "gives a reasoning behind it after you run it" [2]. They also report on specifically trying out some functionalities of the tool to gain an understanding of them [1]. Moreover, researchers say that "you always have sort of a clear methodological thing, that you can reference and people can look at it after" [2]. They appreciate that the tool provides them with references to academic literature sources, that contextualize the various metrics and functions one can apply with the tool. Gephi is not perceived as an inscrutable black box; but of course, inscrutability is not the only way to be blackboxed.

Users do not necessarily understand how Gephi works

Discussing their use of Gephi, one of the interviewees reflects about using one of the different layout algorithms by saying that it "is sort of you use [it] and then try your best, which feels sort of unmethodological or unscholarly, because I feel like the advice I have gotten is just 'try them all and see what fits', which is just like rolling a dice. Feels sort of a bit random." Another interviewee reflected on how she learned to use Force Atlas 2. She said that another student explained it to her via text messaging by telling her to "put gravity on 10 and enable

the ‘prevent overlap’ setting”, however they were not able to provide rationale for it and when the interviewee asked them why, they replied “nevermind, just do it”.

Some users reflect on often using the tool without making the decisions consciously [1, 2, 4] and not being aware of what and why they should question in the process of implementing the tool-use in their research [2, 3]. They articulate that the instrument is “complicated and there’s a lot of things that you have no idea what is happening and what is going on” [3]. Moreover, one of the interviewees also argued that the interface does not necessarily help the user understand what is happening method-wise [4]. These observations can be complemented by Bruns’ (2013) argument that even though there is quite a lot of information about the technical aspects behind various algorithms in Gephi, understanding them to a needed level is often out of the skillset of humanities and social sciences scholars. Not everyone has the knowledge necessary to grasp what Gephi performs. Likewise, “beginners” are of particular concern for Lemerrier and Zalc, in supplemental material^{viii} to their book *Quantitative Methods in the Humanities: An Introduction* (2019). They warn against software that “presents some analysis tools like ‘black boxes’” and comment that “contrary to many, we are no fans of Gephi, for this type of reasons; ... we do not find it well-suited to beginners.” The mediation performed by Gephi is not visible or understandable to all users at all moments.

Rieder and Röhle (2017, p. 118) also highlight the interplay between the tool and its public: “The problem ... comes from the fact that tools such as Gephi have made network analysis accessible to broad audiences that happily produce network diagrams without having acquired robust understanding of the concepts and techniques the software mobilizes. This more often than not leads to a lack of awareness of the layers of mediation network analysis implies.” For them, the problem is at the same time about visibility (“awareness”) and “understanding” of the method. But as we have seen, Gephi is not perceived as inscrutable; most users are aware that it involves algorithms described in academic papers, and gradually improve their understanding of the “layers of mediation” involved (Rieder and Röhle, 2017). Interviewed researchers say, for example, that “through the time of working with the actual tool you ... get to know what it ... represents. So, it gives you more insight into what it is ... that you are working with” [1]. Moreover, several of the interviewees express a wish to do more related research to learn more about the tool [2, 4]. Gephi is perceived as an entry point to academic literature. However, this knowledge is not required to use it.

This situation is in line with Gephi’s original design intent (documented in e.g., Jacomy *et al.*, 2014). It was initially developed for the needs of the *e-Diasporas Atlas* project (Diminescu *et al.*, 2012), a collaborative exploration of the web of diasporas (on the apparatus, see Diminescu *et al.*, 2011; on the project, see Diminescu, 2012). The participating scholars were specialists of migrations inexperienced in network analysis. Gephi was largely designed for the persona of the *scholar-going-digital*: someone trained to the social sciences in an analog setting, and willing to extend their skills to digital methods. It was designed to be functional for a user aware of the existence of an underlying methodology, but without the ability to understand it. The design intent assumed that users would be willing to acquire the necessary knowledge, thus it was not rendered necessary to using the instrument in practice.

Our empirical material rules out a major problem of blackboxing by inscrutability, and leaves us with the logical conclusion, substantiated by several observations, that Gephi suffers from blackboxing by embodiment. Yet it is also clear that the visibility and understandability of Gephi's layers of mediation also depend on who is using it, and in which context. Which begs the question: to whom, and in which situation, is Gephi a black box?

How the black box is materialized in the artifacts of Gephi's epistemic subculture

Enacting self-evidence through "storyletting"

In a research seminar at the IT University of Copenhagen, a few PhD students have the occasion to present their work in 10 minutes and one of them has a slide with a network visualization. It shows a colorful screenshot of a decently sized Gephi network on a black background. It is hard to read because many labels overlap but it seems to have a clustered structure. The student just comments: "We also do these kinds of fancy visualizations" and immediately moves on to the next slide without providing any explanation of what we are supposed to see, or how it was useful to them.

One does not have to find a network visualization useful, but then why showcase it without providing the context necessary to its interpretation? We gave this situation a name: *storyletting*, the act of circulating an image without an explanation, letting it tell its own story. Indeed, a Gephi screenshot is prone to tell stories such as: my data is complex and massive; my research is computational and quantitative; my skillset includes advanced analysis techniques... Visualization plays a clear role in the politics of method (Ruppert and Scheel, 2019; Savage and Burrows, 2007). Naming the practice allows us to refer to it without assuming malice, or even intent; Gephi images may circulate context-free for many reasons. As we have seen, most academic papers featuring Gephi do not provide any discussion or justification of the methodological decisions related to it [e.g., 8, 12, 13, 14, 15, 17, 21] (see also Bruns, 2013). Several researchers presented Gephi as "only one tool they used" [4; see also 13, 14] or "just a data visualization" tool [3; also in 10, 12, 14, 15, 17, 19, 20, 21, 22, 24], and some interviewees [2, 4] reflect that scrutinizing Gephi is not necessary because network analysis is deemed peripheral to their research. Yet regardless of the justification and intent, the act of storyletting signifies that the visualization does not require explanation, that it is "self-evident" (Ruppert and Scheel, 2019). As critiques have argued, this absence of context prevents us from re-examining our relation, as knowing subjects, to the known object (the network), failing to situate the knowledges we produce and/or circulate (Haraway, 1988).

Brooke Foucault Welles and Isabel Meirelles (2014) account for the circulation of the network visualization from Lada Adamic and Natalie Glance's (2005) influential paper *Divided they Blog*. Adamic and Glance had situated their claims, yet other authors stripped the visualization of its context when they circulated it, transforming it into a (claimed!) self-evident illustration of the

polarization of US politics. As Latour (1986) argues, the materiality of visualization may matter more than its semiotics, notably through its ability to circulate while maintaining its consistency (what he calls “immutable mobiles”). Situatedness is about more than the production process of a visualization. To understand Gephi’s blackboxing by embodiment, we need to widen the focus and look at the practices with it, in the context of its “laboratory” (Knorr Cetina, 1999).

Shortcut practices within Gephi’s laboratory

When asked about what situation they were in when they were first learning to use Gephi, one of the interviewees reflects on participating in the fast-paced data sprints of the Digital Methods Initiative (presented in Berry *et al.*, 2015): “Well, usually it’s in our Winter or Summer Schools. There is a project and suddenly there is a network, I need to open it and do things. And then it’s more like, OK, learn what you need to do [in] this specific project to solve this kind of issue you have. So, it’s more like learn how to do a specific thing with the software, related to the project and less about how to use all the specific features and all the settings.” As we observed many times, in the heat of a data sprint the use of scientific instruments gets less reflexive and more pragmatic.

Gephi is often used in a time-restricted environment. Interviewees explain that they learned how to use the tool during data sprints [1], short introductory classes [4] or project-based university courses [4, 6] and other project-to-project cases. That is, while the researchers often arrive at the use of Gephi without former knowledge of network analysis, their educational path from the beginning can be seen as largely focused on producing findings in short-time projects. They often learn from step-by-step tutorials and “speed courses” [1] or “best practices” that increase productivity when using the tool. Following that, the researchers report being encouraged to “just use it” [3] or instructed by their peers to take some methodological steps without providing further information on why and what the implications might be [1, 4]. Consequently, some researchers refer to their use of Gephi as *automatic*, for instance “if I want a specific end result, I know how to get there, but it’s very automatic, I don’t really understand why this is what I do, it’s just that I go through the moves, because that’s what you have to do.”

This situation is typical of a know-how teaching, where reproducing the actions of the trainer or tutorial is the first step of the learning process, but there is more to it. This researcher accepts to trade methodological awareness for efficiency. Shortcuts are problematic but ordinary, and Gephi is relatively aligned with such utilitarian uses. It can be mobilized purposefully to produce a visualization, and once this goal is met, it will offer no resistance to the repurposing of that image, for instance as a dubious marketing asset in some slides. This situation matches Rieder and Röhle’s (2017) observation of “broad audiences that happily produce network diagrams without having acquired robust understanding of the concepts and techniques the software mobilizes.” With a nuance, however: it does not only point at the users’ lack of “digital bildung” (Rieder and Röhle, 2017) but also at Gephi’s lack of resistance to the most utilitarian practices.

Gephi’s design only resists one kind of blackboxing

Gephi’s design addressed blackboxing, but in the sense of inscrutability, not embodiment. Gephi’s proxy paper (Bastian *et al.*, 2009) is too short to discuss its design, but another paper

developed by the same team (Jacomy *et al.*, 2014) offers some elements. The authors state: “Our goal was to provide some network analysis methods to social scientists, that would not require learning graph theory.” This commoditization goal implies blackboxing by embodiment: concealing complexity to offer more practicality. When the authors^{ix} explain why their algorithm displays its process in real time they emphasize the importance of embodiment, but when they criticize black boxes, they limit the notion to inscrutability: “Visualizing the ‘live’ spatialization is a key feature of Gephi. It provides a very intuitive understanding of the layout process and its settings. ... Social scientists cannot use black boxes, because any processing has to be evaluated in the perspective of the methodology.” Gephi’s design aims at resisting blackboxing by inscrutability but does not address blackboxing by embodiment.

Should we then conclude that Gephi’s blackboxing problem roots in its design? The situation is not so simple, at least because its design implementation differs from the stated intent. First, many aspects of the instrument are unpolished if not unfinished, and most users find it frustrating at times. Second, Gephi’s interface occasionally exercises intentional resistance. As an emblematic illustration, let us look at how the *Force Atlas 2* algorithm (Jacomy *et al.*, 2014), commonly construed as Gephi’s default layout, is contextualized in the software. It features as a mere item in a list that you must unfold to see; it is not the first item of the list; it is not selected by default; it is not highlighted (see figure 3). The rationale behind this design choice was to resist blackboxing by inscrutability: forcing users to acknowledge the existence of multiple algorithmic alternatives, none of which is intrinsically better. But notice that by imposing friction (more clicks and choices) it also resists blackboxing by embodiment. What makes it the default algorithm is not inscribed in the software, but in cultural artifacts: the *Force Atlas 2* paper presents it as Gephi’s “default layout algorithm”; teaching material presents it as the go-to solution; its popularity^x in other papers reinforces its status of standard within the Gephi subculture. In this case, the practices have counteracted the design and institutionalized a default option despite its absence from the material semiotics of the instrument.

[INSERT_FIGURE_3]

Latour (1994) highlights the symmetry of technical mediation: the tool shifts the goals of its users and, symmetrically, users shift the goals of the tool. We have observed both sides of it but in different situations. First, Gephi yields to utilitarian pressure in time-restricted situations where users are the most willing to trade methodological awareness for efficiency. In this case, users shift Gephi’s goals, cutting through intentionally featured resistances with popular and/or heuristic recipes. Second, by facilitating the production of images (outputs or screenshots), it produces an “epistemic surplus” (Rieder and Röhle, 2017) susceptible to get repurposed through *storyletting*. Here it is Gephi that shifts the goals of users.

Conclusion

We have seen that SSH scholars criticize data visualization in general, and network visualization in particular, including Gephi, for being a “black box.” This argument blames network maps for appearing self-evident, which leads to forgetting that they come with

methodological commitments. By analyzing empirical material from participant observation, interviews, and a publication review, we endeavored to diagnose Gephi's blackboxing problem. We clarified that blackboxing may refer to two different things, and we advocated for dropping the confusing opacity/transparency dichotomy. We called *blackboxing by inscrutability* the problem of limited or nonexistent access to relevant information about the method performed, and *blackboxing by embodiment* the problem of invisibilization of the mediation. Each problem calling for a different solution, we assessed each possibility. We observed that Gephi is not perceived as inscrutable, and that many scholars put it to use before they fully grasped its methodological effects. This led us to conclude that Gephi is primarily a black box *by embodiment*. We observed that some users circulated network images produced by Gephi without providing any interpretative context, a situation we refer to as *storyletting*. Letting such images tell their own narrative, allows them to get received as self-evident by the audience, as if they did not need to be explained, which contributes to invisibilizing the layers of mediation involved. We also observed that Gephi is often used in time-constrained situations where users are willing to trade methodological awareness for efficiency. On this ground, we argued that Gephi's blackboxing happens for two reasons: because of utilitarian practices in its epistemic subculture, and because it facilitates storyletting by generating an epistemic surplus: network visualizations that cannot be understood outside of the laboratory.

Ultimately, what makes Gephi a black box? Gephi's design philosophy explicitly acknowledged blackboxing as an issue, but it conceived it as a matter of opacity (by inscrutability) to fight with transparency. It therefore missed that transparency could also lead to blackboxing (by embodiment). The same approach is common in algorithm accountability, but as the case of Gephi shows, it fails to address a criticism about the concept of technical mediation, rooted in STS and phenomenology of perception. Indeed, the context of the laboratory is necessary to understand how Gephi gets blackboxed. The time-constrained environment contributes to it; the cultural artifacts of the community, like tutorials promoting a utilitarian use, contribute to it; and of course, Gephi's design choices contribute to it. Users shape their instrument through interactions and inscriptions that extend far beyond its source code. Their repurposing of the tool can gradually get institutionalized through the artifacts of its epistemic culture, and contribute to reframing it. This makes it clear that it takes more than software design to succeed in unblackboxing Gephi. Intervening through critical technical practice requires taking into account the multiplicity of the situations where blackboxing arises, which leads to two major questions. First, how can instrument design contribute or be adapted to caring for a largely self-determined epistemic subculture? Second, how to resolve the inherent tension between usability and embodiment? As Latour (1994) reminds us, blackboxing is necessary for a mediation to succeed, but it is also reversible. The black box can be opened and embodied, although not at the same time. That point may be worth reminding to SSH scholars: through the place they give to their apparatus in the knowledge they produce, they too can engage in critical technical practice, and contribute to unblackboxing their digital instruments.

References

- Adamic, L. A., and Glance, N. (2005) The political blogosphere and the 2004 US election: divided they blog, *Proceedings of the 3rd international workshop on Link discovery*, pp. 36-43.
- Agre, P. E. (1997) Toward a Critical Technical Practice: Lessons Learned in Trying to Reform AI, In G. C. Bowker, et al. (eds), *Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide* (Mahwah, NJ: Lawrence Erlbaum Associates), pp. 131-57.
- Akrich, M. (1993) *Les formes de la médiation technique*, Réseaux, 60, pp. 87-98.
- Barabási, A. L. (2016) *Network science*. (Cambridge: Cambridge university press).
- Barnes, J. A. and Harary, F. (1983) *Graph theory in network analysis*, Social Networks, 5, pp. 235-244
- Barnett, G. A. and Benefield, G. A. (2017) *Predicting international Facebook ties through cultural homophily and other factors*, New Media & Society, 19(2), 217-239.
- Bastian, M., Heymann, S. and Jacomy, M. (2009) *Gephi: an open source software for exploring and manipulating networks*, Third international AAAI conference on weblogs and social media.
- Berry, D. M. (2011) *The computational turn: Thinking about the digital humanities*. Culture Machine, 12.
- Berry, D. M., Borra, E., Helmond, A., Plantin, J.-C. and Rettberg, J. W. (2015) *The data sprint approach: exploring the field of Digital Humanities through Amazon's application programming interface*, Digital Humanities Quarterly, 9(4).
- Blondel, V. D., Guillaume, J. L., Lambiotte, R. and Lefebvre, E. (2008) *Fast unfolding of communities in large networks*, Journal of statistical mechanics: theory and experiment, 10.
- Bonini, T., Caliendo, A. and Massarelli, A. (2016) *Understanding the value of networked publics in radio: employing digital methods and social network analysis to understand the Twitter publics of two Italian national radio stations*, Information, Communication & Society, 19(1), pp. 40-58.
- Borgatti, S. P., Everett, M. G. and Freeman, L. C. (2002) *Ucinet for Windows: Software for social network analysis* (Harvard, MA: Analytic technologies)
- Borgatti, S. P., Mehra, A., Brass, D. J. and Labianca, G. (2009) *Network analysis in the social sciences*, Science, 323(5916), pp. 892-895.
- Börner, K., Sanyal, S. and Vespignani, A. (2007) *Network science*, Annual review of information science and technology, 41(1), pp. 537-607.
- Boyd, D. and Crawford, K. (2012) *Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon*, Information Communication and Society, 15(5), pp. 662-679.
<https://doi.org/10.1080/1369118X.2012.678878>
- Bruns, A. (2013) *Faster than the speed of print: Reconciling 'big data' social media analysis and academic scholarship*, First Monday, 18(10). <http://firstmonday.org/ojs/index.php/fm/article/view/4879>
- Bruns, A., Moon, B., Münch, F. and Sadkowsky, T. (2017) *The Australian Twittersphere in 2016: Mapping the follower/followee network*. Social Media + Society, 3(4), 2056305117748162.

- Butts, C. T. (2009) *Revisiting the foundations of network analysis*. Science, 325(5939), pp. 414-416.
- Callon, M. (2002) Writing and (re) writing devices as tools for managing complexity. *Complexities: social studies of knowledge practices*, pp.191-217.
- Celli, F. and Rossi, L. (2012) *The role of emotional stability in Twitter conversations.*, Proceedings of the workshop on semantic analysis in social media, pp. 10-17. Association for Computational Linguistics.
- Chiesi, A. M. (2015) *Network Analysis*, International Encyclopedia of the Social & Behavioral Sciences, pp. 518–523. <https://doi.org/10.1016/B978-0-08-097086-8.73055-8>
- Diminescu, D. (2012) *Diasporas on the Web*, Social science information, 51(4).
- Diminescu, D., Bourgeois, M., Renault, M. and Jacomy, M. (2011) *Digital Diasporas Atlas Exploration and Cartography of Diasporas in Digital Networks*, Fifth International AAAI Conference on Weblogs and Social Media, pp. 657–658.
- Diminescu, D., Asal, H., Balalovska, K., Ben-David, A., Berthomière, W., Bruslé, T., Carsignol, A., Gangloff, S., Graziano, T., Knaute, D., Kumar, P., Le Bayon, S., Leclerc, É., Ma Mung Kuang, E., Marchandise, S., Mazzucchelli, F., Morgunova (Petrunko), O., Moua, M. H., Percot, M., Renault, M., Reyhan, D., Scioldo Zürcher, Y., Severo, M., Therwath, I., Varrel, A., Venier, P., Watin, M., Wolff, E. and Zuolo, E. (2012) *e-Diasporas Atlas*. Available at <http://www.e-diasporas.fr> (accessed 18 August 2020).
- Dourish, P. (2004) *Where the action is: the foundations of embodied interaction*. MIT press.
- Dourish, P., Finlay, J., Sengers, P. and Wright, P. (2004) *Reflective HCI: Towards a critical technical practice*, CHI'04 extended abstracts on Human factors in computing systems (pp. 1727-1728).
- Dunne, C., Ross, S. I., Shneiderman, B. and Martino, M. (2015) *Readability metric feedback for aiding node-link visualization designers*, IBM Journal of Research and Development, 59(2/3), pp. 11–14. <https://doi.org/10.1147/JRD.2015.2411412>
- Drucker, J. (2011) *Humanities approaches to graphical display*, Digital Humanities Quarterly, 5(1), pp. 1-21.
- Drucker, J. (2012) Humanistic Theory and Digital Scholarship, In: Matthew K. Gold (Ed.) *Debates in the Digital Humanities* (Minneapolis: University of Minnesota Press), pp. 85-95.
- Drucker, J. (2013) *Performative Materiality and Theoretical Approaches to Interface*, DHQ: Digital Humanities Quarterly, 7(1).
- Drucker, J. (2014) *Graphesis: Visual Forms of Knowledge Production* (Cambridge, MA: Harvard University Press).
- Erdős, P. and Rényi, A. (1960) *On the evolution of random graphs*. Publ. Math. Inst. Hung. Acad. Sci, 5(1), pp. 17-60.
- Erikson, E. (2013) *Formalist and relationalist theory in social network analysis*, Sociological Theory, 31(3), pp. 219–242. <https://doi.org/10.1177/0735275113501998>
- Foucault Welles, B. and Meirelles, I. (2014) *Visualizing Computational Social Science: The Multiple Lives of a Complex Image*, Science Communication, pp. 1–25. <https://doi.org/10.1177/1075547014556540>
- Freeman, L. (2004) *The development of social network analysis*, A Study in the Sociology of Science, 1, 687.
- Gibson, H., Faith, J. and Vickers, P. (2012) *A survey of two-dimensional graph layout techniques for information visualisation*, Information Visualization, 12(3–4), pp. 324–357. <https://doi.org/10.1177/1473871612455749>

Grandjean, M. and Jacomy, M. (2019) *Translating Networks: Assessing Correspondence Between Network Visualisation and Analytics*, Digital Humanities Conference. Available at <https://halshs.archives-ouvertes.fr/halshs-02179024> (accessed 2020-09-17)

Grundberg, M. D. and Lindgren, S. (2015) *Translocal frame extensions in a networked protest: Situating the #IdleNoMore hashtag*, IC Revista Científica de Información y Comunicación, (11).

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F. and Pedreschi, D. (2018) *A survey of methods for explaining black box models*. ACM computing surveys (CSUR), 51(5), pp.1-42.

Hagen, L., Keller, T., Neely, S., DePaula, N. and Robert-Cooperman, C. (2018) *Crisis communications in the age of social media: A network analysis of Zika-related tweets*, Social Science Computer Review, 36(5), pp. 523-541.

Haraway, D. (1988) *Situated knowledges: The science question in feminism and the privilege of partial perspective*, Feminist studies, 14(3), pp. 575-599.

Heidegger, M. (1962) *Being and time*. 1927. Trans. John Macquarrie and Edward Robinson. New York: Harper.

Herdağdelen, A., State, B., Adamic, L. and Mason, W. (2016) *The social ties of immigrant communities in the United States*, Proceedings of the 8th ACM Conference on Web Science, pp. 78-84.

Hidalgo, C. A. (2016) *Disconnected, fragmented, or united? a trans-disciplinary review of network science*, Applied Network Science, 1(1), pp. 1–19. <https://doi.org/10.1007/s41109-016-0010-3>

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. PloS one, 9(6).

Johnson, J. C. and Weller, S. C. (2002) Elicitation techniques for interviewing, In: Gubrium J. F. and Holstein J. A. (Eds.) *Handbook of interview research: Context and method* (Thousands Oaks, CA: SAGE Publications), pp. 491-514.

Jokubauskaite, E. (2018) *Gephi and its Context: Three Oppositions in The Epistemic Culture of Network Analysis And Visualisation* (Amsterdam: University of Amsterdam).

Kaptelinin, V. and Nardi, B., 2012. Activity theory in HCI: Fundamentals and reflections. *Synthesis Lectures Human-Centered Informatics*, 5(1), pp.1-105.

Kelty, C. M. (2012). Geeks and recursive publics: How the internet and free software make things public. In *Beyond Habermas: Democracy, Knowledge, and the Public Sphere* (pp. 99–115). Recursive Publics.

Knoke, D. and Yang, S. (2019) *Social network analysis* (Thousands Oaks, CA: SAGE Publications).

Knorr Cetina, K. D. (1992) The couch, the cathedral, and the laboratory: On the relationship between experiment and laboratory in science.

Knorr Cetina, K. D. (1999) *Epistemic cultures: How the sciences make knowledge* (Cambridge, MA: Harvard University Press).

Knorr Cetina, K. and Reichmann, W. (2015) *Epistemic Cultures*, International Encyclopedia of the Social & Behavioral Sciences: Second Edition (Amsterdam: Elsevier), pp. 873–880. <https://doi.org/10.1016/B978-0-08-097086-8.10454-4>

Kranzberg, M. (1986) Technology and History: "Kranzberg's Laws". *Technology and culture*, 27(3), pp.544-560.

- Krzywinski, M., Birol, I., Jones, S. J. and Marra, M. A. (2012) *Hive plots—rational approach to visualizing networks*, *Briefings in bioinformatics*, 13(5), pp. 627-644.
- Larsson, A. O. (2013) *Tweeting the viewer—use of Twitter in a talk show context*, *Journal of broadcasting & electronic media*, 57(2), pp. 135-152.
- Larsson, A. O. and Moe, H. (2014) *Triumph of the underdogs? Comparing Twitter use by political actors during two Norwegian election campaigns*. *Sage Open*, 4(4), 2158244014559015.
- Latour, B. (1986) *Visualisation and Cognition: Drawing Things Together*. In: Kucklick, H. (Ed) *Knowledge and Society Studies in the Sociology of Culture Past and Present* vol. 6 (Greenwich, CT: Jai Press), pp. 1-40.
- Latour, B. (1994) *On Technical Mediation*, *Common Knowledge*, 3(2), pp. 29-64.
- Latour, B. (1999) *Pandora's Hope: Essays on the Reality of Science Studies* (Cambridge, MA: Harvard University Press).
- Latour, B. (2000) *Opening Pandora's Black Box*, In: Dawson, P., McLoughlin, I., and Preece, D. (Eds.) *Technology, Organizations and Innovation: Critical Perspectives on Business and Management* (Abingdon: Routledge), pp. 679–698.
- Law, J. (2011) *Collateral realities*. In: Dominguez Rubio, F. and Baert, P. (Eds.) *The Politics of Knowledge* (London: Routledge), pp. 156–178.
- Law, J. (2019). *Material Semiotics. Heterogeneities*, 1–19.
www.heterogeneities.net/publications/Law2019MaterialSemiotics.pdf (accessed 2020-04-20)
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A. L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D. and Van Alstyne, M. (2009) *Computational social science*, *Science*, 323(5915), pp. 721-723.
- Lemerrier, C. and Zalc, C. (2019) *Quantitative Methods in the Humanities: An Introduction*, (Charlottesville: University of Virginia Press).
- Lévy-Strauss, C. (1973) *Anthropologie structurale deux* (Paris: Plon), pp. 33-34.
- Lindgren, S. (2013) *The potential and limitations of Twitter activism: Mapping the 2011 Libyan Uprising*, *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society*, 11(1), pp. 207-220.
- Magnani, M., Montesi, D., and Rossi, L. (2010) *Friendfeed breaking news: Death of a public figure*, 2010 IEEE Second International Conference on Social Computing, pp. 528-533. IEEE.
- Manovich, L. (2013) *Software Takes Command* (New York: Bloomsbury Academic).
- Marres, N. (2012) *On some uses and abuses of topology in the social analysis of technology (or the problem with smart meters)*, *Theory, Culture & Society*, 29(4-5), pp. 288-310.
- Marres, N. (2017). *Digital Sociology* (Cambridge: Polity Press).
- Masson, E. (2017) *Humanistic Data Research: An Encounter between Epistemic Traditions*. In: M. T. Schäfer & K. van Es (Eds.) *The Datafied Society: Studying Culture Through Data* (Amsterdam: Amsterdam University Press), pp. 109–124.

- Merleau-Ponty, M. (1962) [1945]. *Phenomenology of Perception*. transl. Colin Smith (New York: Humanities Press).
- Metaxas, P. and Mustafaraj, E. (2013) *The rise and the fall of a citizen reporter*, Proceedings of the 5th Annual ACM Web Science Conference, pp. 248-257.
- Meyer, E. T. and Schroeder, R. (2015) *Knowledge Machines: Digital Transformations of the Sciences and Humanities* (Cambridge: MIT Press).
- Milgram, S. (1967) *The small world problem*, Psychology today, 2(1), pp. 60-67.
- Moreno, J. L. (1934) *Who Shall Survive?* (Washington, DC: Nervous and Mental Disease Publishing Company).
- Munzner, T. (2015) *Visualization Analysis and Design* (Boca Raton: CRC Press).
- Newman, M. (2018) *Networks* (Oxford: Oxford university press).
- O'Callaghan, D., Greene, D., Conway, M., Carthy, J. and Cunningham, P. (2013) *Uncovering the wider structure of extreme right communities spanning popular online networks*, Proceedings of the 5th Annual ACM Web Science Conference, pp. 276-285.
- Paßmann, J. and Boersma, A. (2017) Unknowing Algorithms: On Transparency of Unopenable Black Boxes, In: Schäfer, M. T. and van Es, K. (Eds.) *The Datafied Society. Studying Culture through Data* (Amsterdam: Amsterdam University Press). <https://doi.org/10.5117/9789462981362>
- Rieder, B. (2013) *Studying Facebook via Data Extraction: The Netvizz Application*, Proceedings of the 5th Annual ACM Web Science Conference (WebSci'13). Pp. 346–355. <https://doi.org/10.1145/2464464.2464475>
- Rieder, B. and Röhle, T. (2012) Digital Methods: Five Challenges. In: Berry, D. M. (Ed.) *Understanding Digital Humanities* (New York: Palgrave Macmillan), pp. 67-8.
- Rieder, B. and Röhle, T. (2017) Digital methods: From challenges to Bildung, In: Schäfer, M. T. and van Es, K. (Eds.) *The Datafied Society. Studying Culture through Data* (Amsterdam: Amsterdam University Press), pp. 109-124. <https://doi.org/10.5117/9789462981362>
- Rogers, R. (2013) *Digital methods* (Cambridge: MIT Press).
- Ruppert, E., Law, J. and Savage, M. (2013) *Reassembling social science methods: The challenge of digital devices*, Theory, culture & society, 30(4), pp.22-46.
- Ruppert, E. and Scheel, S. (2019) *The Politics of Method: Taming the New, Making Data Official*, International Political Sociology, 13(3), pp. 233-252.
- Savage, M. and Burrows, R. (2007). *The coming crisis of empirical sociology*, Sociology, 41(5), pp. 885–899.
- Simondon, G. (1958) *Du mode d'existence des objets techniques* (Paris: Aubier).
- Star, S. L., Bowker, G. C. and Neumann, L. J. (2003) Transparency beyond the Individual Level of Scale: Convergence between Information Artifacts and Communities of Practice, In: Peterson-Kemp, A., Van House, N. A. and Battenfield, B. P. (Eds.) *Digital Library Use* (Cambridge: MIT Press), pp. 241–269. <https://doi.org/10.7551/mitpress/2424.003.0014>
- Tukey, J. W. (1977) *Exploratory Data Analysis* (Reading, MA: Addison-Wesley).

van Es, K., Wieringa, M. and Tobias Schäfer, M. (2018) *Tool Criticism: From Digital Methods to Digital Methodology*, Proceedings of the 2nd International Conference on Web Studies. <https://doi.org/10.1145/3240431.3240436>

van Es, K. and Schäfer, M. T. (2017). *The datafied society. Studying culture through data* (Amsterdam: Amsterdam University Press), Section 2, pp. 109-182.

van Geenen, D. and Wieringa, M. (2020) *Approaching data visualizations as interfaces: An empirical demonstration of how data are imag(in)ed*, Data Visualization in Society, 141.

Veltri, G. A. and Atanasova, D. (2017) *Climate change on Twitter: Content, media ecology and information sharing behaviour*, Public Understanding of Science, 26(6), pp. 721-737.

Venturini, T., Jacomy, M. and Jensen, P. (2019) *What do we see when we look at networks?* arXiv preprint arXiv:1905.02202 (accessed 2020-09-17).

Venturini, T., Jacomy, M. and Pereira, D. (2014) *Visual network analysis*. Available at: <https://pdfs.semanticscholar.org/67e4/8238146119ea4081da4f3f0d0dc3aa1cb4b6.pdf> (accessed 2020-09-17)

Wertsch, J. V. (1991) *Voices of the Mind*. Harvard University Press.

White, H. (1963) *An Anatomy of Kinship: Mathematical Models for Structures of Cumulated Roles* (Engelwood, NJ: Prentice Hall).

Wieringa, M. (2020). What to account for when accounting for algorithms: A systematic literature review on algorithmic accountability. *FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 1–18. <https://doi.org/10.1145/3351095.3372833>

Wieringa, M., Geenen, D., van Es, K. and van Nuss., J. (2019) The Field Notes Plugin: Making Network Visualization in Gephi Accountable, In: Daly, A., Devitt, K. and Mann., M. (Eds) *Good Data. INC Theory on Demand*. Available at <http://networkcultures.org/blog/publication/tod-29-good-data/> (accessed 2020-09-17)

ⁱ According to Google Scholar at the time of writing (June 2020)

ⁱⁱ <https://www.facebook.com/groups/gephi/> All the URLs featured in footnotes have been last checked on 18 August 2020.

ⁱⁱⁱ See for instance:

<http://www.martingrandjean.ch/gephi-introduction/>

https://www.pauloldham.net/gephi_patent_network/

^{iv} <https://gephi.org/>

^v <https://github.com/gephi/gephi/wiki>

^{vi} *Mastering Gephi Network Visualization* and *Network Graph Analysis and Visualization with Gephi*, by Ken Cherven; and *Gephi Cookbook*, by Devangana Khokhar; all three edited by Packt.

^{vii} <https://wiki.digitalmethods.net/>

^{viii} Available online: <https://quantum.hypotheses.org/176>

^{ix} For clarity: the main author of that paper, Mathieu Jacomy, is also author of this one. The change of perspective between the two papers corresponds to his updated views on the notion of black box.

^x 1255 citations according to Google Scholar, June 2020