Sally Wyatt, Staša Milojevič, Han Woo Park, and Loet Leydesdorff, Quantitative and Qualitative STS: The Intellectual and Practical Contributions of Scientometrics. Pp. 87-112 in: Ulrike Felt, Rayvon Fouché, Clark Miller, and Laurel Smith-Doerr (Eds.), Handbook of Science and Technology Studies (4th edition). Boston, MA: MIT Press, 2017.

Quantitative and Qualitative STS:

The intellectual and practical contributions of scientometrics

Sally Wyatt, Staša Milojević, Han Woo Park and Loet Leydesdorff

(18 September 2015)

As an interdisciplinary specialty, Science and Technology Studies (STS) deploys a wide variety of methods, including (participant) observation, interviews, focus groups, and close reading of textual and visual materials. These draw upon anthropology, sociology and history, three of the core constituent disciplines of STS. Quantitative methods, based on numerical data and/or statistical analysis of large-scale surveys, experiments, national censuses and, more recently, data and information visualizations, are less visible within STS. The application of quantitative methods to study the formal and semantic aspects of the scientific literature as information and communication processes is known as scientometrics. The relation between scientometrics and qualitative approaches within STS has been distant and sometimes even fraught, given the different epistemological assumptions, forms of representation and skills needed. Yet the two specialties share a common origin, even if they have grown apart over the past decades in terms of research practices, norms and standards. For example, they have always shared a deep commitment to the empirical study of science and technology, even if they have different approaches to the ways in which the "empirical" is done, and practitioners of both can be reflexive about their own knowledge production practices.

In this chapter, we argue for greater conciliation between qualitative and quantitative methods within STS, broadly defined. We suggest that such collaboration, by enriching the repertoire of methods available to STS scholars, provides opportunities for exploring new and old research questions. There are also two more pragmatic reasons for this plea for further integration of quantitative and qualitative approaches in STS. The first is the growing attention for "big data", computational methods, and digital forms of representation of data and knowledge in the humanities and the social sciences (Mayer-Schonberger and Cukier 2013; Borgman 2015). The second reason, especially for those STS scholars based in the academy, is that a deeper understanding of scientometrics is necessary for making sense of and for formulating informed criticism about university rankings, evaluations and the audit culture within which academics work (Dahler-Larsen 2011; Hicks *et al*. 2015 (also known as 'Leiden Manifesto'); Halffman and Radder 2015; Strathern 2000).

In the next section, we provide an overview of the common origins of qualitative and quantitative forms of STS, offering a discursive account of this history. We then demonstrate how scientometric techniques can be used to address substantive research questions, and we provide examples relevant both to the origins of STS and its state of the art. Our purpose is not to provide an exhaustive review of either qualitative or quantitative methods, as there exist many methods textbooks for both (e.g. Moed *et al*. 2004; Franklin 2012), although contemporary STS has tended to neglect methods.[1] The final substantive section picks up the themes of "big data" and "reflexivity", and also provides some reflection on the current use of indicators.

A very brief history STS and scientometrics

The origin of STS is often placed in the 1970s with the development of the "sociology of scientific knowledge" (SSK), and with political activism, especially against nuclear power (Bijker 1993). However, we suggest that a fuller understanding of both qualitative and quantitative STS needs to look back to the 1940s and the work of Robert K. Merton, one of the giants[2] of 20th century sociology. Merton was a remarkably prolific scholar, and made major contributions not only to the sociology of science, but also to functionalism and the sociology of deviance. In his *œuvre*, he combined elements from the philosophy and history of science with the sociological tradition (e.g., Durkheim, Mannheim, and Parsons). At the same time, he insisted on an empirical approach and on the value of middle-range theorizing as most appropriate for developing STS as an interdisciplinary field (Merton 1973; Wyatt and Balmer 2007). One of Merton's most influential papers (1942) formulated the "CUDOS norms of science": Communalism, Universalism, Disinterestedness, and Organized Skepticism. He argued that together these norms distinguish science from other areas of activity.

Mertonian sociology of science, also known as the institutional sociology of science, focused mainly on topics such as the stratification of science and the development of specialties, using quantitative methods prevalent in the empirical American sociology of the 1960s (e.g., Mullins 1973). Stratification studies used measures of productivity for the evaluation of the sciences and also drew attention to formal and informal networks. These resonated with the work of Derek de Solla Price, an historian of science and one of the founders of scientometrics. He published extensively in the 1960s and 1970s (e.g. Price 1965, 1970), laying the foundations for the newly emerging field of quantitative science studies. Price's contributions to measuring the productivity of scientists were of interest to the stratification of science group, while his operationalization of the notion of "invisible college" (Crane 1972) was of interest to researchers studying the formation and development of scientific specialties. A major impetus for the development of institutional studies of science and scientometrics was provided by the development of the Science Citation Index (SCI) during the 1960s (Elkana *et al*. 1978).

The SCI was conceptualized and developed as a tool to ameliorate a well-known information science problem of retrieving relevant documents. (Information retrieval continues to guide much information and computer science research.) The SCI was based on the realization that "indexing" does not have to be done only in the form of keywords and descriptors, but that the references and citations present in published journal articles can be used to bring related documents together. Eugene Garfield, an information scientist, promoted this idea in an article published in *Science* in 1955, with the title, "Citation indexes for science: A new dimension in documentation through association of ideas". To materialize his approach, Garfield founded a company, the Institute of Scientific Information (ISI) in Philadelphia that started producing such indexes under the name "Science Citation Index" (Wouters 1999).[3] Price (1965) made early and novel usage of the tool to measure networks of scientists, whereas others used it primarily to develop citation and co-citation measures that could be used to understand the emergence, growth and decline of scientific specialties and the diffusion of ideas. Thus, early usages of citations were not so much focused on citation as a currency or as an indicator of quality and reputation, but on the symbolic functions of citations (Small 1978).

Most studies within an empirical and institutional sociology of science could be described as scientometrics *avant-la-lettre*. However, quantitative studies of science soon began to attract researchers from natural and life sciences and from information and library sciences, the former primarily concerned with measuring productivity of different groups, the latter primarily interested in information retrieval; but both concerned with tracing genealogies of concepts and theories. During the 1970s, scientometrics itself started to exhibit the properties of a specialty, with a new journal, *Scientometrics* launched in 1978 (Milojević and Leydesdorff 2013).

The 1970s also marked the beginning of the constructivist tradition in the studies of science and technology, which developed as a response not only to analytic philosophy of science but also to the normative perspective of Mertonian sociology. Barry Barnes and Alex Dolby (1970), who were working within the emerging “sociology of scientific knowledge” (SSK), also known as the “Strong Program” or the Edinburgh School (cf. Bloor 1976), proposed a constructivist alternative to Merton’s “normative” orientation, and argued for a focus on scientific practices. Adherents of the Strong Program consider science as community-based belief systems, and the units of analysis are actors or collectives driven by a blend of socio-epistemic interests.[4] Merton’s norms can from this perspective be considered as *professed* norms, and citation becomes an instrument for winning a cause in a rhetorical argument (Gilbert 1977; Latour 1987). A constructivist interpretation of citation thus became juxtaposed to the Mertonian or normative interpretation in what has sometimes been called “the citation debate” (Cozzens 1989; Luukkonen 1997).

Historical methods predominated during the further development of SSK. At the same time, an approach that came to be known as the “Bath School” focused on studying controversies in science using observational methods (Collins 2004). This marked the shift of attention away from structural or macro-level phenomena to micro-sociological processes that remained the focus of "laboratory studies" which observed the process of science in making mostly in research laboratories (Latour and Woolgar 1979; Knorr Cetina 1999). Later work by Bruno Latour (1987, 2005) and others, what came to be known as actor-network theory (ANT), rejected notions of scale implied by the use of "macro" and "micro", and highlighted the networks of associations between humans and non-humans to create more or less stable assemblages.

In the 1980s, Michael Mulkay *et al.* (1983) argued that priority should be given to the analysis of constructions in scientific discourse. Nigel Gilbert and Mulkay (1984) then distinguished between a contingent repertoire in scientific practices and the rationalized ("empiricist") discourse in formal publications. In their study of the scientific debate about oxidative phosphorylation, they concluded that scientists tend to use a "contingent repertoire" when explaining error versus an "empiricist repertoire" when hindsight proved them right. In other words, the epistemic contexts of discovery versus validation (Popper [1935] 1959) matter in social practices. Translations between these discourses are mediated and can be studied empirically using, for example, co-word analysis (Callon *et al.* 1983).

The early disputes between STS and scientometrics were around the nature of citations. From a Mertonian perspective, citation analysis is considered primarily as a methodology for the historical and sociological analysis of the sciences (e.g., Cole and Cole 1973). For example, David Edge, a member of the Edinburgh School, criticized citation analysis for ignoring the construction of knowledge claims in "the soft underbelly of science" (1979, 117). In his view, one should "give pre-eminence to the *account from the participant's perspective*, and it is the *citation analysis* which has to be 'corrected'" (1979, 111, italics in the original).

Steve Woolgar (1991) reached "beyond the citation debate"—between normative and constructivist interpretations of citations—by emphasizing that scientometric measurement instruments are used in science policy. Thus, the processes of translations and codifications in and between scholarly and political discourses merit sociological attention from both normative and constructivist perspectives, because of their use in policy and their performative implications. Ever since, however, STS has moved beyond SSK to the post-constructivist, or "critical and cultural studies of science and technology" (Hess 1997, 85), drawing heavily on qualitative and interpretative theories and methods from queer, feminist and post-colonial theories, and literary and cultural studies. During the same period, the scientometrics literature experienced an exponential growth with contributions from researchers from very diverse disciplinary backgrounds. This growth can in part be attributed to the increased reliance on quantitative metrics in research evaluation by funding agencies and academic administrators, as well as advances in computational capabilities, availability of data sets and new analytic techniques. As a result, despite their common origins, "there has been relatively little interaction between scientometrics and STS since the late 1980s. Each … has its own conferences and journals, and only a handful of researchers operate at the interface — most individuals would identify themselves as either 'scientometricians' or 'STS' scholars." (Rotolo, Hicks and Martin, 2015, 1838). We return to this later, but first want to give some examples of how scientometric methods can be used.

Using scientometric methods to answer research questions

The account provided above about the development of qualitative STS and quantitative scientometrics is based on our own experience and reading of the literature. It is not only brief but, like all histories, it is partial, identifying a few crucial moments relevant to our broader narrative. As yet, there is no canonical origin story of STS, perhaps as a result of the different histories of the contributing disciplines. For example, Sergio Sismondo (2008) does provide an overview, starting with Thomas Kuhn, and emphasizing the construction of scientific and technical knowledge, however he does not mention Merton. In the opening chapter of the 1995 *STS Handbook*, Edge provides his own perspective on the history of the field, in which he does acknowledge the influence of Merton, and in which he makes a plea for a "creative reconciliation" (1995, 15) between what he calls normative and reflexive approaches to STS, and what we label here as quantitative and qualitative approaches. Elsewhere, Edge and Roy MacLeod (1986), then editors of *Social Studies of Science*, acknowledged that submissions using quantitative methods were routinely diverted to *Scientometrics*.[5] This practice has had long-lasting effects on both fields, as we shall see later in this chapter. In this section, we provide more input for a history by providing some examples of scientometric methods, also as a means of providing guidance of when they can be helpful in addressing particular kinds of research questions. Throughout the discussion, we pay particular attention to visualizations, how they are produced, and how they can be interpreted.

Our first example is based on co-word analysis, a technique that captures the frequency of pairs of words or phrases in and between documents. It is used to trace links between texts, and to understand the development of scientific fields. We have applied this to bibliographic data from *Scientometrics* and *Social Studies of Science*, key journals in, respectively, scientometrics and

STS in order to show both topics studied and knowledge lineages used. There is an important caveat to be made: because the subset of publications in *Scientometrics* is more than twice as large as the qualitative set, one can expect the former to dominate the latter in the representations.[6]

We map the contents of the field in terms of the co-words used in the titles and abstracts of the articles in these two journals, providing an overview or "distant reading" (Moretti 2013). "Distant reading" is the term increasingly applied to analyzing large quantities of text in order to identify patterns and anomalies, and can be used in distinction to the "close reading" central to the hermeneutic tradition. Although computers cannot read and understand text, they can organize text in quantities that humans cannot read ("big data"); for example, into semantic maps. Semantic maps are based on clustering algorithms, but translate the results into visualizations that can provide an orientation (Latour 1986). For example, one is able to map different repertoires and vocabularies confronting each other in controversy studies (Rogers and Marres 2000).

In the set of texts under study, four main discourses were algorithmically distinguished (Figure 1). On the right, the red nodes indicate the vocabulary of qualitative studies of science and technology with an emphasis on words such as "theory," "practice," "debate," "sociology," "market," and "controversy." On the left, the program distinguishes among a group of terms led by "journal" (in blue) representing studies of structural developments in the sciences as revealed by or measured in terms of the journal literature, and a third group (green) devoted to indicators of research output using words such as "institute," "output," and "China" (the latter reflecting the growth of attention to research evaluation in China in recent years) among the dominant terms. The fourth group (yellow) is focused on "patents," "industry," and "innovation," and borders on the qualitative repertoire, but on the opposite side from the structural cluster with blue nodes, using terms such as "technological change" (qualitative) versus, for example, "innovation" (quantitative), as innovations can be counted in terms of the number of patents involved.[7]

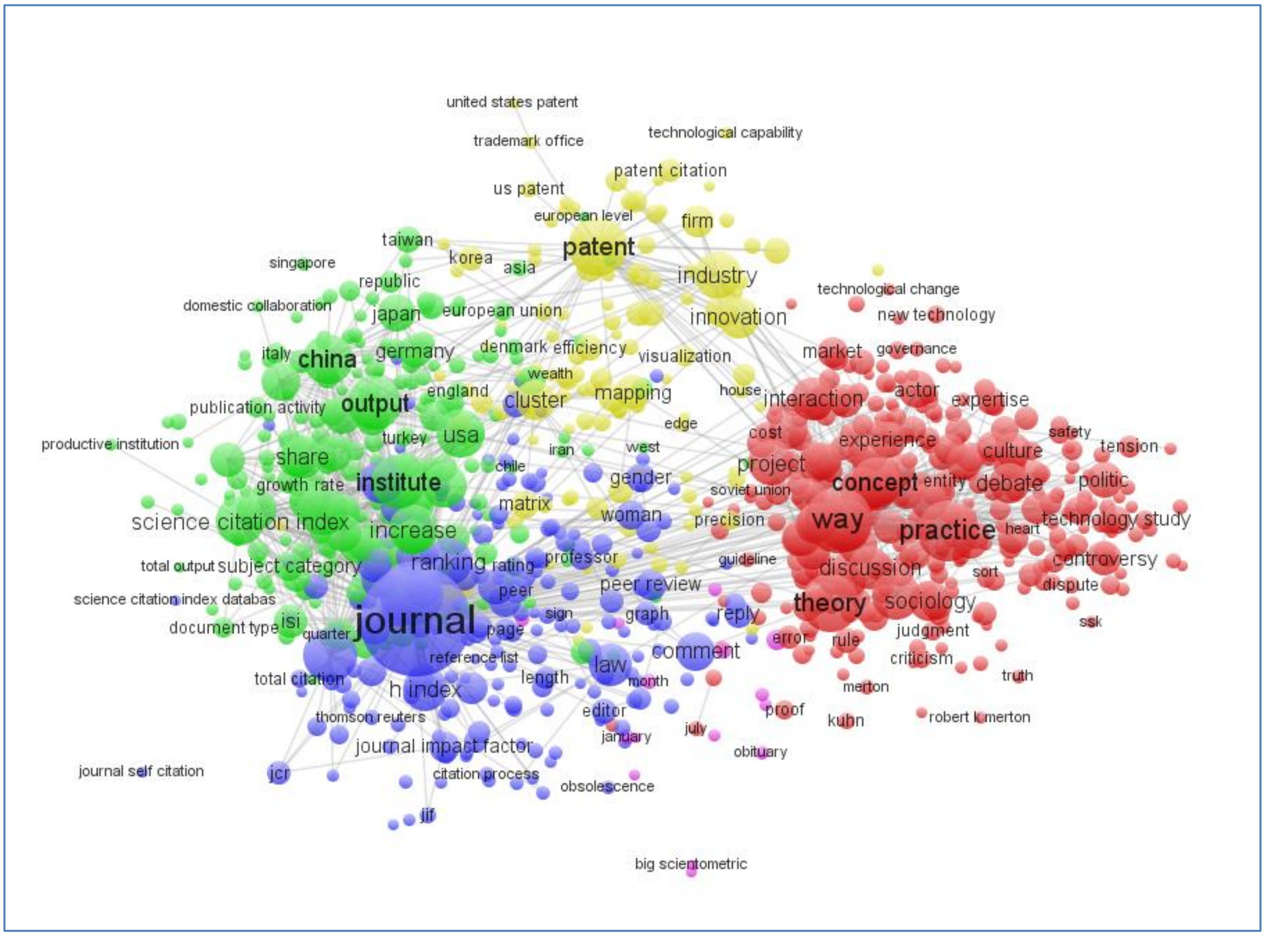

Figure 1: Semantic map of 896 most frequently used words in titles and abstracts of 5,677 publications in *Scientometrics* and *Social Studies of Science*. The map is available for web-starting at http://www.vosviewer.com/vosviewer.php?map=http://www.leydesdorff.net/sts_hbk/fig4map.txt&network=http://www.leydesdorff.net/sts_hbk/fig4net.txt&label_size_variation=0.5&n_lines=1000&zoom_level=1

Note to be inserted: The titles and abstracts contain 50,533 terms, of which 1494 occur more than ten times. VOSviewer reduces this set using a relevance score (van Eck and Waltman 2010). Abstracts are available in the database since 1991.

Although the transitions among the repertoires are gradual and sometimes fuzzy, the analysis points to distinctions that are relevant for understanding the literature that cannot easily be demarcated otherwise because of the interdisciplinary character of the field, and that are difficult for a single individual to capture through "close reading", one of the current definitions of "big data". Consequently, the indicated distinctions are also not easy to validate; but one can use them as heuristics. Since there is no baseline of what is represented—as in a geographic map—one can also use other representations. Figure 2 shows the results of using another common technique with this same data, namely co-citation analysis, based on the frequency with which two documents are cited together in other documents. This works from the assumption that if two authors are citing the same material then they are working from the same knowledge base.

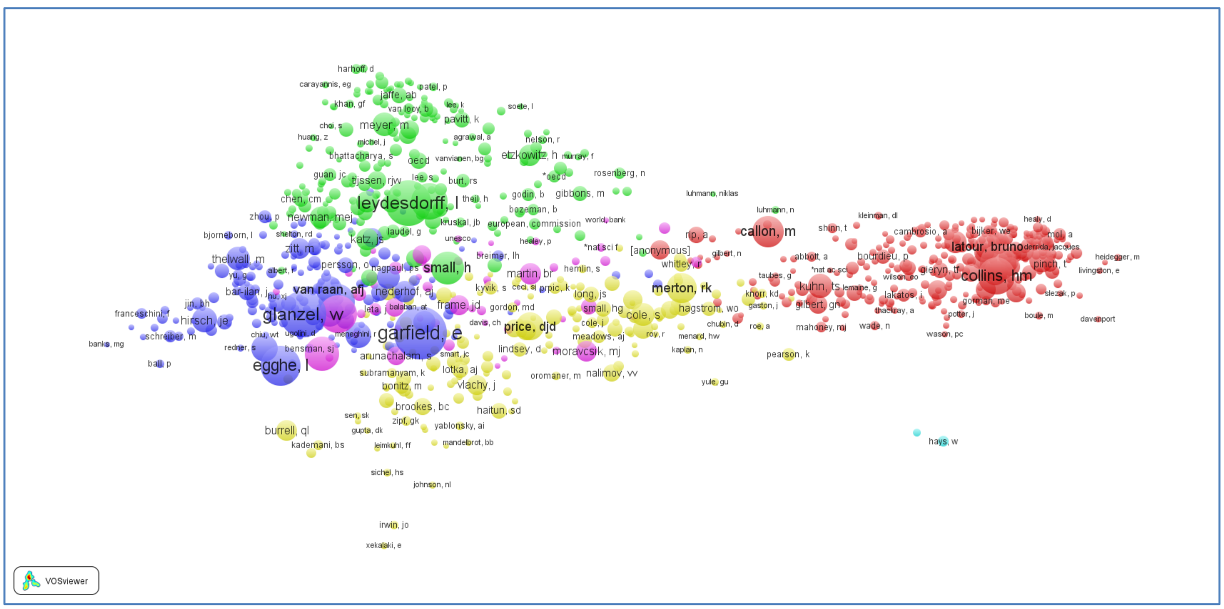

Figure 2: Co-citation analysis of 879 authors in *Scientometrics* and *Social Studies of Science* who are cited more than twenty times (on October 6, 2014). The map is available for web-starting at http://www.vosviewer.com/vosviewer.php?map=http://www.leydesdorff.net/sts_hbk/fig6map.txt

&network=http://www.leydesdorff.net/sts_hbk/fig6net.txt&label_size_variation=0.4&n_lines=1000&visualization=1

In Figure 2, Merton is represented as a node in the center among a group of yellow nodes at the interface between qualitative STS (red nodes on the right) and three groups of authors in quantitative STS (blue, green and pink nodes on the left). Merton shares this position with, among others, Price (mentioned earlier) and Vasily Nalimov, the founding father of Russian scientometrics (*nauchometria*). Richard Whitley (1984/2000), known for his work about the intellectual and social organization of the sciences, is positioned next to Merton in the yellow-colored domain. Note that citation is more frequent and probably more disciplined on the quantitative side when compared with the qualitative one (Larivière *et al*. 2006), which also provides us with a different kind of evidence for the differences between epistemic cultures.

Among the "qualitative" authors on the right side of Figure 2 (red nodes), the co-citation analysis highlights authors who have engaged in discussions with their "quantitative" colleagues, such as Michel Callon, Harry Collins, Nigel Gilbert, and Bruno Latour. Kuhn is positioned among these "qualitative" authors because his work is more co-cited in this constructivist context than in the structuralist (that is, Mertonian) context indicated in yellow. On the left side of Figure 2, three clusters are distinguished: one with Eugene Garfield, Wolfgang Glänzel, and Leo Egghe, prominent founding members of scientometrics (blue nodes). A second cluster indicated in green is led by Henry Small and Loet Leydesdorff who share an interest in visualization, but this group also includes authors interested in university-industry relations and patenting, such as Martin Meyer and Robert Tijssen. Authors who focus on technology studies are also indicated with green dots (e.g., Nathan Rosenberg and Michael Gibbons). These authors combine scientometric methods with economic and sociological theorizing, more so than the methodologically oriented authors in the pink and blue clusters.

Figures such as those above can be compared for different years so that one can visualize developments as animations (Börner 2010; van Eck and Waltman 2014a). One can also combine the two maps because both words and references are attributes of documents, and may thus co-occur or not. One can extend this to other characteristics such as author names, address or country information, and thus construct what can be considered as a "heterogeneous network". Analytically, different possible perspectives are then combined into a more complex construct. Very different maps are possible using combinations of attributes, other algorithms or, more technically, different parameters. However, categories that are meaningful for humans cannot be expected to match algorithmic constructs. Thus, figures such as those above offer a platform for further discussion and more detailed reconstructions.

Just as co-word analysis allows us to engage in "distant reading", the technique of "algorithmic historiography" (Garfield *et al*. 1964, 2003; Van Eck and Waltman 2014b).[8] facilitates a form of computational history by using the dates of publication, and projecting and clustering the citations among them as arrows along the time axis. A historiogram of the set of documents we have used (not shown here) demonstrates that, with the exception of Kuhn (1962), authors in the qualitative tradition are cited only incidentally on the quantitative side. Furthermore, there is almost no citation traffic, or no flow of knowledge, *from* the quantitative *to* the qualitative tradition.

The central articles in the qualitative tradition are tightly knit together by citation relations (see Figure 3 for the citation relations among 35 core papers), whereas the quantitative group in

Figure 2 is divided (by the clustering algorithm of the routine) into two groups: a core group who developed the specialty and a group with broader research interests who use scientometric methods. In summary, the historiographic analysis suggests that, in terms of citation relations, qualitative STS functions as a source of knowledge, but is not itself informed by the scientometric analysis to the extent that qualitative STS scholars actually refer to this literature.

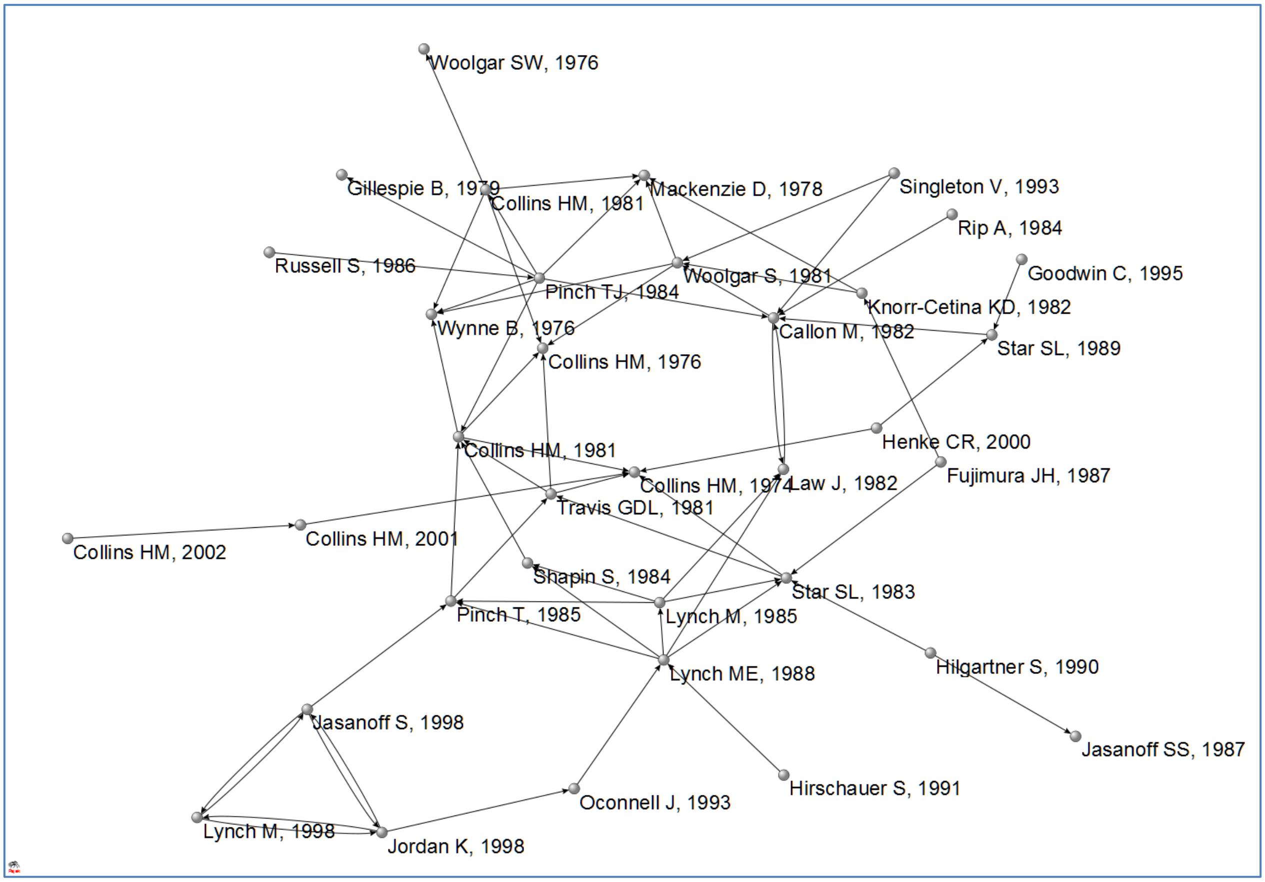

Figure 3: Citation relations among 35 core documents sorted into the qualitative set.

One conclusion to be drawn from Figure 3 is that the core documents in the set are relatively old. This is partly an effect of the accumulation of citation if there is no fixed citation window, but is also typical for fields without a “research front” (Price 1970; Whitley 1984/2000).

However, sciences with research fronts tend to “obliterate [literature] by incorporation” (Garfield 1975; Cozzens 1989, 438). Older literature is incorporated in a process of periodic reviews (Price 1965) or in the production of textbooks.

Scientometrics: From discourse analysis to policy tool

In the previous section, we illustrated how scientometric instruments and techniques can be used to inform theoretical and historical debates at the core of STS, however defined. The scientometric map is performative because it seems to be objective. But the assumptions are hidden in parameter choices that may lead to interaction effects that cannot be overseen by the mapmakers. Thus, the status is that of a machine reading of the literature under study that should not be reified as “objective” or “intersubjective” because the scientometric constructs remain grounded in discursive decisions about how to best represent latent structures in the data.

As shown above, the main units of analysis in scientometrics are documents, and these documents provide a wealth of data for analysis about the development of fields, the emergence of ideas and institutions, and the socio-epistemic organization of knowledge. The scientometric focus on the textual domain provides new perspectives for further developments of STS. In addition to social and epistemic dynamics, traditionally named “context of discovery” and “context of justification”, communication networks and mediation increasingly co-structure the empirical advances of science and technology in society (Luhmann 1995). Codes of communication do part of the work in science as a symbolic system of signs that can reflexively be (re)invoked and modified (Latour 1987, 2005; Wouters 1998).

During the 1980s, the analysis of scientific discourse was increasingly prioritized in STS (e.g., Amsterdamska and Leydesdorff 1989; Callon 1986; Myers 1985; Pinch 1985). Translations between discourses are mediated and can be studied empirically using, for example, co-word analysis. When studying scientific texts, however, one retrieves not only words but also references (Braam *et al*. 1991). References enable authors to mobilize resources for making a persuasive argument (this is the constructivist interpretation) and to shortcut an argument reflexively. In other words, references code the communication which no longer has to be elaborated in words (Bernstein 1971; Coser 1975), and thus enable authors to be more precise. References also enable authors to construct potential audiences and readers. This can be particularly important in interdisciplinary work, where mutual referencing can help to construct bridges between disparate epistemic communities (Fujigaki and Leydesdorff 2000). For example, we have deliberately invoked authors from the qualitative STS canon in order to try to communicate with those readers more familiar with that tradition.

The communication dynamics cannot be reduced to the sociology of the authors or the epistemic dynamics of knowledge claims and validation. Analysts do not know if citations are some form of strategic behaviour, nor when a citation reflects a cognitive debt and when it is mainly a reflection of a social hierarchy within the scientific community. The translation between the social and cognitive aspects of a citation remain problematic (Leydesdorff and Amsterdamska 1990). The three dimensions of texts, authors, and epistemics can be expected to shape a triple helix with a *complex* dynamics. Densities in networks of documents and authors can be recognized as areas of shared interest. The epistemic contents, however, are only reflexively available to both participants and analysts enrolled in the respective discourses, and one can move the discourse forward by contributing. For example, Barnes and Dolby's (1970) "deviant viewpoint" restructured the meaning of the citation from a countable reference to a potentially strategic instrument in an argument (Latour 1983). Mulkay *et al*.'s (1983) emphasis on the analysis of discourse made it relevant to study the meaning of references in documents, as different from words and other textual marks. In other words, the challenge could be framed in terms of understanding how the sciences develop in a semiosis of texts and thus become amenable to measurement (Callon *et al*. 1983; Fujigaki 1998).

The shortcut of the translation between the cognitive and the social dimensions provide scientometric instruments and techniques with their performativity. By counting textual units in terms of cognitive ones, one seems able to make inferences about social units of analysis. The resulting constructs have become increasingly important in the policy arena because of the gradual transition from industrial to knowledge-based economies in some parts of the world (Bell 1973; Castells 1996-98; David and Foray 2002; Godin 2006; OECD 1996). In addition to input indicators such as the numbers of scientists and engineers (and their gender), and funding for research and development (R&D) (OECD 1963; OECD/Eurostat 1997), output indicators such as numbers of publications and patents became policy relevant, and not only signs of the dynamics of academic specialties. Greater attention was given to the measurement of science and technology, and the theoretical (e.g., statistical) foundations for the measurement (Martin and Irvine 1983; Moed *et al*. 1985; Narin 1976).

What were first conceptualized as relevant variables in STS have become indicators used in science policy and research management. Qualitative STS has informed and continues to inform science and technology policies albeit often mediated by quantitative approaches (Collins

1985; Martin 2012; Martin *et al.* 2012; Wagner *et al*. 2009). The Citation Indices first provided the sociology of science and scientometrics with new measurement instruments for understanding social and epistemic relationships between people, groups and ideas, across time and space. At the same time, however, they became the major source for measuring research performance.

In 1972, the US National Science Board launched the biannual series of Science Indicators using, among other things, the SCI. Since then, funding agencies and research advisory bodies have increasingly taken up STS insights, both qualitative and quantitative, in their science and technology policies and R&D management. The construction and deconstruction of the indicators has become a literature in itself (e.g. van Raan 1988; Moed 2005), but at arm's length from STS (Martin 2012). Nevertheless, a leading producer of university rankings and purveyor of institutional and disciplinary evaluations continues to call itself the Centre for Science and Technology Studies (CWTS) at Leiden University, and similar spin-offs from STS departments can be found at other universities (e.g., in Montréal; Granada; Atlanta, GA). At the institutional level it is possible to integrate qualitative and quantitative STS, even though they address different audiences, and deploy different epistemic norms, standards, and codes of communication. In many instances, "multi-disciplinarity" may be more suited as a format for combining quantitative and qualitative STS than an ambition of "interdisciplinarity".

Research evaluation has become a specialty in itself with sophisticated discussions about issues such as different normalizations. Francis Narin (1976) first coined the name "evaluative bibliometrics" for this field (cf. Pritchard 1969). In the scientometric literature, however, the purpose of a deconstruction is mostly the further improvement of an indicator (e.g., Waltman *et al*. 2011). In the meantime, journal impact factors,[9] university rankings, and the h-index[10] have become an everyday part of academic life. In such policy contexts, these measures are used without scholarly articulation in terms of STS. In the next and final section, we turn to the question of how the development of scientometric methods can be expected to further contribute to the development of STS from the perspective of qualitative STS.

Conclusions and future directions

In this chapter, we have highlighted some of the common origins of (qualitative) STS and (quantitative) scientometrics, and their common commitment to the empirical study of the sciences. Documents and text broadly conceived are also an important, if not the most important resource for STS researchers, whatever their theoretical and methodological commitments. We argued that scholarly publications can be analyzed both as constructs and as factual data amenable to text, data and network analytics. For the latter, as illustrated above, agents are reconstructed as author names (*actants*); institutions as institutional addresses; and cognitions can be coded as references. The process of knowledge production can thus be considered as one of argumentative purification and coding of texts at the network level.

In the long history of the dissemination of knowledge in the form of manuscripts, documents were first scarce and transcription was error-prone. By the end of the 20th century, the selection mechanisms have changed. Text is everywhere, and texts are abundant as variation; word processing induces working in versions, where each new version is intended to improve the content and the formulation. A variety of selection environments is envisaged by authors. Scholarly manuscripts, for example, are processed and selected by journal editors before and after review. The composition of review boards remains contingent, and change in editorship may affect the selection mechanisms operating at the journal level. These selection processes can be studied empirically, using both quantitative and qualitative methods (Lamont 2012). However, the mechanisms develop at the supra-individual level, with their own dynamics, but they are not closed off from agency. As Latour (1988) noted, one has access to them “infra-reflexively”: as authors one can access the network through their own publication activities. For example, an author can read a journal’s mission statement and author guidelines, but individually or locally a single author cannot control the emerging network dynamics. In other words, only knowledge of selection mechanisms operating in the networks can sometimes inform authors’ expectations or specify options for change.

From this perspective of uncertain selection environments, performativity and reflexivity are no longer localized and attributable to individuals, but distributed in relations and network configurations, and thus a potential object of sociological analysis. Communication in network relations as units of analysis different from agency provide another frame of reference for what the concepts may mean. While reflexivity and performativity have mainly been attributed to (individual or institutional) agency in the sociological tradition, the reflection of concepts in documents can be expected to lead to different meanings at other ends and can thus open or reinforce a discourse (Bourdieu 2004). A sociology of communication emerges alongside, in relation to, but not isomorphic to the sociology of human agency and groups of individuals in social organizations (Luhmann 1995). Actor-network theory noted that this newly emerging sociology could be expected to include also non-human *actants*.

Our conclusions are of two types. The first relates to the nature of the scientometric enterprise itself. We have argued that scientometrics can be useful for understanding, among other things, the emergence and transformation of concepts and theories, the dynamics of the communication of science, and mapping the socio-epistemic organization of knowledge. However, we have warned against the other arrow: the unreflexive use of such techniques by policy makers and management, and thus the export of presumed insights in the dynamics of science and technology. The further refinement of theoretically uninformed categories, especially if they are then used to produce highly consequential rankings of universities, departments and individuals, needs to be resisted, and scientometricians have the necessary arguments. The data, methods and techniques are as artefactual and constructed as any other kind of social scientific fact or method, and need to be treated with this care for one's theoretical backgrounds and foundations.

Let us repeat that the scientometric maps and tables are algorithmic constructs that remain heavily dependent on the parameters chosen for generating them, and this is true for all computational methods. Refining the parameter choices does not validate the constructs. By presenting our own scientometric results in this chapter as algorithmic constructs, we have tried to re-open the discussion between quantitative and qualitative STS. Can the scientometric endeavor be made useful to the study of STS? Can the two specialties be mutually infused with each other's heuristics? Following Woolgar (1991), we have argued that one way of reconciling the apparent epistemological contradictions between quantitative and qualitative methods and thus quantitative and qualitative STS is to recognize that both types of data require interpretations in contexts of use, not only by researchers but also by social actors, including respondents, policy makers, and other audiences.

This brings us to the second type of conclusions, related to the future of STS more broadly, and the benefits of incorporating scientometric methods. As so-called "big data" become more prevalent in the social science and the humanities (leading to fields such as computational social science, network science, and digital humanities), we wonder whether the usage of massive data and associated statistical, mathematical, and computational methods of analysis, could lead to significant changes within STS that would make the exchange of knowledge between qualitative and quantitative STS more feasible.

Since the early 1990s, with the widespread diffusion of the internet and other changes in the research environment, publication and communication dynamics in the sciences have changed. Open access requirements to make data transparently available when submitting a manuscript, and at a later stage for the publication itself, add additional dimensions to the selection environments. This more complex environment also generates expectations and standards that STS cannot afford to ignore. Inspired by the research program of scientometrics, we suggest that the socio-epistemic has irrevocably been extended by a textual dimension that provides us with "big data" about, among other things, developments in science and technology. To address this complexity, STS needs to be able to build bridges, and engage in multiple processes of translation between qualitative and quantitative approaches.

The possibilities offered by digitally mediated communication have brought to the fore an awareness that the dynamics of communication in scholarly discourses are both specific and interacting with other forms of communication. The study of communications (e.g., in discourse analysis) invites questions about sample choices and statistical testing of significance (Krippendorff and Bock 2009). Communication and the means for measuring it can always be improved and thus one can expect standards, i.e., quality measures, to emerge. From this perspective, scientometric indicators hitherto have mainly captured communication within the sciences, whereas future challenges include the measurement and analysis of communication across scientific domains, and at the interfaces between science and society.

While the scientometric results presented in this chapter tend to confirm a division and lack of communication between quantitative and qualitative STS, the opening of a conversation between the two is essential for the successful utilization of novel data sources to advance STS. Büscher and Urry (2009) suggest that the lack of exchange and transfer of knowledge is not surprising given the different epistemological foundations underlying the construction of the data, and the very different skills researchers need in order to analyze and interpret the findings based on such different types of data. Our argument is slightly different in that we emphasize the always constructed nature of data, methods and algorithms. But we do agree that the skill sets differ, and that greater dialog could enrich the skills and methodological repertoires of both qualitative and quantitative researchers. This would open up more questions about "big data". Of course, STS can contribute to "big data" debates by analyzing what big data means for other research fields,[11] but we also need to foster thinking about how STS itself can remain relevant to on-going debates, by making statistical, reflexive, and policy-relevant use of its own large archives of data (not only bibliographic databases and citation indices, but also oral history archives and ethnographic field notes). STS could embrace data science, not only as an object of critique but also as a source of inspiration for future work. In this context Venturini, Jensen and Latour (2010) have proposed to consider digital methods as "quali-quantitative methods" . New *actants* such as blogs, Wikipedia, and mark-up languages are and will be generated in the digital domain, and require the development of new methodologies (Light *et al*. 2014; Rogers 2013).

Our focus on the textual dimension of the sciences has theoretical implications, including the attribution of performativity to networks of documents, instead of to individuals as would be common in the sociological tradition (e.g. Giddens 1979). We have criticized the ways in which performance indicators are attributed as rankings of individual authors or institutes, with performative and exclusionary consequences. But we are aware that our own reconstruction of STS can also be considered as performative and exclusionary. For example, we did not address issues of gender and ethnicity in science, nor did we explicitly address publishing in national languages and in book or other formats. The history of scientometrics and the usage of indicators and rankings in science and technology policies and research management make a coupling to qualitative STS urgent because algorithmic reifications cannot tell the stories. The stories of policy makers and indicator suppliers may be based on sophisticated and innovative statistics, but discussions and debates about science and technology remain intellectually indebted to STS.

Acknowledgements

We thank Thomson Reuters for providing us with relevant data. Diana Lucio-Arias and Selma Sabanović offered valuable support and input during early stages of preparation. We are also grateful to the editors of the Handbook for their patience and support, and to the anonymous reviewers of an earlier draft for their challenging and constructive remarks.

References

Amsterdamska, O., and Leydesdorff, L. 1989. “Citations: Indicators of Significance?” Scientometrics 15 (5-6): 449-471.

Barnes, S.B., and Dolby, R.G.A. 1970. “The Scientific Ethos: A Deviant Viewpoint.” European Journal of Sociology 11 (1): 3-25.

Bell, D. 1973. The Coming of Post-Industrial Society. New York, NY: Basic Books.

Bernstein, B. 1971. Class, Codes and Control, Vol. 1: Theoretical Studies in the Sociology of Language. London: Routledge and Kegan Paul.

Bijker, W.E. 1993. “Do not despair. There is life after constructivism.” Science, Technology & Human Values 18 (1): 113-138.

Bloor, D. 1976. Knowledge and Social Imagery. London: Routledge and Kegan Paul.

Borgman, C. 2015. Big Data, Little Data, No Data. Scholarship in the Networked World. Cambridge, MA: The MIT Press.

Bourdieu, P. 2004. Science of Science and Reflexivity. Chicago, IL: University of Chicago Press.

Börner, K. 2010. Atlas of Science: Visualizing What We Know. Cambridge, MA: The MIT Press.

boyd, d. and Crawford, K. 2012. “Critical questions for big data: Provocations for a cultural, technological and sociological phenomenon.” Information, Communication & Society 15 (5): 662-679.

Braam, R.R., Moed, H.F., and van Raan, A.F.J. 1991. “Mapping of science by combined co-citation and word analysis. I. Structural aspects.” Journal of the American Society for Information Science 42 (4): 233-251.

Büscher, M. and Urry, J. 2009. “Mobile Methods and the Empirical.” European Journal of Social Theory 12 (1): 99-116.

Callon, M. 1986. “The sociology of an actor network: the case of the electric vehicle.” In Mapping the Dynamics of Science and Technology, ed. M. Callon, J. Law and A. Rip, 19-34. London: Macmillan.

Callon, M., Courtial, J.-P., Turner, W. A., and Bauin, S. 1983. “From Translations to Problematic Networks: An Introduction to Co-word Analysis.” Social Science Information 22 (2): 191-235.

Castells, M. 1996-98. The Information Age: Economy, Society and Culture. Oxford: Blackwell. Vol. I: The Rise of the Network Society (1996); Vol. II: The Power of Identity (1997); Vol. III: End of Millennium (1998).

Cole, J.R., and Cole, S. 1973. Social Stratification in Science. Chicago, IL: University of Chicago Press.

Collins, H.M. 1985. “The possibilities of science policy.” Social Studies of Science, 15: 554-558.

Collins H.M. 2004. Gravity’s Shadow. The Search for Gravitational Waves. Chicago, IL: University of Chicago Press.

Coser, R.L. 1975. The complexity of roles as a seedbed of individual autonomy. In The Idea of Social Structure. Papers in Honor of Robert K. Merton, ed. L.A. Coser, 237-264. New York/Chicago: Harcourt Brace Jovanovich.

Cozzens, S.E. 1989. “What do citations count? The rhetoric-first model.” Scientometrics 15 (5): 437-447.

Crane, D. 1972. Invisible Colleges. Chicago, IL: University of Chicago Press.

Dahler-Larsen, P. 2011. The Evaluation Society. Stanford, CA: Stanford University Press.

David, P. A., and Foray, D. 2002. “An introduction to the economy of the knowledge society.” International Social Science Journal 54 (171): 9-23.

Edge, D. 1979. “Quantitative measures of communication in science: A critical overview.” History of Science 17: 102-134.

Edge, D. 1995. “Reinventing the wheel.” In Handbook of Science and Technology Studies, ed. S. Jasanoff, G.E. Markle, J.C. Petersen and T. Pinch, 3-23. Thousand Oaks, CA: Sage.

Edge, D., and MacLeod, R. 1986. “Editorial.” Social Studies of Science 16 (1): 3-8.

Edwards, P. *et al*. 2011. “Science frictions: Data, metadata, and collaboration.” Social Studies of Science 41: 667-690.

Elkana, Y., Lederberg, J., Merton, R.K., Thackray, A. and Zuckerman, H. 1978. Toward a Metric of Science: The Advent of Science Indicators. New York: Wiley.

Franklin, M.I. 2012. Understanding Research. Coping with the Quantitative-Qualitative Divide. London: Routledge.

Fujigaki, Y. 1998. “Filling the Gap Between Discussions on Science and Scientists' Everyday Activities: Applying the Autopoiesis System Theory to Scientific Knowledge.” Social Science Information 37 (1): 5-22.

Fujigaki, Y. and Leydesdorff, L. (2000). “Quality control and validation boundaries in a triple helix of university-industry-government: ‘Mode 2’ and the future of university research.” Social Science Information 39 (4): 635-655.

Garfield, E. 1964. “Science Citation Index-A new dimension in indexing.” Science 144 (3619): 649-654.

Garfield, E. 1975. “The ‘obliteration phenomenon’ in science—and the advantage of being obliterated.” Current Contents nos. 51/52 (22 December): 396–398.

Garfield, E., Pudovkin, A. I. and Istomin, V. S. 2003. “Why do we need algorithmic historiography?” Journal of the American Society for Information Science and Technology 54 (5): 400-412.

Garfield, E., Sher, I.H. and Torpie, R.J. 1964. The Use of Citation Data in Writing the History of Science. Philadelphia, PA: Institute for Scientific Information.

Giddens, A. 1979. Central Problems in Social Theory. London: Macmillan.

Gilbert, G.N. 1977. “Referencing as persuasion.” Social Studies of Science, 7: 113-122.

Gilbert, G.N. and Mulkay, M.J. 1984. Opening Pandora's Box. A Sociological Analysis of Scientists' Discourse. Cambridge, UK: Cambridge University Press.

Godin, B. 2006. “The Knowledge-Based Economy: Conceptual Framework or Buzzword?” Journal of Technology Transfer 31 (1): 17-30.

Hackett, E.J., Amsterdamska, O., Lynch, M. and Wajcman, J. 2008. “Introduction.” In Handbook of Science and Technology Studies, ed. E.J. Hackett, O. Amsterdamska, M. Lynch and J. Wajcman, 13-31. Cambridge, MA: The MIT Press.

Halffman, W. and Radder, H. 2015. “The Academic Manifesto. From an occupied to a public university.” Minerva 53 (2): 165-187.

Hess, D.J. 1997. Science Studies: An Advanced Introduction. New York, NY: NYU Press.

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., and Rafols, I. 2015. “The Leiden Manifesto for research metrics.” Nature 520: 429-431.

Hirsch, J.E. 2005. “An index to quantify an individual’s scientific research output.” Proceedings of the National Academy of Sciences of the USA 102 (46): 16569-16572.

Knorr Cetina, K. 1999. Epistemic Cultures: How the Sciences Make Knowledge. Cambridge, MA: Harvard University Press.

Krippendorff, K. and Bock, M. A. 2009. The Content Analysis Reader. Los Angeles: Sage.
Kuhn, T.S. 1962. The Structure of Scientific Revolutions. Chicago, IL: University of Chicago Press.

Lamont, M. 2012. “Toward a comparative sociology of valuation and evaluation.” Sociology, 38 (1): 201-221.

Larivière, V., Archambault, É., Gingras, Y. and Vignola-Gagné, É. 2006. “The place of serials in referencing practices: Comparing natural sciences and engineering with social sciences and humanities.” Journal of the American Society for Information Science and Technology 57 (8): 997-1004.

Latour, B. 1983. “Give me a laboratory and I will raise the world.” In Science Observed, ed. K. Knorr Cetina and M.J. Mulkay, 141-170. London: Sage.

Latour, B. 1986. “Visualisation and cognition: Drawing things together.” Knowledge and Society: Studies in the Sociology of Culture Past and Present 6: 1-40.

Latour, B. 1987. Science in Action. Milton Keynes: Open University Press.

Latour, B. 1988. “The politics of explanation: An alternative.” In Knowledge and Reflexivity: New Frontiers in the Sociology of Knowledge, ed. S. Woolgar and M. Ashmore, 155-177 London: Sage.

Latour, B. 2005. Reassembling the Social: An Introduction to Actor-Network-Theory. Oxford: Oxford University Press.

Latour, B. and Woolgar, S. 1979. Laboratory Life: The Social Construction of Scientific Facts. London: Sage.

Leonelli, S. 2014. “What difference does quantity make? On the epistemology of Big Data in biology.” Big Data & Society 1 (1): 1-11.

Levallois, C., Steinmetz, S. and Wouters, P. 2013. “Sloppy data floods or precise social science methodologies?” In Virtual Knowledge, Experimenting in the Humanities and the Social Sciences, ed. P. Wouters, A. Beaulieu, A. Scharnhorst and S. Wyatt, 151-182. Cambridge, MA: The MIT Press.

Leydesdorff, L. and Amsterdamska, O. 1990. “Dimensions of Citation Analysis.” Science, Technology & Human Values 15 (3): 305-335.

Light, R. P., Polley, D. E. and Börner, K. 2014. “Open data and open code for big science of science studies.” Scientometrics 101 (2): 1535-1551.

Luhmann, N. 1995. Social Systems. Stanford, CA: Stanford University Press.

Luukkonen, T. 1997. “Why has Latour's theory of citations been ignored by the bibliometric community? Discussion of sociological interpretations of citation analysis.” Scientometrics, 38 (1): 27-37.

Martin, B. and Irvine, J. 1983. “Assessing Basic Research: Some Partial Indicators of Scientific Progress in Radio Astronomy.” Research Policy 12: 61-90.

Martin, B. R. 2012. “The evolution of science policy and innovation studies.” Research Policy, 41 (7): 1219-1239.

Martin, B.R., Nightingale, P. and Yegros-Yegros, A. 2012. “Science and technology studies: Exploring the knowledge base.” Research Policy 41 (7): 1182-1204.

Mayer-Schomberger, V. and Cukier, K. 2013. Big Data: A Revolution that will Transform how we Live, Work, and Think. Boston, MA: Houghton Mifflin Harcourt.

Merton, R.K. 1942. “Science and Technology in a Democratic Order.” Journal of Legal and Political Sociology 1: 115-126.

Merton, R.K. 1965. On the Shoulders of Giants: A Shandean Postscript. New York, NY: The Free Press.

Merton, R.K. 1973. The Sociology of Science: Theoretical and Empirical Investigations. Chicago: University of Chicago Press.

Milojević, S. and Leydesdorff, L. 2013. “Information Metrics (iMetrics): A Research Specialty with a Socio-Cognitive Identity?” Scientometrics 95 (1): 141-157.

Moed, H.F., Burger, W.J.M., Frankfort, J.G. and Van Raan, A.F.J. 1985. “The Use of Bibliometric Data for the Measurement of University Research Performance.” Research Policy 14: 131-149.

Moed, H.F., Glänzel, W. and Schmoch, U. 2004. Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S & T Systems. Dordrecht: Springer.

Moed, H.F. 2005. Citation Analysis in Research Evaluation. Dordrecht: Springer.

Moretti, F. 2013. Distant Reading. London: Verso.

Mulkay, M., Potter, J. and Yearley, S. 1983. "Why an Analysis of Scientific Discourse is Needed." In Science Observed: Perspectives on the Social Study of Science, ed. K.D. Knorr and M.J. Mulkay, 171-204. London: Sage.

Mullins, N.C. 1973. Theories and Theory Groups in Contemporary American Sociology. New York, NY: Harper and Row.

Myers, G. 1985. "Texts as knowledge claims: The social construction of two biology articles." Social Studies of Science 15: 593-630.

Narin, F. 1976. Evaluative Bibliometrics: The Use of Publication and Citation Analysis in the Evaluation of Scientific Activity. Washington, DC: National Science Foundation.

OECD. 1963, 1976. The Measurement of Scientific and Technical Activities: "Frascati Manual". Paris: OECD.

OECD. 1996. The Knowledge-Based Economy. Paris: OECD; retrieved on March 30, 2015 from www.oecd.org/dataoecd/51/8/1913021.pdf .

OECD/Eurostat. 1997. Proposed Guidelines for Collecting and Interpreting Innovation Data, "Oslo Manual". Paris: OECD.

Pinch, T.J. 1985. "Towards an Analysis of Scientific Observation: The Externality and Evidential Significance of Observational Reports in Physics." Social Studies of Science 15: 3-36.

Popper, K.R. [1935] 1959. The Logic of Scientific Discovery. London: Hutchinson.

Price, D. de Solla 1965. "Networks of scientific papers." Science 149 (3683): 510- 515.

Price, D. de Solla 1970. "Citation Measures of Hard Science, Soft Science, Technology, and Nonscience." In Communication among Scientists and Engineers, ed. C.E. Nelson and D.K. Pollock, 3-22. Lexington, MA: Heath.

Rogers, R. 2013. Digital Methods. Cambridge MA: The MIT Press.

Rogers, R. and Marres, N. 2000. "Landscaping climate change: A mapping technique for understanding science and technology debates on the World Wide Web." Public Understanding of Science 9 (2): 141-163.

Rotolo, D., Hicks, D. and Martin, B. 2015. “What is an emerging technology?” Research Policy, 44 (10): 1827-1843.

Sismondo, S. 2008. “Science and Technology Studies and an Engaged Program.” In The Handbook of Science and Technology Studies, 3rd ed., ed. E.J. Hackett, O. Amsterdamska, M. Lynch and J. Wajcman, 13-31. Cambridge, MA: The MIT Press.

Small, H. 1978. “Cited documents as concept symbols.” Social Studies of Science 8 (3): 113-122.

Strathern, M. 2000. Audit Cultures. Anthropological Studies in Accountability, Ethics and the Academy. London: Routledge.

van Eck, N.J. and Waltman, L. 2010. “Software survey: VOSviewer, a computer program for bibliometric mapping.” Scientometrics 84 (2): 523-538.

van Eck, N.J. and Waltman, L. 2014a. “Visualizing bibliometric networks.” In Measuring Scholarly Impact, ed. Y. Ding, R. Rousseau and D. Wolfram, 285-320. Cham: Springer.

van Eck, N.J., and Waltman, L. 2014b. “CitNetExplorer: A new software tool for analyzing and visualizing citation networks.” Journal of Informetrics 8 (4): 802-823.

van Raan, A.F.J. (Ed.). 1988. Handbook of Quantitative Studies of Science and Technology. Amsterdam: Elsevier.

Venturini, T., Jensen, P. and Latour, B. 2015. “Fill in the gap. A new alliance for social and natural sciences.” Journal of Artificial Societies and Social Simulation 18 (2): 11.

Waltman, L. et al. 2011. “Towards a New Crown Indicator: Some Theoretical Considerations.” Journal of Informetrics 5 (1): 37-47.

Wagner, C.S. et al. 2011. “Approaches to Understanding and Measuring Interdisciplinary Scientific Research (IDR): A Review of the Literature.” Journal of Informetrics 5 (1): 14-26.

Whitley, R.J. 1984/2000. The Intellectual and Social Organization of the Sciences, 2nd ed. Oxford: Oxford University Press.

Woolgar, S. 1991. “Beyond the citation debate: towards a sociology of measurement technologies and their use in science policy.” Science and Public Policy 18 (5): 319-326.

Wouters, P. 1998. “The signs of science.” Scientometrics 41 (1): 225-241.

Wouters, P. 1999. The Citation Culture. Unpublished Ph.D. Thesis. Amsterdam: University of Amsterdam.

Wyatt, S. and Balmer, B. 2007. "Home on the Range. What and Where is the Middle in Science and Technology Studies?" Science, Technology & Human Values 32 (6): 619-626.

Notes

[1] As Sergio Sismondo argued in the opening chapter of the 2008 *Handbook*, "STS looks to **how** the **things** it studies are constructed." (2008, 13, emphasis added) He goes on to show that those "things" have expanded in scope to include not only scientific and technical knowledge practices but also practices associated with policy, politics, governance of science and technology. But the "how" is not elaborated, and indeed the 2008 *Handbook* does not have a section on method, nor even a mention in the Index, though the editors lament this lack in their Introduction (Hackett *et al*. 2008, 5).

[2] "Stand on the shoulders of giants" is the motto adopted by Google Scholar. It invokes the letter that Isaac Newton (1642-1727) wrote to Robert Hooke on February 5, 1676, in which he wrote: "If I have seen further, it is only by standing on the shoulders of giants". Merton borrowed the aphorism for the title of one of his books (1965).

[3] From its beginnings, scientometrics and bibliometrics has been characterized by the involvement of private and for-profit university-based groups in the provision of data, indices, rankings and evaluations. The political economy of these activities is beyond the scope of this chapter but is nonetheless extremely important.

[4] There was a debate about "interests" in *Social Studies of Science* in the early 1980s, involving MacKenzie, Woolgar, Barnes and Callon. It was agreed that the formation of interests is a slow historical process, that interests cannot always be read from subject positions, and that interests are both an actor's and an analyst's category.

[5] Tibor Braun, then editor of Scientometrics, confirmed this in personal communication with one of us (Leydesdorff) in 1985. The arrangement worked in both directions.

[6] The analyses are based on the full sets of publications in *Scientometrics* (since 1978) and *Social Studies of Science* (since 1971) downloaded from the installation of the Web of Science (WoS) on 6 October 2014.

[7] Although this is common practice in technology and innovation studies, patents are not indicators of innovation, but of invention.

[8] This was further made accessible for general use as a program HistCite™, developed by Garfield in collaboration with Alexander Pudovkin (Garfield *et al.* 2003).

[9] The impact factor of a journal is the average number of times that articles published in the journal during the preceding two or five years are cited in the current year.

[10] The *h*-index was proposed by the physicist J.E. Hirsch (2005). An author has an index of *h* where *h* is the number of papers that have each been cited at least *h* times. It has had a huge influence on scientometrics, but it is not so much reflected in the STS literature.

[11] STS-informed critiques and analyses of "big data" are already available in biology (Leonelli 2014), in social science (Levallois *et al*. 2013), more generally by boyd and Crawford (2012) and Edwards *et al*. (2011), and in the journal *Big Data & Society*, launched in 2014.