



HAL
open science

Beyond networks: Aligning qualitative and computational science studies

Alberto Cambrosio, Jean-Philippe Cointet, Alexandre Hannud Abdo

► **To cite this version:**

Alberto Cambrosio, Jean-Philippe Cointet, Alexandre Hannud Abdo. Beyond networks: Aligning qualitative and computational science studies. *Quantitative Science Studies*, 2020, 1 (3), pp.1017-1024. 10.1162/qss_a_00055 . hal-03895618

HAL Id: hal-03895618

<https://sciencespo.hal.science/hal-03895618v1>

Submitted on 13 Dec 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.



Distributed under a Creative Commons Attribution 4.0 International License



RESEARCH ARTICLE

Beyond networks: Aligning qualitative and computational science studies

Alberto Cambrosio¹ , Jean-Philippe Cointet² , and Alexandre Hannud Abdo³ 

¹Department of Social Studies of Medicine, McGill University, Canada

²Sciences Po, Médialab, Paris, France

³Université Paris Est, LISIS, UMR 1326 INRA, Champs sur Marne, Marne-la-Vallée, France

an open access  journal



Citation: Cambrosio, A., Cointet, J.-P., & Abdo, A. H. (2020). Beyond networks: Aligning qualitative and computational science studies. *Quantitative Science Studies*, 1(3), 1017–1024. https://doi.org/10.1162/qss_a_00055

DOI:
https://doi.org/10.1162/qss_a_00055

Corresponding Author:
Alberto Cambrosio
alberto.cambrosio@mcgill.ca

Handling Editors:
Loet Leydesdorff, Ismael Rafols,
and Staša Milojević

Copyright: © 2020 Alberto Cambrosio, Jean-Philippe Cointet, and Alexandre Hannud Abdo. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.



Keywords: actor-network theory, computational science studies, hypergraphs, network analysis, stochastic block models

ABSTRACT

This article examines the thorny issue of the relationship (or lack thereof) between qualitative and quantitative approaches in Science and Technology Studies (STS). Although quantitative methods, broadly understood, played an important role in the beginnings of STS, these two approaches subsequently strongly diverged, leaving an increasing gap that only a few scholars have tried to bridge. After providing a short overview of the origins and development of quantitative analyses of textual corpora, we critically examine the state of the art in this domain. Focusing on the availability of advanced network structure analysis tools and Natural Language Processing workflows, we interrogate the fault lines between the increasing offer of computational tools in search of possible uses and the conceptual specifications of STS scholars wishing to explore the epistemic and ontological dimensions of techno-scientific activities. Finally, we point to possible ways to overcome the tension between ethnographic descriptions and quantitative methods while continuing to avoid the dichotomies (social/cognitive, organizing/experimenting) that STS has managed to discard.

1. IN THE BEGINNINGS: A FEW “HISTORICAL” REMARKS

From the very outset, quantitative analyses of scientific and technological activities have accompanied the growth of the field presently known as Science & Technology Studies (STS). One has just to recall the pioneering work by Derek De Solla Price (1963, 1965, 1975), and the development of citation indexing and related methods such as cocitation analysis (Small, 1973). Yet, intersections between quantitative approaches and conceptual developments in STS have been more the exception than the rule. In fact, these two domains have not always coexisted peacefully. Early skirmishes include strongly worded criticism (“Why I am not a co-citationist”) by David Edge (1977, 1979), the cofounder of STS’s flagship journal, *Social Studies of Science*. While, in part, the debate revisited the opposition between qualitative (ethnographic) and quantitative approaches raging for decades within general sociology, in the case of STS the situation is more ambivalent. For instance, one of the foundational texts of the ethnographic turn in STS, *Laboratory Life* (Latour & Woolgar, 1979), mobilized citation analysis, albeit informed by semiotics, as part of its investigation of laboratory practices. Latour (1976) had in fact broached this topic in an early paper presented at the First Annual Meeting of the Society for Social Studies of Science in 1976. Yet, one of the most telling examples of the complex relationship that STS entertains with quantitative methods is provided by an alternative (some would say complementary) approach to citation analysis, namely co-word analysis (Callon, Courtial, et al., 1983).

Championed by Michel Callon, co-originator with Bruno Latour of Actor-Network Theory (ANT), co-word analysis was an attempt to develop a computational approach—provocatively termed “qualitative scientometrics” (Callon, Law, & Rip, 1986)—that would be consistent with the sociology of translation (as ANT is also known). Co-word analysis was grounded in a model of scientific practices as flexible activities likely to undergo rapid change (Callon, 1995). As both the outcome of research activities and artifacts purposefully designed to enroll other scientists, scientific texts were analyzed as peculiar assemblages of terms leading to specific ways of problematizing issues by defining the roles assigned to relevant entities (researchers, substances, tools, and technologies). By tracing new associations between words, one could thus track the emergence of research fronts and capture their dynamics. In contrast to the program of building a “science of science”—initially sketched by Price (1963) and recently respecified by Fortunato, Bergstrom, et al. (2018) as an endeavor that “places the practice of science itself under the microscope, leading to a quantitative understanding of the genesis of scientific discovery, creativity, and practice and developing tools and policies aimed at accelerating scientific progress”—co-word analysis strived to provide a coherent conceptual and methodological framework to overcome the tension between ethnographic descriptions and quantitative approaches often devised to answer different questions (Callon, 2001; Callon et al., 1983).

This attempt at developing a coherent framework was not prompted by an abstract yearning for methodological consistency. Rather, developments in the substantive domains investigated by STS practitioners raised a new challenge as the natural and biomedical sciences underwent a profound transformation of their working patterns. Large research structures such as consortia and networks embodied a “collective turn” that, as ANT practitioners warned, should not be reduced to a mere increase in the number of authors cosigning papers but, rather, encompasses the heterogeneous nature of these emerging collectives. Heterogeneity, in this context, refers to specific activities that associate a variety of practitioners with a motley of emerging bio-clinical entities, such as genes and mutations, as part of a continuing process of stabilization and (re)qualification (Cambrosio, Bourret, et al., 2014). A search for new methods was prompted by the following conundrum: While thick descriptions of selected sites missed the configurational dimensions of the collectives, resorting to a few quantitative indicators to account for configurational complexity destroyed for all practical purposes the very phenomena under investigation (Callon, 2001).

A number of developments channeled methodological innovation in quantitative approaches to the analysis of techno-scientific activities. First, the collective turn was accompanied by a massive growth in electronic publications and related digital archives that opened the way for “metaknowledge” investigations of the heterogeneous components of scientific activities (Evans & Foster, 2011). On the conceptual side, natural scientists (in particular physicists, biologists, and computer scientists) became interested in investigating large-scale real-world networks. Since the 1970s, traditional social network analysis—often relying on the topological analysis of small-size ego-networks sampled through surveys and interviews—had become an accepted subfield of sociology (Freeman, 2004). In contrast, the aforementioned natural scientists who migrated to the field of complex network analysis promoted a new understanding of large-scale real-world networks, showing that they shared a number of properties (e.g., small-world effects, scale-free nature) and could be analyzed using dedicated tools (e.g., network morphogenesis models and innovative network topology metrics; Barabási & Réka, 1999; Watts, 2004; Watts & Strogatz, 1998). Most relevantly for our present purpose, they introduced new computational methods for the analysis of very large data sets, allowing the generic modeling of the structure of multiplex networks (featuring various kinds of edges) and their dynamics (Mucha, Richardson, et al., 2010; Palla, Barabási, & Vicsek, 2007; Rosvall & Bergstrom, 2010).

One could argue, as suggested by a reviewer of this text, that the contrast between earlier citationist methods and network analysis was more a matter of conceptual background and

research horizons than of technical differences. Indeed, while the former approach was largely defined by its focus on research policy and evaluation, compounded by lack of reflexivity, the latter was centered on modeling and mapping the dynamics of knowledge production. But technical differences clearly mattered. In particular, in addition to prestructured categories such as citations and authors' names mobilized by bibliometrics, researchers engaged in network analysis also began exploring semantic approaches, using Natural Language Processing (NLP) to extract new sets of variables from texts. They replaced unidimensional data analyses based on limited amounts of data or highly aggregated data sets, with broader analyses of heterogeneous data, resorting to relational approaches and complex data visualization tools. Obviously, this contrast is an ideal-typical one, as there are several examples of hybrid approaches that, while rooted in the citation tradition, borrowed insights from network analysis, embracing its analytical visualization techniques and introducing innovative solutions such as overlay maps (Rotolo, Rafols, et al., 2017; Van den Besselaar & Heimeriks, 2006). Still, it does capture significant trends in the field.

The new analytical approaches mobilizing advanced network structure analysis tools (Blondel, Guillaume, et al., 2008; Palla et al. 2007) and NLP workflows (e.g., Van Eck & Waltman, 2011) were quickly embedded in a growing number of publicly available software platforms, such as ReseauLu (e.g., Jones, Cambrosio, & Mogoutov, 2011), VOSviewer (Van Eck & Waltman, 2009), or CorTexT (e.g., Tancoigne, Barbier, et al., 2014; Weisz, Cambrosio, & Cointet, 2017). These platforms allow investigators to explore relationships between heterogeneous entities (e.g., diseases, institutions, substances) within a single map. They use dimension scaling (Van Eck & Waltman, 2007) or force-directed graph layout algorithms (Fruchterman & Reingold, 1991) that depend only on the topology of the network. This latter family of algorithms models each entity as an object connected to other objects by more or less strong elastic springs, depending on the strength of the links between entities. Repulsive forces are simultaneously used to prevent pairs of nodes from getting too close. A dynamic positioning algorithm optimizes the position of all of the nodes in order to minimize the stress of the layout. The proximity of two entities is not directly representative of the specific strength of relationship between them but instead represents the overall set of relationships of that entity and the other entities to which it is specifically linked.

Unsurprisingly, several ANT practitioners, in particular those working at the médialab established at SciencesPo Paris in 2009 under the leadership of Bruno Latour (<https://medialab.sciencespo.fr/>) but also in Amsterdam around the media studies program put forward by Richard Rogers, quickly embraced the analytical opportunities offered by these new developments, albeit not without reservations. No doubt, new computational approaches did change the conversation, redefining the relation between conceptual developments and quantitative approaches, but they did not dispense with the aforementioned essential tension between conceptual developments and quantitative approaches within STS. How so, and why?

2. FROM PAST TO PRESENT: ISSUES AND PROBLEMS

What are the main criticisms raised against the new computational approaches? A first issue has to do with the fact that although people tend to conflate actor-networks with digital networks (even more so in the era of social media), actor-networks and social networks are two very different kinds of things (Latour, 2011). The fact that the term *network* can refer to four possible meanings, namely “a conceptual metaphor, an analytic technique, a set of data, a sociotechnical system” (Venturini, Munk, & Jacomy, 2019, p. 513) contributed to this confusion. Moreover, as argued in more detail in Cambrosio et al. (2014), network-like representations generated by the aforementioned software platforms tend to promote structural and strategic interpretations, rather than dynamic analyses. The dynamics of an actor-network have heterogeneous roots, cannot be equated

to changes in the morphology of network representations, and are often accounted for by the subsequent intervention of entities that were not included in the initial descriptions. Similarly, equating clusters on a map to collectives is questionable insofar as the latter cannot be reduced to the presence of collaborative ties but, rather, are defined by the reorganization of work around emerging entities (Rabeharisoa & Bourret, 2009). Arguably, notions such as assemblage (DeLanda, 2016) or “agencement,” with its distributed-agency undertones (Callon, 2017), are better suited to capture the dynamics of actor-networks than the network metaphor. In short, network mapping comes with built-in epistemological models that are not necessarily compatible with STS research agendas.

And yet, in spite of these openly acknowledged shortcomings, some ANT practitioners have argued that because of its “figurative power,” network mapping using the aforementioned force-directed layouts still represents a useful, viable strategy for exploring scientific and technological activities, provided that one deploys them as components of an argument encompassing the heterogeneous nature and the mutually defining elements that account for the dynamics of actor-networks (Latour et al., 2012). While we have also implicitly adopted a similar position in our own work, arguing for instance that one could use networks to produce something different from networks, we now believe that the time has come to go beyond (digital) networks with their simple dyadic graph representations connecting homogeneous entities (i.e., graphs that can only model interactions as links connecting pairs of entities of the same type). In the remainder of this paper we would like to discuss a few promising alternatives, such as hypergraphs, that more closely adhere to the specifications of a sociology of translation. Hypergraphs offer means of exploring the specific assemblages of persons, skills, technologies, entities, institutions, organizations, and claims that define a given collective and its distributed agency, as well as of accounting for its dynamics and thus for the heterogeneous ways in which its diverse components are linked, forming higher order groups and hierarchies that can eventually be qualified as part of different regimes of engagement (Thévenot, 2006).

To our knowledge, the closest equivalent to an ANT-informed quantitative approach is provided in an article by Shi, Foster, and Evans (2015). Explicitly mentioning the sociology of translation as its main conceptual referent, the article models science as a dynamic hypergraph, contrasting single-mode standard networks (e.g., coauthorship or co-word networks) with the multimode character of scientific activities and related “wanderings” across different kinds of things, namely people, methods, diseases, and substances. A case study of millions of abstracts from MEDLINE provides strong empirical support for the viability and heuristic value of this approach (see Figure 1 in Shi et al. [2015] for a visualization of a random sample of the MEDLINE hypergraph). Conceptually, the article also resorts to James March’s well-known “garbage-can model” (Cohen, March, & Olsen, 1972) to formalize techno-scientific assemblages and assembly processes as hypergraphs “in which articles are hyperedges and contain nodes of several distinct types.”

To understand the originality of the model developed by Shi et al. (2015) it is useful to contrast it with traditional mapping approaches that focus on a single analytical dimension, such as coauthorship ties as proxies for collaborative endeavors (Moody, 2004), the semantic structure of a given field as elicited, for instance, by topic modeling (Griffiths & Steyvers, 2004), or cocitation networks of the core references of a given field (Small, 1999). Bourret, Mogoutov, et al. (2006) have used network analysis to investigate the interplay between human and nonhuman actors, but they did so by visually inspecting maps produced with a traditional dyadic model of interaction rather than hyperedges. Other scholars have investigated the possible articulation of descriptions and aggregates produced by different approaches, for instance by combining cocitation and co-word analysis (Braam, Moed, & Van Raan, 1991; Zitt & Bassecoulard, 2006). However, they simultaneously appraised only two dimensions.

In contrast, Shi et al.'s (2015) model—which partly capitalizes on previous efforts to use hypergraphs to capture the joint sociosemantic dynamics of team formation (Taramasco, Cointet, & Roth, 2010; see also Falk-Krzesinski, Börner, et al., 2010)—integrates a number of different components of a scientific publication within a joint framework, thus allowing for the proper measurement and modeling of the heterogeneous topological relationships between distinct types of entities, such as authors and chemicals. One should note, however, that insofar as the sets of entities deployed by individual articles are reduced to a number of dyadic interactions between these constituents, the resulting metaknowledge network does not provide an entirely satisfactory model of the processes characterizing assemblages, such as upward causality (i.e., emergent properties), downward causality (i.e., the effect of the whole on its constituents), or the degree of homogenization (DeLanda, 2016). Indeed, one wonders whether or to what extent sophisticated qualitative theorizing of this kind of dynamic assembling processes can in fact be translated into quantitative analyses. Arguably, rather than a conflation of qualitative and quantitative analyses, one should develop a trading zone where they could cross-fertilize each other. The multidimensional correlations that may be found, for instance, within manifolds consisting of more than two entities (Baudot, Tapia, et al., 2019) certainly call for additional modeling efforts to identify heterogeneous complex patterns as an alternative to tracing coarser correlations between phenomena elicited from different analytical dimensions. In this respect, combining hypergraphs and stochastic block models may offer a promising option insofar as the latter can be used to measure the likelihood of any configuration of heterogeneous entities, as seen in recent inquiries that adopt a purely quantitative approach (Shi & Evans, 2019) or include qualitative considerations (Abdo, Cointet, et al., 2019).

The conversion of textual content into analytical data that can be mobilized for specific purposes is a challenging process that largely accounts for the difficulties encountered by scholars who have been attempting to develop quantitative models of the dynamics of scientific activities compatible with the conceptual specifications of qualitative STS models. Although the content of scientific databases is extremely rich, several issues have long plagued their exploitation. First, the principle of generalized symmetry (e.g., Callon, 1984), has contributed to a “flat” treatment of language that only takes into account associations between words irrespective of the qualities attributed to them. In co-word analysis, for instance, articles are modeled as sets of keywords with no hypothesis about their distinctive role in a given publication. As a result, experimental devices, biological entities, or statistical apparatuses all receive equal treatment and attention. Yet, it is crucial to distinguish between these different entities according to the (shifting) ontological categories to which they are assigned within a given experimental system or at a given point in time—as noted by Rheinberger (2009), an epistemic entity can turn into a technological object. Their propensity to mutually interact will depend critically on those broader attributes and categories, each defining a different “mode of existence” (Simondon, 1958). Again, Shi et al. (2015) and also Li, Zhu, and Chen (2009) offer an alternative to such flat modeling. The approach they champion, however, also needs to take into account domain-dependency; in other words, the “same” entities mobilized in a given subfield differ from domain to domain, by the same token espousing multiple ontologies¹. Named-entity recognition algorithms (see Nadeau and Sekine (2007) for a review) could be very useful for establishing the situated roles played by different textual markers within scientific texts. Recent developments in NLP such as word embeddings that resort to neural architectures for training (Devlin et al., 2018) have pushed further the state-of-the-art performance in terms of drugs, genes, prescriptions, and other kinds of named-entities recognition (Lample, Ballesteros, et al., 2016). The automatic detection of relations between entities is increasingly

¹ But see Ribes, Hoffman, et al. (2019) on “domain independence.”

effective (Bekoulis, Deleu, et al., 2018), paving the way to more accurate ways to model how these entities are enrolling each other beyond the joint co-occurrence of terms on which co-word analysis relies. Put it differently, fine-grained NLP methods may allow us to better characterize what the hyphen in *co-word analysis* stands for.

In addition to extracting specific entities (drugs, techniques, etc.) from scientific texts, machines can capture the context in which hypotheses, claims, or pieces of evidence are deployed. Scientific articles have long been described by STS analysts as literary devices aiming at enrolling readers and funneling their attention by carefully distributing claims and statements of different nature and degree of generality in different sections of the text (Law, 1983). Past experiments and general considerations are often staged in introductory sections, evidence claims located in the results section, promissory notes and future developments consigned to the concluding remarks. Contextual information of this kind could be incorporated in the modeling of texts (Guo, Korhonen, & Poibeau, 2011). Similarly, it is now possible to integrate the context or the general purpose for which a given reference is being cited (Jurgens, Kumar, et al., 2018). For instance, is a citation used to solidify an argument or as a contentious source of knowledge? Advanced dependency methods and more largely AI-powered text analysis methods, including semantic hypergraphs, allow for the detection and characterization of claims (Menezes & Roth, 2019) beyond the sole co-occurrence of words within the same sentence or paragraph. These methods open the way to a more refined analysis of evidence construction modalities.

3. CONCLUDING REMARKS

In her book on “digital sociology,” Noortje Marres (2017) raises a number of interesting issues that apply, *mutatis mutandis*, to STS. Faced with the question of whether the investigation of digital regimes requires the development of new methods, she points to the presence of *both* continuities and discontinuities between traditional and recent quantitative approaches. She therefore pleads for the adoption of “interface methods” that specifically interrogate the relation between different methodological traditions, including qualitative ones. Her proposal is a useful antidote to the careless way in which “much computational social science projects simply go along with whatever ontology, epistemology or methodology is wired in to the platforms, packages or tools they use to capture, analyze and visualize data, without querying whether and how they are appropriate to the research project at hand” (p. 187; see also Cambrosio et al., 2014 for a similar argument). Given that many STS scholars have insisted on reflexivity as a key aspect of their domain, it is actually surprising that this issue has not attracted more widespread attention despite recent calls for a more balanced collaboration between ethnographers and computational sociologists (Evans & Foster, 2019; Goldberg, 2015).

Beyond the lack of reflexivity lurks the thorny issue of the adequacy between quantitative science studies and conceptual aspects of leading STS approaches, in particular ANT, with its admittedly ambiguous terminological reference to “networks.” As mentioned in the introductory section, while calculations, broadly understood, played an important role in the beginnings of STS, quantitative and qualitative approaches subsequently strongly diverged, leaving an increasing gap that only a few scholars have tried to bridge. Although developed for a different purpose, namely to add judgment to the notion of calculative agency or, in other words, to “think in the same terms about (quantitative) calculations and (qualitative) judgments,” the notion of “qualculation” (Callon & Law, 2005), if applied reflexively to the present discussion, could provide a useful heuristic for interrogating the fault lines between the increasing offer of computational tools in search of possible uses, and the conceptual specifications of those brands of science studies interested in exploring the epistemic and ontological dimensions of techno-scientific activities.

ACKNOWLEDGMENTS

We would like to thank Loet Leydesdorff and a second, anonymous reviewer for their extremely useful comments and suggestions.

COMPETING INTERESTS

The authors have no competing interests.

FUNDING INFORMATION

Research for this paper was made possible by a grant from the Canadian Institutes of Health Research (MOP-142478).

DATA AVAILABILITY

This is a position paper, not an article based on original data.

REFERENCES

- Abdo, A. H., Cointet, J. P., Bourret, P., & Cambrosio, A. (2019). Domain-topic models with chained dimensions: Modeling the evolution of a major oncology conference (1995–2017). *arXiv preprint*, <http://arxiv.org/abs/1912.13349>
- Barabási, A. L., & Réka, A. (1999). Emergence of scaling in random networks. *Science*, *286*, 509–512.
- Baudot, P., Tapia, M., Bennequin, D., & Goillard, J. M. (2019). Topological information data analysis. *Entropy*, *21*(9), 869.
- Bekoulis, G., Deleu, J., Demeester, T., & Develder, C. (2018). Joint entity recognition and relation extraction as a multi-head selection problem. *Expert Systems with Applications*, *114*, 34–45.
- Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008.
- Bourret, P., Mogoutov, A., Julian-Reynier, C., & Cambrosio, A. (2006). A new clinical collective for French cancer genetics: A heterogeneous mapping analysis. *Science, Technology, & Human Values*, *31*(4), 431–464.
- Braam, R. R., Moed, H. F., & Van Raan, A. F. J. (1991). Mapping of science by combined cocitation and word analysis. I. Structural aspects. *Journal of the American Society of Information Science*, *42*(4), 233–251.
- Callon, M. (1984). Some elements of a sociology of translation: Domestication of the scallops and the fishermen of St Brieuc Bay. *The Sociological Review*, *32*(1_suppl), 196–233.
- Callon, M. (1995). Four models for the dynamics of science. In S. Jasanoff, G. E. Markle, J. C. Petersen, & T. Pinch (Eds.). *Handbook of science and technology studies* (pp. 29–63). Thousand Oaks, CA: Sage Publications.
- Callon, M. (2001). Les méthodes d'analyse des grands nombres peuvent-elles contribuer à l'enrichissement de la sociologie du travail? In A. Pouchet (Ed.). *Sociologies du travail: Quarante ans après* (pp. 335–354). Paris: Elsevier.
- Callon, M. (2017). *L'emprise des marchés. Comprendre leur fonctionnement pour pouvoir les changer*. Paris: La Découverte.
- Callon, M., Courtial, J. P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information*, *22*(2), 191–235.
- Callon, M., Law, J., & Rip, A. (1986). Qualitative scientometrics. In M. Callon, J. Law, & A. Rip (Eds.). *Mapping the dynamics of science and technology* (pp. 103–123). Houndmills: Macmillan.
- Callon, M., & Law, J. (2005). On qualculation, agency, and otherness. *Environment and Planning D: Society and Space*, *23*, 717–733.
- Cambrosio, A., Bourret, P., Rabeharisoa, V., & Callon, M. (2014). Big data and the collective turn in biomedicine: How should we analyze post-genomic practices? *Tecnoscienza*, *5*(1), 13–44.
- Cohen, M. D., March, J. G., & Olsen, J. P. (1972). A garbage can model of organizational choice. *Administrative Science Quarterly*, *17*(1), 1–25.
- DeLanda, M. (2016). *Assemblage theory*. Edinburgh: Edinburgh University Press.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. B. (2018). Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint*, arXiv:1810.04805.
- Edge, D. (1977). Why I am not a co-citationist. *Society for Social Studies of Science Newsletter*, *2*, 13–19.
- Edge, D. (1979). Quantitative measures of communication in science: A critical review. *History of Science*, *17*, 102–134.
- Evans, J. A., & Foster, J. G. (2011). Metaknowledge. *Science*, *331*, 721–725.
- Evans, J. A., & Foster, J. G. (2019). Computation and the sociological imagination. *Contexts*, *18*(4), 10–15.
- Falk-Krzesinski, H. J., Börner, K., Contractor, N., Fiore, S. M., Hall, K. L., ... Uzzi, B. (2010). Advancing the science of team science. *Clinical and Translational Science*, *3*(5), 263–266.
- Fortunato, S., Bergstrom, C., Börner, K., Evans, J., Helbing, D., ... Barabási, A.-L. (2018). Science of science. *Science*, *359*(6379), eaao0185.
- Freeman, L. C. (2004). *The development of social network analysis. A study in the sociology of science*. Vancouver: Empirical Press.
- Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software. Practice & Experience*, *21*(11), 1129–1164.
- Goldberg, A., 2015. In defense of forensic social science. *Big Data & Society*, *2*(2), 2053951715601145.

- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(suppl 1), 5228–5235.
- Guo, Y., Korhonen, A., & Poibeau, T. (2011). A weakly supervised approach to argumentative zoning of scientific documents. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 273–283). Association for Computational Linguistics.
- Jones, D., Cambrosio, A., & Mogoutov, A. (2011). Detection and characterization of translational research in cancer and cardiovascular medicine. *Journal of Translational Medicine*, 9, 57.
- Jurgens, D., Kumar, S., Hoover, R., McFarland, D., & Jurafsky, D. (2018). Measuring the evolution of a scientific field through citation frames. *Transactions of the Association for Computational Linguistics*, 6, 391–406.
- Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., & Dyer, C. (2016). Neural architectures for named entity recognition. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 260–270).
- Latour, B. (1976). Including citation counting in the system of actions of scientific papers. Paper presented at the First Annual Meeting of the Society for Social Studies of Science, 4–6 November, Cornell University.
- Latour, B. (2011). Networks, societies, spheres: Reflections of an actor-network theorist. *International Journal of Communication*, 5, 796–810.
- Latour, B., & Woolgar, S. (1979). *Laboratory life. The social construction of scientific facts*. Beverly Hills: Sage.
- Latour, B., Jensen, P., Venturini, T., Grauwin, S., & Boullier, D. (2012). The whole is always smaller than its parts—A digital test of Gabriel Tarde's monads. *The British Journal of Sociology*, 63(4), 590–615.
- Law, J. (1983). Enrôlement et contre-enrôlement: Les luttes pour la publication d'un article scientifique. *Social Science Information*, 22, 237–251.
- Li, J., Zhu, X., & Chen, J. Y. (2009). Building disease-specific drug-protein connectivity maps from molecular interaction networks and PubMed abstracts. *PLOS Computational Biology*, 5(7), e1000450.
- Marres, N. (2017). *Digital sociology*. Cambridge (UK): Polity Press.
- Menezes, T., & Roth, C. (2019). Semantic Hypergraphs. *arXiv preprint*, arXiv:1908.10784.
- Moody, J. (2004). The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American Sociological Review*, 69(2), 213–238.
- Mucha, P. J., Richardson, T., Macon, K., Porter, M. A., & Onnela, J. P. (2010). Community structure in time-dependent, multiscale, and multiplex networks. *Science*, 328(5980), 876–878.
- Nadeau, D., & Sekine, S. (2007). A survey of named entity recognition and classification. *Linguisticæ Investigationes*, 30(1), 3–26.
- Palla, G., Barabási, A. L., & Vicsek, T. (2007). Quantifying social group evolution. *Nature*, 446(7136), 664.
- Price, D. J. de Solla (1963). *Little science, big science*. New York: Columbia University Press.
- Price, D. J. de Solla (1965). Networks of scientific papers. *Science*, 149(3683), 510–515.
- Price, D. J. de Solla (1975). *Science since Babylon*. New Haven: Yale University Press.
- Rabeharisoa, V., & Bourret, P. (2009). Staging and weighting evidence in biomedicine: Comparing clinical practices in cancer genetics and psychiatric genetics. *Social Studies of Science*, 39, 691–715.
- Rheinberger, H. J. (2009). Recent science and its exploration: The case of molecular biology. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 40, 6–12.
- Ribes, D., Hoffman, A. S., Slota, C. S., & Bowker, G. C. (2019). The logic of domains. *Social Studies of Science*, 49, 281–309.
- Rosvall, M., & Bergstrom, C. T. (2010). Mapping change in large networks. *PLOS ONE*, 5(1), e8694.
- Rotolo, D., Rafols, I., Hopkins, M. M., & Leydesdorff, L. (2017). Strategic intelligence on emerging technologies. Scientometric overlay mapping. *Journal of the Association for Information Science and Technology*, 68, 214–233.
- Shi, F., & Evans, J. (2019). Science and technology advance through surprise. *arXiv preprint*, arXiv:1910.09370.
- Shi, F., Foster, J. G., & Evans, J. A. (2015). Weaving the fabric of science: Dynamic network models of science's unfolding structure. *Social Networks*, 43, 73–85.
- Simondon, G. (1958). *Du mode d'existence des objets techniques*. Paris: Aubier.
- Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24(4), 265–269.
- Small, H. (1999). Visualizing science by citation mapping. *Journal of the American Society for Information Science*, 50(9), 799–813.
- Tancoigne, E., Barbier, M., Cointet, J. P., & Richard, G. (2014). The place of agricultural sciences in the literature on ecosystem services. *Ecosystem Services*, 10, 35–48.
- Taramasco, C., Cointet, J. P., & Roth, C. (2010). Academic team formation as evolving hypergraphs. *Scientometrics*, 85(3), 721–740.
- Thévenot, L. (2006). *L'action au pluriel. Sociologie des régimes d'engagement*. Paris: La Découverte.
- Van den Besselaar, P., & Heimeriks, G. (2006). Mapping research topics using word-reference co-occurrences: A method and an exploratory case study. *Scientometrics*, 68(3), 377–393.
- Van Eck, N. J., & Waltman, L. (2007). VOS: A new method for visualizing similarities between objects. In H.-J. Lenz & R. Decker (Eds.). *Advances in data analysis: Proceedings of the 30th Annual Conference of the German Classification Society* (pp. 299–306). Heidelberg: Springer.
- Van Eck, N. J., & Waltman, L. (2009). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538.
- Van Eck, N. J., & Waltman, L. (2011). Text mining and visualization using VOSviewer. *arXiv preprint*, arXiv:1109.2058.
- Venturini, T., Munk, A.K., & Jacomy M. (2019). Actor-network versus network analysis versus digital networks. Are we talking about the same networks? In J. Vertesi & D. Ribes (Eds.). *digitalSTS. A field guide for science & technology studies* (pp. 510–523). Princeton: Princeton University Press.
- Watts, D. (2004). The “new” science of networks. *Annual Review of Sociology*, 30(1), 243–270.
- Watts, S. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature*, 393, 440–442.
- Weisz, G., Cambrosio, A., & Cointet, J. P. (2017). Mapping global health: A network analysis of a heterogeneous publication domain. *BioSocieties*, 12(4), 520–542.
- Zitt, M., & Bassecouard, E. (2006). Delineating complex scientific fields by an hybrid lexical-citation method: An application to nanosciences. *Information Processing & Management*, 42(6), 1513–1531.