Modelling and measuring the dynamics of scientific communication

Lucio-Arias, D.

Citation for published version (APA):
Chapter 1. Introduction

Knowledge is increasingly considered one of the main resources for the development of contemporary societies. Concepts such as knowledge-intensive industries or knowledge-based innovations exemplify the importance of knowledge as a potential source of economic growth. Different policy measures aiming at increasing the knowledge base of our societies indicate that knowledge can also improve the quality of our social existence. Knowledge is fundamental since it allows us to understand and relate better to the reality in which we are immersed and to each other.

There are different ways of knowing: through experience, through the experience of others, through learning and interpreting. We know about our social and natural environment, and we also know in the form of skills and competences (Johnson, Lorenz, & Lundvall, 2002). There are also different types of knowledge. On one hand, some knowledge is tacit, it is rooted in our own individualities, and it is difficult to share or transfer. On the other hand, codified knowledge has been regarded as “formal” knowledge. It can be written down, read, shared. More specifically, discursive knowledge emerges from the exchange of meaning in interactions; it emerges as a property of a social system.

In this dissertation I provide a framework for modelling and measuring the emergence of scientific knowledge—which I argue, is discursive by nature. I take a communication-systems perspective to model the production and organization of scientific knowledge which emerges from discursive practices among scientists. Science as a social institution is integrated and articulated to our societies (Blume, 1974). Scientific knowledge is manufactured inside laboratories and research groups which are contingent upon socio-institutional and temporal factors (Whitley [1984], 2000; Knorr-Cetina, 1981). This makes scientists’ individual actions contextually embedded.

How do these social institutions transcend their local contingencies to integrate their scientific findings and argumentations into disciplinary repertoires? Results obtained inside laboratories, through field work or experimentation, collectively or through the reflexive enterprise of individual scientists, need to be

---

1 just in Europe: the Lisbon Strategy, The six and seventh Framework Programme. In the United States, “an agenda that will require us first and foremost to train and educate our workforce with the skills necessary to compete in a knowledge-based economy”; (Obama, 2008) see http://www.barackobama.com/2008/06/09/remarks_of_senator_barack_obama_76.php for full speech
communicated and shared among the relevant scientific communities. This is a requirement not only for the contributor to be accepted as a member of the scientific community, but also because scientific contributions acquire meaning when they are effectively tested and criticized (Fröhlich, 1996). In addition, the process of standardizing conventional interpretations of contributions in the past provides them with new meaning (Small, 1978).

Discursive practices make scientific communication crucial for the most irreversible process in contemporary science, the progressive accumulation of knowledge. Scientists engage in dialogue with peers; they communicate their own results and findings, as well as their interpretations of the results and findings of scholars in the past. This constant dialogue contributes to the public wealth of scientific knowledge and validates the contributions of peers. Scientific knowledge is thus not only conditioned by the social and institutional contexts; it is intellectually conditioned by the conversations in which it is embedded as well.

Scientific knowledge is produced by discursive interactions among scientists, this justifies the use of the concept of autopoiesis to understand the processes of production and further organization of scientific knowledge. Maturana & Varela (1992, at p. 43) define autopoiesis as the “property of a system to produce its components which make up the network of transformations that produces the system through ongoing interactions.” Autopoiesis was introduced in biology to describe the nature of living systems (Varela, Maturana, & Uribe, 1974). Its application in social systems has been subject of many debates. Nevertheless, its metaphoric use in the social sciences is fruitful to understand the social operations that characterize social systems (Mingers, 2002; Morgan, 1986). Scientific knowledge, which is attained collectively, emerges from the network of discursive interactions between individual scientists. Conversations among scientists support the ongoing coordination that emerges at the systems level (Maturana, 1988).

This perspective based on communications facilitates the specification of macro-structure as emerging from the micro-level of contextually embedded actions. The macro-structure emerges as a consequence of the recursive nature of knowledge which makes its organization crucial. The intellectual organization in the form of disciplines and cognitive categories conditions the submission of further scientific findings. When scientific communications are assumed as the units of analysis,
preference is given to the constructive social interactions instead of the distinctions between social and cognitive dimensions (Cozzens, 1985).

A common wealth of scientific knowledge is nourished by the communication of scientific results. It is publicly available to members of a scientific community who, in their turn, contribute to its growth and take from its stock. The importance of communicating scientific results sustains a publication culture among scientists: contributions provide novel claims that act as stepping stones for further advances.

Alongside the communication of original knowledge claims, scientists refer back to archives of scientific knowledge in order to link their knowledge claims to previously established bodies of knowledge. Citation cultures have evolved from this communication practice (Wouters, 1999b). Through citations, the intellectual property of individual contributions is recognized when used by other scientists (Merton, 1968). Furthermore, scientists’ interest to make their contributions publicly available is regulated by their expectations to get credit and recognition for so doing.

When scientists make their own findings available to other scientists, the protection of her/his intellectual output is assured by the recognition of peers—“only when a scientist’s contribution becomes part of the public domain of science can they truly lay claim to it as theirs” (Merton, 1979, italics mine). The recognition by peers of having used an individual contribution acts as an acknowledgment that a valuable “scientific contribution” has been made (Gilbert, 1976). The value of a scientific contribution depends on the way in which it is received, perceived and used by other scientist (Amsterdamska & Leydesdorff, 1989). Previous contributions constitute a repertoire of ideas that scientists use—or not—making the development of science contingent on the selection, perception and utilization of scientific contributions.

Science can thus be understood as a system of knowledge production where the competitive pursuit of reputations for contributions plays an important part (Whitley [1984], 2000). As scientists’ communicate their scientific experiences and feed on the experiences of their peers, networks of expectations are shaped constraining the allocation of reputations. Although such networks emerge from the scientists communicative acts, they remain latent to the scientists (Leydesdorff, 1998). The networks shaped by these communicative interactions have social and cognitive meaning. Since scientists formalize their communication through publications the networks have a textual dimension as well.
Written communication is crucial for the production and organization of scientific knowledge. This enables us to build models of these processes using scientific literature as data. The use of the literary footprint of scientific communication to study the dynamics of science was proposed by Derek de Solla Price in his paper “Networks of Scientific Papers” (1965). This use of scientific literature to model developments in the sciences follows the previous claims that communication is essential for the functioning of the system of scientific knowledge production and control. Accordingly, publications can be considered as the crucial events in the model because the circulation of scientific knowledge is coupled to the circulation of scientific texts (Fujigaki, 1998a).

The purpose of this dissertation is to measure and model developments in the system of knowledge production. A representation based on scientific texts facilitates this task. When reporting their findings, scientists emphasize some cognitive features compromising others (Gilbert & Mulkay, 1983; Knorr-Cetina 1981). Studies in the sociology of scientific knowledge have emphasized the importance of socio-cultural contexts when articulating scientific arguments (Collins, 1983). In this dissertation I suggest that socio-cultural contexts are recursively constrained as well by the emergent properties of the evolutionary operations of the system producing scientific knowledge—and could thus be also analyzed using the proposed perspective.

The properties and characteristics of the system emerge from the scientists’ communications. Even though they can be reinforced by the continuous communication, these properties are time specific and subject to change. Systems can be expected to evolve to higher forms of complexity resulting in functional differentiation which allows the system as a whole to process more variety (Collier, 2004). The socio-cultural contexts brought forward by studies in the Sociology of Scientific Knowledge, are affected by the intellectual organization of scientific texts into cognitive dimensions. ²

Two parallel procedures are fundamental to understand the system of knowledge production from an evolutionary perspective: variation and selection. While variation provides the system continuity following the arrow of time, selection operates as a feedback mechanism. The existence of a feedback operation is related to

² The social-cultural contexts are social implications of the intellectual organization. This illustrates that the cognitive, social and textual can also be considered as coupled and intertwined like in a triple helix (Leydesdorff, 1995).
the autopoietic property of the system. The system uses the information retained in the selections as input for its own development creating recursive loops of selections.

The recursive feedback means that the past is retained and can be reinforced by the present. This implies that the operation of the selection mechanisms may begin to structure the variation potentially reinforcing historical trajectories. Scientists select validated knowledge and differentiate their own results to submit them for further publication. The recursive process of selection and codification makes the development of scientific knowledge non-linear (Fujigaki, 1998b).

Since I am using a literary simplification, the properties of the system are functionally simplified as interactions among scientific texts. The memory of the system is carried by the citations and the dynamics of the system are determined by the trade-off between the introduction of new knowledge claims and the recursive selection of previous knowledge claims to support new findings.

New publications modify the network structures updating them along trajectories. The recursive selection of significant knowledge claims updates the system of knowledge production and it provides it with subsymbolic mechanisms of control. In the literary model, control is exercised by the network structures generated by the publications. In the system of knowledge production, this network represent agreements and disagreements about relevant theories, choices of methods and objects of inquiry that emerge from the communications and that can be expected to influence the formulation of new scientific claims.

Change is carried by the diffusion of literature along trajectories. The recursive selection of supporting claims in the intellectual dimension can reverberate in differentiation inside the system causing non-linear change. Functional differentiation causes the emergence of a new discipline or scientific specialty. The literary representation of the system of scientific knowledge production is constantly changing. Linear change responds to the variation carried by the streams of new publications. Non-linear change is carried by the recursive operation of the selection mechanisms. When the selection of knowledge claims to support new findings behaves semi-deterministically, the system will expand according to its trajectories. The re-interpretation of previous findings might be such that the history of the specialty is partially re-written, in this case, the knowledge claims used to support new findings are subject to change. This change can be translated in the network as a new cluster representing a differentiated social community and cognitive trajectory.
References to previous literature and contained in the publications formalize the selection process. The historical operation of the selection mechanisms allows the system to organize along intellectual dimensions. Codes of communication emerge from this organization. The codes are defined in terms of, for example, relevant journals to communicate findings in specific topics as well as in terms of specific “jargons” used by the specialty (Leydesdorff, 2002). The perception of and recognition by citing authors is subject to change (Small, 1978; Leydesdorff, 1998). Scientists’ socio-cultural context has to correspond to these changes—to the differentiation generated at the systems level—in order to integrate their scientific contributions to the bodies of knowledge.

In essence, the codes of communication act as organizing principles: they structure variation over time. They emerge as a result of the refinement and elaboration of theoretically relevant arguments that organize scholarly discourse. While the emergence from the communicative interactions is bottom-up, they exercise (top-down) control in future publications. This refers to the earlier introduced notion of autopoiesis: the system produces the networks of interactions that produce and reproduce the system. The system produces the conditions that regulate the production of new scientific knowledge but these conditions are subject to change and are updated as a consequence of the production of new scientific knowledge.

Using the model based on scientific literature, the dynamics, development and structure of science can be simulated in terms of the resulting networks that link scientific papers (Price, 1965). The networks are an emergent property of the scientists’ interactions. Scientific fields and specialties emerge from the organization of intellectual arguments following the aggregation of the citation practices of individual authors (Small, 1973; Small & Griffith, 1974; Griffith, Small, Stonehill, & Dey, 1974). In other words, citation practices define clusters that can be expected to represent socio-cognitive domains.

The processes of production and organization of new scientific knowledge is operationalized in this dissertation using the literary model. The role of citations as the retention mechanisms in the literary simplification, allows their use to analyze the intellectual organization that results from the operation of the codes of communication—i.e. from codification. Some of the tensions generated by the emergence of new knowledge reinforce existing codes while others trigger the system to differentiate.
Research specialties, as intellectually organized cognitive categories, can thus be analyzed in terms of the citation relations among documents (i.e. Price, 1965; Leydesdorff, 1987; Hjorland, & Albrechtsen, 1995; White & McKain, 1998; Chen, 2003). The relations among documents shape an archive with self-referentially structuring dynamics which set the conditions for authors to submit new knowledge claims. While the submissions provide variation, the evolving networks have a tendency to select deterministically.

Variation and codification shape two different types of literatures. A classical literature—in the archive—belongs to the validated knowledge base of a specialty. It reinforces the codes of communication that rule the codification. A transient literature—at the research front—might become obsolete after its publication; its importance might be temporal and its role is to provide variety to the system. This later type of literature is characteristic of the rapidly changing dynamics at research fronts (Chen, 2006).

Both the classical literature and the literature at the research front affect the evolutionary development of the system. The extent to which codification exerts disciplinary and intellectual control over the production of new knowledge at the research front is a varying phenomenon that needs to be explained and not assumed (Whitley [1984], 2000). In this dissertation I provide a methodology to measure intellectual organization in scientific specialties applying entropy statistics to the literary model of the system of scientific knowledge production.

From an information-theoretical perspective, entropy statistics originated from the mathematical theory of communication. Based on probabilistic distributions, it measures “how much choice is involved in the selection of an event or how uncertain we are of the outcome” (Shannon & Weaver, 1949, at p. 18, italics mine). Although developed to respond to an engineering problem about the transmission of information from a transmitter to a receiver through a communication channel, entropy statistics can generally be used as a measure of variety—of randomness. Entropy statistics can be used to measure the uncertainty contained in probability distributions.

Two specific measures of entropy are applied in this dissertation. The Kullback-Leibler divergence is used to assess critical transitions using publications as events and comparing the expected information value of an event to that value of preceding documents. Configurational information is used in later chapters to measure
how selection and variation might co-evolve in time reducing uncertainty at the research front of a scientific specialty.

Messages contained in the publications can be aggregated at a specialty level and then decomposed as frequency distributions of any set of textual attributes. This property of the literary model facilitates the use of entropy statistics. I assume the distributions of cited references and citing texts in order to assess critical transitions in the evolution of research fronts and specialties.

From an evolutionary perspective, critical transitions have been defined as gradual processes of societal changes in which social systems structurally change their character (Rotmans, Kemp & van Asselt, 2001). Critical transitions occur when dynamic systems suddenly shift to a contrasting regime (Sheffer, Bascompte, Brock, Brovkin, Carpenter, Dakos, Held, van Nes, Rietkerk, & Sugihara, 2009). The shift or gradual change does not discontinue the process of development of the system but indicates a change in the evolving historical trajectories.

Research specialties are expected to develop along trajectories defined by their past. But this past is overwritten continuously as new publications integrate the research front. Parts of the history can be obliterated or forgotten in hindsight resulting in shifts of the historically shaped trajectories. The concept of critical transitions is used to pinpoint those documents where a shift of the intellectual trajectories occurs. These points could be indicating discontinuity in the process of codification.

Publications will be considered as events and aggregated to measure the uncertainty contained in yearly sets of documents. I have argued that the development of scientific specialties is contingent upon selection and variation mechanisms. The configuration between variation and selection can be such that the codification resulting from the selections structures the variation. In this case, scientists at the research front are faced with less uncertainty about the codes of communication that guide the validation of their contribution. The reduction of uncertainty happens when the organization provided by the intellectual dimension reduces uncertainty generated by the variation introduced by new publications.

The properties of aggregation and decomposition of the literary model are used to simplify knowledge claims as configurations of title words and cited

---

1 Frequency distributions are transformed into probability distributions by calculating the probability of occurrence of each value in the distribution.
Introduction

references. Whereas novel combinations among words provide variation, the selection of context specific cited references leads the system towards stabilization reducing uncertainty. While this reduction of uncertainty will not be evident to the researcher, it will facilitate a straightforward positioning of his/her contribution along cognitive relevant dimensions.

Modelling the development of science in terms of evolving scientific literature makes the tensions between the processes of top-down codification and bottom-up diffusion of the variation amenable to measurement. In this sense, the contribution of this dissertation is primarily methodological with some empirical applications.

1. Case studies

Three different cases are used in this dissertation. They provide examples of the simplification of specialties using the literary model and the possibilities to analyze them using network algorithms and entropy statistics. The first two cases are related to nanoscience research; the third uses as an example of a more social-science specialty, namely, scientometrics itself.

The specific case of fullerenes and fullerene-like structures nanotubes was chosen at a moment when the nanotechnology hype was perhaps at its highest point. Fullerenes constituted a significant research front in nanoscience with its own terminology beyond the usage of the prefix “nano” (Schummer, 2004). A Nobel price was awarded for the discovery of fullerenes in 1984. The discovery of fullerene-like structures carbon nanotubes in 1991 followed the discovery of fullerenes. Sumio Iijima regarded his discovery as serendipitous while researching fullerenes (Iijima, 2005). Fullerenes pioneers regard the 1991 discovery as “fullerinizing” carbon nanotube research (Colbert & Smalley, 2002). A more detailed account of research in fullerenes and nanotubes and their relation follows in Chapter Three.

Although words can change in meaning depending on their context (Leydesdorff & Hellsten, 2005), one can expect that in the natural sciences, specific words can pinpoint discoveries that give constitutive rise to research specialties (as in the case of fullerenes and nanotubes). The social sciences behave differently having their own publication and citation cultures; new developments in this case are generated from metaphors. I chose the specialty of scientometrics as a case study akin to the social sciences. Scientometrics belongs to the social studies of science (Wouters, 1999b) but has increasingly differentiated its communications
Modelling and measuring the dynamics of scientific communication

(Leydesdorff, 2007a; Van den Besselaar, 2000; 2001). Furthermore, the importance of the introduction of the Citation Indexes as well as the older library practice of citation analysis allows contextualizing the development of the specialty within some important references.

2. Outline of the dissertation

Following this chapter, a theoretical framework is presented in Chapter Two. In this chapter, the autopoiesis of the system of scientific knowledge production and the implications for the growth and change of the system are further explained as well as the simplification of the system to a literary model. The literary model is used to provide a descriptive picture of the two related specialties of fullerenes and nanotubes in Chapter Three. Subsequent chapters illustrate the use of entropy statistics to characterize the development of scientific specialties for the three cases mentioned.

The Second Chapter—“Dynamic of exchanges and references among scientific text and the autopoiesis of discursive knowledge”—provides the theoretical framework of this dissertation. The definition of an autopoietic system of scientific communication and the role of publications as the basic operation of this system is explained. Three selection mechanisms are introduced in detail together with the implications of their operation for the system of knowledge production. The possibility to use entropy statistics in the literary model of the system of scientific communication is introduced in this chapter as well.

The Third Chapter—“From fullerenes to nanotubes: knowledge emergence in scientific communication”—provides my first empirical study. Using a model based on scientific literature, the development of fullerenes and the related specialty of nanotubes is detailed. The aim of this chapter is to provide an illustration of the disrupting effects of new discoveries in the networks of scientific communication. For this, the journal space of the journal created for the diffusion of research in this topic—Fullerenes Science and Technology—was examined considering the changes in the cited and citing environment through time. Changes in the semantic codification of both discoveries in a scientific context and in terms of its technological diffusion are explored using co-word analysis. The processes of the production of knowledge at the frontiers of science and technology are explored in this chapter including how existing knowledge converges into new discoveries to be later transmitted through the existing structures.
In Chapter Four, research in fullerenes and nanotubes is examined from the perspective of codification and diffusion. Different algorithms that build on the frequencies of citation relations are used to enhance the concept of an algorithmic historiography.

The notion of algorithmic historiography follows the introduction of HistCite™ into the scientometric community which aids the process of uncovering transmissions of knowledge that lead to scientific breakthroughs (Pudovkin & Garfield, 2002). It relies on citation data to describe historically scientific fields, specialties and breakthroughs (Garfield, 1979). The software creates a mini-citation matrix for any set of documents retrieved from the ISI Web of Science facilitating historical reconstructions based on a literary simplification of science (Garfield, Pudovkin & Istomin, 2002, 2003a; 2003b). Because citations carry the memory of the system, they can be used to build historiograms which can be further enhanced using algorithms from network and information theory. The algorithmic approach to the historical reconstructions enables us to include more variety in the perspective than the reconstruction based on a single narrative (Kranakis & Leydesdorff, 1989).

In Chapter Four, I use an algorithmic approach to reconstruct the history of fullerenes and nanotubes. The alternatives presented in this chapter extend the possible interpretations from historical reconstructions based on scientific literature. Main path algorithms are used to detect the structural backbones of the citation networks (Hummon & Doreian, 1989; Carley, Hummon, & Harty, 1993). These are constructed using the connectivity of the documents; i.e. the citations they receive and the citations they make. Algorithms based on entropy statistics are applied to distinguish potentially critical transitions in the networks. Critical transitions can only be detected in hindsight but they signal moments when the citation traditions might have changed in the discipline.

Configurational information is applied to measure the intellectual organization in scientific specialties in chapter five. The trade-off between codification and variation can be such that uncertainty is reduced at the research front. The specialty thus can process the surplus of information and absorb more variety. The configurational information stems from entropy statistics. It measures the synergy at the systems level when more than two sources of variance interact. The measure captures uncertainty prevailing from the interaction of more than two probability distributions. Three interacting dimensions can potentially co-evolve synergetically.
Reductions of uncertainty thus are a property of the system as a whole and can not be attributed to its parts (McGill, 1954).

The measured uncertainty is used as an indicator of intellectual organization for the specialties of fullerenes, nanotubes and scientometrics. For the case of scientometrics, the relevant discourse is operationalized as the dialogue between the publications in selected journals (Scientometrics, Journal of Information Science, Journal of Documentation, and Information Processing and Management). Furthermore, the indicator is also used for topics of research illustrating that these are not codified enough to reduce uncertainty for researchers at the research front.

The Sixth Chapter is the last chapter of this dissertation. It contains a summary and conclusions as well as some formulations on limitations and topics for further research.