Modelling and measuring the dynamics of scientific communication
Lucio Arias, D.P.
Chapter 6. Summary and Conclusions

Is the process of intellectual organization in the sciences amenable to measurement?

I have argued that self-organization in the intellectual dimension emerges from the communicative interactions between scientists and can be measured as reductions of uncertainty in the textual domain. The conceptualization of the development of science generated by the interdependent interactions between social, cognitive and textual factors follows from the assumption that the main function in science is the development of discursive knowledge. Assuming the perspective of communication-systems theory, which emphasizes the discursive nature of scientific knowledge, the development of the sciences can be modelled as co-evolutions between cognitive and social contexts which are retained in the textual dimension.

This communicative turn allows simplifying the system producing scientific knowledge to the literary footprint of scientific communication, commonly used in the field of scientometrics but seldom reflected upon. In the field of scientometrics, the metaphor of a “a tail wagging a dog” has been used recently to warn of the dangers of the use and design of scientometric indicators without appropriate theoretical reflections (Garfield, 2009). The contextualization of a literary model (Price, 1965; Garfield, 1979) together with the operationalization of the dynamics of the system in terms of attributes of scientific publications requires the consideration of sociological and philosophical reflections on the mechanisms of growth and change.

The circumstances surrounding knowledge production are concerns in sociological accounts of science (Hackett, Amsterdamska, Lynch, & Wajcman, 2008, at p. 4). But science is not only a social institution, and the predominant theories should respond to more than social factors (Newton-Smith, 2001, at p. 6). In the philosophy of science, concerns focus on the epistemological nature and the validity of knowledge claims. Emphasizing the discursive dimension as the driver of scientific progress allows bridging between people and cognitions in their contexts, between the sociological queries that characterize the Sociology of Scientific Knowledge and the philosophical concerns about methods and knowledge production. Accordingly, in this dissertation the social organization of science is not isolated from the knowledge produced (Whitley [1984], 2000, at p. 2).
A more in-depth reflection on the sociological and philosophical considerations regarding the production of scientific knowledge are contained in the first two chapters of this dissertation. They provide the methodological and theoretical contributions that were elaborated in the empirical studies. I proceed by summarizing the main theoretical arguments of the dissertation—Chapters One and Two—as well as the empirical findings which are contained in Chapters Three, Four, and Five. Some methodological as well as theoretical reflections follow the summary of the chapters. I conclude with a discussion of some limitations of the study and the formulation of further research questions.

1. Summary

In the first chapter, science was introduced as a system producing discursive knowledge. While approaches to the scientific enterprise from the Sociology of Scientific Knowledge have emphasized the context of production, the emphasis on communications used in this study allows us to conceptualize science as a system in permanent change where the socio-cognitive and the textual are intertwined constantly conditioning and enabling each other in their co-evolutions.

The generation of a functioning structure out of the interaction and interrelation of communications among scientists allows us to specify science as a communication system: a communication system producing discursive knowledge. The latter can be considered as a further refinement—codification—of meaning in scientific communications (Luhmann, 2002b; Leydesdorff, 2007b). Luhmann’s (1986; 1990a) suggestion to place emphasis on communications in order to understand how meaning is processed in social systems, was introduced in the second chapter to understand science as a system producing meaningful information.

Given the relationship between the interaction of communications and the functioning structure one can characterize the system producing scientific knowledge as autopoietic, that is, as the network of productions of components which realize the network that produced them and constitute the system (Maturana & Varela, 1980, p. 80). The use of autopoiesis to understand the dynamics in science required defining the output of the system as a consequence of the interactions between the elements of the system. For this reason, a further refinement was the assumption that scientific publication is the basic—autopoietic—operation of the system. The process of producing and validating knowledge claims was thus acknowledged as discursive by
Summary and conclusions

nature; it is generated from the communications among scientists. Furthermore, the
discursive practices are reflexive: scientists communicate according to and deviating
from what has been communicated in the past.

Earlier scientometricians had a similar conceptualization of science without
these semantic formalizations. Nalimov (1991) justified that science was a system
because the flow of publications had a structure. He argued that science had a

“a structure. It is structured by citations. Citations show, how
separate publications are related to one another, how science is
developing as a result of interaction of publications in a
particular domain and even in several domains. Thus what
citation characterizes is not the achievement of a separate
scientist but a contribution to the information process” (Nalimov,
in Wouters, 1999a, at p 90, italics mine).

I formalized the information process mentioned in this quotation as the
development of discursive knowledge. Following the definition that communications
are the main carriers of this knowledge, the system can be defined as the networks of
scientific publications and their interactions. Networks provide not only a
representation of the emergent intellectual structures but also the means to link the
structural properties to the dynamic process of producing scientific knowledge.
Discursive knowledge emerges from a process of refinement and elaboration of
arguments. Furthermore, this discursive characteristic of scientific knowledge
provides the means for the organization of the knowledge produced.

The definitive attribute that typified the production of scientific knowledge as
systematic is expressed in the relationships between the publications, which define the
identity of the system at any given moment, as well as its dynamics. These
relationships are the result of the operation of the selection mechanisms that are
detailed in Chapter Two. Three selection mechanisms were indicated as operating in
the production of new knowledge claims and providing the systems dynamics: (i) the
instantaneous positioning of a publication in the network in terms of its (e.g., citation)
relations with other texts, (ii) the positioning of this publication in relation to other
publications along the time axis, and (iii) the reflexive re-positioning and rewriting of
the knowledge content of the paper in the further process of developing discursive
knowledge (Lucio-Arias & Leydesdorff, 2009a).
The first two selection mechanisms relate to the mediation between the context of discovery and the context of justification that was introduced in Chapter One. An author performs the first selection when constructing the paper as an output of the context of discovery. Subsequently, the peer review processes that validate the knowledge claims articulate the context of justification in the systematic production of discursive knowledge. The third selection mechanism incorporates the intellectual feedback into the non-linear dynamics of this complex process. The operation of these three selection mechanisms makes the development of science reflexive with respect to its history. While the first and second selection mechanisms can socially be organized, the operation of the third selection mechanisms inflicts self-organization in a next-order dynamics of the intellectual—that is, relatively delocalized—dimension that recursively feeds back into the system.

Recursivity operates as a consequence of reflexivity in the communications. This reflexivity is enhanced by the textual dimension which allows scientists to communicate their findings as well as their interpretations of previous findings. This also implies that a contribution made by a scientist is subject to (re-)interpretation in the future. The historical developments of discursive knowledge stabilize communications along disciplinary trajectories. Meanwhile, reconstructions from the perspective of hindsight—following the recursive operation of the third selection mechanism—provide communications with codified meaning.

In other words, while the production of scientific knowledge provides the system with a first-order dynamics, the produced knowledge is organized through second-order dynamics. This second-order intellectual (self-)organization provides codes of communication that feedback on the first-order production. The development of discursive knowledge thus needs to consider (a) its instantiation in terms of publications, and (b) its memory traces that recursively feedback into the system (Giddens, 1981, at p. 171)

The specification of three selection mechanisms enables us to bridge the communication systems perspective with operationalizable network structures. The link between the theoretically defined system of communication, publications as the basic autopoietic operation, and the emergent network structures justified the methods used in the subsequent chapters.

The purpose of Chapter Three was to understand the emergence of knowledge as a consequence of the discovery of fullerenes—and the subsequent discovery of
nanotubes—and their effects on the structures of scientific communications. Different network analyses suggested that even though a unified research specialty covering both fullerene and nanotube topics emerged from the discoveries, nanotube research acquired its own dynamics and became differentiated from fullerene research during the 1990s.

The production of scientific knowledge that follows a scientific discovery can be very dynamic in the beginning: new institutions and codes of communication are expected to emerge as the topic gets attention on research agendas. As a result, new hot topics may emerge, as in the case of nanotubes which seemed to emerge from the original scientific breakthrough of fullerenes.

Further analysis of the relationship between research on fullerenes and on nanotubes was presented in the Fourth Chapter. The operationalization of science using the literary model allowed an algorithmic reconstruction of the cognitive history for both specialties. The historical reconstruction based on publications with “fullerenes” and “nanotubes” as part of their titles was enhanced algorithmically to characterize the intellectual bases of both these specialties. This supported the findings from Chapter Three. The specialties became differentiated in terms of their cognitive historiography. Furthermore, research on nanotubes seemed to be more dynamic, as cognitive transitions among documents suggested that the intellectual base was not stable.

In Chapter Five, self-organization of a research specialty was measured as reduction of uncertainty in different periods of time (Leydesdorff & Fritsch, 2006). Self-organization was measured as the information processing capacities of a specialty based on the tensions that arise between processes of codification and variation at the research fronts. Theories, scholars and texts are mutually co-evolving in the production of discursive knowledge. For example, text and cognitions are combined with other social resources such as funding and infrastructure. New theories are sometimes the result of new instruments as in the discovery of fullerenes after the introduction to the scientific community of high definition electro microscopes like the Scanning Tunneling Microscope.

In sum, the social, textual and cognitive dimensions are mutually dependent and interacting. The importance of publications as the carriers of discursive knowledge makes the hypothesis of self-organization of discursive knowledge amenable to testing using the textual dimension, but the textual layer itself can also be...
expected to develop its own dynamics. Configurational information was used to indicate self-organized specialties in Chapter Five. The results suggested that the research fronts behave more dynamically—in terms of their property to absorb new contributions—when the intellectual dimension was self-organizing. This follows arguments introduced in the earlier chapters about the properties that the intellectual dimension has to structure the variation provided in new publications.

Results of configurational information for fullerenes and nanotubes confirmed the differentiation of these specialties, as indicated in Chapters Three and Four. While for fullerenes uncertainty was reduced until 1994 and behaved erratically thereafter, for nanotubes a constant reduction of uncertainty from 1994 suggests a self-organizing research specialty that is no longer part of the specialty of fullerenes.

Moving on to the social sciences, the indicator was used to measure the uncertainty prevailing in documents with “paradigms” or “citations” as part of their title words. On the one hand, for paradigms, the result was erratic and difficult to interpret, suggesting that the topic did not represent a self-organized specialty. This was confirmed when the results were disaggregated in terms of indices: while the remaining uncertainty for documents with paradigms as part of their title words was reduced for the case of documents contained in the Science Citation Index (and less so in the Social Science Citation Index); the erratic behavior for the remaining uncertainty in documents contained in the Arts and Humanities Citation Index confirmed that the word is not codified enough to represent a specialty or topic of research. On the other hand, the results for documents with “citation” as part of their title word indicated that the topic was increasingly self-organized.

Discourses in the social sciences require different specifications than in the natural sciences in terms of their operational delineation. While for fullerenes and nanotubes, the title words provided a meaningful number of closely related documents, in the social sciences, delineations of discourse are better operationalized in terms of journals (Leydesdorff & Cozzens, 1993). To gain further insight on the results obtained for citation, a discourse on scientometrics was represented by documents contained in the Journal of Documentation, the Journal of the American Society of Information Science and Technology, Scientometrics, and, Information Processing and Management.

In terms of a scientometric discourse, the indicator reinforced the previous conclusion regarding dynamic research fronts. Processes of codification resulting in
reduction of uncertainty at the research front stimulate the growth of the scientometric specialty—again in terms of contributions. Increased intellectual organization can be expected to enable a journal to absorb more contributions. This was the case for *Scientometrics*, the *Journal of the American Society of Information Science and Technology*, and *Information Processing and Management*.

While in Chapter Four, the Kullback & Leibler’s (1951) relative entropy measure was used to indicate evolutionary turning points in a network of documents, in Chapter Five, configurational information was used to measure the absolute entropy in a set of documents. Entropy measures uncertainty in a probability distribution; reductions of configurational information for the distribution of cited references and title words in publication years was associated with self-organization of the intellectual structures: with a synergetic co-evolution of the intellectual, social and textual dimension.

### 2. Methodological reflections

The literary simplification of the system of scientific knowledge production allows us not only to use networks to model scientific developments but also to use probabilistic entropy to measure uncertainty contained in the networks when these are decomposed as probabilistic distributions.

Entropy statistics stem from the Mathematical Theory of Communication (Shannon, 1948) where information was defined as uncertainty contained in a probability distribution. Earlier in this concluding chapter, the reflexivity of scientific discourse was explained; this reflexivity implies a selection on Shannon-type information. When a scientist submits a contribution and when its knowledge claim is validated through the peer review—operation of the first and second selection mechanisms—variation is introduced in the system. This variation can be expected to increase uncertainty in the system. The operation of the codes of communication in a next-order dynamics potentially reduces this uncertainty by structuring the variation in terms of the codes of communication.

Publications can be decomposed into distributions of attributes containing Shannon-type information. The system of science—the system producing scientific knowledge—has the capacity to organize this information into meaningful discourses. Thus the information produced by new publications needs to be defined with
reference to this system in order to determine how the selection mechanisms condition each other resulting in potential decreases of the uncertainty contained.

Intellectual structures emerge from the constant stream of publications, as a consequence of the operation of the selection mechanisms. But these intellectual structures are in constant flux; their cognitive trajectories are continuously updated by new knowledge claims. Using scientific publications as the basic operation in the production of discursive knowledge, the variation introduced by new knowledge claims can be understood as probabilistic entropy. In Chapter Three, network algorithms were used to identify change produced by variations in scientific discourse. Probabilistic entropy—Chapters Four and Five—allows us to define communications mathematically. Structural change can be indicated when the probabilistic entropy—i.e. uncertainty—is reduced at the systems level.

Uncertainty is the result of the variations produced in social and cultural contexts. In other words, it can be expressed in terms of the expected information content of one event in relation to a previous one. In our case, publications are defined as the events and their information content is measured in terms of the probability distributions of cited references and title words in citing documents.

This definition of uncertainty in terms of the information content of messages was used in Chapter Three. The Kullback-Leibler divergence was used to measure critical transitions in terms of the information content of three sequential documents. The meaning of this transition in evolutionary systems was introduced in the Introductory Chapter. For the case of our model, the probabilistic distributions were defined in terms of citing and cited relations.

In Chapter Five, reductions of uncertainty due to intellectual self-organization were measured using the combination of title words (to represent variation) and cited references (to represent intellectual structuring). The concept of configurational information or mutual information was used to measure the uncertainty remaining after the co-evolution of the textual, social and cognitive dimensions.

3. Theoretical reflections

In this summary I have argued hitherto that the dynamics of science need to be specified in terms of linear and nonlinear dynamics: in terms of variation and structuring selections. Publications provide variation and linear (additive) change to
the system. Selection of valid knowledge claims provides recursivity to the system. Codification emerges from the further articulation of knowledge claims in new publications by considering some knowledge claims as more relevant than others. The intellectual dimension is contingent upon the continuous stream of publications, but at the same time acts as a coordination mechanism.

This formalization offers a complementary perspective to scientometric approaches to the study of patterns and mechanisms of change in scientific literature. The existing approaches measure change from the perspective of hindsight as a disruption in the intellectual dimension (Chen, 2006). The heuristics of this study have been to measure stabilization and change as a consequence of analytically specified subdynamics. While variation is pertinent to the context of discovery and socially organized at the lab bench, stabilization provided by the intellectual dimension is relevant to the context of justification and cognitively organized in the form of discursive knowledge (carried by the textual dimension.)

The measurement of the sciences proposed in this dissertation stems from a view of science in which the intellectual, social and textual dimensions are intertwined: they interact conditioning and enabling each other in co-evolutions and mutual shapings. Their co-evolution shapes the structures that enable and constrain the development of the sciences as a result of the individual scientists’ communications.

In the Structure of Scientific Revolutions, Kuhn (1962) juxtaposed the accumulative role of scientific knowledge to the idea of paradigmatic revolutions. Progress in science was defined in terms of three different stages with paradigm shifts as the central process. Pre-paradigmatic science was characterized by scientists inductively creating hypotheses but without a dominant theory to organize the communications among them. Following the pre-paradigm stage, normal science was characterized by a consensus among scientists on the relevant concepts and theories of the field (ibid, p.5). This phase required a comprehensive body of knowledge that scientists agreed upon (ibid, p. 10).

The conceptual tools in Kuhn’s paradigms reduced the uncertainty for the production of new knowledge claims. Kuhn’s definition of science as incommensurable paradigms depicted scientific developments as the continuous elaboration, revision and refinement of successful scientific theories guiding research in specific problem areas, until it is superseded by a new paradigm. The reflexivity of
the communications is essential in Kuhn’s explanation of the shifts in paradigms—
“the resolution of revolutions is by selection within the scientific community of the
fittest way to practice future science” (Kuhn, 1962, at p. 171, italics mine)— which is
consistent with the operation of the selection mechanisms specified in this
dissertation.

Similarly, in this dissertation the hypothesis of a paradigm shift could be made
measurable as reorganization in the communication structures. Communications
within a paradigm can be expected to self-organize; each new publication provides
variation but at the same time re-produces the paradigmatic structure. The self-
organizing structures will tend to differentiate as a consequence of the variation when
the structure is not reproduced but disturbed. Previous sections of this chapter
summarized the case of research on nanotubes as a spin-off effect of research on
fullerenes which differentiated soon thereafter (Chapters Three, Four, and Five).
Differentiation of scientific specialties formalizes Derek de Solla Price idea of science
as a big jigsaw puzzle “where strategically important pseudopods” were made to
erupt (Price, 1978, italics mine).

In summary, the development of the sciences was modelled as evolving
communication structures in scientific discourses. In this model, the social and
cognitive dimensions are represented in scientific texts. While the social dimension
provides variation to the system, the reflexivity of the communications allows us to
measure the intellectual structures from the perspective of hindsight. These
intellectual structures operate through the reflexive selection (the third selection
mechanisms as defined in Chapter Two and introduced earlier in this text).

The dynamics in discursive knowledge are determined by the co-evolution
between variation and selection. While in theories of biological evolution variation
and selection are separate process, reflexivity of the communication system defines
them as interdependent and structurally coupled: the constant streams of publications
create and (re)create the structures that select from variations. In the development of
discursive knowledge, variation interacts with the selective structures in each
instantiation, i.e. in each publication.
4. Limitations and further work

In this dissertation, scientific communication was assumed as the main driver of scientific development. In accordance with the importance of scientific texts, scientific communication was reduced to its literary footprint. The accuracy of representing science using scientific literature has been argued before (Price, 1965; Garfield, 1979; Wouters, 1999b). Nevertheless, the model built using the information inscribed in scientific literature falls short in recognizing particularities at the local level. Additionally, in this representation of science the citation preferences among scientists—which are neither homogeneous nor unambiguous—determine the networks structures that represent the intellectual organization.

The use of citation data in empirical studies should take into consideration possible biases that respond to overestimation of the contributions from elite scientists at the expense of an underestimation of the contributions of average scientists (MacRoberts & MacRoberts, 1987; 1989). Nevertheless, it has been noted that error in the measurement is common when analyzing social phenomena quantitatively, as a consequence of the aggregation of behavior (Zuckerman, 1987). Furthermore, the use of probabilistic entropy to measure self-organization in the intellectual dimension is about relative rather than absolute frequencies (Stigler, 1987).

Nevertheless, knowledge is relational by nature; it is situated and should be studied in its context. In this sense, validation in terms of expert interviews was desirable. A first attempt to validate the results obtained from the study of ‘fullerenes’ and ‘nanotubes’ proved difficult as researchers did not easily recognize a quantitative reconstruction of their specialty.

Shannon’s concept of information is defined independently of context and meaning. It provides measures of similarity in probability distributions and as such can be applied to understand many of the dynamics of producing scientific knowledge. In this dissertation I used title words and cited references to test the hypothesis of self-organization in the intellectual dimension as organizing discursive knowledge and to understand science as an evolutionary system where variation produced at the social dimension (operationalized as title words) co-evolves with the selective structures that operate in the intellectual dimension (operationalized as cited reference). A follow-up of this study, considering variation in the distribution of citing documents could provide complementary results; in that case, the cited references
would inform about the structures while the citing documents would indicate variation.

Furthermore, traditional ways of communication between scientists have been affected by the inclusion of the Internet in the research practices (Heimeriks & Vasileiadou, 2008). In this dissertation, a representation of research on the selected cases was built using the ISI Web of Knowledge as the primary source of information. However, it can be expected that the inclusion of publications outside this mainstream database as well as open access journals will not change the overall results of the suggested indicators. Nevertheless, using Google Scholar might provide a more accurate representation of contemporary ways of communication among scientists.