

Socio-epistemic analysis of scientific knowledge production in little science research

Alberto Pepe

Department of Information Studies, University of California, Los Angeles, apepe@ucla.edu

Abstract: *The processes that drive knowledge production and dissemination in scientific environments are embedded within the social, technical, cultural and epistemic practices of the constituent research communities. This article presents a methodology to unpack specific social and epistemic dimensions of scientific knowledge production using, as a case study, the Center for Embedded Networked Sensing (CENS), a National Science Foundation “little science” research center involved in theoretical and applied work in the field of wireless communication and sensor networks. By analysis of its scholarly record, I construct a social network of co-authorship, linking individuals that have co-authored scholarly artifacts (journal articles and conference papers), and an epistemic network of topic co-occurrence, linking concepts and knowledge constructs in the same scholarly artifacts. This article reports on ongoing work directed at the study of the emergence and evolution of these networks of scientific interaction. I present some preliminary results and introduce a socio-epistemic method for an historical analysis of network co-evolution. I outline a research design to support further investigations of knowledge production in scientific circles.*

Keywords: social studies of science, laboratory studies, social networks, epistemic networks, socio-epistemic analysis, scientific knowledge production

Acknowledgement: The author would like to thank Christian Fuchs of the University of Salzburg, Leah Lievrouw and Christine Borgman of the University of California, Los Angeles and an anonymous reviewer of the tripleC journal for the helpful comments on this article. Some of the work presented in this article was performed in collaboration with Gamer Kesheshe and Ashley Davis during the 2008 edition of the Summer@CENS High School Scholars Program.

1. Introduction

The social, technical and cognitive processes that drive knowledge production and dissemination in scientific circles have been the focus of intense study for centuries in domains as diverse as philosophy, logic, sociology, psychology and cognitive science. Analyzing the “making of science” from these different methodological lenses has enabled a comprehensive framework for the study of scientific formation and its dissemination process. Different disciplines have brought forward very different perspectives. Social studies of science, for example, have analyzed the processes of scientific activity in the social, cultural and political context in which science takes place (Latour & Woolgar, 1986; Hess, 1997). Another example is the

field of cognitive science, which has focused on the study of the scientific mind and the mental processes underlying scientific reasoning (Dunbar, 2005). Although theoretically grounded in their own domains, these approaches to the study of science do not exist in isolation; they have often borrowed theories and concepts from each other.

An example of this disciplinary cross-fertilization is the domain of psychology that deals with scientific creativity and problem solving. Traditional studies of scientific reasoning were concerned mostly with the study of the scientific mind as an individual entity and as the sole source of new ideas, insights and discoveries. The process of knowledge generation was studied in terms of cognitive capability, such as scientists' ability to generate and test hypotheses (Wason,

1968), and the ability to use analogical reasoning to construct novel links between known and unknown scientific facts (Schiano et al., 1989). In other words, many subfields of psychology and related disciplines have traditionally approached the study of “science making” in purely subjective terms. This follows a trend of studies that regards the generation of new ideas as the result of contemplations, insights and social experiences that are difficult to communicate to others or to describe in any systematic procedure. Remarkably, some knowledge is so intimate that it remains untold: “we can know more than we can tell” (Polanyi, 1966). Tacit knowledge is knowledge that resides within the personal intellectual sphere and cannot be explicated in propositional terms. Besides one’s internal perception and elaboration of knowledge, its presentation and dissemination can also be very intimate in nature. Writing a book, carving a sculpture or setting up a lab experiment are examples of scholarly activities that might be the result of solo work. Yet, these activities may also involve cooperation, at different stages: the elaboration of new knowledge, as well as its presentation, is an unstructured, creative process that draws both from self-reflection and interpersonal collaboration (Hutchins, 1995; Sawyer, 2003). In this view, knowledge production is the result of both social and cognitive activities. For this reason, many traditional studies of science grounded in psychology and cognitive science have progressively been implemented with theories, drawn from the fields of sociology and anthropology, which locate the process of knowledge production in the social and cultural context in which it takes place.

In a review of the history of scientific creativity, Simonton (2004) refers to an *internal zeitgeist*, defined by the subjective individual capability of scientific thinking, and an *external zeitgeist*, defined by the context in which scientific thinking takes place shaped by broader social, temporal, cultural and political dimensions. He notes that:

“Galileo became a great scientist only because he had the fortune of being born in Italy during the time when it became the center of scientific creativity.

Similarly, Newton’s creative genius could appear only because he lived in Great Britain when the center had shifted there from Italy. If Galileo and Newton had switched birth years without changing national origins, then neither would have secured a place in the annals of science.” (Simonton, 2004, p.134)

This example, although specifically grounded in the field of psychology, illustrates this broader academic trend – that a theoretical framework for the study of scientific knowledge production has been reformulated in terms of the collective dimension in which science takes place. Studies of science and its knowledge production mechanisms focused on the *individual* have been progressively been implemented with studies that account for the *collective*. For these latter investigations, science making is not only an individual endeavor; it is also a collective, distributed process, involving the close interaction among a number of social, cultural, technological, economical, and political dimensions.

The present article intends to approach the production and dissemination of scientific knowledge in a research center from a similar variety of observational lenses and perspectives. In particular, I investigate the relationship between two different dimensions of scientific interaction among researchers: (i) social activity, as evinced by their bibliographic record of co-authorship, and (ii) epistemic activity, as evinced by the intellectual interaction expressed in the aforementioned bibliographic record. In this article, I utilize a young “little science” research institution, the Center for Embedded Networked Sensing (CENS), as a case study to construct and analyze the aforementioned networks of scientific interaction.

This article is organized as follows. In section 2, I introduce the theoretical underpinnings of a socio-epistemic approach for the study of scientific knowledge production. In section 3, I provide a general introduction to CENS, the focus of the present study. In section 4, I present the methods for the construction and analysis of social and epistemic networks of scientific knowledge production. In section 5, I discuss my

preliminary research results, tracing out the outline for future research.

2. Socio-epistemic analysis of scientific knowledge production

How can we blend methods and theories that emerge at the convergence of social and epistemic notions of “science making”? A number of studies that address both the social and epistemic aspects of the generation of knowledge have recently appeared in specialized literature. These studies are often found at the intersection of sociology, cognitive science, information studies and social studies of science. An example is Birger Hjørland's *domain analysis* that proposes a platform of investigation in which individuals are studied within the knowledge-domains and the discourse communities to which they belong (Hjørland & Albrechtsen, 1995). By acknowledging the different roles that people play in the division of labor in society, this method develops on the conceptual and intellectual fragmentations (knowledge structures, cooperation patterns, and relevance criteria, among others) that exist within the academic and research fields. Domain analysis leverages a domain-based growth, driven by a “rich manifold of discourses, domains and documents” (Hjørland, 2002). Hjørland does not posit a continuous, homogeneous growth, but rather one characterized by “shifts”: the acceptance and rejection of conflicting paradigms by the involved discourse communities — very much in line with Kuhn's perspective on the structure of scientific revolutions (Kuhn, 1962).

My present investigation of “little science” interdisciplinary collaboration follows a scheme that is conceptually based on Hjørland's domain analysis. I mine knowledge production patterns emerging from communities of scientific researchers. In particular, I concentrate on two dimensions: the social interaction, evinced by the production of joint scholarly output and the epistemic interaction, evinced by the production of scholarly content around the same research topics.

The epistemic mechanisms of scientific knowledge production and the social

processes of scientific collaboration are tightly coupled notions that are hard to disassemble and operationalize. Despite the difficulty and heterogeneity of the study of idea formation and collaboration, some distinct patterns can be identified if both scientific thinking and collaboration are studied in terms of quantifiable indicators. A number of ad-hoc indicators to study the nature of scholarly and scientific production have emerged in specialized literature.

Co-authorship, for example, is an evident indicator of collaborative knowledge production and, in turn, it elicits the collective, distributed dimension of scientific thinking (Cronin, 2003). In scholarship, the interplay of individual and collaborative scholarly endeavors differs significantly across different disciplines. In the arts and the humanities, for example, single-authored work is, by and large, the most common vehicle of scholarly communication. This is because the nature of these disciplines requires an extensive amount of personal introspection and intuition. Yet, the fact that most books, manuscripts and artifacts in the humanities are single-authored does not imply that intensive collaboration does not take place (Cronin, Shaw, & Barre, 2003; Sacco & Milana, 1984).

In contrast, the physical and life sciences are, historically, domains in which collaboration among scholars often materializes in joint printed works (Clarke, 1964). High energy physics is a popular example of this: research collaborations often involve human and technical cooperation of faculties, scientists and engineers from several countries worldwide, culminating in works co-authored by tens (if not hundreds) of scholars (Traweek, 1992; Knorr-Cetina, 1999). Unlike the humanities, in scientific endeavors technical expertise and theoretical knowledge are often propagated using an incremental approach, that requires continuous intellectual exchange as well as the joint use of large-scale technology, which could not be possibly fabricated and employed within a small research group.

Despite these differences, in recent years, all scholarly domains have progressively become more diversified in nature (Qin & Lancaster & Allen, 1997). It has become

increasingly more common for scholars from different disciplines to collaborate and write papers together and new methods are needed to investigate such cooperative environments. Interdisciplinarity has affected the scholarly communication process, as a whole. In this new paradigm, scholars do not only exchange and collectively create knowledge within their research group and academic domain, but they deliver and shape their ideas beyond the boundaries of traditional academic fields (Pierce, 1999). The mutual, direct engagement among previously uncorrelated disciplines has advantages not only for the researchers, that are able to draw from a wider, diverse intellectual environment, but also for the nature of research performed, that is circulated, validated and enriched by contact with new research and social circles. Moreover, interdisciplinary investigation brings disciplines closer and may lead to new discoveries and insights (Swanson & Smalheiser, 1997). The shift towards interdisciplinary research is resulting in a proliferation of interdisciplinary centers for research and innovation. Interdisciplinarity has become the *raison d'être* of many emerging scholarly centers for innovation. One of such centers, the Center for Embedded Networked Sensing (CENS), is the subject of the study presented here. As discussed later in the article (section 4.1), co-authorship is a valid indicator of social collaboration in a small-scale research environment like CENS. In section 4.2, I present the method by which I use the bibliographic record of CENS to construct a social network of co-authorship, i.e. a network composed by human actors (individual authors), linked by a collaborative relationship of co-authorship¹.

Co-authorship depicts very well the extent and arrangement of collaborative activities among researchers within an institution. Yet, it does not provide a description of the content of such collaborative activity, i.e., it captures the contributions of researchers in co-

authored scholarly material, but it does not take into consideration the content of those scholarly contributions. For this reason, I implement the aforementioned analysis of co-authorship activity with an investigation of the concepts utilized by researchers in their scholarly output. By connecting similar concepts available in the scholarly items, I construct a network linking related topics and knowledge constructs extracted from the bibliographic record, i.e. an epistemic network. I borrow this definition from early work in the field of qualitative bibliometrics by Callon *et al.* (1986) and Leydesdorff (1991) who construct networks describing associations between the many heterogeneous entities that lead to the construction of scientific facts into an array that is "strong and durable". Closely analyzing a set of scientific texts, they build, for every analyzed scholarly article, an epistemic network composed of "powerful words" (keywords and other extracted key phrases). In section 4.3, I discuss how I build an epistemic network to represent manifestations of intellectual interactions within the context of CENS research.

As discussed thus far, a socio-epistemic approach for the study of scientific knowledge production involves the creation of two networks: (i) a social network, depicting co-authorship activity among researchers, and (ii) an epistemic network, depicting the conceptual connection among research topics in the scholarly production. In section 4.4, I outline an analytical method for the investigation of the co-evolution of these networks. Can social and epistemic dimensions of scientific knowledge production, extracted from a single data source (the bibliographic record), be studied in conjunction to detect the emergence and evolution of specific patterns? For example, can the appearance of new topic clusters in the epistemic network be explained in terms of novel social ties in the co-authorship network?

The purpose of the present study is to provide a theoretical and computational framework to enable the analysis of multiple indicators of scientific knowledge production. I expect to carve out specific social and epistemic traits of the "epistemic culture" - "the

¹ Besides being a "social" network, I might also infer that the co-authorship network is an "acquaintanceship" network, i.e. that coauthoring researchers are acquainted with each other (Newman, 2004). Although not relevant to the present study, this is a fair assumption, given the size of the analyzed network (see Section 4.1).

amalgam of arrangements and mechanisms [...] bonded through affinity, necessity and historical coincidence" (Knorr-Cetina, 1999) – in which scientific knowledge is produced and disseminated.

3. Present study: a little science research center

The subject of the study presented here is the Center for Embedded Networked Sensing (CENS), a National Science Foundation venture involved in the development and application of wireless sensing systems to critical scientific and societal pursuits. CENS consists of scholars at all levels (faculty, scientists, engineers, graduate and undergraduate students) from five member institutions (University of California at Los Angeles, University of Southern California, University of California Riverside, California Institute of Technology, and University of California at Merced). CENS features a headquarter base located at UCLA, yet work is conducted at all member institutions in California. These institutions (and sometimes even departments) are sufficiently distant to prevent continuous physical interactions among researchers: computer-supported communication is at the basis of their collaborative work. The type of research conducted at CENS spans across a wide spectrum of disciplines and applications (from biology to seismology, from wireless telecommunications to statistics) requiring continuous cooperation among individuals that, otherwise, would probably not interact beyond the walls of traditional university departments and faculties. In such a scholarly and scientific environment, distributed collaboration on multi-disciplinary subjects constitutes a fundamental leverage for scientific research. (Wallis & Borgman & Mayernik & Pepe, 2007).

Unlike "Big science" collaborations (De Solla Price, 1963), such as those typically found in many high energy physics and observational astronomy laboratories, CENS is a typical "little science" research centre: CENS research does not require continuous use of massive amounts of human and computational resources. Yet, similar to big science endeavors, CENS research relies

heavily on information and communication technologies. The cultural and cognitive processes that drive knowledge production and dissemination at CENS are embedded within the social fabric of the interpersonal communication and in the technical systems employed by the research community.

4. Preliminary results and analysis

The presented socio-epistemic approach aims at describing two separate, yet interdependent indicators of scientific knowledge production: social and epistemic activity. Collecting data relative to these indicators for a given research centre yields two networks of scientific interaction: a social network, depicting the temporal evolution of collaboration among authors of scholarly items, and an epistemic network, depicting the related evolution of topics and concepts in the aforementioned scholarly corpus.

As discussed in the remainder of this section, the proposed method utilizes data from a single source: the scholarly output of the research centre under study, as evinced by its scholarly record. In particular, I collect and analyze only bibliographic data relative to published journal articles and conference papers authored by CENS researchers and archived in the institution's repository. Thus, the present study leaves out other indicators of scientific interaction that result in products that are not necessarily recorded in the traditional format of a scholarly work. In the field of sensor networks, examples are co-production and co-authorship of software products, sensor devices, raw and analyzed data. Yet, for a disciplinary diverse environment like CENS, scholarly publication of journal articles and conference papers is the *de facto* indicator of scientific production and dissemination. Scholarly publication is entrenched in the *modus operandi* of engineers, natural scientists, statisticians, computer scientists and life scientists, among others.

4.1. Data collection

The first part of the data collection process consisted of gathering the entire scholarly record of CENS, from its inception (in 2002) to

present day. Bibliographic data relative to CENS journal articles and conference papers were harvested in BibTeX format² from the research center's institutional repository³. At the time the analysis was performed, the bibliographic database consisted of 552 manuscripts (369 conference papers and 183 journal articles) published between 2002 and 2007 by a total of 291 authors. Some fundamental statistics relative to the collected dataset are displayed in Tables 1, 2 and 3 which depict the distribution of scholarly items in the bibliographic dataset per publication year, number of authors per publication, and venue of publication (name of conference or journal), respectively.

year	count
2002	46
2003	106
2004	128
2005	115
2006	78
2007	38
2008	49

Table 1: Distribution of CENS scholarly items per publication year

# authors	count
1	53
2	131
3	160
4	91
5	51
6	25
7	16
8	11
9	3
10+	11

Table 2: Distribution of CENS scholarly items per number of authors

It is particularly interesting to note that the vast majority of CENS publications are

² BibTeX: <http://www.bibtex.org/>

³ California Digital Library's eScholarship repository available at <http://repositories.cdlib.org/cens/>

authored by either 2, 3 or 4 authors. This suggests that authorship at CENS is a very collaborative endeavor that takes place among a fairly small number of individuals that are very likely to be personally acquainted with each other.

venue	Count
IEEE ICRA	31
ACM/IEEE IPSN	22
ACM Sensys	18
IEEE/RSJ IROS	14
IEEE MEMS	14
ACM EMNETS	11
ACM Mobicom	8
ACM/IEEE DCROSS	8
AGU Conference	7
IEEE TMC	6

Table 3: Distribution of CENS scholarly items per venue of publication (top 10).

4.2. Co-authorship network

The collected bibliographic data was used to construct a network of co-authorship in which each author of a scholarly item, represented by a node, is connected to another author, represented by another node, if a co-authorship event exists. Moreover, the network is weighted and edge weights are established by partitioning an arbitrary value of 1.0 for every publication. For example, if a scholarly item is authored by two individuals, the edge connecting the two authors has a weight of 0.5 (i.e. $1.0/2$). This mechanism allows assigning weights accounting for the fact that publications authored by a smaller number of individuals indicate stronger interpersonal collaboration than multi-authored publications. The network can be sliced by its temporal component (the date of publication of the scholarly items) into a number of sub-networks, to allow an historical study of the evolution of co-authorship activity. In the present article, I subdivide the co-authorship network into two 3-year long time slices: 2002-2004 and 2005-2007, displayed in Figures 1 and 2, respectively. In the images, author names have been replaced with numbers and the diameter of the nodes is

proportional to the author's eigenvector centrality⁴ (Bonacich, 1987).

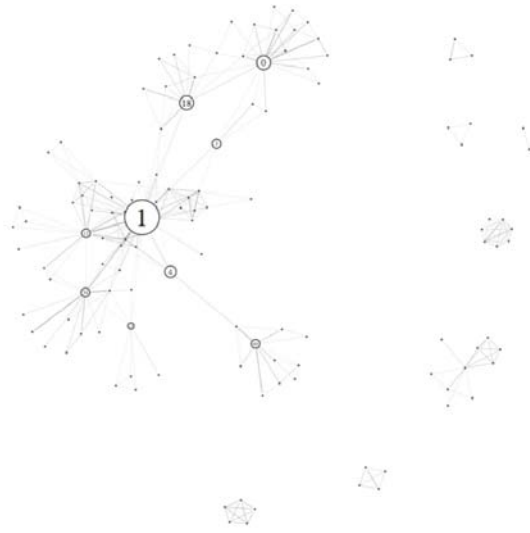


Figure 1: Co-authorship network, 2002-2004



Figure 2: Co-authorship network, 2005-2007

4.3. Epistemic network

The bibliographic data collected via the procedure presented in section 4.1 can be used to construct an epistemic network, i.e. a network that depicts the co-occurrence of concepts, topics and knowledge constructs in

⁴ For the scope of this article, it is unnecessary to provide identification of the authors included in the co-authorship network.

the items of the given bibliographic record. For the purpose of this article, concepts were extracted from scholarly items parsing the full text and the bibliographic metadata for index keywords (provided either by authors or indexers)⁵. Thus, in the CENS epistemic network, a concept, represented by a node, is connected to another concept, represented by another node, if they co-occur as keywords of the same scholarly item (journal article or conference papers). As for the co-authorship network, the epistemic network is also weighted – edge weights are established by partitioning an arbitrary value of 1.0 for every item by the number of concepts associated with each item. Similarly to the co-authorship network, I subdivided the epistemic network into two 3-year long time slices: 2002-2004 and 2005-2007, displayed in Figures 3 and 4, respectively. The font size of the node labels (i.e. the concepts) is proportional to their eigenvector centrality.

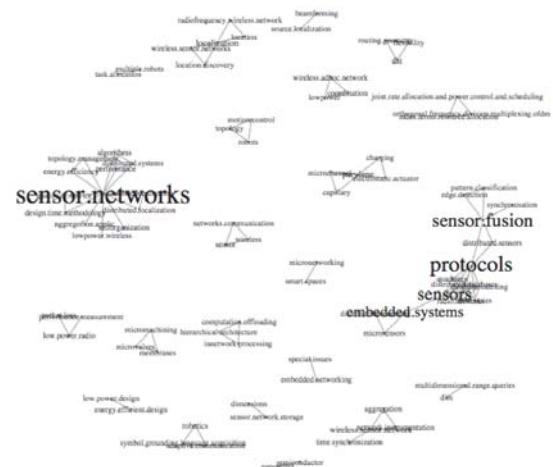


Figure 3: Epistemic network, 2002-2004

4.4. Socio-epistemic analysis

The co-authorship and epistemic networks presented in sections 4.3 and 4.4 can be employed for a number of comparative analyses. Even a quick visual inspection of these networks reveals clear patterns between different networks at different times.

⁵ The following assumption was made at this stage: that keywords assigned by both authors and indexers belong to a controlled vocabulary or taxonomy (such as those provided by ACM and IEEE). This is the case for most publications in the bibliographic database, but not all. This assumption represents a limitation of the study.

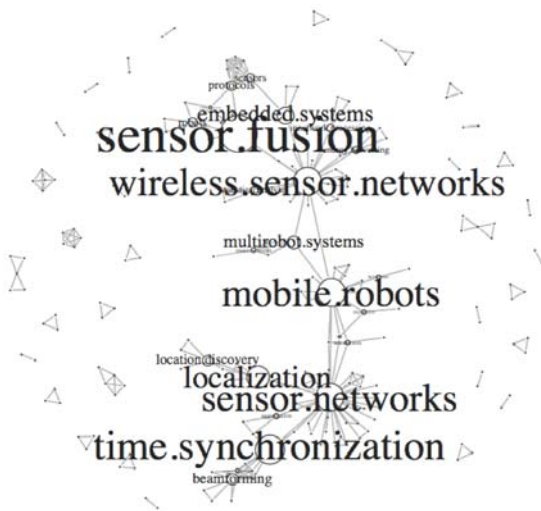


Figure 4: Epistemic network, 2005-2007

For example, the co-authorship networks diagrammed in Figures 1 and 2 show that CENS shifted from a dispersed to a more collaborative co-authorship model, as evinced by the low connectivity of Figure 1 and the large central connected component of Figure 2. One could infer that during the first years of CENS activity, research still relied heavily upon the specific domain, department and institution of the researchers, and that CENS acted as leverage for interdisciplinary, multi-institutional research.

Similarly, a quick analysis of the epistemic networks in figures 3 and 4 reveals that the focus of CENS research became more homogeneous (at least in its descriptive content) with time. Up to 2004 (Figure 3), keywords used to describe CENS research were either very high level (e.g. “sensor networks”, “sensors”, “embedded systems”) or overly descriptive. The more recent epistemic network (Figure 4) features more concise, well-established keywords (such as “sensor fusion”, “wireless sensor networks”, “localization”, “time synchronization”). Also, note that very general keywords, such as “sensors”, are only present in the early network. This suggests that the research field of embedded wireless networked sensing, still in its infancy phase in the early 2000s, grew into a more solid, widely recognized research area in the course of the years.

These light comparative analyses offer various insights into the evolution of co-

authorship and epistemic models in the research center under study. However, the intent of a socio-epistemic analysis is to investigate emerging patterns in the joint evolution of both models, i.e. the co-evolution of selected social and epistemic indicators of scientific knowledge production. For this reason, I propose to utilize a quantitative framework for the study of network evolution, to uncover specific patterns of interaction among the social and epistemic networks under study. Roth (2006) describes three different patterns that are important to study the evolution of networks: progress vs. decline, enrichment vs. impoverishment, and reunion vs. scission. The first two, in particular, reveal a dichotomy that is very relevant to the present analysis. I define these evolutionary patterns as follows:

- *Progress* and *decline* are marked, respectively, by the growth and decrease of certain interactions in existing connections;
- *Enrichment* and *impoverishment* are marked, respectively, by the birth of novel connections and the decrease of existing ones, for an existing node.

In terms of graph theory, these patterns can be conceptualized as follows. Progress is manifested as the reinforcement of a link among two nodes over time, e.g., in a co-authorship network, progress corresponds to the co-authoring of a new article among two authors. Decline is the opposite process, by which the weight of the link among two nodes lowers from one time period to another. Enrichment is manifested when new nodes join existing patterns of interaction over time, e.g., in a co-authorship network enrichment might correspond to the appearance of a new author within an established circle of collaborating authors. Impoverishment is the opposite process, by which an existing connection ceases to exist from a time period to the next. These patterns are depicted in Figure 5 (progress and decline) and 6 (enrichment and impoverishment) by the use of a simple, arbitrary 3-node network at time steps t_1 and t_2 .

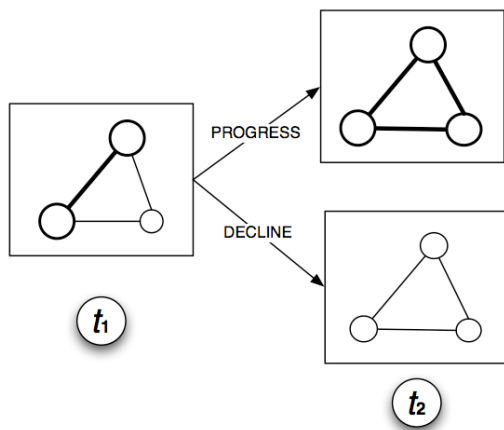


Figure 5: Evolutionary patterns in a simple network: progress and decline.

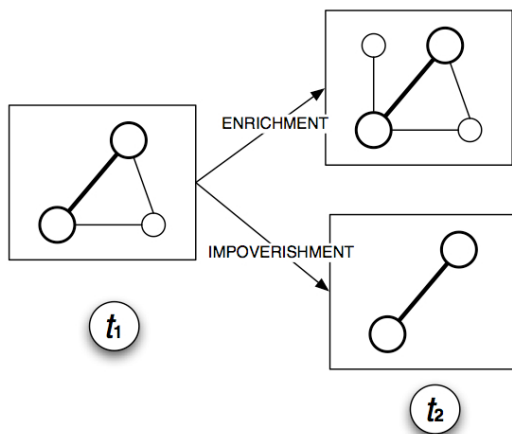


Figure 6: Evolutionary patterns in a simple network: enrichment and impoverishment.

I intend to utilize this framework of evolution to analyze the networks under study. By performing a topological comparison among the networks at different time steps, I plan to extract evolutionary patterns and associate them with each network at each time interval. For example, looking at both co-authorship and epistemic networks between t_1 (2002-2004) and t_2 (2005-2007), this analysis might uncover the appearance of a number of co-authorship ties in specific clusters of collaboration (enrichment), and a parallel increase of epistemic activity around specific topics (growth). These evolutionary patterns will be used to annotate every network at specific time-steps. By merging these annotations, from the first time-step to the

last, an historical record of evolution will be compiled for each network. These records represent the aggregated incidence of specific social and epistemic scientific interactions that have taken place at the CENS collaboratory over a given period of time.

At this point, a comparative analysis between the two historical records will be performed. Annotations evinced from each record (social and epistemic networks) will be grouped according to the specific type of evolutionary pattern. For example, I would gather all the social interactions exhibiting enrichment and compare them to the relative annotations in the epistemic record for the same time interval. Let us suppose that in a given time interval the epistemic network of CENS was “enriched” by a novel set of research topics (e.g. “participatory sensing” and “urban sensing”). Is such enrichment reflected in the historical record of the co-authorship network? Can the appearance of an entirely novel research topic be attributed to specific evolutionary patterns in a parallel network of social and scholarly interaction? These types of qualitative comparisons will determine the relationship and interdependence between evolutionary patterns in both social and epistemic traces of scientific knowledge production.

5. Discussion

In this article, I outline a methodological framework for the construction and analysis of social and epistemic networks of scientific knowledge production. I operationalize knowledge production specifically in terms of scholarly output (the co-authorship of scholarly items) and intellectual relationship (represented by the co-occurrence of research topics in scholarly material). I utilize CENS, a “little science” research centre involved in sensor network research, to demonstrate the possible application of this method in the context of a realistic scientific environment. In this section, I discuss the limitations of my study and how I plan to address them in future work.

The construction of the co-authorship network, presented in section 4.2, is subject to a first important limitation. At CENS, and in many related scientific endeavors, activities

and projects oftentimes culminate in the publication of specialized articles, conference papers, books, posters, technical reports and related scholarly material. In this article, I concentrate my research solely around the production of two types of scholarly items: articles published in academic journals and papers published in conference proceedings. I exclude posters, books, and technical reports from the present discussion for a number of reasons. Posters are crucial vehicles of scientific dissemination, but their author lists do not always reflect the true arrangement of a given collaboration; oftentimes all members of a project or team are indicated as authors of a poster. Moreover, they provide less content to perform textual analyses and their topics and title sometimes overlap with journal or conference papers. Books do not represent a fundamental mechanism of scholarly communication in the field of sensor networks and related areas. Finally, technical reports can be difficult to obtain and classify, compared to traditional publications; as for posters, their topics and titles often overlap with journal articles and conference papers as well. Yet, posters, books and technical reports constitute an important fraction of the scientific knowledge production of a research community and should thus be included in the study of co-authorship in future research.

A second limitation of this study concerns the construction of the epistemic network, presented in section 4.3. The epistemic network is generated gathering and associating author- and indexer-contributed keywords extracted from the scholarly items. As different indexing systems and controlled vocabularies to describe sensor network research exist, this method is subject to inconsistencies. In future work, I plan to perform more detailed textual analyses of the scholarly corpus under study. In particular, I will use a technique of topic indexing by Medelyan & Witten (2008), to extract and associate both general purpose and domain-specific concepts from the corpus under study.

Third, by operationalizing knowledge production solely in terms of specific social (co-authorship) and epistemic (topic co-occurrence) dimensions, my study does not take into consideration other possible

indicators. As discussed in section 4, there are other scientific interactions that result in intellectual products that are not necessarily recorded in the traditional format of a scholarly work and that could be mined both for social and epistemic interactions. In future work, networks depicting sharing and co-authorship of additional artifacts (e.g. raw sensor data) will be considered.

Finally, the proposed socio-epistemic analysis addresses the topological co-evolution of the presented networks, but does not fully account for the organizational, disciplinary and institutional position of the individuals and groups represented in these networks. For example, in section 4.4, I speculate that CENS research became more inter-disciplinary and multi-institutional by performing a rough analysis of the connectedness of the co-authorship network at different time intervals. In order to produce more detailed and reliable results concerning the evolution of organizational, disciplinary and institutional arrangements of CENS, I plan to collect and analyze a number of additional parameters, such as academic position, scholarly expertise, institutional and departmental affiliation of the individuals in the population under study. Some related preliminary research results are contained in a recent article that compares CENS' "structural communities" of co-authorship (Girvan & Newman, 2002; Newman, 2006) with their "socio-academic communities", i.e. the social and academic groupings of their constituent members (Rodriguez & Pepe, 2008).

The socio-epistemic approach outlined in this article has the potential to be extended to other research institutions. In particular, young inter-disciplinary research centers like CENS are especially suited for analytical studies of this kind for at least two reasons: size and interdisciplinarity. With less than 300 research participants, CENS is a fairly small institution; its small size allows for social and epistemic patterns to be detected and studied fairly easily by the proposed qualitative analyses. Also, highly interdisciplinary research centers like CENS, are constantly subject to novel influences from external disciplines – the authors and topics involved in scientific interactions are ever changing and are most interesting for the proposed analysis.

Particularly to CENS, socio-epistemic analyses can be integrated into the study of the institution's scholarly communication and scientific data practices. The results of the socio-epistemic analysis could be converted to a structured data format, to allow later reuse in other analyses. For example, the socio-epistemic data could be integrated within the present CENS data ecology that comprises contextual information about sensor deployments and experiments. This will enable a number of large-scale

quantitative analyses, as well as practical applications assisting in making scientific data more readily intelligible and interoperable among CENS researchers. Finally, from an institutional viewpoint, the results of this study might be used to make specific policy recommendations for CENS. These might include developing prediction models and promoting organizational changes to leverage interdisciplinary and multi-institutional work as well as specific socio-epistemic interactions.

References

- Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology*, 92 (5), 1170-1182.
- Callon, M. & Law, J. & Rip, A. (1986). Qualitative scientometrics. In M. Callon & J. Law & A. Rip (Eds.), *Mapping the dynamics of science and technology*. London: Macmillan.
- Clarke, B. L. (1964). Multiple Authorship Trends in Scientific Papers. *Science*, 143, 822-824.
- Cronin, B. (2003). Scholarly communication and epistemic cultures. *New Review of Academic Librarianship*, 9(1), 1-24(24).
- Cronin, B. & Shaw, D. & Barre, K. L. (2003). A cast of thousands: Coauthorship and subauthorship collaboration in the 20th century as manifested in the scholarly journal literature of psychology and philosophy. *Journal of the American Society for Information Science and Technology*, 54(9), 855-871.
- De Solla Price, D. (1963). *Little Science, Big Science*. Columbia University Press.
- Dunbar, K. (2005). Scientific thinking and reasoning. In K. J. Holyoak & R. Morrison (Eds.), *The Cambridge Handbook of Thinking and Reasoning*. Cambridge University Press.
- Girvan, M. & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99, 7821.
- Hess, D. J. (1997). *Science Studies. An advanced introduction*. New York: University Press.
- Hjørland, B. (2002). Principia informatica. Foundational theory of information and principles of information services. In *Proceedings fourth international conference on conceptions of library and information science*. Greenwood Village, Colorado, USA: Libraries Unlimited.
- Hjørland, B. & Albrechtsen, H. (1995). Toward a new horizon in information science: domain-analysis. *Journal of the American Society for Information Science*, 46(6), 400-425.
- Hutchins, E. (1995). *Cognition in the wild*. MIT Press.
- Knorr-Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge*. Harvard University Press.
- Kuhn, T. S. (1962). *The structure of scientific revolutions*. University Of Chicago Press.
- Leydesdorff, L. (1991). In search of epistemic networks. *Social Studies of Science*, 21(1), 75-110.
- Latour, B. & Woolgar, S. (1986). *Laboratory life: the social construction of scientific facts*. Princeton, New Jersey: Princeton University Press.
- Medelyan, O. & Witten, I. (2008). Domain independent automatic keyphrase indexing with small training sets. *Journal of the American Society for Information Science and Technology*. 59(7), 1026-1040.
- Newman, M. E. J. (2004). Who is the best connected scientist? A study of scientific coauthorship networks. In E. Ben-Naim & H. Frauenfelder & Z. Toroczkai (Eds.), *Complex Networks*. Heidelberg: Springer
- Newman, M. E. J. (2006). Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, 74(3), 1-22.
- Pierce, S. J. (1999). Boundary crossing in research literatures as a means of interdisciplinary information transfer. *Journal of the American Society for Information Science and Technology*, 50(3), 271-279.
- Polanyi, M. (1966). *The tacit dimension*. New York, NY: Doubleday & Co.
- Qin, J. & Lancaster, F. W. & Allen, B. (1997). Types and levels of collaboration in interdisciplinary research in the sciences. *Journal of the American Society for Information Science and Technology*, 48(10), 893-916.
- Rodriguez, M. A. & Pepe, A. (2008). On the relationship between the structural and socioacademic communities of a coauthorship network. *Journal of Informetrics*, 2(3), 195-201.
- Roth, C. (2006). Co-evolution in epistemic networks - reconstructing social complex systems. *Structure and Dynamics: eJournal of Anthropological and Related Sciences*, 1(3), 1-150.

- Sacco, W. P. & Milana, S. (1984). Increase in number of authors per article in ten APA journals: 1960-1980. *Cognitive Therapy and Research*, 8(1), 77-83.
- Sawyer, R. K. (2003). *Group creativity: Music, theater, collaboration*. Mahwah, NJ: Erlbaum.
- Schiano, D. J. & Cooper, L. A. & Glaser, R. & Zhang, H. C. (1989). Highs are to lows as experts are to novices: Individual differences in the representation and solution of standardized figural analogies. *Human Performance*, 2(4), 225-248.
- Simonton, D. K. (2004). *Creativity in science: Chance, logic, genius, and zeitgeist*. Cambridge, UK: Cambridge University Press.
- Swanson, D. R. & Smalheiser, N. R. (1997). An interactive system for finding complementary literatures: a stimulus to scientific discovery. *Artificial Intelligence*, 91(2), 183-203.
- Traweek, S. (1992). *Beamtimes and lifetimes: The world of high energy physicists*. Cambridge, MA: Harvard University Press.
- Wallis, J. C. & Borgman, C. L. & Mayernik, M. S. & Pepe, A. (2008). Moving Archival Practices Upstream: An exploration of the life cycle of ecological sensing data in collaborative field research. *International Journal of Digital Curation*, 3(1), 114-126.
- Wason, P. C. (1968). Reasoning about a rule. *The Quarterly Journal of Experimental Psychology*, 20(3), 273-281.

About the Author

Alberto Pepe

Alberto Pepe is a doctoral candidate in Information Studies at the University of California, Los Angeles and a graduate research assistant in the statistics and data practices group of the Center for Embedded Networked Sensing. His research interests revolve around scholarly, scientific and social data. In particular, he currently studies scientific data practices in highly collaborative interdisciplinary research environments via network theory and statistical mechanics. Prior to starting his Ph.D., he developed digital library software and promoted open access among particle physicists at CERN, Switzerland. He also worked in the scientific visualization department of CINECA, the Italian Scientific Consortium, based at the University of Bologna. He holds a M.Sc. in Computer Science and a B.Sc. in Astrophysics, both from University College London, UK.