The Transformation of Scientific Communication Systems in the Digital Age –

Towards a Methodology for Comparing Scientific Communication Cultures

Theresa Velden, Carl Lagoze
tav6, cjl2 @cornell.edu
Information Science, Cornell University

August 25, 2008

Abstract: We argue that a key question for the comparison of the field-specific shaping of scientific communication systems in the Digital Age is the identification and characterization of the unit of research: scientific collectives and their communication cultures. We address this conceptual and methodological challenge using a combination of ethnographic methods and network analysis. In this paper we present a basic methodological approach that we have developed in a pilot study of research groups in chemistry. Our first results underline the value of combining quantitative and qualitative methods.

1 Introduction

The rise of the WorldWideWeb and recent advances in information and communication technologies (ICTs) provide new opportunities to transform the scientific communication system. These transformations go beyond increased speed and efficiency and include new communication regimes that change the manner in which scientific results are shared, validated, and re-used [Borgman, 2007; de Sompel et al., 2004; Hilgartner, 1995]. Disciplines and fields differ in the extent to which they have adopted these new opportunities as demonstrated by the success of preprint servers[1] in many areas in physics, and their failure in the life sciences and chemistry [Kling et al., 2004]. Such differences in the manner and pace in which scientific communities transform their communication systems into the Digital Age are poorly understood, and challenge technological-deterministic assumptions about the impacts of the information technology revolution on research and scientific communication [Kling et al., 2003].

Increasing evidence (as well as theory, such as Social Construction of Technology [Bijker et al., 1989]) suggests that social and cultural arrangements of research fields play a major role in the shaping and adoption of new information and communication technologies [Walsh and Bayma, 1996; Kling and McKim, 2000; Cronin, 2003; Nentwich, 2005; Fry and Talja, 2007; Hine, 2008], but systematic analysis and comparisons between scientific fields and their subfields are lacking.

A major stumbling block for the systematic investigation of these differences is determining the unit of analysis corresponding to these research cultures. This presents a conceptual and a methodological challenge: What is the appropriate unit of analysis to describe collective differences in scientific communication practices, i.e. at what level of aggregation (discipline, field, specialty, community of practice...) can communication practices be effectively compared within and across scientific domains? And how can we combine quantitative methods that capture structural patterns of collective communicative behavior with qualitative methods that capture local meaning and interpretative context of communication acts (including acts of information withholding) to support the comparison of communication cultures [2] across science?

Our research takes an empirical approach to clarifying the fundamental issue of an appropriate unit of analysis for a comparative study of communication cultures in science. To do this we pull together qualitative, in particular ethnographic methods from social studies of the sciences, and quantitative approaches from the newly emerging specialty of network science. In the following we discuss the research problem in the context of related

---

[1] Web-based databases used by scientists to publicly share electronic versions of their papers before, and in parallel to formal publication in a scientific journal. Pioneering has been www.arxiv.org, serving communities in physics, mathematics and computer science.

[2] We here refer by ‘communication cultures’ to collective patterns of communication practices.

research (section 2), describe our research approach in more detail (section 3), and report and discuss results of a pilot study that we have conducted to test and refine our approach (section 4).

2 Research Problem

2.1 Unit of Analysis

The most thorough attempt at developing a systematic understanding of field specific differences in the adoption of new ICTs and the social shaping of scholarly communication practices has been conducted by Fry [Fry, 2003], and Fry and Talja [Fry and Talja, 2007] in a series of qualitative case studies. They make use of a taxonomy of scientific fields developed by Whitley [Whitley, 2000] that takes into account the functional and strategic dependencies between researchers and the degree of technical or strategic uncertainty of their research tasks. Fry and Talja argue that certain taxonomic types of fields shape their communication systems in particular ways. For example, they link high functional and strategic dependence in a field with a high degree of interdependence between scientists and a high degree of concentration of research efforts and goals, which they claim increases the degree of trust and accountability - shown by earlier research to be a crucial factor for the design and uptake of preprint servers.

As appealing as Whitley's taxonomy is as an explanatory framework for a systematic comparative analysis, it is problematic in its application to actual research collectives. At what level of aggregation is it appropriately applied? As Fry and Talja critically remark, Whitley has developed his taxonomy at the level of well-established disciplines (e.g. '20th century physics'). It falls short of capturing more dynamic, possibly transdisciplinary research collectives. Also, it has been observed that smaller, sub-disciplinary entities such as research specialties expose significant differences in social organization and research culture relative to their parent discipline or neighboring subfields [Becher and Trowler, 2001; Galison, 1997; Mulkay, 1977], and that they vary substantially in their cohesiveness, with some collectives being tightly integrated, while others remain rather diffuse [Crane, 1972; Mulkay, 1977].

To complicate matters further, scientific communication networks are multi-plex (institutionally, socially, epistemically driven), and overlapping, allowing for multiple membership of researchers in different communities. Where does one field end and another start? Even in bibliometrics and scientometrics, which have a decade long tradition of analyzing statistical properties of data sets of scientific publications [Borgman and Furner, 2002], the delineation of research fields, the starting point of any comparative analysis, presents an open research challenge [Zitt and Bassecoulard, 2008, 2006; Zitt et al., 2005; Adams et al., 2008; Mogoutov and Kahane, 2007]

We intend to approach this question empirically, combining qualitative and quantitative methods to help us arrive at notions of scientific community that can be meaningfully compared to characterize collective communication behaviors.

2.2 Combining Quantitative and Qualitative Approaches

There exist two strictly demarcated academic camps that contribute to research into scientific communication: 1) the qualitative, critically or relativistically-inclined Science and Technology Studies, and 2) the quantitative, positivistically-inclined scientometrics and bibliometrics [Van Den Besselaar, 2001]. Science and Technology Studies conceives of scientific communication holistically, considering its full range from informal communication such as 'technical gossip' [Knorr Cetina, 1999] to rhetorical moves in formally published scientific papers [Latour and Woolgar, 1986]. It has demonstrated in historical and ethnographic case studies how tightly interwoven communication practices are with the material and epistemic culture of a research field, see e.g. [Knorr Cetina, 1999; Hicks, 1992; Traweek, 1988; Shapin and Schaffer, 1985]. Because the focus is mostly on local practices of knowledge production, analysis of accumulated scientific action at a meso-level is underdeveloped [Glaser and Laudel, 2001].

Scientometrics on the other hand restricts itself to the formal, measurable output of research activities (such as publications and patents) at a high level of aggregation, analyzing the statistical properties or structural properties of the derived coauthor or citation networks. It posits its research objects (publications, citations) as generic and unproblematic, and suggests straightforward quantitative comparability - possibly up to some normalization factor - to account for field specific productivity or citation levels. It works with a standard model of scientific research and disregards field and context-specific interpretations of the underlying processes and communication acts [Lievrouw, 1990; Edge, 1979].

Very recently, a related, highly interdisciplinary area of research is emerging. It has been called '*network science*' [Börner et al., 2006], and builds on the availability of large digital data sets on the Web, and the development of algorithms for the analysis of large, complex networks built from these sets. Its goal is to increase our understanding of natural and man-made networks. It parallels scientometrics in its focus on numerical measures and structural analysis. The analysis of social networks (such as citation and coauthorship networks) is one of its application domains, and detection of community structures is an active branch of research. To date though there are few empirical studies - most empirical results on coauthor networks are published only as secondary results in method papers to demonstrate the applicability of a new algorithm, e.g. [Newman, 2001, 2004; Palla et al., 2005]. Consequently these results lack empirical studies to validate findings.

In isolation, we regard either methodology as insufficient for addressing the identification, characterization, and comparison of scientific communication cultures. A purely quantitative approach disregards context and motivations of participants and thereby risks comparing incommensurable entities, producing results lacking generality. On the other hand a purely qualitative approach does not scale-up to understanding collective phenomena due to its failure to identify a unit of analysis for comparison. It is therefore in danger of contributing only anecdotal insight to our research problem. Hence in our research we set out to critically investigate the measurable traces that communication cultures leave, both in formal literature and on the Web. Acts and patterns of communication need to be interpreted within their underlying processes and the specific cultural research context. To gain insight into the latter we are conducting ethnographic studies of research groups and their wider communities, and obtain participant feedback to evaluate structural representations.

Our decision to integrate the ethnographic and quantitative is further motivated by Callon who also challenges traditional scientometric approaches, as well as the new network science, for positing a unified social space, and a unicity of action regimes [Callon, 2006]. To account for the subjective, ego-centered creation of networks and their overlay, he instead suggests aiming for*"the most faithful representation of the configurations resulting from eternally unfinished work of composition and combination of egocentric networks."* So one may describe our research as critically exploring this space of a mostly descriptive representation of actors and relations, and assessing the potential for bridges to a network scientific approach that supports a quantitative comparison of scientific collectives and their communication patterns.

3 Research Approach

As argued above a key question for the comparative analysis of the field-specific shaping of scientific communication systems in the Digital Age is the identification and characterization of the *unit* of research: scientific collectives and their communication cultures.

To answer this question, we have started a comparative study in the larger context of two neighboring scientific disciplines: chemistry and physics. We chose this concentration for a number of reasons: selecting a set of fields where the journal article is a dominant form of formal communication ensures accessibility to network data of comparable quality and comprehensiveness (from the Web of Science database by Thomson Reuters); the surface similarity between the chosen fields challenges the methodological development and ensures adequate resolution; finally, chemistry presents a particularly interesting case of conservatism and resistance to change in scientific communication in comparison to more 'progressive' fields in physics. In spite of its decade-long tradition in producing and using digital information systems, the chemistry community, and notably its major professional societies and publishers, has resisted movements towards publishing reform and open access. Consequently, in comparison to physics or life sciences, chemistry can be perceived as a 'laggard' in adopting the Internet for new forms of scholarly communication.

From this larger scientific domain we have select three focal points for case studies: a subfield that can be seen as representing a traditional communication model (organic chemistry) [3], a hybrid field where different research cultures need to cooperate (physical chemistry), and a field from physics that has adopted the use of preprint servers, i.e. a rather progressive communication model of the Digital Age (such as subfields of condensed matter physics). Using an ethnographic approach we explore notions of community and scientific communication practices in these subfields. Based on this contextual input we analyze

---

[3] In our ongoing pilot study in physical chemistry participants repeatedly referred to organic chemistry as being particularly traditional in its practices, and *'different'* in their collaborative and communicative behavior.

community structures and behavioral patterns that are revealed in scientific networks that are derived from secondary data such as publication networks or extracted from web pages or other supplementary documentation (e.g. conference attendance lists). We iteratively evolve and validate our methods with the help of specific feedback from participants to our network analytic results.

Specifically, we pursue the following questions:

- Assuming that not all data types have the same relevance in every field, and not all structural elements the same meaning, what is the context-specific meaning of elements in scientific networks (e.g. co-citation links, a co-author grouping or co-author links between coauthor groupings)?
- How do the structures we extract from these networks relate to participants' views of their research environment and community connections? What kind of collectives are represented in scientific network structures? What are their roles and meaning for our participants?
- Can we distinguish between cohesive and rather diffuse communities: that is, areas with large inter-subjective agreement between participants on the definition of a specific community they participate in, and those fuzzy ones where subjective variation is rather high?
- To what extent can we identify communication cultures ethnographically in the collectives studied? How do they differ, e.g. do we find confirmation on differences in truth regimes and attitudes towards commercialization of information vs. socialization, as suggested in the literature [Walsh and Bayma, 1996; Kling and McKim, 2000; Bohlin, 2004]?
- How, and to what extent, are those characteristics of a communication culture reflected in features of these scientific networks? What differences in communication culture are *not* captured?
- Consequently, at what level of aggregation and with what generality or specificity of interpretation can quantitative methods contribute to a characterization of communication patterns of research collectives?

- What are implications of our findings for assessing the role of communication cultures in the shaping of new communication regimes?

In the next section we report results from a pilot study of this larger comparative study that we undertook to test and refine our methodological approach.

4 Results

We here report work-in-progress, with results from a pilot study that investigates scientific collectives as units of analysis for comparison of communication cultures. One of us has conducted a 12-week field study at two chemistry labs, one in physical chemistry in Europe, one in materials chemistry in the U.S.A. Both labs, led by their principal investigator (PI), are involved in several research specialties, hence complex ties exist to different scientific communities.

4.1 Ethnographic observations

We observed daily interactions within the lab and among members of one of its research teams during experimentation at a synchrotron facility. We further carried through semi-structured interviews on a range of topics relating to research practices, collaboration, and scientific communication. Below we describe preliminary observations from a physical chemistry lab that will guide our further research for validation.

4.1.1 Research Cultures and Implications for Scientific Communication Cultures

Most participants view the lab as divided into one chemistry subgroup and several physics subgroups. We assume that this division into chemistry and physics subgroups corresponds to distinct research cultures that we will tentatively identify as *preparative chemistry* and *experimental physics*.

The coexistence of these two research cultures within the lab is neither static nor in polar opposition.The configuration of people and projects in the lab is constantly evolving. The

PI describes his research program as guided by an overarching interest in the effect of size on physical and chemical properties of substances (from atoms or molecules through clusters and colloids to nanoparticles). Only a few years back his group has acquired preparative capabilities by hiring a postdoc to build up a subgroup for nano particle synthesis. In the local context of the lab, a potential opposition between a preparative chemistry and an experimental physics subculture is modulated by their need for cooperation with one another in order to be a functional part of a physical chemistry lab. What may distinguish the lab‘s chemistry subgroup‘s orientation from a prototypical preparative chemistry research group is their strong emphasis on analytics and on the systematic understanding of particle size effects. According to the PI, the publication of a journal article by his lab that focuses exclusively on synthetic methods is a rare exception.

In the preparative chemistry subgroup. daily activities are focused on creation and characterization of substances, here in particular nanoparticles (metal, semi-conductor, multilayered/structured). One of our informants from this lab, a physicist with a life-long career in a physical chemistry lab, observes that chemists take pride in creating substances themselves (rather than ordering them by catalogue as a physicist might do):

“Then chemists always say: this a physicist can do as well, somehow open the Olbridge catalogue and tick a box there [...] Then he [the PI] said, I must expect a graduated chemist to be able to produce the substance by himself [] Foremost, it is also dishonorable. So a chemists sees himself, even if he only wears a white lab coat and manipulates instruments, he can do that. It is always said so, he is no chemist, if he only turns knobs. In case of need he must be able, like as a student, to create the substance himself.”

Instruments on the other hand, are bought rather than built. Consequently knowledge of their inner workings is limited, and they are used as ready-made tools (‘black box). Our informant describes the difference between physicists and chemists as follows:

“A physicist has [...], in my experience, has for example a large-scale instrument that he has spent a lot of time on, say a mass spectrometer or a photon-electron-spectrometer, or something similar, and then he asks: what can I measure? A chemist does this exactly the other way around. He has a compound and asks, what methods will provide me with data for the conclusion I need, to find out for example, how big is this or how pure [...] many people are already in an advanced stage of their University studies [in chemistry], they have passed three quarters of their courses, and they dont know what a spectrometer consist of. They know these instruments, lets say, as a suitcase, where

you open a flap, and then you put in a sample, and what turns and moves, somehow a detector, they dont know. And I think neither will they know later."

In stark contrast to the chemistry subgroup, the building of instruments to measure properties of substances and observe physical processes is at the centre of activity of the physical subgroups. Even if large shared experimental facilities like synchrotrons are used, the instrument at the end of a beamline that represents the 'experiment' is build by the research group itself. For each beamtime of typically two weeks' duration the experiment is transported to the site and set-up as an end station of the beamline.

Building these instruments and setting-up the experiments is a difficult task and a major time investment. The lab owns instruments that have been developed, adapted and extended by generations of diploma and PhD students. Recollections by fellow co-workers, as well as theses that document technical details of the apparatus, become important information resources. The deputy lab head, points to three strategies for successful development of instruments: one is to learn from others. The other is an evolutionary cascading approach to instrument development, such that for each experiment only part of the instrument is changed, rather than having to construct a new instrument from scratch. Finally, in order to avoid inbreeding, or 'autism', a lab member can be sent abroad as a post-doc to a new experimental technique learn at another lab. Consequently scientific meetings and workshops are valued for the opportunity they provide to meet other people and learn about their experiences with certain methods. Similarly collaboration links to other groups through jointly funded projects provide access to critical technical knowhow. The predominance of instrument building and importance of collectively and not individually owned experimental capabilities is reflected in the following remark of one of the physicists in the lab, who identifies (lab) membership as an important property of individuals encountered at meetings

"You don't necessarily have to know him, but I know the research group, so you have to know, ah this guy belongs to this research group, well then one knows, yes, this kind of experiment, and they can do this, or they cannot do that."

The research cultures of preparative chemistry and experimental physics seem to differ in how they value open exchange and critical scientific discourse. One of the senior scientists interviewed expresses how he values and enjoys critical discussions in a specific community that he is involved in:

> *"It can well happen that your dear colleague says, 'something is wrong with this, I did this measurement too, this is not right', see? Naturally, this is very valuable when it happens. Because this happens in a community where you can simply , ah... nothing is embarrassing to us anymore, we are well beyond any notion of embarrassment, see? When someone says 'you have measured crap', 'Why did I measure crap, come over and explain to me', and then we start on the blackboard, and then there is a proper duel. That's fun, this really is.... the most beautiful moments. [with emphasis]"*

In contrast a senior researcher of the chemistry subgroup reveals a much more hesitant attitude towards scientific exchange at meetings. He feels a constant need to be alert about competition and cautious with regard to how open to be in communication with researchers from outside the lab. For example he describes the dilemma of what to put into an invited scientific talk. Whereas getting invited as a speaker is desirable as it looks good on a cv, it requires presenting new material as you cannot repeat the same story over and over again. This conflicts though with the perceived need of not being able to disclose much before publication for fear of someone else may pick up critical information and scoop your latest research.

The lab head confirms such differences between instrument oriented physical research and preparative chemistry with regard to openness in scientific communication. He points out that even published articles in chemistry may omit essential details or deliberately lead the reader astray - if this happens in physics he suggests, this is due to error or sloppiness but not malintent. He suggests that these differences are due to the rather different material conditions for generating scientific results: the former require substantial investment into instrument building and access to crucial facilities such as beamtime at synchrotrons, whereas the latter allow an individual scientist to produce publishable results within the time span of a few days. This affects how competitive advantages can be gained:

> *"To do one such experiment here, you need two years' preparation. It's much more secure, as you need access to a synchrotron source. [...]. This takes time. [In chemistry] it is pouring something together and sending [the article] off. There are research group that can produce material for two publications within a single week. Not much effort needed."*

This indicates that such differences in communication culture may correlate with Whitley's [Whitley, 2000] distinction between degrees of technical dependency in a field.

4.1.2 Notions of scientific communities

As explained by the PI, because research activities of the lab range across a number of specialties (such as x-ray cluster physics, nano particle synthesis, and environmental physics), there is not one unique scientific community with which al lab members identify. This is confirmed by interviews with further lab members. Some individuals identify with more than one distinct research community, whereas for others the idea of belonging to and participating in a specific research community seems to be very vague and only weakly developed. Our impression is that the very term community is more often introduced in conversations with members of the physics subgroups, while never actively used or elaborated on by members of the chemistry subgroup.

The very notions of scientific community that come across in the interviews are multifarious and no clear picture emerges at this point. It is apparently diverse and complex. There are small, coherent communities of researchers who actively seek opportunities for regular exchange and discussion, there are communities who find their forum at larger conferences as intersections of research fields, there are ad-hoc communities that form under the auspices of a common government research funding program and dissolve as the funding ends. Furthermore there are the user communities of large experimental facilities such as synchrotrons and the new free electron lasers. Two scientists we interviewed pointed to the high entry barrier into these communities. They suggest that the way applications for beamtime are handled makes it difficult for newcomers to get access to a synchrotron facility, as they lack a track record of successful experiments conducted at this specific facility. They characterize such user communities as closed, conservative communities. Finally, at a higher level of aggregation, there are the disciplinary communities organized in scientific societies at national level such as the German Physical Chemistry Society (Bunsengesellschaft) or the German Physical Society, DPG. Visiting the annual meetings of those scientific societies is desribed by most lab members as a standard exercise and duty. Having to show up to prove ones activity (both as a group, but also as an individual), and maintaining visibility at the national level is given as major motivation for attending these meetings.

In summary, lab members perceive themselves as members (or non-members) of a variety of research collectives or scientific communities that differ in their size, cohesiveness, and social organization, and that provide access to different resources such as critical scientific discourse, government funding, visibility critical for repuation and careers, material resources such as beam time, etc.

4.2 Network Analysis

In our pilot study we have explored publication networks at two different levels of aggregation:

At the research group level: A research group's publication output over a few decades typically amounts to several hundred publications. We obtain their bibliographic data from WoS[4]. The analysis of this data is directed at a) getting a better picture of the group's range of research interests, its collaboration patterns, and history, to complement the qualitative accounts of group members, and b) characterizing and comparing the research profiles of the most active groups in a community network, to learn e.g. about the degree of their investment of this community's field.

At the level of larger scientific collectives: We are exploring different ways of extracting data sets that represent a specific research specialty. To delineate sets of publication data that can be identified with collectives at various levels of aggregation we pursue bottom up and top down strategies. Bottom-up we rely on participants perspectives of the research networks they are involved in. With their support we formulate lexical queries that are iteratively refined or extended to capture a representative set of publication data from WoS. Top down, we are exploring an empirical approach to detecting community structures in publication networks using meta clustering. For this we combine article data from several entire journal years, amounting to several 10,000 articles for a 5 year period. Using a meta clustering approach [Caruana et al., 2006] we are investigating the variety of community substructures [5].

To visualize and analyze publication networks we use freely available tools, such as *Citespace* [Chen, 2005], and *pajek* [Batagelj and Mrvar, 2003]. Our focus has been on coauthor networks at this point, though we have looked as well at additional information gained

---

[4] WoS is reasonably comprehensive, though far from complete due to seemingly random omissions even for journals and time periods that have been indexed.

[5] Due to the high computation demand of this technique, results are not yet available for this paper

by including inter-citation links, and links that represent co-presence at workshops, see figure 5.

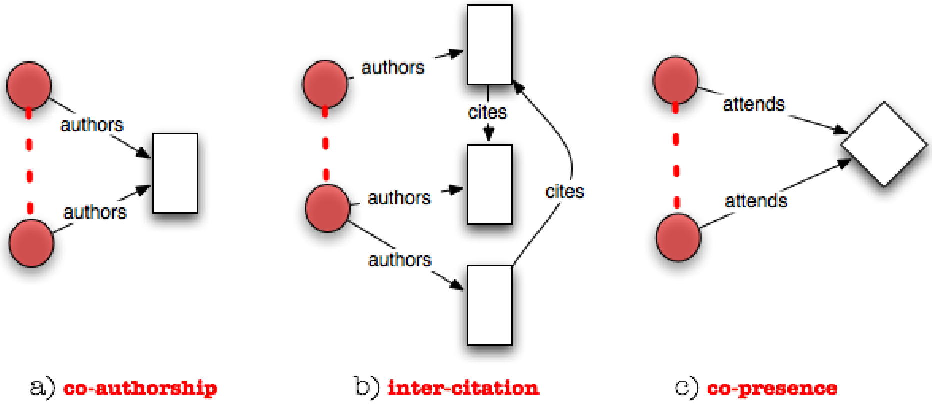


Figure 1: Social network relations (red dotted lines) derived from relations between social actors and artefacts. Circles represent scientific authors, rectangles represent scientific articles, and the square represents a scientific meeting such as a workshop or conference. The social network relations are symmetric (undirected links), whereas the underlying networks they are derived from have directed links.

4.2.1 Research Group Analysis

To study a research group's collaboration network we build the coauthor network from the group's publication output [6]. We found a tree ring map presentation of networks from *Citespace* most intuitive for reviewing with participants the range and history of collaborative ties, see figure 2). It color-codes the various time-slices during which a collaboration has been active, while the size of circles representing coauthors indicate the number of jointly published articles.

We used the term extraction features of *Citespace* to generate e.g. term networks to represent a group's scientific output and activities with limited success. No clear picture of the different research areas and their connections emerged from those term networks. As an

---

[6] Effectively we identify the group, and its publication output with the PI. This seems justified at least for those groups that we have been studying since the PI has a strong role in determining the group's research directions and appears as a coauthor on all publications of group members during the time they are part of the lab.

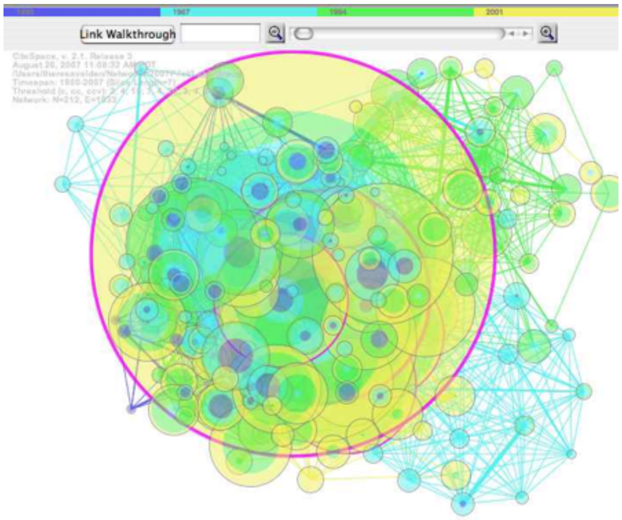


Figure 2: Co-author Network of a PI over three decades (vis. *Citespace*). The central circle with a pink rim represents the PI of the research group, all other circles represent individuals who have coauthored papers with the PI. Links represent joint co-authorships between individuals. The 28 year period has been sliced into 7 year slices and ring colors represent co-authorship events in those periods: dark blue -1980-1986, turquoise - 1987-1993, green 1994-2000, and yellow - 2001-2007. Hence circles filled with rings of all four colors point to long term collaborators. The size of rings represents the number of co-authored publications.

alternative, we analyzed the structure of the self-citation network that can be derived from the group's publications (where each node represents a publication, and directed links represent citations.) Hellsten et al. [2007] have shown the hidden value of self-citations for the analysis of scientists' movement between different research fields (so called field mobility). The more densely linked subgroups of nodes in this network (document 'clusters') identify research areas that the creators of theses publications have been active in. Due to the lack of a publicly available version of the code used for extracting those clusters in [Hellsten et al., 2007], we used the *dirgreedy* code provided by Rosvall and Bergstrom [2007] with encouraging results. As shown in the series of pictures in figure 3 those visible substructures of the network that were identified and labeled by the PI as representing relevant fields in his research were also extracted by the clustering algorithm (but for one exception, where the PI identified a group of documents as originating from his PhD thesis work while the algorithm could not distinguish this subgroup based on its citation link structure alone).

For a scientific œvre represented by ~ 99 publications that span 22 years we obtained nine different subareas. This level of granularity may seem to contrast with that found by Hellsten et al. [2007] in a study of a senior physicist whose œvre as retrieved from WoS spans sixty years and 315 publications. They restrict their analysis of the self-citation network to its largest subset of articles that are interlinked by self-citations. For these 251 articles they find three clusters. In addition to this largest subset, we considered also unconnected smaller components. The large connected component of our network is partitioned by the *dirgreedy* algorithm into five clusters. It remains to be seen whether this seemingly greater granularity is due to the different clustering algorithms used or whether this is an inherent property of the data sets studied.

Finally, following the example of Hellsten et al. [2007], we also generated a 'bar code' type representation of the group's publications that shows a chronologically ordered sequence of publications belonging to these different research areas. This allows a field mobility analysis of the scientist(s) in question. If the subfield analysis returns rather fine-grained results as in our case, the barcode representation can be rather difficult to analyze. We present in figure 4 the comparison of the field mobility bar codes of the PIs of the physical chemistry lab (PI 1), and the materials chemistry lab (PI 0) that we studied. As reported by Hellsten et al. [2007], the move from one field into another is seldom abrupt and definite. Instead 'old' topics are revisited during a scientific career. Nevertheless though we

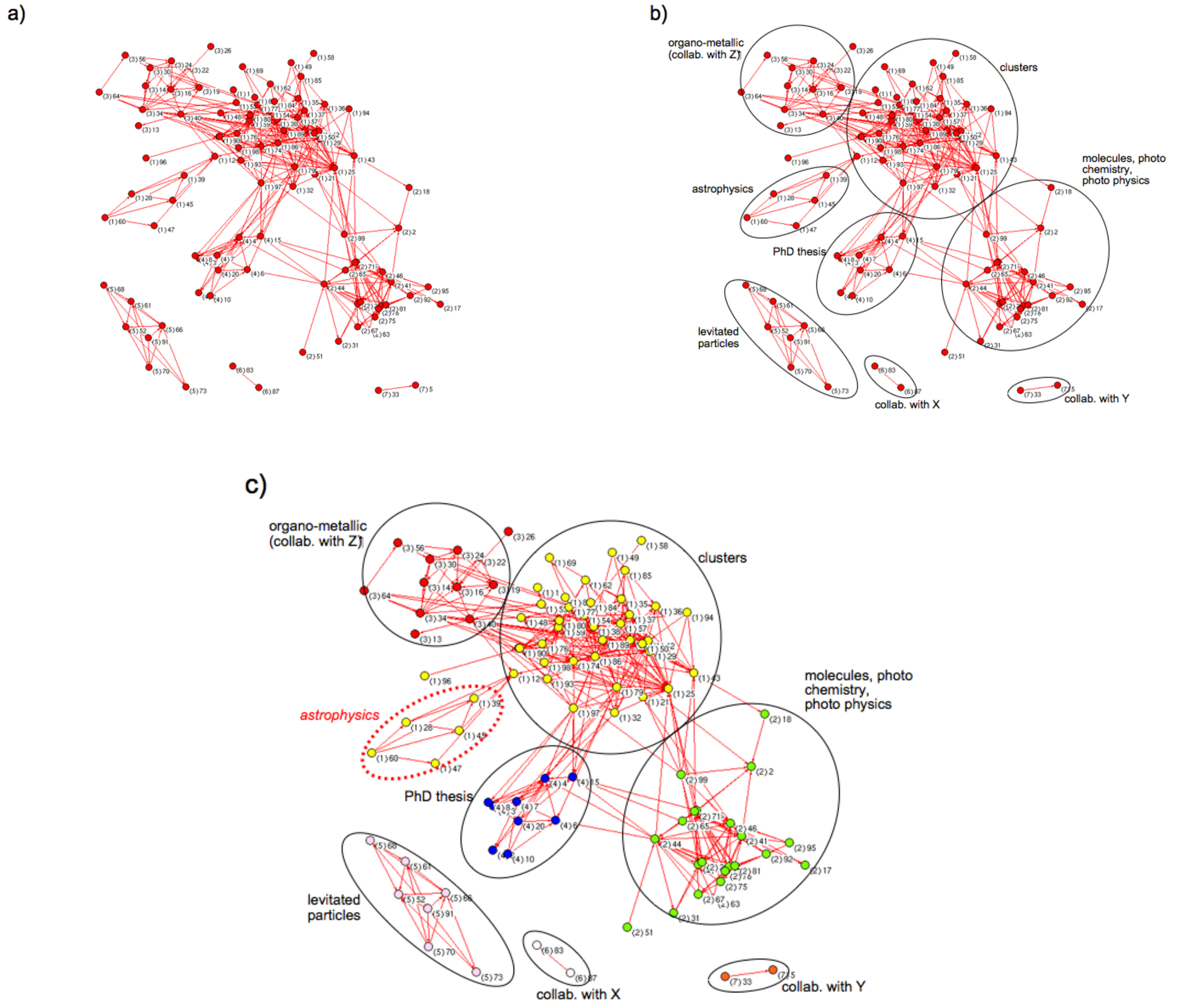


Figure 3: Self-citation network of a research groups covering a period of 22 years. a) visual representation of self-citation network using *pajek*. Nodes represent articles, arrows between nodes represent citation of one article by another. b) Identification of research areas by PI of the group based on the visual representation of the graph. The nodes are numbered to allow lookup of an article's title in a list of the group's publication output. c) The coloring of the nodes specifies the partitioning of the network into clusters as found by the clustering algorithm for directed networks *dirgreedy* [Rosvall and Bergstrom, 2007]

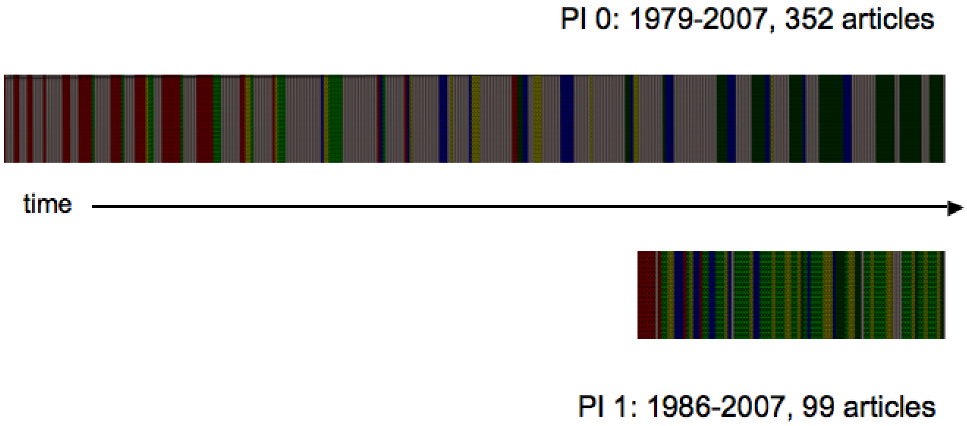


Figure 4: The temporal evolution of two research group leader's research interests gained from their published work in a so called bar code representation Hellsten et al. [2007]. Each stripe in a barcode represents an article, colors indicate which specific subject area an article belong to. The barcode of PI 0 shows a migration from one larger field (red and light green stripes) to another (dark blue and dark green stripes). The barcode of PI 1 shows greater continuity (but for the early PhD work, red stripes), i.e. that research topics are revisited.

can recognize in the bar code of PI 0 that he migrated in the early 90's from one larger topic area (chemistry) to another (materials science), see figure 4. This matches well with his own account of his career where he describes his move from chemistry to materials science as a significant development.

4.2.2 Community Structures

For mapping the structure of one of the scientific communities in which this lab is involved, we selected bibliographic data from Web of Science (WoS) using lexical queries. To ensure that we gather a representative set for this community, we repeatedly gather feedback from participants to validate and interpret our results. The series of graphs displayed in figure 5 exemplifies this process. We use the lexical query:

```
TS=((clusters OR "molecular* cluster*" OR "der-Waal* cluster*"
OR "free cluster*" OR "noble gas cluster*" OR "rare-gas cluster*"
OR "Argon cluster*" OR "AR cluster*") AND (inner-shell* OR
core-level* OR NEXAFS OR EXAFS OR XANES OR "soft x-ray*" OR
"interatomic coulombic decay"))

NOT TS=(ligand* OR zeolite* OR supports OR supported)
```

and further filter the records we retrieve for relevancy using the subject field of the WoS records. From the remaining 833 articles we build a coauthor network where nodes represent authors, and links represent joint coauthorship of two authors. In the construction of the network we exclude all authors that have authored only one single paper in the entire data set to simplify the resulting network. This simplified network has 632 articles and 642 authors. Links are weighted using a simple scheme that is the weight of a link between two authors equals the number of joint coauthorship incidents for these two authors, i.e. the number of papers those two have published together (possibly together with further coauthors). Part a) of figure 5 displays the largest component of this coauthorship network that consists of 419 authors. In the next step we use Rosvall's *undirgreedy* algorithm for undirected networks to partition the network into subcommunities of tightly linked authors. In part b) of figure 5 node colors indicate the clusters resulting from the partitioning. Some subgroups seem to expose a clear hierarchical structure with one large node that represents a research group leader.

From the underlying publication data we extract a list of all the author names for each cluster. We presented this list to participants that have been involved in this specific research field for many years and asked them to identify those author groups that they consider to be part of the community of researchers in this field. The subset resulting from this selection is shown in part c) of figure 5. To further analyze the composition of this small community network we produced an abstracted version, i.e. a community network that consists of nodes that represent an entire coauthor cluster and links between nodes represent coauthorship relations between any member of a cluster with any member of the other cluster, see part d) of figure 5. The size of circles corresponds to the number of individuals in the respective cluster, and links are weighted with the number of coauthor relationships of authors in different clusters. It turns out that the largest three clusters correspond to three large working groups in this area, each dominated by a senior scientist and PI (one is the lab we are studying, the other two are well known from accounts of lab members). These three groups are labeled in the diagram. Those three are connected to one another and seem to represent a kind of core for this community. Further groups are annotated based on participant commentary which specify either the type of research group (theory, method centered) or geographical classification (of the kind: “there must be some groups in Japan that show up in your data...”). It seems striking that (collaborative) connections between the various groups in this ‘community’ are rather weak.

Based on a similar data set (extracted using a more specific variant of the original query) we have further analyzed the lack of cohesion in this ’community’. So far we have built the coauthor network from data spanning the time period 1991-2008. Figure 6 shows how this cumulated view over large time-span can easily lead to misinterpretation of the strength and type of interactions between groups. The network generated from data covering the entire time period over 18 years suggests strong collaborative ties in particular between the groups of P2 and P3, see part a) of figure 6. A more detailed analysis that considers only the last nine years shows no coauthor ties at all between the three major research groups P1, P2, and P3, see part b) of figure 6 . Our informants explained this abrupt change as follows: in the late 90’s P3 was postdoc of P2 and hence they published together a substantial number of joint papers. Then P3 got a professorship in his home country and since then has been building his own research group embedded in strong collaborations with his former PhD supervisor and other colleagues in his home country. We conclude that inter-group collaborations between PI’s are rare to non-existent in this ‘community’ and collaboration links based on formal joint coauthorship are rather ephemeral.

Only when we consider further types of relations, such as inter-citation links[7] or co-presence at a biannual workshop series focusing on this specialized research area we see the faint footprint of an interacting ’community’ of researchers re-emerging, see series of graphs in figure 7.

---

[7] A link between two authors exists when they independently cite each others work in at least one publication each. We take this as an indicator of mutual awareness and recognition of one another as scientific peers.

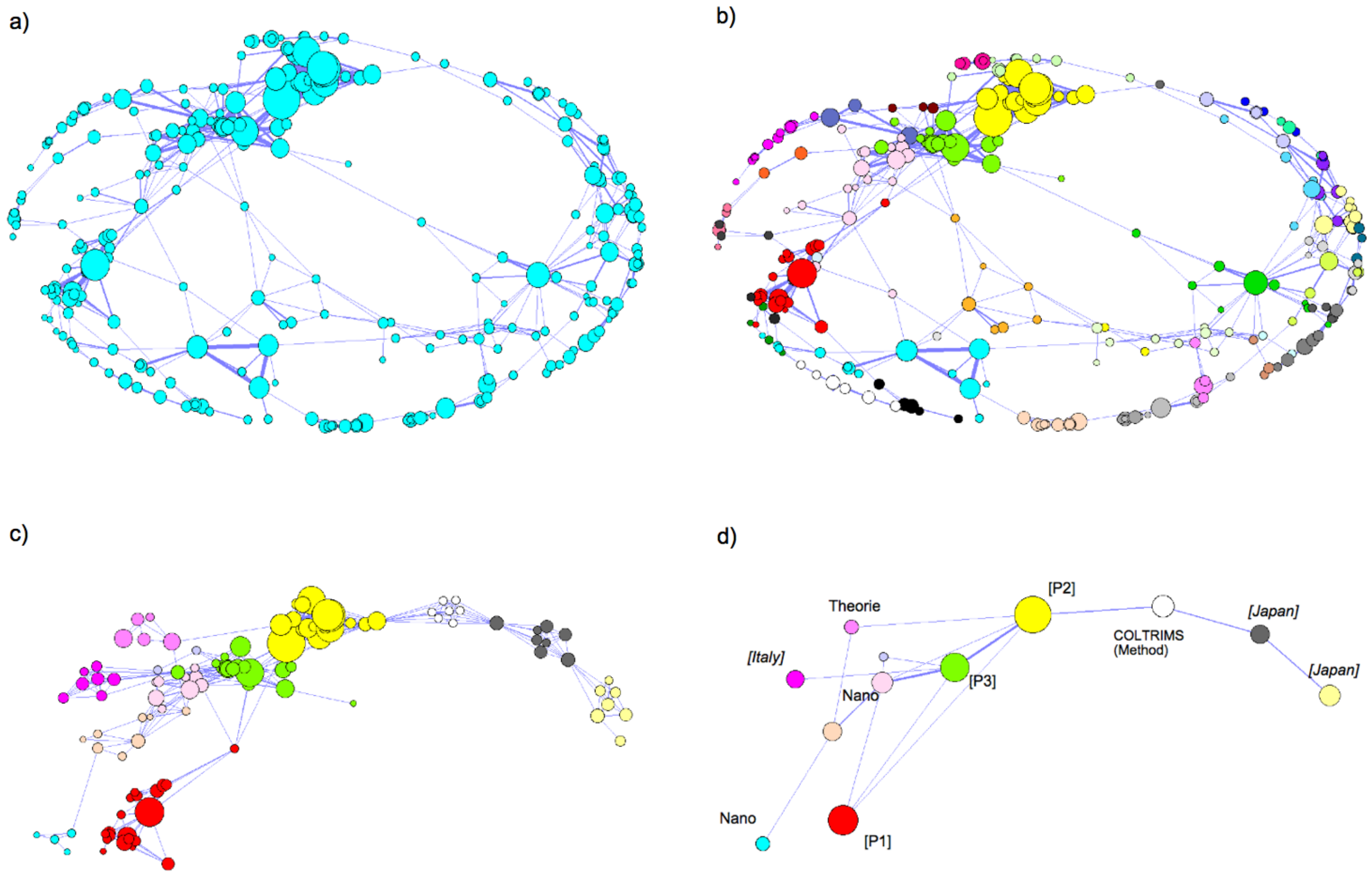
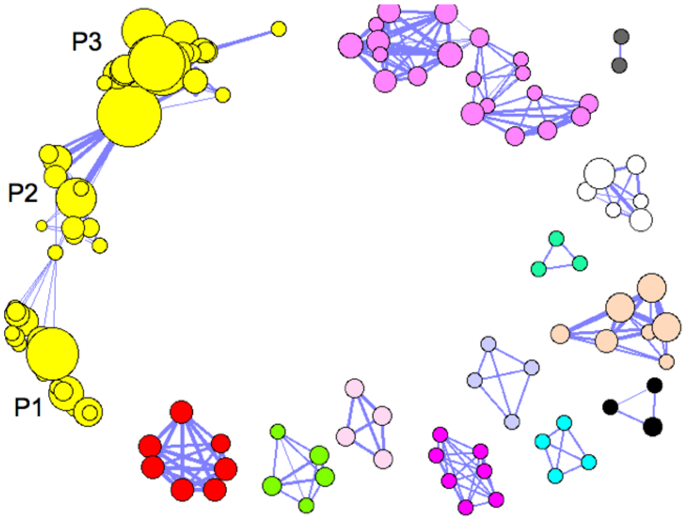


Figure 5: Exemplified structural analysis of a co-authorship network usinf *pajek* for viualization. a) largest connected component of co-authorship network, nodes represent authors that have coauthored at least two papers. The size of nodes reflects the number of papers included in this data set that an author has coauthored. b) Node color identify the node cluster retrieved by the graph clustering algorithm Rosvall and Bergstrom [2007]. c) subset of clusters selected by participant as belonging to his specialized scientific community. d) reduced network - author nodes of the same color are merged into a single dot. The size of the circles corresponds to the number of authors. The annotations are based on participant commentary, labels in square brackets our own annotations. See text for interpretation.

a) *1991-2008*



b) *2000-2008*

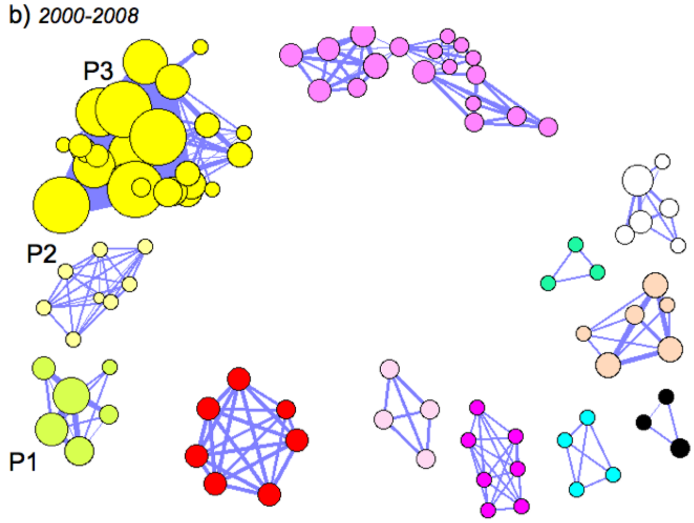


Figure 6: Co-authorship network of a very specialized scientific community, visualized using *pajek*. Nodes represent individual authors, their size the number of papers that they have coauthored, and node colors distinguish connected components of the network. a) The cumulated data set over 18 years exposes a large connected component (yellow nodes) that suggests collaboration between three research groups (P1, P2, P3). b) Restricting the time frame to the last 9 years results in a break-up of this largest component, pointing to the ephemerality of coauthorship links.

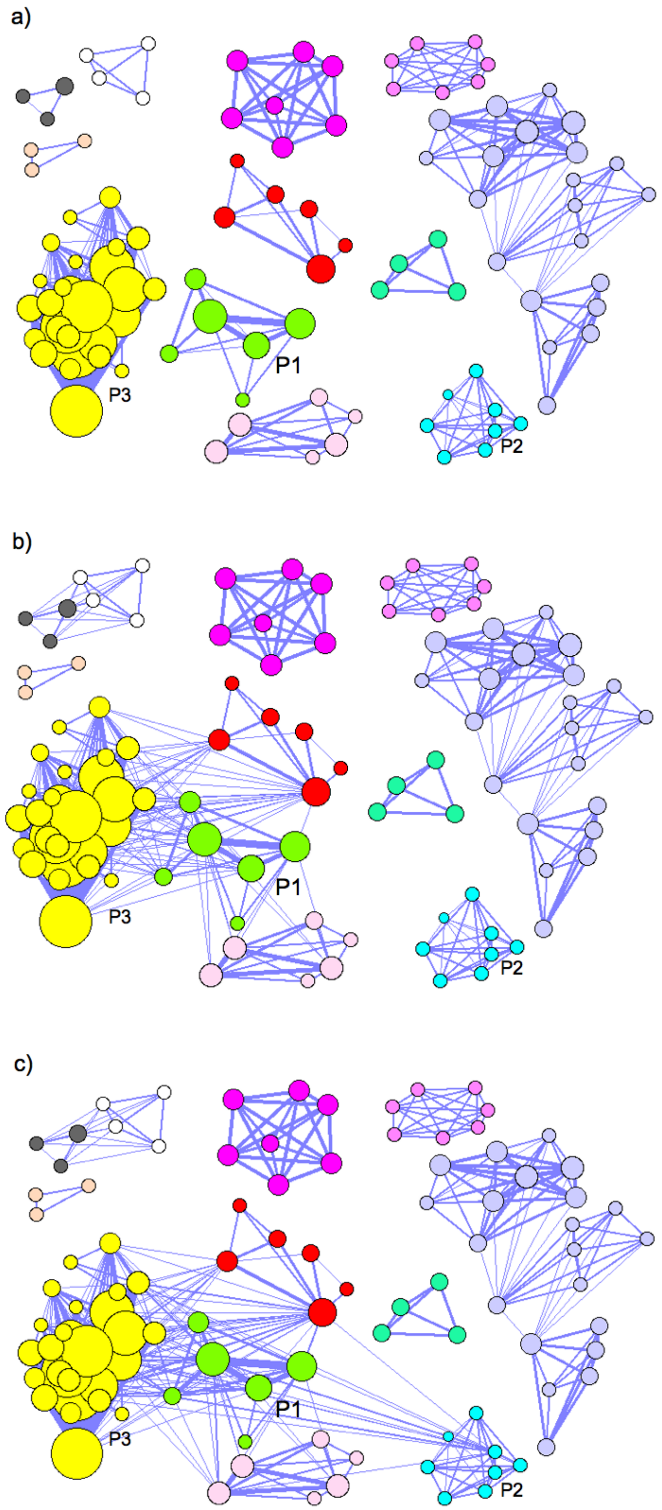


Figure 7: Nodes represent individual authors, links different types of relationships: a) co-authorship network, b) coauthorship network extended by inter-citation links. c) coauthorship network extended by inter-citation and co-oresence at workshop links. Node sizes indicate number of publications coauthored by an author, colors distinguish the connected components of the initial coautorship network.

4.3 Discussion

We have learned a number of things from this analysis and the discussions of the resulting networks with our informants. We have gained substantive insight into structural properties of this community that displays extremely weak collaboration links at the level of formal publications, but is well connected through co-participation in small, specialized workshops. This points to the potentially strong role of such events for communication and community building. We acknowledge that this perspective on this community is one informed so far only from one of the major players involved. We are curious to learn how accounts of other major players (such as P2, and P3), or members of groups on the periphery of this community, differ.

Methodologically we recognize the criticality of temporal factors in our analysis as well as the value of participant commentary for interpretation. We further observe the difficulty of applying 'community' as a guiding notion in our discussions with informants - they repeatedly ask back what we mean to capture, and are unsure of whether to regard a particular group as in or out. So community, and in particular its boundaries, present themselves as rather fluid concept in this research environment.

As described above we are assembling a tool box for structural analysis and comparison of networks of scientific networks. Our next steps will be to complete our basic set of analyses by looking in more detail at different types of nodes and their role within the larger network. Then we will move on to conduct those kind of analysis for several fields to compare features. We expect that this will trigger further questioning of network features and may reveal pointers to community specific social organization and communication patterns. So far, our ethnographic observations strengthen our assumption about different communication cultures at some level of aggregation. The diversity of community notions remains challenge for matching such communication characteristics to scientific collectives.

5 Conclusions

In this paper we have argued that the differences in the manner and pace in which different groups of scientists transform their communication systems into the Digital Age are poorly understood. As a major challenge for a systematic analysis of such differences between research domains we have identified the lack of clarity about an appropriate unit of analysis for comparative study of scientific communication cultures. In our research we take an empirical approach to resolving this issue. During a pilot study we have developed a basic methodological approach for the investigation of communication cultures. First results underline the value of a rich methodology that combines network analysis with qualitative, ethnographic research.

References

J. Adams, K. Gurney, and L. Jackson. Calibrating the zoom - a test of Zitts hypothesis. *Scientometrics*, 75 (1):81–95, 2008.

V. Batagelj and A. Mrvar. Analysis and visualization of large networks. In *Graph Drawing Software*, pages 77–103. Springer, Berlin, 2003.

T. Becher and P. R. Trowler. *Academic Tribes and Territories*. The Society for Research into Higher education, 2001.

W. Bijker, T. Hughes, and T. Pinch. *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. MIT Press, 1989.

I. Bohlin. Communication regimes in competition: The current transition in scholarly communication seen through the lens of the sociology of technology. *SOCIAL STUDIES OF SCIENCE*, 34(3):365–391, 2004.

C. L. Borgman. *Scholarship in the Digital Age*. The MIT Press, 2007.

C. L. Borgman and J. Furner. Scholarly communication and bibliometrics. *Annual Review Of Information Science And Technology*, 36:3–72, 2002.

K. Börner, S. Sanyal, and A. Vespignani. Network Science: A Theoretical and Practical Framework. *Annual Review of Information Science and Technology*, 40, 2006.

M. Callon. Can methods for analysing large numbers organize a productive dialogue with the actors they study? *European Management Review*, 3:7–16, 2006.

R. Caruana, M. Elhawary, N. Nguyen, and C. Smith. Meta clustering. In *Proceedings of the Sixth International Conference on Data Mining (ICDM'06)*, 2006.

C. Chen. Measuring the movement of a research paradigm. In *Proc. of SPIE-IS&T , San Jose, CA, 17-18 January, 2005. Visualization and Data Analysis 2005 (VDA 2005)*, volume 5669, pages 63–76. SPIE and IS&T., 2005.

D. Crane. *Invisible Colleges - Diffusion of Knowledge in Scientific Communities*. The University of Chicago Press, 1972.

B. Cronin. Scholarly communication and epistemic cultures. In *Scholarly Tribes and Tribulations: How Tradition and Technology Are Driving Disciplinary Change. ARL, Washington, DC, October 17, 2003*, 2003.

H. V. de Sompel, S. Payette, J. Erickson, C. Lagoze, and S. Warner. Rethinking scholarly communication — building the system that scholars deserve. *D-Lib Magazine*, 10(9), 2004.

D. Edge. Quantitative measures of communication in science: a critical review. *Hist Sci*, 17(36 Pt 2):102–34, 1979.

J. Fry. *The Cultural Shaping of Scholarly Communication Within Academic Specialism*. PhD Dissertation (unpublished), University of Brighton, 2003.

J. Fry and S. Talja. The intellectual and social organization of academic fields and the shaping of digital resources. *Journal of Information Science*, 33(2):115, 2007.

P. Galison. *Image and Logic: A Material Culture of Microphysics*. University Of Chicago Press, 1997.

J. Glaser and G. Laudel. Integrating scientometric indicators into sociological studies: methodical and methodological problems. *SCIENTOMETRICS*, 52(3):411–434, 2001.

Hellsten, Lambiotte, Scharnhorst, and Ausloos. Self-citations, co-authorships and keywords: A new approach to scientists' field mobility? *Scientometrics*, 72:469–486, 2007. doi: 10.1007/s11192-007-1680-5. URL `http://dx.doi.org/10.1007/s11192-007-1680-5`.

D. Hicks. Instrumentation, Interdisciplinary Knowledge, and Research Performance in Spin Glass and Superfluid Helium Three. *Science, Technology & Human Values*, 17(2):180, 1992.

S. Hilgartner. Biomolecular databases: New communication regimes for biology? *Science Communication*, 17(2):240–263, Dec. 1995. URL `http://scx.sagepub.com/cgi/content/abstract/17/2/240`.

C. Hine. *Systematics as Cyberscience: Computers, Change, and Continuity in Science*. MIT, 2008.

R. Kling and G. McKim. Not just a matter of time: Field differences and the shaping of electronic media in supporting scientific communication. *Journal of the American Society for Information Science*, 51(14): 1306–1320, 2000.

R. Kling, G. McKim, and A. King. A bit more of it: Scholarly communication forums as socio-technical interaction networks. *Journal of the American Society of Information Science and Technology*, 54:47–67, 2003.

R. Kling, L. B. Spector, and J. Fortuna. The real stakes of virtual publishing: The transformation of E-Biomed into PubMed central. *Journal of the American Society for Information Science and Technology*, 55(2):127–148, 2004. URL `http://dx.doi.org/10.1002/asi.10352`.

K. Knorr Cetina. *Epistemic Cultures - How the Sciences Make Knowledge*. Havard University Press, 1999.

B. Latour and S. Woolgar. *Laboratory Life: The Construction of Scientific Facts*. Princeton University Press, 1986.

L. A. Lievrouw. Reconceiling structure and process in the study of scholarly communication. In *Scholarly Communication and Bibliometrics*. Sage Publications, 1990.

A. Mogoutov and B. Kahane. Data search strategy for science and technology emergence: A scalable and evolutionary query for nanotechnology tracking. *Research Policy*, 36(6):893–903, 2007.

M. Mulkay. Sociology of the scientific research community. In Spiegel-Rösing, editor, *Science, Technology, and Society: a cross-disciplinary perspective*. 1977.

M. Nentwich. Cyberscience: modelling ICT-induced changes of the scholarly communication system. *Information, Communication & Society*, 8(4):542–560, 2005.

M. Newman. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2):404, 2001.

M. Newman. Coauthorship networks and patterns of scientific collaboration. *PNAS*, 101:5200–5205, 2004.

G. Palla, I. Derenyi, I. Farkas, and T. Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–8, 2005.

M. Rosvall and C. Bergstrom. An information-theoretic framework for resolving community structure in complex networks. *Proceedings of the National Academy of Sciences*, 104(18):7327, 2007.

S. Shapin and S. Schaffer. *Leviathan and the Air-Pump*. Princeton University Press, 1985.

S. Traweek. *Beamtimes and Lifetimes: The World of High Energy Physicists*. Harvard University Press, 1988.

P. Van Den Besselaar. The cognitive and the social structure of STS. *Scientometrics*, 51(2):441–460, 2001.

J. P. Walsh and T. Bayma. Computer networks and scientific work. *Social Studies of Science*, 26(3):661–703, 1996. ISSN 03063127. URL `http://www.jstor.org/stable/285703`.

R. Whitley. *The intellectual and social organization of the sciences*. Clarendon Press, 2nd edition, 2000. (1st published 1984).

M. Zitt and E. Bassecoulard. Delineating complex scientific fields by an hybrid lexical-citation method: An application to nanosciences. *Information Processing & Management*, 42(6):1513–1531, Dec. 2006. URL `http://www.sciencedirect.com/science/article/B6VC8-4K42DTF-1/2/63e170974f277bb3cbb0d4c7877f56f7`.

M. Zitt and E. Bassecoulard. Challenges for scientometric indicators: data demining, knowledge-flow measurements and diversity issues. *Ethics In Science And Environmental Politics*, 8:49–60, 2008.

M. Zitt, S. Ramanana-Rahary, and E. Bassecoulard. Relativity of citation performance and excellence measures: from cross-field to cross-scale effects of field-normalisation. *Scientometrics*, 63(2):373–401, 2005.