

# Mapping Scientific Communities to Scale-up Ethnographies

**Theresa Velden**  
Cornell University  
Department of Information Science  
301 College Avenue  
Ithaca, NY 14853  
tav6@cornell.edu

**Carl Lagoze**  
Cornell University  
Department of Information Science  
301 College Avenue  
Ithaca, NY 14853  
cjl2@cornell.edu

## ABSTRACT

We will present the multistep process for generating bibliometric mappings of research fields and their community structure, a process that we have developed using a combination of network analysis and ethnographic field studies of scientific communities. We suggest that such maps are useful to support the strategic sampling of ethnographic field sites and the transparent scaling-up of ethnographic findings for the comparative study of collaboration and communication practices across scientific fields.

## Categories and Subject Descriptors

J.4 [Computer Applications]: Social and Behavioral Sciences, E.1 [Data Structures]: Graphs and Networks

## General Terms

Human Factors

## Keywords

Comparative research design, ethnographic field studies, network analysis, field specific scientific communication practices.

## 1. INTRODUCTION

There is growing interest in understanding research field-specific communication and knowledge sharing practices (Cronin 2003) to guide policy design and the development of cyberinfrastructure services. However, an in-depth theoretical understanding of what field specific factors are relevant to explain differences between fields, and how they shape communication and knowledge sharing practices, is lacking. Developing such a theoretical understanding requires input and validation from empirical studies that systematically explore and compare field specific influences. Such studies need to combine nuanced context sensitive understandings of local practices with the ability to scale up such observations to an appropriate field-level for comparison.

While the ethnographic genre of ‘laboratory studies’ that focuses on research practices at the localities of scientific work is well-established, ethnographic field studies that analyze the collective production of scientific knowledge in distributed research communities are more challenging. A multi-sited ethnography approach (Marcus 1995) along with long-term immersion in the field can produce a rich understanding of research cultures, however such studies require extensive resources and time commitment. Research designs that include

comparisons across fields (e.g. Knorr Cetina 1999) are extremely rare, as are systematic explorations of community boundaries and areas of overlap that would support the scaling-up of findings by specifying domains of validity.

Comparative field-level or discipline-level research designs are much more common in bibliometric network studies. The analysis of large numbers of publications can provide an important macro-level perspective onto the collective production process in the sciences. However, bibliometric data is limited in the kind of behaviors it captures. Co-authorship is a useful indicator of collaboration, but many other forms of cooperation remain undocumented (Laudel 2002). Although bibliometric studies can point to striking field differences, they suffer from a lack of explanatory power, as they provide only limited insight into the research context in which publications are created and the motivations of the scientists involved (Lievrouw 1990).

Recognizing the limitations in the sole application of either of these methodological approaches, we have developed and had initial success with a combination of ethnographic and network analytic methods building on their complementary strengths.

## 2. METHODOLOGY

We have been experimenting with a methodology that combines ethnographic field studies with large-scale network analysis to undertake a comparative case study of communication practices and openness and sharing in two research specialties in the physical and chemical sciences (Velden 2011). The value of combining network analysis with ethnographic field studies has also been demonstrated by Howard’s work (2002). We can iteratively use network analysis and ethnographies with results of one feeding dynamically into the other. In addition, by combining both methods we can leverage the strategic sampling supported by network analyses of a community. Finally, network analysis permits tracking temporal evolutions of community structures that can be hard to capture by ethnography alone.

Our approach to network analysis builds on recent advances in the algorithmic analysis of complex networks to extract structures of collective organization from co-author networks. Our work is innovative due to its focus on mesoscopic network features (Guimera 2007), that is the analysis of co-author clusters and their interlinking patterns, and in its combination with ethnographic observations to interpret these features. The focus on mesoscopic network features accounts for the increasing role of scientific research in teams (Wuchty et al. 2007) and supports the field-level comparison of team structures and patterns of collaboration.

## 3. RESULTS

The three key components of our methodology as developed at this point are:

Copyright is held by the author/owner(s).

*iConference 2012*, February 7–10, 2012, Toronto, Ontario, Canada.  
ACM 978-1-4503-0782-6/12/02

*Field Delineation* The use of a combination of ethnographic field studies and bibliographic data analysis to delineate a research specialty and scientific community as the analytic unit for comparison is a crucial and challenging step. We retrieve data spanning a 20-year period using a lexical query of the Web of Science database, and disambiguate author names algorithmically to reduce the distortion of co-author networks by name homonymy (Velden et al. 2011). To ensure that the data retrieved represents a research field well, precision and recall need to be optimized. To check recall, we make use of the self-citation network of renowned researchers (Hellsten et al. 2007) to assess overlap of their published work with the target data set. To check precision we determine a subdivision of the data into topic areas by clustering the field-wide citation network, and assess social cohesion in terms of author overlap between these topic areas.

*Group Structures and Inter-Group Collaboration* The iterative analysis and interpretation of network features, informed by understandings developed during field studies, results in refinements of the network analytic instruments, as well as new questions to be pursued in field studies. This complimentary informing and refinement of the two methods continues to iterate. We have had initial success with this approach, extracting and comparing across fields the smallest collective units of research in a research specialty, and extracting global inter-group collaboration networks, filtering out inter-group co-author links that signify other forms of exchange (Velden et al. 2010).

*Mapping of Sub-Community Structures* The global collaboration network is combined with data on the topical subdivision of the research field derived from document citation networks to generate community structure maps. We argue that the resulting network visualizations of community structures such as figure 1 are informative because they provide insight into the complexity of research fields as analytic unit for comparative studies. They provide valuable information to the field researcher on where local observations may hold, and where further field sites or interview partners may be sampled from to support the scaling-up of ethnographies.



**Figure 1. Activity pattern of a sub-community in a global research specialty. Nodes represent research groups or tight research networks, and links inter-group collaboration extracted from co-authorship data. Node colors indicate publishing activity (dark: high, light: low) in a major topic area within a research specialty.**

## 4. CONCLUSION

Our poster will describe the multistep process of how to produce mappings of research fields and their community structure. Such maps are intended to support the strategic sampling of ethnographic field sites for the comparative study of collaboration and communication practices across scientific fields. The work presented is part of an ongoing wider methodological research program on integrating publication network analysis and ethnographic methods for research on the collective production of knowledge in the sciences. Future work will add a temporal dimension to the analysis.

## 5. ACKNOWLEDGMENTS

Supported by National Science Foundation grant OCI-1025679.

## 6. REFERENCES

- [1] Cronin, B. (2003). Scholarly communication and epistemic cultures. *New Review of Academic Librarianship*, 9, 1-24.
- [2] Guimera, R., Sales-Pardo, M., and Amaral, L. (2007). Classes of complex networks defined by role-to-role connectivity profiles. *Nature Physics*, 3(1):63-69.
- [3] Hellsten, Lambiotte, Scharnhorst, and Ausloos (2007). Self-citations, co-authorships and keywords: A new approach to scientists' field mobility? *Scientometrics*, 72:469-486
- [4] Howard, P. N. (2002), *Network Ethnography and the Hypermedia Organization: New Media, New Organizations*, *New Methods New Media & Society*, 4(4): 550-574
- [5] Knorr Cetina, K. (1999). *Epistemic Cultures - How the Sciences Make Knowledge*. Harvard University Press.
- [6] Laudel, G. (2002). What do we measure by co-authorships. *Research Evaluation*, 11, 3-15.
- [7] Lievrouw, L. A. (1990). Reconciling Structure and Process in The Study of Scholarly Communication. In C. L. Borgman (Ed.), *Scholarly Communication and Bibliometrics* (59-72). Sage Publications.
- [8] Marcus, G. (1995). Ethnography in/of the world system: The emergence of multi-sited ethnography. *Annual Review of Anthropology*, 24, 95-117.
- [9] Velden, T., Haque, A., and Lagoze, C. (2010). A new approach to analyzing patterns of collaboration in co-authorship networks: mesoscopic analysis and interpretation. *Scientometrics*, 85(1):219-242
- [10] Velden, T., Haque, A., and Lagoze, C. (2011). Resolving author name homonymy to improve resolution of structures in co-author networks. In *Proceeding of JCDL '11*, (241-250), ACM, New York, NY, USA.
- [11] Velden, T. (2011). A methodology for studying field differences in scientific communication — explaining openness and sharing in two scientific communities in the chemical and physical sciences. PhD Thesis, Cornell University.
- [12] Wuchty S., Jones, B.F., Uzzi, B. (2007) The Increasing Dominance of Teams in Production of Knowledge. *Science* 316 (5827) p. 1036-1039