

Mapping The Temporal Evolution of Scientific Community Structures

Theresa Velden

School of Information
University of Michigan
309 S. State Street
Ann Arbor, MI 48104, USA
tvelden@umich.edu

Kan Yu

School of Information
University of Michigan
309 S. State Street
Ann Arbor, MI 48104, USA
kanyu@umich.edu

Shiyan Yan

School of Information
University of Michigan
309 S. State Street
Ann Arbor, MI 48104, USA
shiyansi@umich.edu

Carl Lagoze

School of Information
University of Michigan
309 S. State Street
Ann Arbor, MI 48104, USA
clagoze@umich.edu

ABSTRACT

We present results from tracking the temporal evolution of community structures within a research specialty. This extends our previous work in which we generated static community maps that combine co-author networks and direct citation networks derived from 20-years of publications in the research specialty. Here, we explore how the temporal evolution of these maps can be used to provide insights into the historical evolution of a field as well as extract more accurate snapshots of the community structures at a given point in time. Such time resolved mappings are an important component in our mixed method approach that integrates network analysis with ethnographic field studies to investigate field-specific communication and collaboration behaviors in scientific communities (Velden & Lagoze 2013). Evaluating these maps in qualitative interviews with field experts will be the next step in our research.

Keywords

Scientific Community, Network Analysis, Temporal Evolution, Science Maps.

INTRODUCTION

The mapping of scientific communities and their

collaboration networks (Morris & Van der Veer Martens 2008) is flourishing for a number of reasons. The increase of services that provide access to bibliographic data on scholarly publications such as dedicated search engines, commercial databases, and subject-specific and institutional repositories, promises the growing availability of a comprehensive, machine-readable corpus for this kind of analysis. Complementing this is the emergence of sophisticated algorithms for the analysis of complex networks (Newman 2003) and the wide availability of advanced user-friendly network analysis and visualization tools like pajek, gephi and VOS Viewer. These technical advances are met by a growing interest of science policy to improve the empirical base for data based decision-making, as evidenced for example by the SciSciP program at NSF¹. Maps of science help to visualize, compare and track through time data on specific entities (such as institutions or research groups) for evaluative purposes (Rafols et al. 2010). In our work we are interested in science maps to support a mixed-method approach that integrates network analysis with the ethnographic study of research specialties to understand field-specific patterns in behaviors such as data sharing and collaborative practices (Velden & Lagoze 2013).

Scientific research specialties are a complex interwoven social and cognitive phenomenon. Sociologically, a research specialty can be defined as a collective production system, that is a constellation of autonomous actors,

¹ Science of Science and Innovation Policy (SciSIP) website:

http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=501084

researchers or research groups, whose activities are indirectly coordinated by a shared knowledge base to which they all strive to contribute (Glaser 2006, Velden 2013). Therefore, in order to capture the dynamics of science, the combined analysis of social and cognitive structures is of particular interest (Ding 2011). In our community mapping approach we focus on the evolution of topics within a research field along with the evolution of collaborative links between research groups in that field. Specifically, to support our empirical field studies we seek to develop tools to facilitate the going back and forth between aggregate community-level structures and the experiences and research trajectories of individual researchers and research groups. In our previous work, we generated static community maps based on data from 20 years of publications in a field. In this paper, we explore how the temporal evolution of these maps can be used to provide insights into the historical evolution of a field as well as extract more accurate snapshots of the community structures at a recent point in time, which then can be used to support the design of ethnographic field or expert interview studies.

METHOD

The data used in this study was extracted from bibliographic records retrieved from the advanced Web of Science web service API in October 2013 using a lexical query to search for topics in the Title, Abstract, Author Keyword, and Keywords Plus fields of publication records. It covers publications in the 22-year period from 1991 to 2012.

The basic entities in our analysis are research groups as represented by co-author clusters in a co-author network (Velden et al 2010), and topic areas, as represented by clusters of document clusters in a direct citation network (Velden & Lagoze 2013). We use the information theoretic clustering algorithm by Rosvall & Bergstrom (2008) for the extraction of clusters from the networks. For those basic entities a topic affinity network and an inter-group collaboration network are constructed, and a topic overlay map is generated that combines information on the publication activity of a research group in a topic area with the group collaboration network.

1. In the topic affinity network, the nodes represent topic areas, i.e. groupings of publications that are closely interlinked by direct citations. Edges represent the relationship between topic areas based on overlap in co-authors or number of inter-citations. To make the analysis tractable we select only the eleven largest topic areas for inclusion in the network. They correspond to 82% of documents in the giant component of the citation network. We develop two annotations to describe and distinguish the subject matter of the topic areas, one derived from the titles of the most popular journals in each topic area, and another derived from specific keywords and their relative frequency in the article titles of the respective.

2. The inter-group collaboration network is a filtered and collapsed version of the underlying co-author network, a process described in (Velden et al 2010). The nodes represent research groups. Links between these groups are filtered to extract only those (more complex and strong links) that indicate actual inter-group collaboration links, filtering out the residuals from author migration from one to another group or the more ephemeral links of one-off collaborations (oftentimes service collaborations).
3. The topic overlay maps are generated by creating a network partition file that indicates, for each node (group) in the group collaboration network and for each of the eleven largest topic areas, what proportion of a group's publications belong to the respective topic area.

The fact that the nodes in our networks are not atomic but represent clusters of documents or authors, introduces a considerable challenge to capturing and visualizing the temporal evolution of these networks. After all, not just the network structure (i.e. the existence and strength of edges between nodes) changes, but also the very composition of the nodes is dynamic. There is still limited work on the evolution of clustered networks and their visualization that addresses these challenges (Giatsoglou & Vakali 2013, Holme & Saramäki 2012), and the applicability of existing work (e.g. Yan et al. 2012, Rosvall & Bergstrom 2010, Lin et al 2009) depends on the specific circumstances, such as the type of network and the analytic purpose of the visualization.

We considered two alternatives for introducing temporal dynamics. In the first, a 'semi-static' approach, the nodes (research groups, topics) are derived from the aggregate data, i.e. by clustering the accumulative 1991-2012 co-author network and document citation network. When using this approach the temporal changes of the network are captured merely by adjusting link strengths and node sizes to reflect the data that falls into the given time window of interest. This approach enables easy identification and tracking of entities (groups, topic) across time but has obvious trade-offs in precision (e.g. one loses information about splits or mergers of entities). In an alternative 'dynamic' approach, the clustering of the underlying network is repeated for each time window to extract dynamic clusters that accurately represent the entity (group, topic) during that specific time window. However, when using this approach, establishing the identity of an entity across time becomes non-trivial. For example, in a co-author network, the continuity and distinctiveness of a research group may be obscured by the noise of temporary collaborations of this group with other groups. Whereas in the accumulative 20-year data set the bibliographic footprint of a research group can be easily extracted, the pattern in a shorter time window becomes much more vague. Hence we decided to use the semi-static approach for the larger inter-group collaboration network where, due

to its size of several hundred groups, tracking the identity of groups across time slices is extremely difficult. For the much smaller topic affinity networks with only eleven nodes, we chose the dynamic approach. In both cases we used 8-year windows to break down the 22-year period into smaller time steps².

RESULTS

In our poster we plan to show visualizations and interpretations of maps of cognitive and collaborative community structures that we produced using the approach described above. Here are some of our main observations:

Evolution of Topic Areas

The accumulative (1991-2012) topic affinity network depicts the research specialty as consisting of an almost linear alignment of research topics that can be associated with specific (sub)disciplines in physics, chemistry and materials science. It stretches from cluster physics, through surface science, materials chemistry to cluster chemistry. The map shows a weak link between the extreme ends of this alignment, between Cluster Physics to Inorganic Chemistry (a new development not yet visible in an earlier 1991-2010 version of this map that is provided in Velden & Lagoze al (2013)).

The temporal resolution of the topic affinity networks provides further insights into the underlying dynamics, including some surprises that are not suggested by the static map. The main developments that are suggested by the dynamic affinity network are the following:

1) The emergence of a separate topic area focused on the interaction of clusters with radiation from advanced light sources such as synchrotrons, x-ray free electron lasers or femto and atto second lasers. It first appears in the 1998-2005 map as fifth largest topic area (DC5) and remains visible in the 2005-2012 map as a distinct topic area (DC6) whereas in the earlier map (1991-1998) this kind of work was inseparable from the large 'cluster science' topic area (DC1).

2) The growth of the surface science orientation within cluster science and its strengthening link with materials chemistry. In the 1998-2005 time window the strengthening of the affinity link between the two areas DC 2 and DC 3 is very visible; In the time window 2005-2012 surface science has become the dominating topic area among the eleven largest topic areas within the field (DC1), and the materials chemistry areas DC 5 and DC 7 have tight affinity links with it.

3) The emergence and strengthening of a link between the extreme ends of the cluster science topic area alignment.

² The size of the time window is a compromise between time resolution and structural cohesiveness of the network to be analyzed, i.e. allowing sufficient time for relevant connections to be made. See also Velden et al. 2013.

This is only hinted at by the growth of topic area AC6 depicted in accumulative affinity network. The dynamic topic clustering shows for the 2005-2012 time window the emergence of DC3, a large (3rd in size) hybrid topic area that combines (Inorganic) Cluster Chemistry and Cluster Physics and connects Cluster Physics with Materials Chemistry without the Surface Science intermediary.

This suggests that these networks capture changes in the prominence, alignment and interrelatedness of research topics in the field. Of particular interest to us, as we design studies of field differences in behavioral patterns among and within scientific communities, is how these changes in the cognitive structures in the field project onto the social network of inter-group collaborations and what we may learn about important actors in the field and how topic focus and collaborative interconnectivity interact.

Overlay Maps on Collaboration Network

For each of the three 8-year windows in our analysis, our overlay maps on the collaboration network show which groups in the overall group collaboration network of the field focus the majority of their work in one of the 11 largest topic areas of that respective time window. An overall feature of the network structure through all three time periods is its division into two parts and a predominance of inter-group collaborations between groups that share a topical focus. Based on the research focus of groups in the two distinguishable parts of the network, this division seems to reflect the disciplinary orientation of groups towards either chemistry or physics. Initially the network has two unconnected large network components that make this separation very obvious. In the later two time periods, an interconnected giant component has formed, however it still exposes a structural subdivision into two connected subgraphs. This division seems to be slowly diminished by growing collaborative connections between the two parts of the network. Groups that connect major parts of the network can be of particular interest for a qualitative follow-up field or interview study to explore reasons for this increased connectivity along with experiences of potential tensions and challenges when trying to bridge subdisciplinary cultures.

DISCUSSION

The picture that emerges is one in which, on the one hand, (sub-)disciplinary orientations of the groups that contribute to the shared knowledge base of a research specialty remain a rather stable feature. A group that is specialized in one sub-disciplinary area rarely changes its disciplinary orientation entirely. This is in agreement with our ethnographic field studies of research specialties in the physical and chemical sciences. We witnessed occasional strategic hires into a group to extend its local skill set, e.g. to add a synthetic chemistry capability into a largely experimental physics group. However, the foundation of the group remains the sub-disciplinary training of its leader that provides continuity to the research trajectory of the group. Further, for some topic areas, large parts of the inter-group

collaboration network are constituted by collaboration links within that topic area. Therefore a significant part of the collaborative work in the field seems to not transcend topic areas that generally correlate with sub-disciplinary orientations.

That said, there are also clear 'connectors', groups that work across two or more areas. Due to their activity we see an overall integration of the collaboration network. Also, there are a couple of topic areas that are hybrid in their (sub)disciplinary orientations. These topic areas and the research groups contributing to them may be of particular interest for the study of interdisciplinary collaboration and exchange in the field.

Hence, time-resolved topic affinity networks and group collaboration network overlay maps provide interesting insights into the socio-cognitive structure of the field and suggest directions for follow-up research. In addition, these overlay maps are valuable resources for planning future ethnographic field studies or interview studies, as they allow us to identify groups in critical positions of the network at specific points in time that may hence be of particular interest to include.

CONCLUSIONS

We have generated time resolved views into the co-evolution of topics and collaborative links between research groups in a research specialty. Time resolved maps suggest trends in the field and allow us to identify actors with critical positions for the evolution of the collaborative community network. The maps provide a promising resource that suggests individuals and groups that would be of particular interest to include in future qualitative studies that aim to develop a deeper understanding of the community structures within this field and their evolution over time. A next step in our research will be to validate and test the utility of these maps in interviews with researchers in the field.

ACKNOWLEDGMENTS

We acknowledge funding support from two grants:

- 1) OCI 1301874 Understanding Conditions for the Emergence of Virtual Orgs, and
- 2) SMA 1258891 EAGER: Collaborative Research: Scientific Collaboration in Time.

REFERENCES

Ding, Y. (2011). Community detection: Topological vs. topical. *Journal of Informetrics*, 5(4):498–514.

Yan, E., Ding, Y., Milojević, S., & Sugimoto, C. R. (2012). Topics in dynamic research communities: An exploratory study for the field of information retrieval. *Journal of Informetrics*, 6(1), 140-153.

Giatsoglou, M. and Vakali, A. (2013). Capturing social data evolution using graph clustering. *Internet Computing, IEEE*, 17(1):74–79, 2013.

Gläser, J. (2006). *Wissenschaftliche Produktionsgemeinschaften - die soziale Ordnung der Forschung*, volume 906 of *Campus Forschung*. Campus Verlag, Frankfurt / New York.

Holme, P., & Saramäki, J. (2012). Temporal networks. *Physics reports*, 519(3), 97-125.

Lin, Y. R., Chi, Y., Zhu, S., Sundaram, H., & Tseng, B. L. (2009). Analyzing communities and their evolutions in dynamic social networks. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 3(2), 8.

Morris, S. and Van der Veer Martens, B.. (2008) Mapping research specialties. *Annual review of information science and technology*, 42(1):213–295.

Newman, M. (2003) The structure and function of complex networks. *SIAM Review*, 45:167–256.

Rafols, I., Porter, A. L. and Leydesdorff, L. (2010), Science overlay maps: A new tool for research policy and library management. *J. Am. Soc. Inf. Sci.*, 61: 1871–1887. doi: 10.1002/asi.21368

Rosvall, M., & Bergstrom, C. T. (2010). Mapping change in large networks. *PLoS one*, 5(1), e8694.

Rosvall, M. and Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123.

Velden, T. (2013) Explaining field differences in openness and sharing in scientific communities. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 445–458. ACM.

Velden, T., Cambo, S., Ahmed, S., and Lagoze C. (2013). Toward a time-sensitive mesoscopic analysis of co-author networks: A case study of two research specialties. In *ISSI 2013*, 15-19 July, Vienna, Austria.

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.

Velden, T. and Lagoze, C. (2013). The extraction of community structures from publication networks to support ethnographic observations of field differences in scientific communication. *Journal of the American Society for Information Science and Technology*, 64(12):2405–2427.