CAIS Paper: Mapping Knowledge Domains to Better Forecast the Future: Challenges at the National Research Council Canada

Jean Archambeault, Catalin Bidian, M. Max Evans
School of Information Studies
McGill University, Montreal, Canada

Abstract: The cross-disciplinary future of knowledge domains mapping requires the fusion of multiple data sources, methodologies, and theories, shifting from descriptive to predictive models. This paper explores some of the National Research Council Canada challenges in using knowledge domain mapping to better forecast the future and advances a call for action.

1. Introduction

Accurately forecasting the trajectories of scientific domains and technologies is essential to the investment strategies of public and private organizations that operate in the area of science and technology. Mapping these trajectories represent a significant challenge because of the rapid pace at which new technologies are introduced, and their transformative impact on global, societal, cultural, and economic environments. While an important part of knowledge domain mapping research has focused on research evaluation and the social studies of science, emerging attention is devoted to the forecasting of science and technology evolution for future planning purposes (Boyack & Klavans, 2014).

The National Research Council Canada (NRC) is Canada's premier Research and Technology Organization. Since 2002, the National Science Library and Knowledge Management branch of the NRC has developed expertise in the visualization of knowledge domains to meet the NRC's competitive intelligence needs and support strategic investment decisions. In producing knowledge maps, the NRC Knowledge Management team faces a series of challenges in the production of plausible futures and technology forecasts. These challenges are particularly relevant for the Canadian information science community, as they relate to data manipulation limitations, the development and diversification of analytical methods, and the production of interpretative and actionable intelligence.

This paper provides a brief overview of the evolution of knowledge domain mapping research and practices, followed by practical challenges from NRC initiatives used to support future R&D investment decisions. In closing, the paper advances a call for action to the Canadian information science community to suggest how information science researchers and professionals could contribute to knowledge domain mapping cross-disciplinary research and practice.

2. Brief Overview of Mapping Knowledge Domains

The recognition of the social function and institutionalization of science (Bernal, 1939), the exponential growth of science publications (Price, 1965), and the advent of computerized citation indexes led to the development of a *science of science* in the late 1950s (Garfield, 2007). Rapidly thereafter, the desire to map the science landscape, graph the boundaries of knowledge domains, and understand the structural dynamics of the science phenomena became a focus area for research in information science. A series of landmark publications during the 1960s through 1990s explored the use of quantitative and basic visualization techniques on science and technical communication artifacts, to reveal the evolution of scientific knowledge domains. The concepts of bibliographic coupling (Kessler, 1963), direct citation network graphs or historiographs (Garfield, 1973), co-citation mapping (Griffith, Small, Stonehill. & Dey, 1974; Small & Griffith, 1974), and co-word analysis (Callon, Courtial, Turner & Bauin, 1983) fueled the emergence of science mapping. While knowledge maps during this period evolved around the institutional concept of scientific disciplines and research fronts (Small, 1999), the advent of the Internet and the science of networks in the late 1990s emphasized the analysis of scholarly social networks and the collaborative nature of science (Barabâsi et al., 2002). In the early 2000s, new capabilities offered by visualization interfaces and data processing techniques propelled the use of information visualization capabilities on large bibliographical datasets (Börner, Chen, & Boyack, 2003). Mapping knowledge domains was then labeled as an emerging interdisciplinary field (Shiffrin & Börner, 2004) and attracted contributions from the computer sciences, R&D management, science policy, and innovation studies. In parallel, a market for specialized commercial software and methodologies emerged to meet the needs of R&D organizations, corporate and legal firms engaged in intellectual property management as well as strategic and competitive intelligence activities (Porter & Cunningham, 2004).

Two interconnected streams can categorize current knowledge mapping initiatives. First, a mostly academic, research-oriented stream geared towards science policy formulation and the understanding of science and scholarly network dynamics. Second, a stream residing in a set of methods, applications, and services seeking to produce actionable insights for the benchmarking of university research programs, R&D strategic planning (Boyack & Klavans, 2014), competitive intelligence, and technology forecasting (Porter et al., 2004). The next section will illustrate this second stream through examples from recent initiatives and projects at the NRC. These examples describe two major challenges the NRC faces in using knowledge mapping practices, methods, and tools to anticipate technological trajectories.

3. NRC Challenges in Mapping Knowledge Domains

The first example highlights challenges in the development of future awareness capabilities used to strengthen investment decisions and the second highlights the modeling and mining of complex knowledge networks.

Challenges in mapping the future: The NRC faces an increased demand for forward-looking, future-oriented knowledge to better manage the uncertainties brought by R&D planning programs, using time horizons of five, fifteen, and even twenty years. Mapping scientific and technological trajectories based exclusively on lagging indicators and data points from scholarly and technical publications is insufficient. These data-driven methodologies have to be combined with future assessment approaches, such as foresight and qualitative methods more attuned to deal with the uncertainty of the future through scenarios, probability, and plausibility frameworks. The shift from descriptive to predictive assessment capabilities raises questions pertaining to how the NRC can best detect and map the emergence and trajectories of new technologies. Other unknowns include the emergence and modeling of the evolution and convergence of knowledge domains, and how data analytics, qualitative methodologies, and futuristic scenarios inform one another.

Challenges in integrating complexity: The NRC's understanding of the interplay of cross-disciplinary science, society, economic development, and global issues is evolving with the complexity of mapping the dynamics of knowledge networks (Börner et al., 2010). Consequently, the NRC faces a challenge in tapping into, fusing, and meshing a wider variety of data sources, beyond the traditional scholarly communication channels. Data analysis and sourcing requirements at the NRC are being redefined by the advent of open science and data policies, as well as the increased participation of multiple social and non-state actors in social media and knowledge production models (European Commission, 2009). Going forward, the NRC is actively interested in discovering how social media conversations may provide more timely access to early insights on the evolution of knowledge domains, versus traditional scholarly communication channels. Important questions pertain to the media of meaningful conversations on science and technology developments and how to integrate them, along with other data sources, into the analysis.

4. Call for Action

The future of mapping knowledge domains expands across disciplines, requiring the fusion of multiple data sources, methodologies, and theories, shifting from descriptive to predictive simulations (Edmonds, Gilbert, Ahrweiler, & Scharnhorst, 2011). Therefore, resolving some of the current challenges of mapping knowledge domains requires the engagement of cross-disciplinary research and teams. Some of the relevant aspects that should be considered by information science researchers and professionals include:

1. Data manipulation limitations – the identification and integration of new and diversified sets of data sources, as well as the transition from applying analytics on static data (databases) to dynamic data (streaming text and data sources);
2. Development and diversification of analytical methods – the development of composite indicators that would include social, political, and economic parameters for the detection and monitoring of emerging technologies, and uncovering new applications for existing ones;
3. Production of interpretative and/or actionable intelligence – the enhancement of capabilities to increase the actionable knowledge value and multidimensional perspectives through the integration of quantitative and qualitative methods.

As our understanding of the interconnectedness of science, society, economic development, global issues and the participation of individuals within the innovation ecosystem increases, so does the complexity and value of mapping the dynamics of networks involved (Börner et al., 2010). Having access to future-oriented knowledge mapping capabilities could provide the Canadian information science community with a better understanding of the trends and forces shaping R&D activities, and offer them a more robust future-aware perspective in these times of high uncertainty.

References

Barabâsi, A.-L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications*, *311*(3), 590-614.

Bernal, J. D. (1939). *The social function of science*. New York: The MacMillan Company.

Börner, K., Chen, C., & Boyack, K. W. (2003). Visualizing knowledge domains. In B. Cronin (Ed.), *ARIST, 37*(5), Medford, NJ: Information Today.

Börner, K., Contractor, N., Falk-Krzesinski, H. J., Fiore, S. M., Hall, K .L., Keyton, J., Spring, B., Stokols, D., Trochim, W., & Uzzi, B. (2010). A multi-level systems perspective for the science of team science. *Science Translation Medicine, 2*(49), 49-54. Retrieved from: http://ivl.slis.indiana.edu/km/pub/2010-borner-et-al-multi-level-teamsci.pdf

Boyack, K. W., & Klavans, R. (2014). Creation of a highly detailed, dynamic, global model and map of science. *Journal of the Association for Information Sciences and Technology, 65*(4), 670-685.

Callon, M., Courtial, J.-P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information, 22*, 191-235.

Edmonds, B., Gilbert, N., Ahrweiler,P., & Scharnhorst, A. (2011). Simulating the Social Processes of Science. *Journal of Artificial Societies and Social Simulation*. *14*(4).14

European Commission (2009). *Global Governance of Science – Report of the Expert Group on Global Governance of Science to the Science, Economy and Society Directorate, Directorate-General for Research, European Commission.* Luxembourg: Office for Official Publications of the European Communities. doi: 10.2777/31776

Garfield, E. (1955). Citation indexes for science: A new dimension in documentation through association of ideas. *Science, 122*(3159), 108-111.

Garfield, E. (1973). Historiographs, librarianship, and the history of science. In C. H. Rawski (Ed.), *Toward a theory of librarianship: Papers in honor of Jesse Hauk Shera* (pp. 380-402). Metuchen, N. J.: Scarecrow Press. Retrieved from http://garfield.library.upenn.edu/essays/v2p136y1974-76.pdf

Garfield, E. (2007). From the science of science to scientometrics: Visualizing the history of science with Histcite software. In D. Torres-Salinas & H. F. Moed (Eds.), *Proceedings of the 11th International Conference of the International Society for Scientometrics and Informetrics (ISSI 2007)* (pp. 21-26). Madrid: CINDOC, CSIC.

Griffith, B. C., Small, H. G., Stonehill, J. A., & Dey, S. (1974). The structure of scientific literatures II: Toward a macro- and microstructure for science. *Science Studies*, *4*(4), 339-365.

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation, 24*,123-131.

Porter, A. L., & Cunningham, S. W. (2004). *Tech mining: Exploiting new technologies for competitive advantage*. Hoboken, NJ: Wiley.

Price, D. J. S. (1965). *Little science, big science*. New York: Columbia University Press.

Shiffrin, R. M., & Börner, K. (2004). Mapping knowledge domains. In R. M. Shiffrin & K. Börner (Eds.), *Proceedings of the National Academy of Sciences of the United States of America*, 101 (Suppl. 1) (pp. 5183-5185). Retrieved from: http://www.pnas.org/content/101/suppl_1.toc.

Small, H. G., & Griffith, B. C. (1974). The structure of scientific literatures I: Identifying and graphing specialties. *Science Studies*, *4*(1), 17-40.

Small, H. G. (1999). Visualizing science by citation mapping. *Journal of the American Society for Information Science*, *50*(9), 799-813.