Research Specialties as Emergent Phenomena: Connecting Emergence Theory and Scientometrics

T. Clay Templeton
University of Texas at Austin
clayt@ischool.utexas.edu

Kenneth R. Fleischmann
University of Texas at Austin
kfleisch@utexas.edu

Abstract

This short paper represents an initial effort to connect the emergence theory literature with the bibliometric, informetric, and scientometic literatures. It begins with a prominent definition of emergence, and then connects each of the components of this definition with the relevant insights about the development of new scientific and technical concepts or research specialties. Finally, it concludes with a discussion of the relationship between these two distinct areas of scholarly inquiry and the need for further exploration of this intersection.

Keywords: emergence theory, emergent phenomena, bibliometrics, informetrics, scientometrics, citation analysis, cyberinfrastructure

Introduction

Emergence refers to a situation in which a qualitatively novel entity, or an emergent, is generated through the interactions of lower-level entities (Sawyer, 2005). Emergence theory thus seeks to describe a wide range of physical, biological, and social phenomena. Since science can be viewed as a social phenomenon (Fleck, 1979), the development of new scientific and technical concepts and research specialties can be seen as a process of social emergence (e.g. Guo et al. 2011, Leitz 2009, Chen et al. 2009). As emergence theory has organized our thinking about the social lives of scientific specialties and concepts, the literature these specialties and concepts left behind beckoned as a realm of data in which to explore its workings. In this paper, we recount this interaction by mapping ideas from a prominent definition of emergence (Goldstein 1999) to ideas in the bibliometric, informetric, and scientometric spaces.

What Is an Emergent Phenomenon?

Goldstein's (1999) work can be seen as part of the third wave of social systems theory, which situates social emergence within a broader framework of interest in systems whose evolution is sensitive to environmental conditions (Sawyer, 2005). Goldstein's focus is on organizational dynamics, and thus, his characterization of emergence is highly compatible with an exploration of science as a social phenomenon. According to Goldstein's definition, emergent phenomena are characterized by five characteristics:

Acknowledgements: This work was supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center contract number D11PC20153. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the U.S. Government.

Templeton, T. C., & Fleischmann, K. R. (2013). Research specialties as emergent phenomena: Connecting emergence theory and scientometrics. *iConference 2013 Proceedings* (pp. 827-830). doi:10.9776/13403

- “Ostensive: Emergents are recognized by showing themselves, i.e., they are ostensively recognized.” (p. 50)
- “Global or macro level: The locus of emergent phenomena occurs at a global or macro level. Observation of emergents is of their behavior on this macro level.” (p. 50)
- “Coherence or correlation: Emergents appear as integrated wholes that tend to maintain some sense of identity over time.” (p. 50)
- “Dynamical: Emergent phenomena are not pre-given wholes but arise as a complex system evolves over time.” (p. 50)
- “Radical novelty: Emergents have features that are not previously observed in the complex system under observation” (p. 50)

Ostensive

Emergents are recognized when they show themselves, which raises the question of to whom and under what conditions they become perceptible. This question is especially pressing for social emergence, since the ontological status of collective entities in the social sciences is controversial (Sawyer 2005). In the case of scientific concepts and specialties, how do we most definitively figure out which ones exist, what their proper names are, and where they begin and end? Arguably, scientific paradigms, fields, subfields, and specialties show themselves most nakedly to scientists working in or near those specialties. Practicing scientists are intimately acquainted with the details of professional scientific practice, and their professional practice depends on an ability to navigate the structure of science. They navigate, however, with the aid of bibliographic tools. Since the advent of citation databases (Garfield 1955), information professionals have augmented the perception of scientists and others with maps of science built from the ground up (Garfield et al. 1964, ISI 1981, Börner et al. 2003). The aim is to allow emergent scientific structures to show themselves earlier, in more detail, to more people.

Global/Macro Level

The locus of emergent phenomena occurs at a global or macro level, even when objective observation may be more easily achieved at a local or micro level. Scientific specialties and concepts emerge from a mangle of journals, authors, papers, keywords, and citations, among other things outside of bibliographic control (Pickering 1995). These basic units of bibliometric analysis are amenable to quantification (Borgman and Furner 2002), which makes them an appealing starting point for generating analysis of higher-level structures. Procedures for generating macro level structures from bibliometric data can be found in review literatures on mapping scientific specialties (Morris and van der Veer Martens 2008) and visualizing knowledge domains (Börner et al. 2003). The boundaries of scientific specialties (as constructed from bibliometric networks) remain open to interpretation, however. At present, information specialists frequently turn to Subject Matter Specialists (SMEs) to validate the results of mapping procedures (Morris and van der Veer Martens 2008).

Coherence or Correlation

Emergents appear as integrated wholes that tend to maintain some sense of identity over time; The entities involved in a research specialty tend to be more densely associated with each other than to entities outside the specialty. This basic assumption informs most approaches to extrapolating research specialties from bibliometric facts. For example, approaches based on co-citation analysis (Garfield 1979) first calculate similarity of papers based on how frequently each pair of papers appears together in reference lists, then cluster papers based on similarity. Approaches based on co-word analysis (Callon, Law, and Rip 1986) calculate similarity of words based on how frequently two words appear together in some part of a paper’s content or metadata, then cluster words based on similarity. Morris and van der Veer Martens review a number of approaches for constructing maps based on how bibliomentric entities cohere in the written record of science (2008).

Dynamical

Emergent phenomena are not pre-given wholes but arise as a complex system evolves over time. The density of associations between entities in an emerging specialty tends to increase over time (Leskovec et al. 2005), and some authors have taken the emergence of a giant connected component in the coauthorship network as a signature feature of emergence (Bettencourt et al. 2009, Lietz 2009). In approaches to mapping scientific specialties based on clustering, a major challenge is tracking clusters from one time period to another. Asur et al. (2007) and Spilopoulou et al. (2006) provide equations for classifying changes in cluster structure over time that are agnostic with respect to clustering method. Research in literature dynamics (Tabah 1999) attempts to describe the processes by which specialties emerge using a variety of mathematical modeling techniques.

Radical Novelty

Emergents have features that are not previously observed in the complex system under observation. It may not be immediately apparent that a new specialty has emerged, and it is challenging to predict what the characteristics of a new specialty might be. Bibliometrics researchers have raised the question of how early a new specialty can be detected (Meadows and O'Conner 1971, Small 2006). Based on information foraging theory (Pirolli 2007), Chen et al. (2009) suggest that an emerging research area might be signaled by citation bursts (Kleinberg 2002) to a paper that bridges existing areas of knowledge (Burt 2004). Ohniwa et al. (2009) describe a method of identifying emerging topics based on keyword increments from one time period to the next. Guo et al. (2011) incorporate word bursts, new author counts, and interdisciplinarity of citations (Porter and Rafols 2009) into their model of emergence.

Connecting Emergence Theory and Scientometrics

As we have tried to illustrate, the development of a new scientific or technical concept or a new research specialty as explored in the bibliometric, informetric, and scientometric research literatures can be viewed as a compelling case study of emergence. In turn, emergence theory helps to put these literatures into a broader context that potentially transcends science as an area of focus, by exploring the connections between science and other human, biological, and natural domains. It will be useful to further explore the connections between these two areas of research, in the interest of bridging this divide and hopefully furthering the research programs of both of these areas.

References

Asur, S., Parthasarathy, S., & Ucar, J. (2007). An event-based framework for characterizing the evolutionary behavior of interaction graphs. *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining* (KDD '07) (pp. 913-922). ACM, New York, NY, USA.

Bettencourt, L.M.A., Kaiser, D.I., & Kaur, J. (2009). Scientific discovery and topological transitions in collaboration networks. *Journal of Informetrics, 3*(3), 210-221.

Borgman, C.L., & Furner, J. (2002). Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology*, 36, 3-72.

Börner, K., Chen, C., & Boyack, K.W. (2003). Visualizing knowledge domains. *Annual Review of Information Science and Technology*, 37, 179-255.

Burt, R.S. (2004). Structural holes and good ideas. *The American Journal of Sociology, 110*(2), 349-399.

Callon, M., Law, J., & Rip, A. (Eds.) (1986). *Mapping the dynamics of science and technology: sociology of science in the real world*. London: Macmillan.

Chen, C., Chen, Y., Horowitz, M., Hou, H., Liu, Z., & Pellegrion, D. (2009). Towards an explanatory and computational theory of scientific discovery, *Journal of Informatics 3*, 191-201.

Corning, P.A. (2002). The re-emergence of "emergence": A venerable concept in search of a theory. *Complexity 7*(6), 18–30.

Fleck, L. (1979). *Genesis and development of a scientific fact*. Chicago: University of Chicago Press.

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

Garfield, E., Sher, I.H., & Torpie, R.J. (1964). *The use of citation data in writing the history of science*. Philadelphia: Institute for Scientific Information.

Garfield, E. (1979). Citation indexing: *Its theory and application in science, technology, and humanities*. New York: Wiley

Goldstein, J. (1999). Emergence as a construct: History and issues, *Emergence: Complexity and Organization 1*(1), 49–72.

Guo, S., Weingart, S., & Börner, K. (2011). Mixed-indicators model for identifying emerging research areas. *Scientometrics 89*, 421-435.

Institute for Scientific Information. (1981). *ZSI atlas of science: Biochemistry and molecular biology, 1978180*. Philadelphia: Institute for Scientific Information.

Kleinberg, J. (2002). Bursty and Hierarchical Structure in Streams. In *Proceedings of the eighth ACM SIGKDD international conference on knowledge discovery and data mining*.

Leskovec, J., Kleinberg, J. & Faloutsos, C. (2005). Graphs over time: densification laws, shrinking diameters and possible explanations. In *Proceedings of the eleventh ACM international conference on knowledge discovery and data mining* (KDD) (pp. 177-187). ACM, New York, NY, USA.

Lietz, H. (2009). Diagnosing emerging science: The cases of the "New Science of Networks" and Scientometrics. *PRIME-ENID Summer School on Science, Technology and Innovation Indicators and Knowledge Dynamics Visualization*, Amsterdam.

Meadows, A.J. and O'Connor, J.G. (1971). Bibliographical Statistics as a Guide to Growth Points in Science. *Social Studies of Science 1*, 95 – 99.

Morris, S.A. and Van der Veer Martens, B. (2008). Mapping research specialties. *Annual Review of Information Science and Technology 42*, 213-295.

Ohniwa, R., Hibino, A. and Takeyasu, K. (2010). Trends in research foci in life science fields over the last 30 years monitored by emerging topics. *Scientometrics 85*, 111-127.

Pickering, A. (1995). *The mangle of practice: Science, society, and becoming*. Durham, North Carolina: Duke University Press Books.

Porter, A.L. & I. Rafols. (2009). "Is science becoming more interdisciplinary? Measuring and mapping six research fields over time." *Scientometrics 81*, 719-745.

Pirolli, P. (2007). *Information Foraging Theory: Adaptive Interaction with Information*. Oxford: Oxford University Press.

Sawyer, R.K. (2005). *Social emergence: Societies as Complex Systems*. Cambridge, UK: Cambridge University Press.

Small, Henry. (2006). Tracking and predicting growth areas in science. *Scientometrics 68*, 595-610.

Spiliopoulou, M., Ntoutsi, I., Theodoridis, Y. & Schult, R. (2006). MONIC: modeling and monitoring cluster transitions. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining* (KDD '06) (pp. 706-711). ACM, New York, NY, USA.

Tabah, A.N. (1999). Literature dynamics: Studies on growth, diffusion, and epidemics. *Annual Review of Information Science and Technology 34*, 249-286.