Monitoring Knowledge Flow through Scholarly Networks

Erjia Yan
School of Library and Information Science, Indiana University, Bloomington, Indiana
eyan@indiana.edu

Ying Ding
School of Library and Information Science, Indiana University, Bloomington, Indiana
dingying@indiana.edu

Xiangnan Kong
Department of Computer Science, University of Illinois at Chicago, Chicago, Illinois
xkong4@uic.edu

ABSTRACT
The production and creation of knowledge is not dependent on any individual and isolated entity; instead, knowledge is diffused, exchanged, and circulated among various entities. To study the knowledge flow and transfer within and across different disciplines can help us better understand science and scientific collaboration. This paper presents a methodological framework to study knowledge flow, including a knowledge hierarchy, the construction of knowledge flow network, and indicators that can be used to measure disciplinarities.

Keywords
knowledge flow, discipline, interdiscipline, scholarly networks

INTRODUCTION
The production and creation of knowledge is not dependent on any individual and isolated entity; instead, knowledge is diffused, exchanged, and circulated among various entities. Knowledge flow, in the past twenty years, has become more inter-sectoral, more inter-organizational, more inter-disciplinary, and more international (Lewison, Rippon, & Wooding, 2005; Wagner & Leydesdorff, 2005; Autant-Bernard, Mairesse, & Massard, 2007; Ponds, Van Oort, & Frenken, 2007; Buter, Noyons, & Van Raan, 2010).

Similar to many important concepts in economics and bibliometrics, the transfer of knowledge is an unobservable phenomenon (Jaffe, Trajtenberg, & Fogarty, 2000). As an alternative, researchers rely on proxies to measure the concepts of interest. The quantitative studies of knowledge flow usually use citations as the research instrument. Citations between scientific articles imply a knowledge flow from the cited entity to the citing entity (Jaffe, Trajtenberg, & Henderson, 1993; Van Leeuwen & Tijssen, 2000; Nomaler & Verspagen, 2008). Using the trading metaphor (Stigler, 1994; Lockett & McWilliams, 2005; Cronin & Meho, 2008), knowledge flow has been explored as the intellectual trading among different disciplines.

The quantitative studies of interdisciplinarity were made available by researching on citation networks aggregated at the field level. Researchers usually choose a subset of representative journals or the full sets of journals from a field based upon the ISI's classification of journals, and then measure the extent to which the chosen field cites the publication of other subject categories. Network-based indicators have also been proposed to measure how interdisciplinary different research fields are. Examples include entropy (Zhang et al., 2010), integration and specialization (Porter, Roessner, Cohen, & Perreault, 2006; Porter, Roessner, & Heberger, 2008), diversity and coherence (Rafols & Meyer, 2010), percentage of multi-assignation (Morillo, Bordons, & Gomez, 2003), and relative openness (Rinia et al., 2002).

Previous endeavors on inter-sectoral, inter-organizational, and interdisciplinary knowledge flows laid sound theoretical and methodological foundations to the inquiry of knowledge flow studies. Nonetheless, these studies only involved a few disciplines as the research target, and consequently were not able to provide a holistic view of the developments and interactions of scientific disciplines. To our best knowledge, currently there is no study on finding the knowledge flows covering all science and social science disciplines. Our study is thus motivated to conduct a more comprehensive examination of the scientific trading, and obtain a bird's-eye view for the developments and interactions of various scientific disciplines.

PROPOSED METHODOLOGICAL FRAMEWORK

Knowledge hierarchy
How to effectively organize scientific knowledge is one of the key concerns of academic databases. Some databases use subject headings to classify papers (such as ACM Digital Library), while some others use subject areas to cluster journals (such as Scopus). The academic database Scopus has a well-defined journal classification schema called All Science Classification Codes (ASJC). The schema is composed of minor subject areas, major subject areas, and top-level divisions. A journal is usually assigned into one or several minor subject areas. In total, there are around 300 minor subject areas. These subject areas are grouped into 27 major subject areas, and these major subject areas are further grouped into 4 top-level
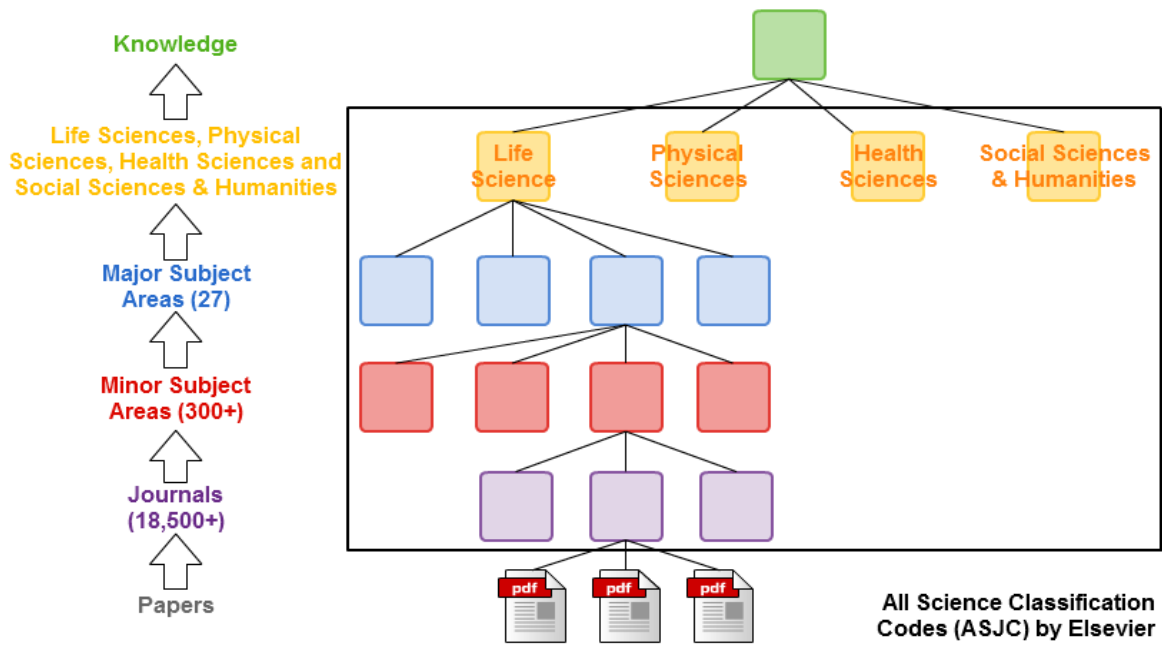
Figure 1. A six-layer knowledge hierarchy

divisions: Life Sciences, Physical Science, Health Sciences, and Social Sciences & Humanities. We refer to this schema as knowledge hierarchy and visualize it in Figure 1. In the proposed study, we will be focusing on the analysis of the middle four knowledge hierarchies: journals, minor subject areas, major subject areas, and top-level divisions. The limitation of this classification scheme is that such scheme may not be purely based on research specialties and topics; factors as practicality and managerial decisions may also contribute to this outcome. Comparisons with other field level clustering results, for instance, Map of Science (Boyack, Börner, & Klavans, 2009), science overlay maps (Rafols, Porter, & Leydesdorff, 2010), are highly recommended.

The construction of knowledge flow network
The raw data input is a journal-to-journal citation matrix (network), with the cell values denoting the number of citations from the citing journal to the cited journal. As each journal is associated with one (or several) minor and major subject categories, two field-to-field citation matrices (Figure 2a) can be aggregated based on the journal-to-journal citation matrix: one is on the minor subject areas and the other is on the major subject areas.

Figure 2. An example of a knowledge flow network

Measurement
Knowledge flow is operationalized by citation flows among various research entities. In order to characterize the knowledge flow between disciplines, we use Dijkstra algorithm (Dijkstra, 1959) to search for shortest path between two fields in the knowledge flow network. The idea of the Dijkstra algorithm is to find a path between two nodes so that the sum of edge weight reaches the minimum in a weighted directed network.

Intuitively, the distance and traffic follow a reverse relationship. Therefore, we propose a measurement to define the distance between two fields in the knowledge flow network (Figure 2b).

$$distance_{i \rightarrow j} = \frac{1}{number\ of\ citations\ from\ j\ to\ i}$$

In Figure 3, two examples (knowledge flow from E to A and knowledge flow from A to E) are given based on the sample knowledge flow network illustrated in Figure 2a. The length of a knowledge path is operationalized as the number of fields involved in a knowledge transfer. Therefore, in Figure 3, the knowledge path from Field E to Field A is two, and the knowledge path from Field A to Field E is four.

	Indicators	Meanings
Life Sciences, Physical Sciences, Health Sciences and Social Sciences & Humanities	Average shortest path length (source) Average shortest path length (Destination) Average shortest path weight (source) Average shortest path weight (destination) Occurrence in knowledge path	How easily its knowledge can be accessed How easily it can access other's knowledge How distant its knowledge is to others How distant its knowledge is from others Patterns of knowledge flows
Major Subject Areas	Average shortest path length (source) Average shortest path length (Destination) Average shortest path weight (source) Average shortest path weight (destination)	How easily its knowledge can be accessed How easily it can access other's knowledge How distant its knowledge is to others How distant its knowledge is from others
Minor Subject Areas	Average shortest path length Average shortest path weight Occurrence in shortest path Visulization of knowledge flow map	How easily its knowledge can be accessed How distant its knowledge is to others How important it is to others knowledge transfer Its position in knowledge flow map
Journals	Number of cited subject areas Number of citing subject areas Incoming citations from varied subject areas Outgoing citations to varied subject areas Visulization of knowledge flow map	How broadly its knowledge are used by others How broadly it receive knowledge from others Interdiscipline indicator as knowledge distributor Interdiscipline indicator as knowledge receiver Its position in knowledge flow map

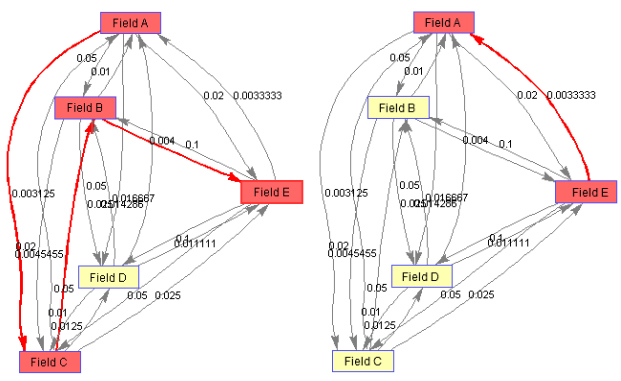**Figure 4. Proposed indicators to measure knowledge flows**

Figure 3. Examples of finding shortest knowledge paths between different fields

Evaluative indicators

For an effective evaluation, we proposed several indicators, including average shortest path length, average shortest path weight, occurrence in shortest path, etc. For each subject area, the average shortest path length measures the average shortest path between the chosen subject area and all other subject areas. This measurement has directions: if it starts from the cited subject area, the average shortest path length denotes how easily its knowledge can be accessed by others (i.e. science/social science, classes, and subject categories); if it starts from the citing subject area, the average shortest path length denotes how easily it can access other's knowledge. The average shortest path weight is the accumulative value of the average shortest paths, and

thus it is a measure of how distant a subject area is from other subject areas. The occurrence in shortest path denotes how important a subject area is to other subject categories' knowledge transfer. It is an indicator related to betweenness centrality. A subject area with higher occurrence in shortest path may have higher betweenness and thus plays a role of interconnecting various knowledge sources. Above mentioned indicators can be applied to minor subject areas, major subject areas, and the top-level knowledge divisions. At journal level, we use indicators number of citing (or cited) subject areas to measure the extent a journal cites (or been cited by) other journals. The indicators and their meanings for the hierarchies are presented in Figure 4.

SIGNIFICANCE AND CONTRIBUTION

Previous work on discipline interactions mainly focused on mapping science, without considering directions of knowledge flow (e.g., Boyack, Börner, & Klavans, 2005), or knowledge path (e.g., Kiss et al., 2010). In the proposed study, we intend to find important knowledge paths among all scientific disciplines, and provide empirical results on the significant patterns of knowledge transfer and dissemination. The proposed knowledge hierarchy is effective in organizing scientific knowledge. This hierarchy can be easily adopted in other related studies by other scholars. The proposed indicators quantify patterns on knowledge flow and dissemination, providing additional insights into interdisciplinary studies. These indicators are also valuable for scientific evaluation and science policy making.

REFERENCES

Autant-Bernard, C., Mairesse, J., & Massard, N. (2007). Spatial knowledge diffusion through collaborative networks. *Papers in Regional Science, 86*(3), 341-350.

Boyack, K. W., Börner, K., & Klavans, R. (2009). Mapping the structure and evolution of chemistry research. *Scientometrics*, 79(1), 45-60.

Buter, R. K., Noyons, E. C. M., & Van Raan, A. F. J. (2010). Identification of converging research areas using publication and citation data. *Research Evaluation, 19*(1), 19-27.

Cronin, B., & Meho, L. I. (2008). The shifting balance of intellectual trade in information studies. *Journal of the American Society for Information Science & Technology, 59*(4), 551-564.

Dijkstra, E.W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik, 1*, 269-271.

Jaffe, A. B., Trajtenberg, M., & Fogarty, M. S. (2000). Knowledge spillovers and patent citations: Evidence from a survey of inventors. *American Economic Review, 90*(2), 215-218.

Jaffe, A. B., Trajtenberg, M., & Henderson, A. D. (1993). Geographical localization of knowledge spillovers by patent citations. *Quarterly Journal of Economics, 108*(3), 577-599.

Kiss, I. Z., Broom, M., Craze, P. G., & Rafols, I. (2010). Can epidemic model describe the diffusion of topics across disciplines? *Journal of Informetrics, 4*(1), 74-82.

Lewison, G., Rippon, I., & Wooding, S. (2005). Tracking knowledge diffusion through citations. *Research Evaluation, 14*(1), 5-14.

Lockett, A., & McWilliams, A. (2005). The balance of trade between disciplines: do we effectively manage knowledge? *Journal of Management Inquiry, 14*(2), 139-150.

Ponds, R., Van Oort, F., & Frenken, K. (2007). The geographical and institutional proximity of research collaboration. *Papers in Regional Science, 86*(3), 423-443.

Porter, A. L., Roessner, J. D., Cohen, A. S., & Perreault, M. (2006). Interdisciplinary research: meaning, metrics and nurture. *Research Evaluation, 15*(3), 187-195.

Rafols, I., Porter, A. L., & Leydesdorff, L. (2010). Science overlay maps: A new tool for research policy and library management. *Journal of the American Society for Information Science and Technology, 61*(9), 1871-1887.

Rafols, I., & Meyer, M. (2010). Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience. *Scientometrics, 82*(2), 263-287.

Rinia, E. J., Van Leeuwen, T. N., Bruins, E. E. W., Van Vuren, H. G., & Van Raan, A. F. J. (2002). Measuring knowledge transfer between fields of science. *Scientometrics, 54*(3), 347-362.

Stigler, S. M. (1994). Citation patterns in the journals of statistics and probability. *Statistical Science, 9*(1), 94-108.

Van Leeuwen, T., & Tijssen, R. (2000). Interdisciplinary dynamics of modern science: analysis of cross-disciplinary citation flows. *Research Evaluation, 9*(3), 183-187.

Wagner, C. S., & Leydesdorff, L. (2009). Network structure, self-organization and the growth of international collaboration in science. *Research Policy, 34*(10), 1608-1618.

Zhang, L., Liu, X., Janssens, F., Liang, L., & Glänzel, W. (2010). Subject clustering analysis based on ISI category classification. *Journal of Informetrics, 4*(2), 185-193.