Maps of Science as Interdisciplinary Discourse: Co-citing Contexts and the Role of Analogy

Henry Small

henry.small@thomsonreuters.com
Thomson Reuters, 3501 Market St, Philadelphia, Pa., 19104 (USA)

Abstract
Interdisciplinary links are studied using a clustering and mapping of documents juxtaposed to a journal-based categorization of document clusters. Strong links between clusters are characterized as interdisciplinary based on the dissonance of their category assignments. To verify the finding and probe more deeply, co-citation context analysis is applied to a selected link on the map and the factors underlying the interdisciplinary connection are revealed, including authors' perceptions of conceptual analogy and scientific promise.

Introduction
Interdisciplinarity in science can be defined as a dissonance between two or more superimposed organizational structures applied to the same scientific objects, whether those objects are individual scientists, papers, tools or materials. As an obvious example, if scientists from different fields come together to work on a collaborative project, two principles of organization come into play: the convention used to assign scientists to fields and the social network or scientific problem that brings them together. In a less obvious sense, the reference list of a scientific paper can be interdisciplinary if it co-cites literature in different disciplines. Here the author provides the initial organization of the references and a field classification scheme applied to the references provides the contrasting structure.

Clearly there are multiple means of organization that may apply, whether human or machine derived, such as academic departments, government funding programs, journal classifications, clustering algorithms, etc. and these can be superimposed on one another and suggest more or less dissonance or heterogeneity. Metrics can be defined to measure this degree of heterogeneity (Adams, 2007; Porter, 2007).

However, problems can arise due to the inaccuracies in the organization or classification schemes themselves. Many discipline definitions may be bound by tradition, convention, or even tribal ritual (Levi-Strauss, 1966), such as academic departments or professional societies. Likewise, human-based classification schemes for journals or papers can be out of date or misleading, and algorithmic methods based on social network detection, clustering, or mapping may be approximate. Such artifacts can interact and undermine the assertion of interdisciplinarity. For example, it is possible that a supposed area of science, deemed to be interdisciplinary, is merely the result of a shift in disciplinary boundaries that the classification scheme or search profile is unable to capture.

Thus it is important to undertake studies to confirm the results of such measurements. One approach is to analyze the content of the discourse and communication that underlies the connections. This paper argues that by conjoining quantitative and qualitative approaches, we can come to a deeper understanding of the nature of interdisciplinarity as well as confirm the quantitative measures.

The present study examines these issues by juxtaposing two modes of classification, document clustering and journal classification. The case study explores the nature of an interdisciplinary relationship using co-citation context analysis (Small, 1980), the first time this method has been applied to a high level map of science, and suggests that such links arise in part from scientists' perceptions of similarity and analogy of problems in different areas.

Proceedings of ISSI 2009, pages 270–274.
Edited by B. Larsen and J. Leta.

Method

Using data from the first bimonthly period of 2008 (ESI, 2008) a map of science was generated using document co-citation and an iterative clustering procedure which is repeated to the point where document clusters have been aggregated to form a large network (Small, 1985). Four iterations of clustering were required to reach convergence. The network of macro-clusters was mapped in two dimensions using a force-directed placement algorithm (Fruchterman & Reingold, 1991). Only the strongest inter-cluster links for each node (using the cosine measure) were selected denoted as solid lines. Weaker links denoted with dotted lines were added to form a minimal spanning tree. We presume that some links on this level 4 map are interdisciplinary since disciplinary groups have to some extent already been aggregated.

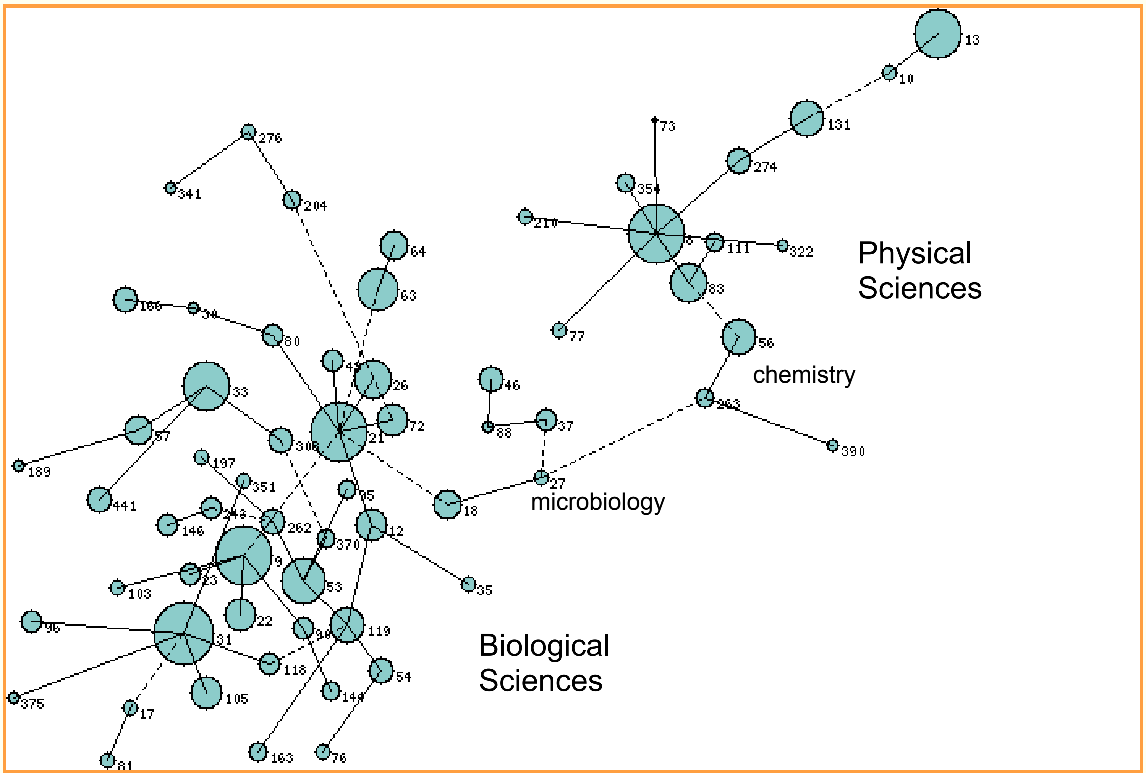

Figure 1: Level 4 map of science

The level 4 document clusters were assigned disciplines based on the Thomson Reuters ESI journal classification scheme (http://in-cites.com/journal-list/index.html). Boyack, Klavans and Börner (2005) have also used journal category schemes to validate maps based on journal linkages. Our approach differs from theirs in that the map is built up from document co-citations rather than journals. We selected the two most frequently occurring categories for each cluster by counting the number of documents in each category. If the top category for a given cluster does not match either of the top two categories for the cluster to which it is linked, then the link is deemed interdisciplinary. The same procedure was applied to maps at two lower levels of aggregation which had been formed in the process of creating the level 4 map, namely levels 2 and 3, and the percentage of interdisciplinary links were similarly computed for those levels. Level 2 corresponds to linkages between the initial clusters of highly cited documents and level 4 corresponds to the global map.
By this method about 53% of the strong links at level 4 are interdisciplinary while at the two lower levels of aggregation the percentage falls sharply to about 5% at level 2 (see Table 1).

This confirms the expected finding that interdisciplinarity increases with the level of aggregation.

Table 1. Interdisciplinarity as a function of clustering level

Level of Aggregation	% of interdisciplinary links
Level 2	4.7%
Level 3	27.0%
Level 4 (global map)	53.4%

Interpretation

While this finding reassures us that clustering is capturing disciplinary divisions, it does not shed light on why the interdisciplinary links have formed and how they can be interpreted. To explore this more deeply we selected a link on the level 4 map and undertook an analysis of the co-citing texts, so-called co-citation context analysis. In the past this method has been applied to document level co-citation maps (Small, 1980; Schneider, 2006) but is here applied for the first time to links between higher level document clusters presumably having more diverse content.

The link selected is between level 3 clusters 27 and 263, indicated as microbiology and chemistry on the map. This link appears to occupy a strategic position connecting physical science clusters on the upper right with biological clusters on the lower left. Because this is denoted as a dotted line, it is a weaker link than others and fits Granovetter's description of a weak tie (Granovetter, 1973). It should be emphasized that this is not the only link between physical and biological science but only a prominent one that was selected in the process of forming the minimal spanning tree representation of Figure 1.

Table 2 shows various attributes of the linked clusters, including the number of highly cited (core) papers, the main journal categories assigned, and the two most frequently occurring journals for each.

Table 2: Attributes of linked clusters on level 4 map

Level3 front	Core papers	Mean year	Mean cites	Top Journal categories	Top 2 Journals
27	37	2005.3	47.4	Microbiology	Science, Journal of Natural Products
263	88	2004.7	69.3	Chemistry	J Am Chem Soc, Chem Reviews

Cluster 27 is classified predominantly in microbiology by virtue of the journals of its highly cited papers and cluster 263 is classified as chemistry. Examining the titles of the co-cited documents involved in the link, we find on the microbiology side a number of papers dealing with drug discovery using microbial genomics and what is called "assembly line enzymology." On the chemistry side, we find a focus on enzyme chemistry involving organometallic complexes, but no mention of genetics or bacteria, suggesting a more purely chemical approach.

A total of 29 co-citing papers created this link, and texts for 26 of these were obtained. One of the co-citing papers joining these clusters has the interdisciplinary sounding title: "Biological inorganic chemistry at the beginning of the 21st century" (Gray, 2003). Analysis of the 26 co-citing texts shows that there is a slight overweighting of cites on the microbiology side with 1.4 references for every reference on the chemistry side, despite the fact that the chemistry cluster has more core papers. In 69% of co-citing texts, reference to the microbiology cluster precedes reference to the chemistry cluster. These findings suggest that authors are initially setting the stage in their texts with references to biological issues and then introducing chemistry at a later stage to perhaps suggest another approach to the problem.

The following passage gives the general flavor of the co-citation contexts:

"It is possible to estimate the biosynthetic potential from a given organism by mining the whole genome sequence because natural-product biosynthetic genes are present in clusters in microbial genomes" [microbiology cluster cited]. ". . . two hybrid polyketide peptides found in different species have recently been shown to be biosynthesized with unexpected halogenation events. . . . This transformation, using cryptic halogenation, represents a unique strategy for biochemical conversion and the enzymatic steps . . . were impossible to predict from DNA sequence alone." [chemistry cluster cited] (Van Lanen, 2006)

Reference to a "cryptic halogenation" event is a recurring theme mentioned in 16 of the 26 co-citation contexts and suggests that chemical studies have disclosed a process that was not previously expected using genetic methods.

To understand the relation of the microbiologically based drug development work to the chemical approach represented by this link, cue word occurrences were coded that might reveal the motivation behind the co-citations (Finney, 1979; Di Marco & Mercer, 2004). For example, the following passage reveals that the link is based on a perceived analogy between the process that goes on in the biological and chemical systems:

"Extending the *analogy* between biological hydroxylation and halogenation . . . when nature carries out hydroxylation at unactivated carbon sites … it turns to iron enzymes and generates high-valent oxoiron species as powerful oxidants." [italics added] (Vaillancourt, 2006)

In this instance the word "analogy" is considered as a cue word, and we group it with other related terms such as "similarity" or "parallel". Twenty of the 26 co-citing contexts contained cue words. Table 3 presents the cue word groupings occurring three or more times.

Table 3: Frequently occurring cue words in co-citing passages

Cue word grouping	Frequency
Discovery, novel, elucidated, remarkable	10
Analogy, similar, parallel, others like, complementary	9
Speculate, unsolved, questions, unclear, postulate	5
Recent, infancy	4
Promise, utility, extend	3

The top two cue words are discovery and analogy and these were often associated with the "cryptic halogenation" process noted above. This finding suggests that natural product and microbiology researchers saw in the chemical discoveries an analogy to the biosynthetic processes they had been studying in bacteria and thus the possibility that this chemical process might play a role in their syntheses and search for new drugs. Thus the interdisciplinary connection detected on a macro scale is in part the manifestation of an underlying perception of analogy in a highly technical micro-context (Holyoak & Thagard, 1996). This supports an earlier finding that also pointed to analogy as an important factor in forging interdisciplinary connections (Small, 1999).

In addition, the "speculate" cue word group with five occurrences suggests that the perceived analogy between biological and chemical processes represented a gap or hole in knowledge (Chen, 2006) and an opportunity for new discoveries. One author stated: "It is likely that other A-T didomains … will be discovered that carry out novel chemical transformations . . . " (Fischbach, 2006). This suggests that in some cases interdisciplinary co-citations can be forward looking and anticipatory, highlighting connections that researchers regard as potentially fruitful and foreshadowing future discoveries, what might be termed turning analogy into reality.

Conclusions

Because interdisciplinarity involves the interaction of two modes of organization in science and inaccuracies in either structure can potentially affect its measurement, it is important to undertake confirmatory studies based on the content of the interactions. This study illustrates that studying the discourse underlying an interdisciplinary connection can enrich our understanding of why and how such connections arise. The coding of cue words can help reveal the nature of the connections that authors are attempting to make between different fields.

Researchers' perception of an analogy between their problem and a discovery in another field and speculation that it might hold promise for their field appear to be important factors in driving interdisciplinary connections. If this finding is confirmed by further interdisciplinary linkage analyses, it may point to effective strategies for accelerating progress across many fields of science by uncovering and disseminating these speculative connections.

Acknowledgment

The author would like to thank Regina Fitzpatrick for assistance in text analysis.

References

Adams, J., Jackson, L., & Marshall, S. (2007). Report to the higher education funding council for England: bibliometric analysis of interdisciplinary research. Leeds: Evidence Ltd.

Boyack, K. W., Klavans, R., & Börner, K. (2005). Mapping the backbone of science. *Scientometrics,* 64(3), 351-374.

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, 57(3), 359-377.

Di Marco, C. & Mercer, R.E. (2004). Hedging in scientific articles as a means of classifying citations. *Working nodes of the American Association for Artificial Intelligence,* https://www.aaai.org/papers/symposia/spring/2004/ss-04-07/ss04-07-009.pdf.

ESI (2008). Essential Science Indicators. Thomson Reuters. http://sciencewatch.com/about/met/rf-methodology/.

Finney, B. (1979). Reference characteristics of scientific texts. Master's thesis. London: The City, University of London.

Fischbach, M.A. & Walsh, C.T. (2006). Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery and mechanisms. *Chemical Reviews*, 106, 3468-3496.

Fruchterman, T.M.J. & Reingold, E.M. (1991). Graph drawing by force-directed placement. *Software - Practice and Experience*, 21(11), 1129-1164.

Granovetter, M.S. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6), 1360- 1380.

Gray, H.B., (2003). Biological inorganic chemistry at the beginning of the 21st century. *Proceeding of the National Academy of Sciences USA*, 100 (7), 3563 – 3568.

Porter, A. L., Alex S. Cohen, A.S., Roessner, D.J. & Perreault, M. (2007). Measuring researcher interdisciplinarity. *Scientometrics* 72(1), 117-147.

Lévi-Strauss, C. (1966). The savage mind. Chicago: University of Chicago Press.

Schneider, J. (2006). Concept symbols revisited: naming clusters by parsing and filtering of noun phrases from citation contexts of concept symbols. *Scientometrics,* 68(3), 573-593.

Small, H. (1980). Co-citation context analysis and structure of paradigms. *Journal of Documentation*, 36(3), 183-196.

Small, H., Sweeney, E., & Greenlee, E. (1985). Clustering the Science Citation Index using co-citations. 2. mapping science. *Scientometrics*, 8(5-6), 321-340.

Small, H. (1999). A passage through science: crossing disciplinary boundaries. *Library Trends*, 48(1), 72-108.

Holyoak, K.J. & Thagard, P. (1996). Mental leaps: analogy in creative thought. Cambridge, Mass. MIT Press.

Vaillancourt, F.H., Yeh, E., Vosburg, D.A., Gameau-Tsodikova, S. & Walsh, C.T. (2006). Nature's inventory of halogenation catalysts: oxidative strategies predominate, *Chemical Reviews*, 106, 3364-3378.

Van Lanen, S.G. & Shen, B. (2006). Microbial genomics for the improvement of natural product discovery. *Current Opinion in Microbiology*, 9, 252-260.