Boundary Objects: Measuring Gaps and Overlap Between Research Areas

Laura Ridenour

University of Wisconsin-Milwaukee, School of Information Studies,
2025 E Newport Ave., Rm 3550, Milwaukee, WI 53211, <ridenour@uwm.edu>

Laura Ridenour is a doctoral student at the School of Information Studies, University of Wisconsin-Milwaukee. She holds a master's degree in information science from Indiana University Bloomington. Her interests include visualization and network theory.

Ridenour, Laura. "Boundary Objects: Measuring Gaps and Overlap Between Research Areas." *Knowledge Organization* 43 no. 1: 44-55. 35 references.

Abstract: The aim of this paper is to develop methodology to determine conceptual overlap between research areas. It investigates patterns of terminology usage in scientific abstracts as boundary objects between research specialties. Research specialties were determined by high-level classifications assigned by Thomson Reuters in their Essential Science Indicators file, which provided a strictly hierarchical classification of journals into 22 categories. Results from the query "network theory" were downloaded from the Web of Science. From this file, two top-level groups, economics and social sciences, were selected and topically analyzed to provide a baseline of similarity on which to run an informetric analysis. The Places & Spaces Map of Science (Klavans and Boyack 2007) was used to determine the proximity of disciplines to one another in order to select the two disciplines use in the analysis. Groups analyzed share common theories and goals; however, groups used different language to describe their research. It was found that 61% of term words were shared between the two groups.

Received: 6 October 2015; Revised: 15 December 2015; Accepted: 16 December 2015

Keywords: boundary objects, research areas, domain analysis, knowledge, science, terminology

1.0 Introduction

Boundary objects (Bowker and Star 1999) are either concrete or abstract objects that have flexible meaning for multiple communities of practice, and can serve as a communication point across these communities. Examples of boundary objects include ontologies, metadata crosswalks, and concepts. Boundary objects can potentially enhance cooperation, coordination, and knowledge management across different disciplines involved in scientific research. Identification of boundary objects can be done using a number of quantitative, qualitative, and mixed methods. These include cognitive work analysis, discourse analysis, and natural language processing. Cognitive work analysis (Marchese and Smiraglia 2013) allows researchers to study members in a domain in their place of work. From observations made within the work environment, taxonomies of domain vocabulary may be created. Corpus-driven natural language processing methods can exploit linguistic and statistical features present in text (Velardi, Fabriani and Missikoff 2001), allowing for the creation of richer ontologies.

Knowledge (Dahlberg 2006) can be shared by means of language through space and time. Furthermore, knowledge must be encoded (Wilson 1978) in order for it to be exploited and controlled. Once recorded, it can be exploited in a number of ways; some that serve to make it easier to control, and others that can result in the creation of new knowledge. This exploitation takes place in Popper's (1979) partially-autonomous third world of objective knowledge, in which published scientific literature resides. In World 3, knowledge is available to be observed, interpreted, and applied to ideas in the minds of other people.

Language, once encoded in written format, is typically stored in information systems as text. Natural language processing technologies can be leveraged to extract features of recorded texts and compare those features to those of other texts or sets of texts. Features of text that are extracted include grammar, syntax, terminology, and to some extent semantics. Terms found in texts can be used to compile community-specific lexicons. Issues of synonymy and polysemy (Fellbaum 1998) arise when analyzing text; such issues require the user or system to disambiguate shared or multiple meanings for symbols found in text. Synonymy occurs when a concept has multiple ways to express its meaning, while issues of polysemy occur when the same symbol can convey multiple meanings. Thesauri such as WordNet (Princeton University 2010) can provide a way to disambiguate meanings for systems and users encountering ambiguous word uses; however, users may use context in order to determine an interpretation that makes sense to them for a term.

We assert that terminology can be examined by various pragmatic contexts, such as spoken, written, or community-specific language, that are found in discourse to determine conceptual similarity in material. Different facets of terminology examined reveal distinct features of that terminological domain. Tomuro (2002) investigated automatic extraction of question terminology from a corpus using two feature sets in order to classify question types. Question terminology was found to be highly lexical, and that the use of words that typically appear in idiomatic phrases would be more effective for categorization than semantics. Semantically rich ontologies categorize both concepts and semantic relationships.

Conceptual and terminological gaps exist in literature of neighboring domains. Domains encompassing groups that share common goals frequently use different vocabulary to explain similar concepts and phenomena. As such, the sharing of knowledge is impaired. Knowledge located in databases where researchers are familiar with their own domain silos provides a challenge, as related disciplines may explore the same or similar research areas but use very different terminology to describe observed phenomena, constructed methodologies, and results within their published work. Classification provides frameworks for understanding within disciplines and reveals underlying epistemological stances, but lacks ties to explain points of information exchange between domains. Categories can be artificially imposed on information landscapes and unintentionally impair information seeking.

Systems such as the Web of Science use categories to facilitate information browsing and retrieval. Current system design does not account for changes in the information landscape, and indexing systems are not updated in a timely fashion to support new terminology. As in the case of Sampalli et al (2010), normalized term matching and expanded term matching can provide an aid for understanding in multidisciplinary settings. Providing an understanding of the conceptual overlap between domains, as well as solutions for mapping between terminologies found in domains, will help address the problem of a lack of interdisciplinary communication in interdisciplinary and multidisciplinary research.

2.0 Literature review

2.1 Boundaries

In determining the concepts central to a field of research, the domain in question must first be defined. According to Smiraglia (2012, 112), "a domain is a group with an ontological base that reveals an underlying teleology, a set of common hypotheses, epistemological consensus on methodological approaches, and social semantics." Communities of some sort are used to define the "who" part of the intension of a domain, and can be variously categorized in Smiraglia's view of domains, discourse communities, invisible colleges, in addition to Bowker and Star's (1999) view of communities of practice. Based on this definition, a domain, for the purposes of analysis, is a group with shared goals. Some domains are not easily analyzed as members of the domain of shared interest are part of an invisible college; actively participating in separate research groups and using communication channels that may be obscured through use of private email or that are otherwise prevented from creating easily traceable digital footprints.

Boundaries are necessary for understanding and interpretation of the world, and allow for easier objective justification of science (Popper 2002), but create artificial divisions, or silos, which lead to the isolation of knowledge. Grouping individual research areas together allows for the analysis of these boundaries. According to Brier (2004), knowledge can be viewed as the phenomenon that can occur when documents are mentally interpreted under "correct circumstances." When individual researchers are involved (Ridenour 2015), documents must be accessed, understood, and reinterpreted into the lens of the individual researcher encountering these documents. As knowledge is encoded in language (Dahlberg 2006), it is important to realize that the intended meaning of recorded knowledge can be changed by the passage of time. Changes in meaning over time (Gulla et al. 2010) have been empirically measured through the shift of meaning in ontologies.

Knowledge organization seeks to group like concepts. Furthering our understanding of what constitutes likeness requires us to create an association between two objects. Describing this association gives us the power to create systems that allow us to leverage the relationships identified and recorded in some way between many objects. Objects exhibit like-characteristics and patterns, which we recognize through processes of understanding; taking advantage of these inherent patterns (Wilson 1978, 3) is one of the ways in which we can exploit and control knowledge.

Knowledge domains (Hjørland and Albrechtsen 1995) influence the formation of individuals' understanding and interpretation of concepts. Bowker and Star (1999) suggest that a period of "learning-as-membership," or indoctrination, into a community of practice is required for an initiate to learn the nuances of their new community. After the new member of the community becomes well versed in the argot of the domain, their fluency improves and their ability to discuss concepts with less specialized terminology diminishes. One way to understand this phenomenon is process modeling (Davenport and Cronin 2000), which creates outlines of tacit knowledge held by individuals within an organization.

Zhang and Jacob (2013) discuss how changes in information environments alter the dimensions that boundaries span, and how they can be crossed. Exploring methods of harmonizing scientific terminology by identifying boundary objects and creating ways to span boundaries will allow for furthering all scientific discoveries through increased ease of research. Choices in classification reflect how individual disciplines view the world, and the interoperability of categorization across domains. Domains establish their own argot, which iteratively creates and affects their unique epistemological standpoints. Congruent with Habermas' pragmatics, in which an individual will create a speech act in a manner so that it is more likely to be seen as true by its intended audience (Habermas 1998), researchers adopt and use terminology to describe concepts with which they are familiar, and which is less likely to be rejected as invalid by the audience they target, be it a journal, discourse community, or their home research group.

2.2 Boundary objects in practice: ontologies and their creation

Strictly formalized knowledge can be found in top-down solutions such as ontologies. Svenonius (Svenonius 2000) asserts that factual claims about the world that are asserted as truths can be encoded in description logic or an XML-based language such as the Web Ontology Language (OWL). Different methods of analysis for boundaries exist, and are applied in various fields of research including KO, where the analysis is frequently used for the creation of ontologies.

Practical applications of boundary objects can be found in multiple fields. In the medical domain of complex chronic conditions, Shepard and Sampalli (2012) created and examined an ontology to evaluate its effectiveness in coordinating care between multidisciplinary communities of practice. They intended for the boundary object approach to enhance communication in the complicated domain of complex conditions. SNOMED CT provided the terminology with which to standardize terms found during the audit of one hundred patient medical charts. Terminological inconsistencies were documented and charted. A list of standardized terms was created, and prototype charts were re-coded by a multidisciplinary team of clinicians. The prototype ontology was created using multidisciplinary classes, tested by clinicians browsing concepts and relationships between them, and evaluated by domain experts. It was found that clinicians strongly agreed the created ontology was useful for the conditions it was intended to address.

Cognitive Work Analysis (CWA), a qualitative methodology for examining members of a community as they work together. Marchese (2012) used CWA to analyze data to create ontologies from emergent vocabulary in a human resources firm. Specificity of the vocabulary depend upon factors such as the actor's fluency in the argot of the domain, their role taken, and the context of the conversation or exchange examined. Terminology can act like pivot points (Marchese and Smiraglia 2013), either facilitating or hindering actors in communicating, based on the actor's understanding of the terminology. Terms found to serve as potential boundary objects in informal communications were frequently verbs that were used as nouns.

Semantic drift can occur within different versions of ontologies, as observed by Gulla et al. (2010) in observations made of individual concepts' evolution as found in a business sector ontology. Methodology used to detect the phenomenon linked concepts to their linguistic representation, or "concept signature," to trace evolution over the course of four years. The "concept signature," instead of being a representation of a concept, functions more like a map and demonstrates how linguistic signs are used in reference to and in discussion of the concept; they are not imposed on mappings to phenomena, but instead relate linguistically to other concepts. Signatures are represented as vectors of linguistic units. While ontologies must evolve to accommodate and reflect the creation and codification of new knowledge, drift from the meaning of concepts as originally described can be found in such a controlled environment. The concept's semantic value may change over time due to societal, domain, or personal revelations related to the concept itself. Ontological evolutionary changes can be observed as existential changes or relational changes. An existential change occurs when a concept that is outdated is omitted from new versions of the ontology, or when a new concept becomes standard, where a relational change occurs when the taxonomic and non-taxonomic relationships between concepts change.

2.3 Corpus approaches to conceptual overlap

Patrick et al. (2003) examined shared use of core terms within the medical domain of opthalmology by analyzing proportions of frequencies of terms in a corpus. Their study involved expert analysis of terms selected from the Unified Modeling Language System (UMLS) Large Scale Vocabulary Test (LSVT) as opthalmic or non-opthalmic, analyzing only the terms both experts agreed to as "core" to the domain. Each term and domain combination was compared for overlapping confidence intervals.

The term "semantic similarity" can have multiple meanings; Lemaire and Denhiere (2006) present it as "an association, that is the mental activation of one term when another term is presented, which is what association norms capture." The association strength of two words can be ascertained through examining the correlation of word co-occurrence, presumed to be high, and word similarity. Latent Semantic Analysis was the cognitive model used to extract and analyze semantic information from children's literature in French. Their simulation demonstrated that semantic similarity was associated with co-occurrence, but that assuming frequency of high co-occurrences as being indicative of semantic similarity could cause people to introduce bias in the interpretation.

2.4 Interoperability of knowledge organization systems

Compatibility of systems is critical for interchange of information between knowledge organization systems, but systems use many different vocabularies that can limit subject access based on a lack of interoperability. Zeng and Chan (2004) analyzed methodologies implemented for creating interoperability between knowledge organization systems. In their view, three general categories of knowledge organization systems exist, ordered from the least complex to the most complex: term lists, classifications and categorization schemes, and relational vocabularies.

When creating systems for representing boundary object concepts between domains, experts in the domain contribute their knowledge to ontologies that represent their understanding of what things in the domain "are," or how they are understood and defined by people within their specialty, and how things in the domain interact with one another. In the biomedical field, UMLS integrates biomedical terminology. Biomedical texts are mapped to the UMLS Metathesaurus, which involves the manual markup of biomedical literature by experts. Features including interactions between conditions and drugs can then be extracted from the texts, helping to enrich available knowledge about research done in this field. Significant effort has been made to further the knowledge represented in UMLS, and updates are made available for newly added terms, as well as terms that have fallen into disuse.

Information outside of a specialty is not considered to be as pertinent when used in support of a hypothesis inside of a field. Wilson's view (Wilson 1993) compliments Kuhn's scientific paradigms (Kuhn 1962), but does not propose a solution for breakdowns in cross-disciplinary communication. Borrowing paradigms from other disciplines encourages the adoption of language used, but according to Wilson (1993), such borrowed terminology may not be regarded seriously by others within a discipline. Additionally, reusing terminology may cause further confusion.

The variety of literature present on analyzing conceptual overlap was published in venues recognized by distinct, self- and publisher-identified domains and research areas. This in of itself illustrates the complexity involved when determining conceptual overlap between domains.

3.0 Methodology

Methodology was developed to answer the following research questions:

1. What are areas of conceptual overlap between two conceptually similar domains studying the same phenomenon?
2. What terms do researchers use when describing their research in scientific abstracts?
   a. How do word and phrase use differ between disciplines when describing similar concepts?

The Web of Science (WoS) was queried for "network theory," and the search was limited articles and conference papers published in English. The Essential Science Indicators file documenting the top-level WoS classification of journals was downloaded from the Thomson Reuters Website (Thomson Reuters 2013). WoS results were merged into a single file, and processed using Sci2 (Sci2 Team 2009, 2). The resulting Web of Science file and the ESI file were imported into Access, and merged on the "Journal Title (Full)" field. The Map of Science (Klavans and Boyack 2007) was used to determine proximity of disciplines, as it mapped connections found between disciplines and articles tracked for five years, from

2001 to 2005, and calculated the likelihood of included disciplines to shift over the next ten years. As such, "Social Sciences, General" (SSG) and "Economics & Business" (E&B) were selected to conduct the analysis. The "Original Keywords" were extracted and analyzed for all disciplines to examine how members of disciplines described their own work.

Abstracts were stemmed, tokenized, and the stop-words were removed to examine topicality in Sci2. WordStat (Provalis Research 2010) was used to perform content analysis and basic statistics on the corpus. Single word occurrences excluding stop list words were analyzed within WordStat by using the number of occurrences of each word as the dependent variable, and the Web of Science assigned high-level category of the journal in which the article was published as the independent variable. Because the two sets were of the same query and contained words surrounding the same conceptual content in published scientific literature, it was assumed that no significant outliers existed between the two sets of words.

4.0 Results

The original dataset consisted of 2,769 results. After merging the Essential Science Indicators Journal Classification file with the original results, 2,259 records remained, the count and distribution of records across the 22 categories are displayed in Table 1. The majority of these records were categorized as "Social Sciences, General," (SSG) which includes several social science fields. As SSG contained the greatest number of records, it was selected as one of the domains for the comparison. From this list, Economics & Business (E&B), which contained 295 records, was selected and compared to 295 randomly sampled records from Social Sciences, General. E&B was selected because of its calculated proximity to SSG in maps of science, and thus, inferred similarity based on location.

The majority of these records were categorized as SSG, which includes several social science fields. As SSG contained the greatest number of records, it was selected as one of the domains for the comparison. From this list, E&B were selected and compared to 295 randomly sam-

Web of Science Category	Total Records Returned
SOCIAL SCIENCES, GENERAL	759
ECONOMICS & BUSINESS	296
ENGINEERING	293
PHYSICS	210
COMPUTER SCIENCE	130
CHEMISTRY	129
ENVIRONMENT/ECOLOGY	78
CLINICAL MEDICINE	70
BIOLOGY & BIOCHEMISTRY	58
PSYCHIATRY/PSYCHOLOGY	50
MATHEMATICS	43
NEUROSCIENCE & BEHAVIOR	37
MATERIALS SCIENCE	34
MULTIDISCIPLINARY	34
IMMUNOLOGY	17
PLANT & ANIMAL SCIENCE	17
GEOSCIENCES	15
MOLECULAR BIOLOGY & GENETICS	8
AGRICULTURAL SCIENCES	6
PHARMACOLOGY & TOXICOLOGY	5
MICROBIOLOGY	4
SPACE SCIENCE	2
Grand Total	**2,295**

Table 1. ESI categories and distribution of records in search results

pled records from SSG. E&B was selected because of its calculated proximity to SSG in maps of science, and thus, inferred similarity based on its mapped location.

Comparing single word frequencies between E&B and SSG revealed the top word occurrences in common between the two categories (Table 2).

Phrases, or n-grams, provided a clearer picture of what was being discussed in abstracts by illustrating noun phrases used in both domains to describe concepts in scientific publications related to network theory. The top 20 n-grams between 2 and 5 words long, displayed in Table 3, were compared between the two WoS categories. The top three results are terms used to describe the units of network analysis in both disciplines.

Both E&B and SSG followed a Zipfian distribution for their respective represented phrases. When displayed in stacked bar charts, the distribution of matching n-grams, such as "actor-network theory," followed a similar

Word	Economics	Social Sciences
NETWORK	894	580
THEORY	428	364
STUDY	265	220
FIRM	255	41
RESEARCH	250	166
SOCIAL	237	273
ACTOR	208	335
PAPER	194	198
MARKET	172	60
KNOWLEDGE	170	61
PROCESS	144	151
INNOVATION	140	30
RELATIONSHIP	138	63
BUSINESS	131	26
BASE	128	82
MANAGEMENT	127	46
APPROACH	115	160
TECHNOLOGY	112	143
ORGANIZATION	112	47
ANALYSIS	105	160
NEW	105	121
DEVELOP	101	100
RESOURCE	99	51
TIE	99	25
MODEL	98	40
EFFECT	95	38
DEVELOPMENT	91	102
ORGANIZATIONAL	91	16
SYSTEM	88	80
RESULT	87	45
EXAMINE	86	61
ARTICLE	84	132
PRACTICE	84	127
PERFORMANCE	82	17
INFLUENCE	81	41
IMPLICATION	80	41
STRUCTURE	79	54
THEORETICAL	77	63
ROLE	76	57
CHANGE	74	76

Table 2. Top 40 single words in both groups

Phrase	% Total	n-length	Economics	Social Science
NETWORK THEORY	74.50%	2	222	217
SOCIAL NETWORK	16.10%	2	67	28
ACTOR NETWORK	11.00%	2	26	39
SOCIAL NETWORK THEORY	11.00%	3	45	20
ACTOR NETWORK THEORY	10.20%	3	23	37
THEORETICAL FRAMEWORK	3.90%	2	15	8
FUTURE RESEARCH	3.70%	2	15	7
NETWORK ANALYSIS	3.20%	2	10	9
PRACTICAL IMPLICATION	3.20%	2	13	6
NETWORK STRUCTURE	2.90%	2	13	4
HUMAN ACTOR	2.90%	2	2	15
SCIENCE AND TECHNOLOGY	2.90%	3	5	12
BRUNO LATOUR	2.50%	2	1	14
SOCIAL SCIENCE	2.50%	2	4	11
EMPIRICAL RESEARCH	2.50%	2	10	5
NETWORK PERSPECTIVE	2.40%	2	11	3
SCIENCE AND TECHNOLOGY STUDY	2.00%	4	3	9
SOCIAL NETWORK ANALYSIS	2.00%	3	6	6
NETWORK APPROACH	2.00%	2	5	7
BASE VIEW	2.00%	2	12	0

Table 3. Phrase frequency comparison, top 20 phrases

distribution per n-gram. Overlapping distributions of unique phrases as found in the dataset are displayed in Figures 1 and 2; Figure 1 focuses on the Zipf-distribution as phrases are ordered focusing on economics, where Figure 2's distribution is focused on social sciences. This type of comparative layout of charts shows n-gram distribution in each domain, as well as the higher occurrence of total phrases found in SSG.

Total phrase distributions for both categories are nearly identical (Figure 3). Phrases occurring four or more times, the cut off for the long tail, occurred with greater frequency in social sciences (Figure 4).

Phrases, or n-grams, between the domains were compared to determine conceptual overlap. Counts were calculated based on the 61% of terms found in the dataset were shared between the two domains, and could be considered boundary object terms, as is shown in Figure 5.

5.0 Discussion

In order to answer both research questions, clear delineation between domains had to be defined for the purpose of analysis. From a practical standpoint, journals provide venues for community discourse. As such, dividing the dataset by journals provided a means to draw boundaries for "domains" for the purpose of this analysis. The Thomson Reuters Essential Science Indicators file provided a widely published, clearly delineated, and recently updated classification of science and social science journals that was compatible with data fields represented in Web of Science data. This high-level classification was used for analysis because it provided a strictly hierarchical classification for journals represented in the dataset. Analyzing sub-fields could return more pertinent classifications, as individual papers could be more or less favored by different communities. This type of analysis was made problematic by the assignment of multiple "Research Areas" to individual journals by Thomson Reuters.

Gathering and analyzing the work of self-identified members of research specialties and the work of colleagues that members of the research specialties identify in allied fields may provide clearer, smaller domain boundaries from which to perform boundary object analysis. This type of data gathering and analysis would be limited by not only the ability to contact members of a research specialty, and their ability to identify themselves as belonging to any particular research specialty. As Wilson (1993) pointed out, communication breakdowns occur between specialties, not individual researchers. The presence of more n-grams occurring more than four times in the SSG dataset may indicate more consistent use of research-specialty specific terminology than in E&B (Table 4), but this requires further

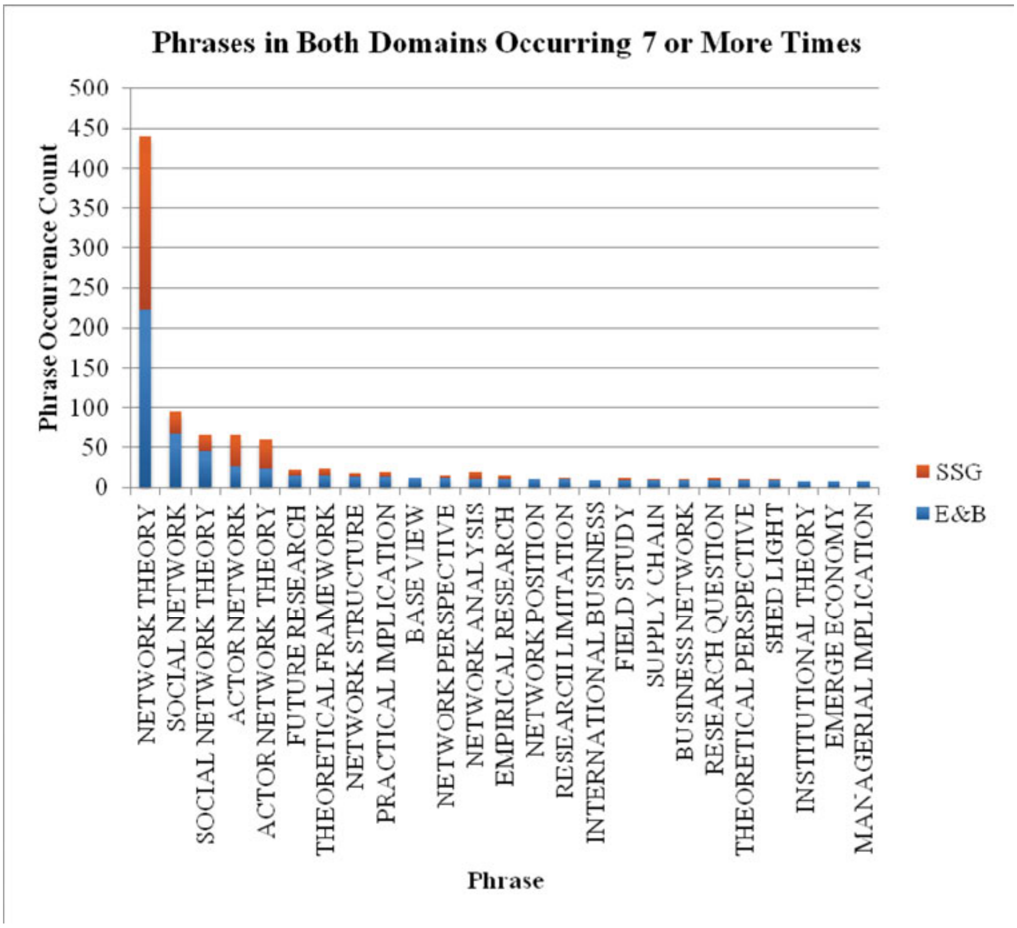

Figure 1. Distribution of phrases between two groups, Zipf focus on economics

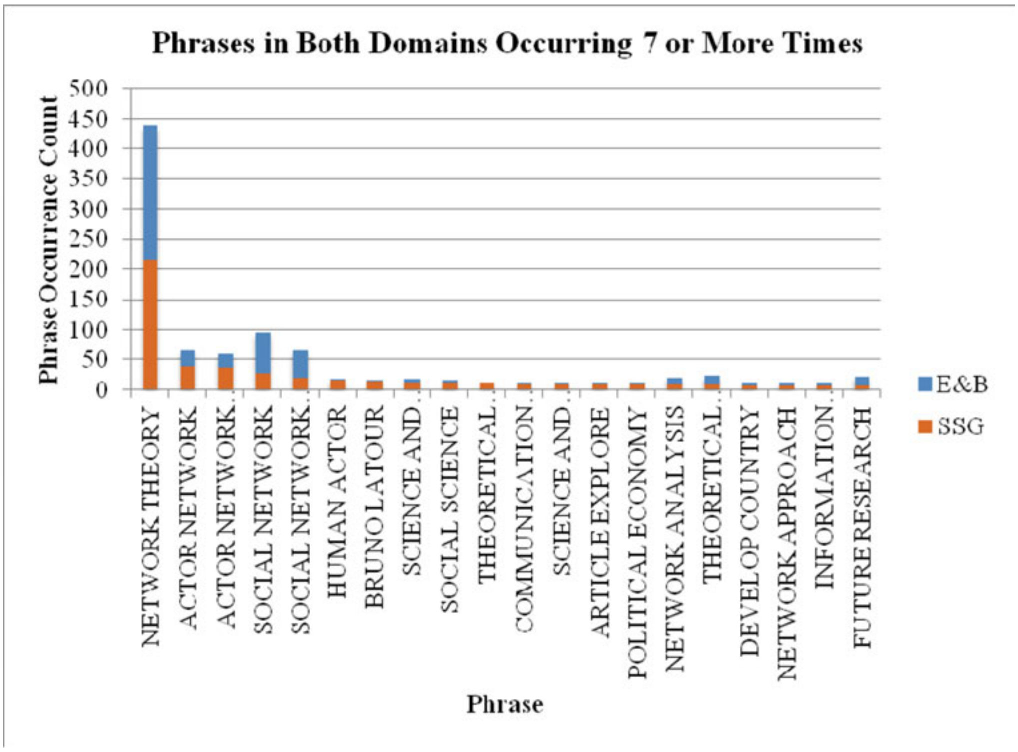

Figure 2. Distribution of phrases between groups, Zipf focus on social sciences

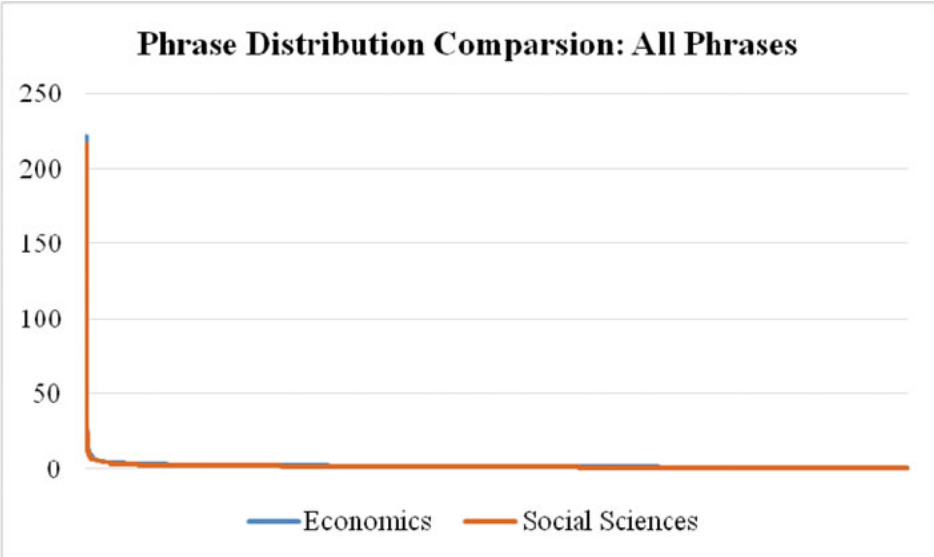

Figure 3. Total phrase distribution for both categories

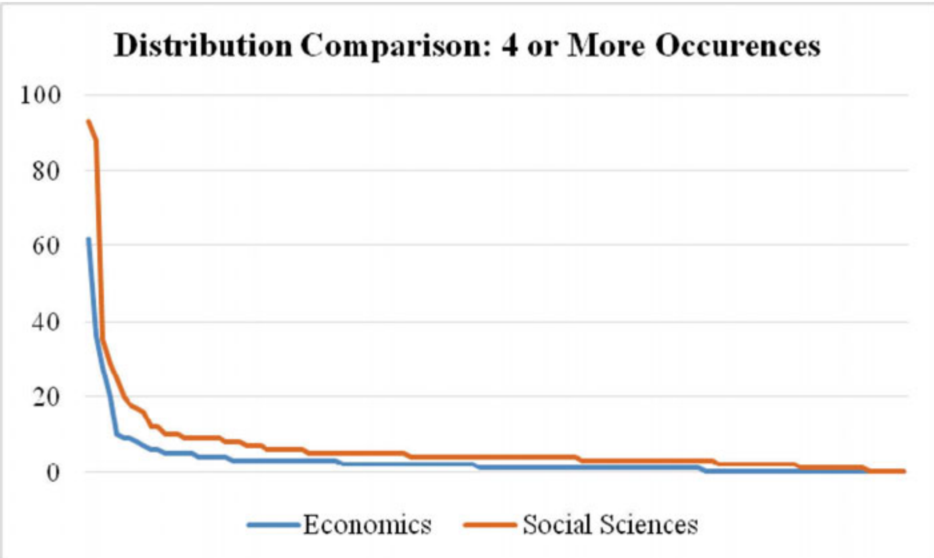

Figure 4. Distribution of phrases occurring four or more times

investigation. That 74% of the phrases used in the domains analyzed were "network theory" is likely an artifact of the way the data was gathered.

Another approach is to examine works cited by members of a clearly defined research specialty to determine which journals they cite from, and perform another step back to analyze commonly shared terms between domains. This type of analysis could be augmented using latent semantic analysis (Landauer, Foltz, and Laham 1998), a technique that analyzes terms against the documents in which they are used. LSA provides a way to automatically extract, contextualize, and represent meaning, as it is more than a measure of co-occurring term counts or usage correlations.

Language can be analyzed in units of varying sizes. Single words can have multiple meanings, whereas n-grams, or phrases, provided a clearer picture of what was being discussed and a better indicator of potential boundary objects between domains. When analyzing textual data, stopwords or words that do not contain enough significance in relation to the content of the text to be indexed such as prepositions and articles, are typically removed. These types of words can skew distribution of meaning-containing terms in the most basic of content analysis. Academic articles contain words and phrases that contribute to the architecture of understanding the content that is to be presented, but do not necessarily contribute to the conceptual aboutness of the article (sometimes terms are key to understanding the article, but are more pragmatic in nature). These academic stopwords and phrases, such as "this paper" and "article explore" occurred in the n-gram

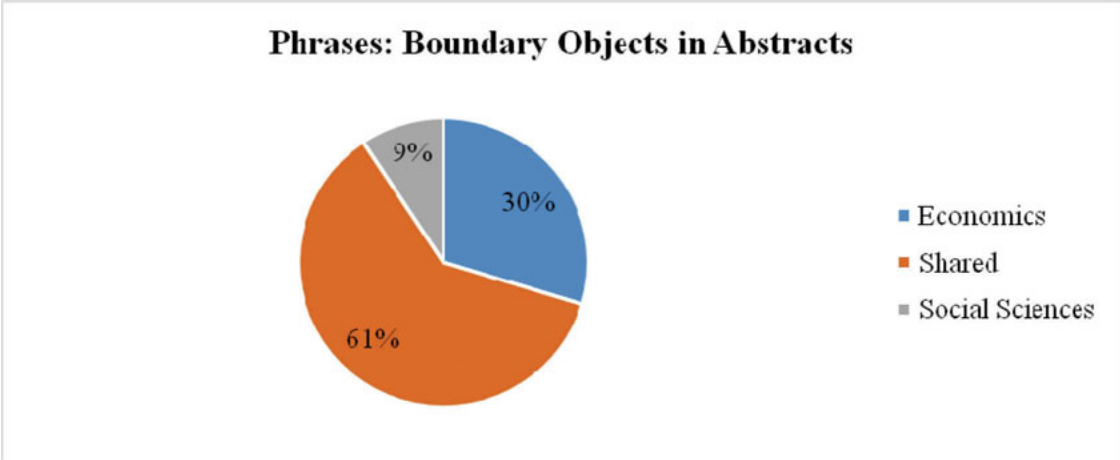

Figure 5. Computed boundary objects between the two domains

analysis, but were not as frequent as stray academic nouns found in the analysis conducted of single nouns.

Categorization reflects the epistemological standpoints held by entities imposing the system of classification; in the Web of Science, journal classification into research areas reflects the view of the works published by those journals from the perspective of a bibliographic database company. Word choice can demonstrate the preferred terminology for each domain, assuming it is consistent. Consistent use of terminology (Hjørland 1997) can demonstrate the maturation of a discipline, as members of more mature disciplines tend to cite similarly. As there appears to be consistent use of terminology describing concepts related to network theory in these domains (Morris and Van der Veer Martens 2008), it is evidence that the disciplines have matured or are in the process of maturing.

Informal communication and invisible colleges in the sciences (Barjak 2006) may influence the use of terminology and citation patterns. It is also possible that one discipline in the analysis influenced the use of terminology in the other, as a less mature discipline draws on the work and conceptual representations of more formalized disciplines. As terminology in this area of research between the two disciplines is fairly consistent, with 61% of multiple word terms shared between the two domains, it is suspected that there is successful interdisciplinary communication. Investigating citation patterns between the two disciplines will provide a means of triangulation, and reveal if they indeed share and draw from a base of common knowledge.

6.0 Conclusion

Boundary objects can be identified using linguistic analysis. Locating conceptually similar documents when using terminology that differs from multiple disciplines that in current information systems is not easy, but the creation of domain-specific corpora for analysis allows for identification of shared terminology between disciplines. Formalized domains tend to have self-similar information seeking habits (Hjørland 1997), and it is reasonable to assume that they are less likely to explore information retrieval systems in a way that would locate conceptually similar documents from other disciplines. Providing means to increase interdisciplinary research is currently a popular trend, and this study establishes term-based boundary objects between two disciplines that share a common area of interest. Next steps include conducting a direct citation analysis of the data set to determine overlap in literature sources influencing both disciplines, and reconciliation of conceptually similar terminology while removing academic stopwords from the analysis, as well as implementation of LSA to provide meaningful context of terms within their respective documents. Citation analysis will reveal if there is active use of the same literature between both research areas; should the dataset reveal citations from both groups between each set, the groups would be engaged in shared discourse.

References

Barjak, Franz. 2006. "The Role of the Internet in Informal Scholarly Communication." *Journal of the American Society for Information Science and Technology* 57: 1350–67.

Bowker, Geoffrey C. and Susan Leigh Star. 1999. *Sorting Things out: Classification and Its Consequences*. MIT Press.

Brier, Søren. 2004. "Cybersemiotics and the Problems of the Information-Processing Paradigm as a Candidate for a Unified Science of Information behind Library and Information Science." *Library Trends* 52: 629–57.

Dahlberg, Ingetraut. 2006. "Knowledge Organization: A New Science?" *Knowledge Organization* 33: 11–9.

Davenport, Elisabeth and Blaise Cronin. 2000. "Knowledge Management: Semantic Drift or Conceptual Shift?" *Journal of Education for Library and Information Science* 41: 294–306. doi:10.2307/40324047.

Fellbaum, Christiane. 1998. *WordNet*. Wiley Online Library.

Gulla, Jon Atle, Geir Solskinnsbakk, Per Myrseth, Veronika Haderlein and Olga Cerrato. 2010. "Semantic Drift in Ontologies." In *WEBIST* 13-20.

Habermas, Jürgen. 1998. "Actions, Speech Acts, Linguistically Mediated Interactions, and the Lifeworld (1998)." In *On the Pragmatics of Communication*, edited by Maeve Cooke. Studies in Contemporary German Social Thought. Massachusetts Institute of Technology, 216-55.

Hjørland, Birger. 1997. *Information Seeking and Subject Representation: An Activity-Theoretical Approach to Information Science*. New Directions in Information Management 34. Westport, Connecticut: Greenwood Press.

Hjørland, Birger and Hanne Albrechtsen. 1995. "Toward a New Horizon in Information Science: Domain-Analysis." *Journal of the American Society for Information Science* 46: 400–25.

Klavans, Richard and Kevin W. Boyack. 2007. "Maps of Science: Forecasting Large Trends in Science." In *Places & Spaces: Mapping Science*, edited by Katy Börner and Julie M. Davis. 3rd Iteration: The Power of Forecasts. http://scimaps.org.

Kuhn, Thomas S. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago press.

Landauer, Thomas K, Peter W Foltz and Darrell Laham. 1998. "An Introduction to Latent Semantic Analysis." *Discourse Processes* 25: 259–84.

Lemaire, Benoit and Guy Denhiere. 2006. "Effects of High-Order Co-Occurrences on Word Semantic Similarity." *Current Psychology Letters. Behaviour, Brain & Cognition* 18.

Marchese, Christine. 2012. Impact of Organizational Environment on Knowledge Representation and Use: Cognitive Work Analysis of a Management Consulting Firm. PhD diss. Long Island University.

Marchese, Christine and Richard P. Smiraglia. 2013. "Boundary Objects: CWA, an HR Firm, and Emergent Vocabulary." *Knowledge Organization* 40: 254–59.

Morris, Steven A. and Betsy Van der Veer Martens. 2008. "Mapping Research Specialties." *Annual Review of Information Science and Technology* 42: 213–95.

Patrick, Timothy B., John C. Reid, Mary Ellen Sievert, Frances E. Rice, James W. Gigantelli, Jade S. Schiffman and Mark E. Shelton. 2003. "A Text Corpus Approach to an Analysis of the Shared Use of Core Terminology." *Proceedings of the American Society for Information Science and Technology* 40: 200–5.

Popper, Karl. 1979. "Three Worlds." Ann Arbor, Michigan. https://www.thee-online.com/Documents/Popper-3Worlds.pdf

Popper, Karl. 2002. *The Logic of Scientific Discoveries*. London; New York: Routledge.

Princeton University. 2010. "About WordNet." Princeton University. http://wordnet.princeton.edu

Provalis Research. 2010. "WordStat." http://provalis research.com/products/content-analysis-software/

Ridenour, Laura. 2015. "Concept Term Repurposing: Framing Shifts in Domains and Terminology." In *2015 North American Symposium on Knowledge Organization June 18-19, 2015, Los Angeles, CA*.

Sampalli, Tara, Michael Shepard, Jack Duffy and Roy Fox. 2010. "An Evaluation of SNOMED CT® in the Domain of Complex Chronic Conditions." *International Journal of Integrated Care* 10: 1-10.

Sci2 Team. 2009. "Science of Science (Sci2) Tool." Indiana University and SciTech Strategies. http://sci2.cns.iu.edu

Shepard, Michael, and Tara Sampalli. 2012. "Ontology as Boundary Object." In *Categories, Contexts, and Relations in Knowledge Organization: Proceedings of the Twelfth International ISKO Conference 6-9 August 2012 Mysore, India*, edited by A. Neelameghan and K.S. Raghavan. Advances in Knowledge Organization 13. Würzburg: Ergon-Verlag, 131–7.

Smiraglia, Richard P. 2012. "Epistemology of Domain Analysis." In *Cultural Frames of Knowledge*, edited by Richard P. Smiraglia and Hur-li Lee, 111–24. Würzburg: Ergon.

Svenonius, Elaine. 2000. *The Intellectual Foundation of Information Organization*. Cambridge, Massachusetts: MIT Press.

Thomson Reuters. 2013. "Thomson Reuters Journal List for Essential Science Indicators." http://sciencewatch.com/info/journal-list

Tomuro, Noriko. 2002. "Question Terminology and Representation for Question Type Classification." *Association for Computation Linguistics*, COLING-02 on COMPUTERM 2002: Second International Workshop on Computational Terminology 14: 1–7.

Velardi, Paola, Paolo Fabriani and Michele Missikoff. 2001. "Using Text Processing Techniques to Automatically Enrich a Domain Ontology." In *Proceedings of the International Conference on Formal Ontology in Information Systems October 17 - 19, 2001 Ogunquit, ME, USA,* edited by Nicola Guarino, Barry Smith and Christopher Welty. New York: ACM, 270-84.

Wilson, Patrick. 1978. *Two Kinds of Power: An Essay on Bibliographical Control*. Vol. 5. University of California Press.

Wilson, Patrick. 1993. "Communication Efficiency in Research and Development." *Journal of the American Society for Information Science* 44: 376–82.

Zeng, Marcia Lei and Lois Mai Chan. 2004. "Trends and Issues in Establishing Interoperability among Knowledge Organization Systems." *Journal of the American Society for Information Science and Technology* 55: 377–95.

Zhang, Guo and Elin K. Jacob. 2013. "Understanding Boundaries: Physical, Epistemological and Virtual Dimensions." *Information Research* 18, no. 3: paper C21. http://InformationR.net/ir/18-3/colis/paperC21.html